PTI

SYSTEMY INFORMATYCZNE ZASTOSOWANIA I WDROŻENIA 2003

TOM III
część 2

Pod redakcją
JANUSZA K. GRABARY
JERZEGO S. NOWAKA

Wydawnictwa Naukowo-Techniczne
Warszawa - Szczyrk 2003

SYSTEMY INFORMATYCZNE ZASTOSOWANIA I WDROŻENIA 2003

TOM III
część 2

Pod redakcją
JANUSZA K. GRABARY
JERZEGO S. NOWAKA

Wydawnictwa Naukowo-Techniczne
Warszawa - Szczyrk 2003

Recenzenci:
Prof. WSM dr hab. Marek Greniewski
Prof. P.Cz. dr hab. inż. Sławomir Iskierka
Prof. dr hab. Aleksander Katkow
Prof. P.Wr. dr hab. Zygmunt Mazur
Prof. P.Cz. dr hab. Henryk Piech
Prof. P.Cz. dr hab. Janusz Szopa
Prof. U.Sz. dr hab. Zdzisław Szyjewski

Wydanie publikacji dofinansowane przez Komitet Badań Naukowych
i Zarząd Główny Polskiego Towarzystwa Informatycznego

ISBN 83-204-2869-6 Całość
ISBN 83-204-2872-6 Tom 3

Indeks autorów opracował mgr inż. Jarosław Łapeta
Redakcja techniczna – mgr inż. Tomasz Lis

Druk wykonano w Zakładzie Graficznym Politechniki Śląskiej w Gliwicach
zam. 184/03 nakł. 400

Polskie Towarzystwo Informatyczne organizuje corocznie konferencję w Szczyrku poświęconą sprawom rozwoju informatyki. W 2001 r. motywem przewodnim był temat efektywności zastosowań systemów informatycznych, wynikający ze współpracy z Instytutem Ekonometrii i Informatyki Politechniki Częstochowskiej. Duże zainteresowanie tą tematyką w aspekcie obserwowanych trudności w szybkich wdrożeniach aplikacji informatycznych spowodowało, że w 2002 r. postanowiono przygotować dla uczestników XIV Górskiej Szkoły PTI – Szczyrk 2002 opracowanie stanowiące reprezentatywny przegląd doświadczeń z zakresu oceny efektywności zastosowań informatyki, wzbogacone o problematykę tzw. *business intelligence* i zarządzania wiedzą. Z kolei w 2003 z okazji XV Jubileuszowej Górskiej Szkoły PTI postanowiono przygotować tom specjalny, obejmujący problemy informatyki w epoce globalizacji z racji wejścia Polski do Unii Europejskiej. Autorzy artykułów poświęconych globalizacji gospodarki nie ukrywają problemów związanych z przygotowaniem polskich systemów informacyjnych państwa w przededniu wejścia Polski do Unii. Zaproszenia wystosowano do przedstawicieli placówek naukowych, przedsiębiorstw i instytucji oraz firm informatycznych.

Nad doborem artykułów czuwała Rada Programowa konferencji w składzie:

Prof. dr hab. Jerzy Kisielnicki – Przewodniczący
Mgr inż. Jerzy S. Nowak – Sekretarz Rady
Prof. dr hab. Witold Chmielarz
Dr inż. Juliusz Czarnowski
Dr Jarosław Deminet
Prof. dr hab. Dariusz Dziuba
Mgr inż. Piotr W. Fuglewicz
Prof. dr hab. Jan Goliński
Prof. dr hab. Jerzy Gołuchowski
Mgr Michał Górski
Dr inż. Janusz K. Grabara
Prof. dr hab. Marek Greniewski
Dr inż. Wacław Iszkowski
Prof. dr hab. Piotr Jędrzejowicz
Prof. dr hab. Mirosława Lasek
Prof. dr hab. Andrzej Marciniak
Prof. dr hab. Zygmunt Mazur
Dr inż. Marek Miłosz
Prof. dr hab. Mieczysław Muraszkiewicz
Dr inż. Krzysztof Nałęcki

Prof. dr hab. Wojciech Olejniczak
Prof. dr hab. Józef Oleński
Dr Bogdan Pilawski
Prof. dr hab. Elżbieta Skrzypek
Dr Jerzy T. Skrzypek
Dr inż. Jacek Stochlak
Prof. dr hab. Janusz Szopa
Prof. dr hab. Zdzisław Szyjewski
Prof. dr hab. Ryszard Tadeusiewicz
Dr inż. Marek Valenta

Powstałe w ten sposób opracowanie zawiera artykuły poświęcone problematyce i wzajemnym relacjom informatyki i globalizacji gospodarki, szacowania efektywności zastosowań systemów informatycznych (tu podkreślamy kwestie szacowania zmian wartości pieniądza w czasie w odniesieniu do inwestycji), strategii informatyzacji przedsiębiorstw, metodyce wdrożeń, kierowania projektem informatycznym. Zwrócono uwagę na problematykę outsourcingu, TCO (Total Cost of Ownership), zastosowania norm ISO 9000 w informatyce, strategicznej karty wyników (Balanced ScoreCard) itp. Szereg autorów przedstawiło oryginalne relacje z wdrożeń systemów informatycznych oraz możliwości i warunki adaptacji systemów klasy MRPII/ERP do wymagań przedsiębiorstwa. O systemach MRPII/ERP wypowiadają się zarówno przedstawiciele nauki jak i praktycy, przy czym redaktorzy z ubolewaniem stwierdzają fakt obawy wielu informatyków przed publikacją doświadczeń z tego zakresu, a dotyczących ich firm. Pozytywnie należy odnotować szereg artykułów poświęconych nowemu zjawisku na polskim rynku informatycznym, a mianowicie rozwiązaniom *business intelligence* i zarządzaniu wiedzą, w tym e-kształceniu (*e-learning*). Dział ten zyskuje coraz większa liczbę autorów, nie tylko ze środowiska naukowego. W br. postanowiono zwrócić uwagę na wpływ rozwiązań zabezpieczeń w systemach informatycznych na kwestie efektywności zastosowań – po raz pierwszy zamieszczamy opracowania na temat podpisu elektronicznego przygotowane przez ośrodki UNIZETO i Sigillum z PWPW.

Szereg artykułów ma charakter przeglądowy i przez to czytelnik uzyskuje możliwość dostępu do aktualnych ocen tych zjawisk w krajowej literaturze informatycznej.

Zdaniem redaktorów tomu należy zwrócić szczególną uwagę na artykuł pod przewrotnym tytułem „Grafika w informatyce", zawierający wybór rysunków absolwenta Politechniki Krakowskiej (o czym mało kto wie),

Pana Andrzeja Mleczki, który precyzyjnie pokazuje zadziwiające zjawiska w nauce i informatyce.

Obfitość nadesłanych artykułów spowodowała, że redaktorzy postanowili przygotować trzy pozycje książkowe – pierwsza – „Informatyka i gospodarka globalna – problemy i metody", druga pod tytułem „Efektywność zastosowań systemów informatycznych – 2003" i trzecia nosząca tytuł „Systemy informatyczne – zastosowania i wdrożenia 2003". Wszystkie opracowania należy traktować łącznie. Redaktorzy wyrażają przekonanie, że opracowania wydane w latach 2001 - 2003 r. sumptem Polskiego Towarzystwa Informatycznego stanowią jeden z najbardziej obszernych przeglądów problematyki zastosowań systemów informatycznych w kraju i zapraszają czytelników do nadsyłania propozycji artykułów do wydania w 2004 r.

Przygotowanie wydawnictwa wymagało znacznego wysiłku organizacyjnego i dlatego redaktorzy opracowania składają podziękowania za pomoc w pracy Paniom Halinie Czarnowskiej i Annie Gembalczyk z Oddziału Górnośląskiego Polskiego Towarzystwa Informatycznego, a także Wydawnictwom Naukowo-Technicznym za znakomitą współpracę.

Redaktorzy

Spis treści

STR.

ROZDZIAŁ 4

ZASTOSOWANIA I WDROŻENIA – OPISY PRZYPADKÓW

WYBRANE ASPEKTY ZARZĄDZANIA INFORMACJĄ NA PRZYKŁADZIE ZINTEGROWANEGO SYSTEMU KOMUNIKACJI ELEKTRONICZNEJ W DUŻEJ ORGANIZACJI

Elżbieta GOLA

Streszczenie: żadna firma nie może sprawnie funkcjonować bez dobrego systemu komunikacji wewnętrznej, który często nazywany jest „systemem nerwowym" przedsiębiorstwa. Dobry i efektywny system gromadzenia i udostępniania dokumentów oraz sprawny przepływ informacji pomiędzy poszczególnymi komórkami firmy ma szczególne znaczenie w okresie transformacji strukturalnych firmy w zmieniającej się sytuacji rynkowej. Artykuł opisuje rozwiązanie jakie wdrożono w ING Banku Śląskim S.A., w organizacji zatrudniającej ok. 7500 pracowników, złożonej z Centrali i blisko 350 oddziałów w całej Polsce. Przedstawiono zintegrowany System Komunikacji Elektronicznej zrealizowany przy wykorzystaniu technologii IBM-Lotus Domino Enterprise Server.

1. PROJEKT KOMUNIKACJI WEWNĘTRZNEJ

Inicjatywa wdrożenia w ING Banku Śląskim S.A. nowoczesnego systemu komunikacji elektronicznej wyszła od Zarządu Banku na początku roku 2000. Powołano Projekt Komunikacji Wewnętrznej, którego celem była poprawa efektywności działania banku poprzez wdrożenie sprawnego systemu komunikacji. System miał stanowić dopełnienie procesu reorganizacji i rozwoju sieci oddziałów banku, w wyniku którego niektóre procesy związane z obsługą klientów zostały przeniesione do tzw. „back office", a niektóre zcentralizowane. Położono duży nacisk na poprawę jakości obsługi klientów, podniesienie wiedzy merytorycznej sprzedawców i doradców klienta, uproszczenie procedur i skrócenie czasu obsługi. Aby to osiągnąć, należało wyposażyć tzw. „front office" w dobre źródło informacji o produktach bankowych, procedurach z nimi związanych, opłatach i prowizjach, zaletach i wadach poszczególnych produktów w zestawieniu z konkurencją. Nie mniej istotny był bieżący dostęp do informacji często zmieniających się, jak np. aktualne kursy walut. Potrzebny był system, który zapewniłby:

- mozliwość stworzenia repozytorium dokumentów bankowych;
- łatwy dostęp do aktualnych informacji z każdej jednostki organizacyjnej banku;
- efektywny i bezpieczny system komunikacji pomiędzy Centralą, a siecią oddziałów;
- ustaloną odpowiedzialność za informacje;
- wdrożone procedury komunikacji.

Aby spełnić narzucone wymagania nalezało zbudować system komunikacji elektronicznej łaczący trzy elementy:

- **intranet**, jako miejsce publikowania bieżących wiadomości oraz zbiór serwisów, ogólnie dostępnych dla pracowników w bankowej sieci teleinformatycznej,
- **poczta elektroniczna**, z wbudowanym podpisem elektronicznym i możliwością szyfrowania przesyłek, powiązana z kalendarzem i planem zadań każdego użytkownika;
- **aplikacje**, przez które są rozumiane specjalnie przygotowane programy i bazy danych dla ściśle określonych zastosowań i ściśle określonych grup odbiorców.

2. WYBÓR ROZWIĄZANIA INFORMATYCZNEGO

Przed kierownictwem projektu informatycznego postawiono zadanie wdrożenia systemu komunikacji elektronicznej, który:

- byłby dostępny we wszystkich jednostkach organizacyjnych ING Banku Śląskiego tzn. we wszystkich 350 oddziałach w całej Polsce oraz dla wszystkich pracowników w departamentach centrali;
- zapewniałby bezpieczną pocztę elektroniczną dla minimum 3500 do 7500 użytkowników,
- dawał możliwość budowy intranetu, baz dokumentów, aplikacji;
- pozwalałby na utrzymanie jednolitej wersji informacji w całej sieci;
- byłby niezawodny i dostępny w trybie 24 godziny / 7 dni w tygodniu;
- ograniczał koszty i zatrudnienie.

W momencie rozpoczęcia zadania ING Bank Śląski posiadał własną sieć rozległą WAN. Brano pod uwagę dwa warianty architektury systemu:

I. Sieć serwerów w oddziałach regionalnych i w centrali:

wady:

- konieczny nadzór nad dostarczaniem poczty,
- ryzyko różnych kopii baz i serwisów w różnych regionach – konieczny nadzór nad procesem replikacji,
- administratorzy regionalnych serwerów (etaty, szkolenia, backup),

zaleta: część ruchu w sieci WAN pozostaje w regionach lub w obrębie centrali.

N I E

I. Jeden centralny serwer Lotus Domino na IBM AS/400 iSeries.

zalety:

- natychmiastowe dostarczanie poczty,
- jedna kopia serwisów informacyjnych i danych,
- centralna administracja systemem,
- centralna archiwizacja, system backupowy i obsługa,

wada: zwiększenie obciążenia sieci WAN.

T A K

Jako docelowy wybrany został wariant II, Serwer Lotus Notes / Domino pracujący na maszynie IBM AS/400 iSeries. System ten dawał możliwość centralnej obsługi kilku tysięcy użytkowników poczty oraz spełniał wszystkie pozostałe wymagania: posiadał wbudowany system podpisu elektronicznego i szyfrowania, pozwalał na budowę intranetu oraz własnych aplikacji. Upraszczał także zarządzanie profilami użytkowników, jedno konto w systemie mogło być wykorzystane w poczcie elektronicznej oraz w listach praw dostępu do aplikacji w środowisku Lotus Domino. Dzięki temu obsługę administracyjną systemu w skali całego banku można było powierzyć niewielkiemu 6-osobowemu zespołowi administratorów.

Wdrożenie postanowiono przeprowadzić siłami pracowników banku, dlatego też rozpoczęto od przeszkolenia grupy administratorów, designerów i specjalistów od eksploatacji systemów, a następnie równolegle prowadzono prace, mające na celu:

- wymianę systemu pocztowego,
- budowę intranetu i aplikacji,
- zmiany w infrastrukturze IT (upgrade łączy, implementacja protokołu IP, priorytetyzacja ruchu w sieci WAN, upgrade stacji PC użytkowników).

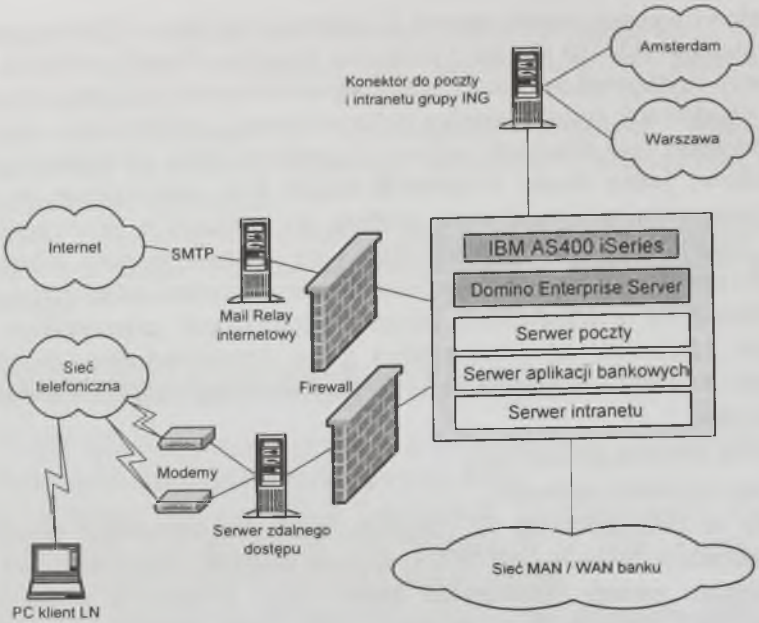
Kolejne zmiany strukturalne banku, m.in. połączenie z warszawskim oddziałem ING Barings S.A. w roku 2001, oraz zacieśnianie współpracy wszystkich jednostek Grupy ING w Polsce i na świecie, spowodowały konieczność rozbudowy architektury systemu Lotus Domino z wersji Application Server do wersji Enterprise Server. Wersja ta daje m.in. możliwość tworzenia wielu partycji serwera Domino w ramach jednej maszyny AS/400.

Obecna architektura systemu Lotus Domino w ING Banku Śląskim przedstawiona została na rys.1 . Trzy partycje Domino obsługują:

1. pocztę elektroniczną,
2. intranet i aplikacje bankowe,
3. witrynę intranetową „Poland Gateway" dostępną dla innych jednostek Grupy ING.

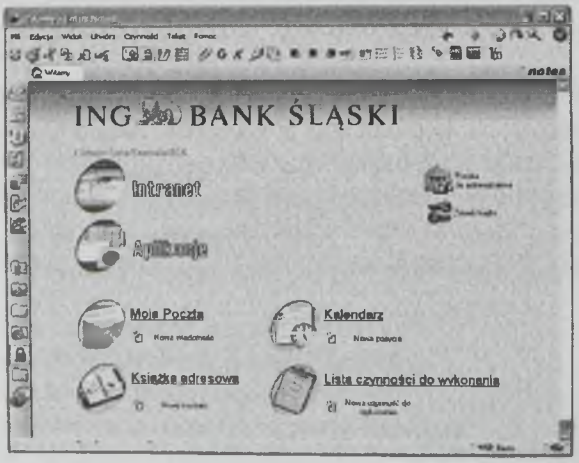
Łączność z innymi jednostkami Grupy ING realizowana jest poprzez odpowiednie konektory, a komunikacja z klientami banku jest możliwa poprzez MailRelay łączący bankowy system poczty z internetem.

Każdy użytkownik systemu może korzystać ze swojego konta z dowolnego miejsca w ramach bankowej sieci MAN / WAN. Możliwy jest także zdalny dostęp z klienta Lotus Notes do sieci bankowej poprzez łącza telefoniczne.



Rys.1 Architektura systemu

Jako standard oprogramowania użytkownika wykorzystano w ING Banku Śląskim oprogramowanie klienta Lotus Notes, z uwagi na podpis elektroniczny, nie obsługiwany przez klienta przeglądarkowego. Dodatkowo wprowadzono standardowy bankowy desktop użytkownika Lotusa – ekran powitalny, który zawiera łącza do najważniejszych elementów systemu komunikacji elektronicznej.

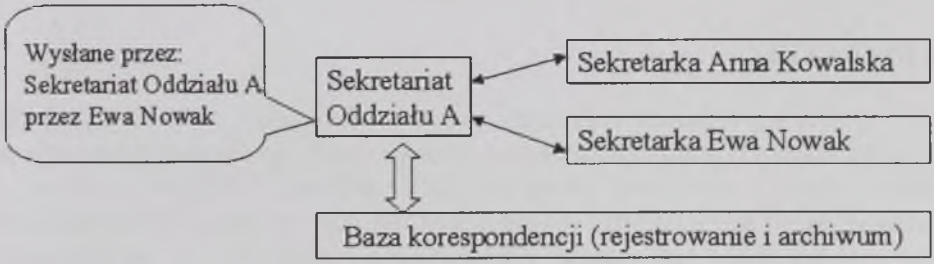


Rys.2 Ekran powitalny systemu komunikacji elektronicznej.

3. ZINTEGROWANY SYSTEM POCZTOWY

System pocztowy w dużej organizacji to niemałe wyzwanie dla jego projektantów i administratorów. Samo założenie kilku tysięcy kont pracowników nie wystarczy, aby połapać się w skomplikowanej strukturze firmy. Nikt nie jest w stanie zapamiętać nazwisk dyrektorów kilkudziesięciu departamentów centrali i dyrektorów kilkuset oddziałów, nie mówiąc już o nazwiskach sekretarek, które najczęściej obsługują pocztę jednostek organizacyjnych. Z drugiej strony wymagania bezpieczeństwa nie dopuszczają bezimiennych nadawców typu *„Sekretariat Oddziału w Sosnowcu"*. Nadawca każdej przesyłki musi być w wiarygodny sposób imiennie identyfikowany w systemie.

W ING Banku Śląskim rozwiązano ten problem poprzez odwzorowanie w pocztowej książce adresowej struktury organizacyjnej banku. Każda jednostka organizacyjna banku, departament centrali i oddział otrzymał tzw. skrzynkę pocztową grupową sekretariatu, o nazwie zgodnej z nazewnictwem bankowym. Dopiero do skrzynek grupowych delegowane są uprawnienia dostępu dla sekretarek. Dzięki temu łatwo zaadresować pocztę do konkretnej jednostki organizacyjnej, natomiast poczta wysyłana ze skrzynki grupowej opatrzona jest podwójnym opisem. Na przykład, jeżeli sekretarka Ewa Nowak wyśle pocztę ze skrzynki Sekretariatu Oddziału A, pojawi się nagłówek nadawcy:
„Wysłane przez Sekretariat Oddziału A, przez Ewa Nowak".



Rysunek 3. Powiązanie pomiędzy skrzynkami pocztowymi jednostki organizacyjnej i obsługujących ją sekretarek.

Dla ułatwienia rozsyłania poczty do wielu jednostek równocześnie utworzono szereg grup adresowych tzw. list dystrybucyjnych, np. *„LISTA Sekretariatów Departamentów Centrali"*, *„LISTA Dyrektorów Pionów"*, itp.. Istnieją także grupy adresowe jednoosobowe np. *„D-DyrOddz Bukowno-Kolejowa-2a"*, do których przypisuje się skrzynkę imienną aktualnego dyrektora. Dzięki temu struktura adresów jednostek organizacyjnych jest niezależna od zmian personalnych na poszczególnych stanowiskach. Składem list dystrybucyjnych i uprawnieniami do obsługi skrzynek grupowych zarządzają Administratorzy

Systemu Komunikacji na podstawie danych kadrowych.
Według stanu na początek maja 2003r na listach adresowych poczty Lotus Notes ING Banku Śląskiego znajdowało się około:

- 4655 adresów indywidualnych,
- 475 skrzynek grupowych,
- 1416 grup adresowych, w tym list dystrybucyjnych.

System pocztowy ING Banku Śląskiego udostępnia użytkownikom kilka usług:

- pocztę wewnętrzną w standardzie Lotus Notes,
- pocztę jednostek Grupy ING na całym świecie, poprzez odpowiednie konektory.
- pocztę internetową, która umożliwia komunikację z klientami oraz firmami współpracującymi z bankiem.

W poczcie wewnętrznej standardowo stosowany jest podpis elektroniczny stanowiący integralny element systemu Lotus Notes. Procedury rejestrowania użytkowników i przekazywania im dyskietek z plikiem ID, zawierającym m.in. klucze elektroniczne, są podobne do stosowanych przez niektóre firmy certyfikacyjne i zostały wdrożone do użytku wewnętrznego w banku już w roku 2000, czyli dwa lata przed Ustawą o podpisie elektronicznym.

Zmiana trybu wewnętrznej korespondencji służbowej z papierowej na elektroniczną została w ING Banku Śląskim wprowadzona oficjalnie odpowiednią instrukcją służbową zatwierdzoną przez Zarządu Banku.

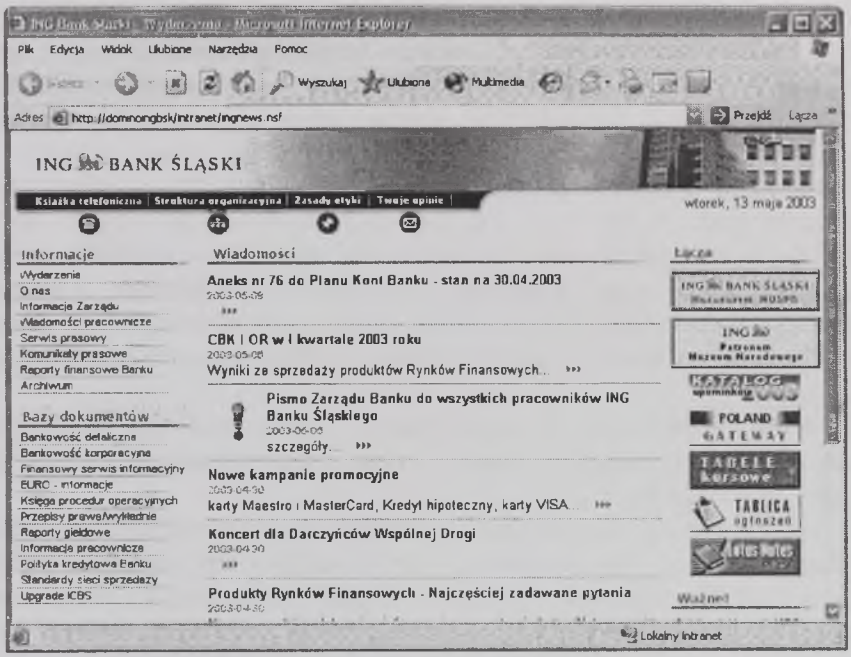
4. INTRANET

Budowę intranetu rozpoczęto od powołania grup problemowych, które w poszczególnych obszarach działalności banku miały zanalizować informacje pod kątem potrzeb i oczekiwań odbiorców tych informacji. Przyjęto założenie, że nadmiar informacji może być równie szkodliwy, jak ich brak. Dlatego też grupy problemowe miały za zadanie zdefiniowanie: jakie dane z poszczególnych obszarów są i powinny być udostępniane, jakim grupom odbiorców i w jakim zakresie. Następnie informacje były segregowane według ich rodzaju i ważności, odpowiednio kategoryzowane i hierarchizowane. W wyniku powstał projekt baz dokumentów i serwisów. W pierwszej fazie zdefiniowano następujące serwisy:

- informacje o banku,
- bankowość detaliczna,
- bankowość korporacyjna,
- polityka kredytowa banku,
- finansowy serwis informacyjny,
- standardy sieci sprzedaży,
- przepisy prawa / wykładnie,
- informacje pracownicze,

- serwis prasowy,
- raporty giełdowe,
- badania i analizy rynku,
- kampanie promocyjne.

Poddano także analizie powiązania pomiędzy poszczególnymi serwisami, np. kampanie promocyjne powiązano z produktami bankowymi - detalicznymi lub korporacyjnymi. Tam, gdzie było to możliwe, przyjęto format dokumentów zgodny z istniejącymi standardami bankowymi np. karta produktu, wzór procedury itp.. Ostateczny projekt stron intranetowych opracowano we współpracy ze specjalistami od marketingu, kładąc jednak nacisk na zawartość informacyjną, szybkość działania baz i serwisów, a standardy graficzne banku i Grupy ING ograniczając do minimum. Tak przygotowany materiał stał się podstawą dla określenia specyfikacji funkcjonalnej dla systemu informatycznego.



Rys.4 Bankowy portal intranetowy.

Po pierwszej fazie wdrożenia nastąpił dalszy rozwój intranetu, który trwa do nadal. Stale powstają nowe serwisy, a stare są modyfikowane i ulepszane. Im bardziej użytkownicy przekonują się o wartości użytkowej intranetu, tym więcej napływa pomysłów i zamówień na nowe elementy.

5. APLIKACJE LOTUS NOTES / DOMINO

System Lotus Notes / Domino daje możliwość stosunkowo prostego

tworzenia aplikacji wykorzystujących rozbudowany system praw dostępu. Zintegrowane środowisko pocztowe i aplikacyjne, a także wbudowany system podpisu elektronicznego i szyfrowania danych stwarzają dodatkowe możliwości, które można wykorzystać dla wielu zastosowań.
Poniżej przedstawiono kilka przykładów, z kilkudziesięciu wdrożonych w ING Banku Śląskim aplikacji o charakterze komunikacyjnym, działających na serwerze Domino.

1. „Transakcje odrzucone" – aplikacja łącząca zalety poczty i baz danych.

Aplikacja ta współpracuje z systemem obsługi masowych płatności banku (Mass Payment System), gdzie skanowane jest codziennie kilkadziesiąt tysięcy papierowych przelewów złożonych w oddziałach przez klientów banku. System MPS generuje raporty o transakcjach odrzuconych (np.brak podpisu, błędna data, nieczytelna kwota itp.). Ponieważ często są to pilne zlecenia na niemałe kwoty, należy jak najszybciej powiadomić klienta o fakcie, że jego zlecenie nie zostanie zrealizowane. W systemie Lotus Notes przygotowano aplikację, która w określonych odstępach czasu pobiera dane o tansakcjach odrzuconych z systemu MPS, dzieli je na porcje przeznaczone dla odpowiednich oddziałów i automatycznie rozsyła do oddziałów według specjalnie przygotowanej książka adresowej.

2. „Prezentacja środków trwałych" – użycie Lotus Notes jako warstwy prezentacyjnej.

Aplikacja, która łączy się z systemem ewidencji środków trwałych funkcjonującym na innej maszynie i w innych standardach, i prezentuje użytkownikowi dane o środkach trwałych, za które jest odpowiedzialny. Dyrektor jednostki ma prawo obejrzeć listę wszystkich środków trwałych przypisanych do jego podwładnych.

3. Różnego rodzaju bazy do raportowania z sieci sprzedaży do centrali – wykorzystanie Lotusa jako znanego użytkownikom interfejsu dostępnego we wszystkich oddziałach.

Jest to funkcja szczególnie przydatna w czasie krótkotrwałych akcji np. sprzedaż specjalnego funduszu inwestycyjnego, która trwa tylko trzy tygodnie. Wtedy nie opłaca się wdrażać skomplikowanych rozwiązań, a dystrybucja oprogramowania do kilkuset jednostek polega tylko na umieszczeniu przez administratorów ikony z linkiem do odpowiedniej bazy na ekranie lotusa.

6. ZARZĄDZANIE INFORMACJĄ ELEKTRONICZNĄ

Niezmiernie ważnym aspektem systemu komunikacji elektronicznej jest ustalenie odpowiedzialności za wiarygodność i aktualność publikowanych w systemie informacji. Na początku wdrożenia systemu trzeba też było przekonać użytkowników, że dokument elektroniczny publikowany w systemie jest ważny na równi z dokumentem papierowym i po jego opublikowaniu, należy go uznać za

obowiązujący. W tym celu wprowadzono do stosowania w banku instrukcję służbową „Organizacja i funkcjonowanie systemu komunikacji elektronicznej", zatwierdzoną przez Zarząd Banku.
Instrukcja wprowadziła definicję „Właściciela" Systemu Komunikacji Elektronicznej jako całości oraz pojęcie „Właściciela aplikacji lub serwisu informacyjnego". Przekładając to na praktykę, Właścicielem Systemu Komunikacji Elektronicznej Banku jest Dyrektor Biura Zarządu Banku, który poprzez wyznaczonych przez siebie Administratorów zarządza bankowym systemem jako całością.

Każdy serwis intranetowy i każda baza lub aplikacja, a nawet pojedynczy dokument w systemie ma swojego właściciela, którym najczęściej jest dyrektor odpowiedniego pionu lub departamentu biznesowego. Na przykład, właścicielem serwisu „Bankowość detaliczna" jest Dyrektor Pionu Detalicznego, a właścicielem „Serwisu prasowego" jest Naczelnik Wydziału Public Relations Biura Zarządu Banku. Właściciele serwisów oczywiście nie muszą osobiście wprowadzać informacji do intranetu, wyznaczają do tego celu specjalistów biznesowych tzw. opiekunów serwisów, czyli osoby bezpośrednio odpowiedzialne za publikowanie i aktualizację dokumentów.

Przyjęto zasadę, że informacje wprowadzane są do systemu w miejscach ich powstawania, a więc bezpośrednio w departamentach odpowiedzialnych za określone obszary działalności banku. Dlatego też przygotowano w systemie narzędzia i formularze do wprowadzania i edycji danych, które nie wymagają od pracowników kwalifikacji informatycznych. Redaktorem intranetu czy bazy dokumentów może zostać każdy użytkownik, posiadający podstawowe umiejętności w zakresie obsługi komputera PC. Mechanizmy nadawania praw dostępu w systemie komunikacji elektronicznej i odpowiednie procedury w tym zakresie gwarantują, że można zidentyfikować autora każdego dokumentu w systemie oraz historię modyfikacji każdego dokumentu.

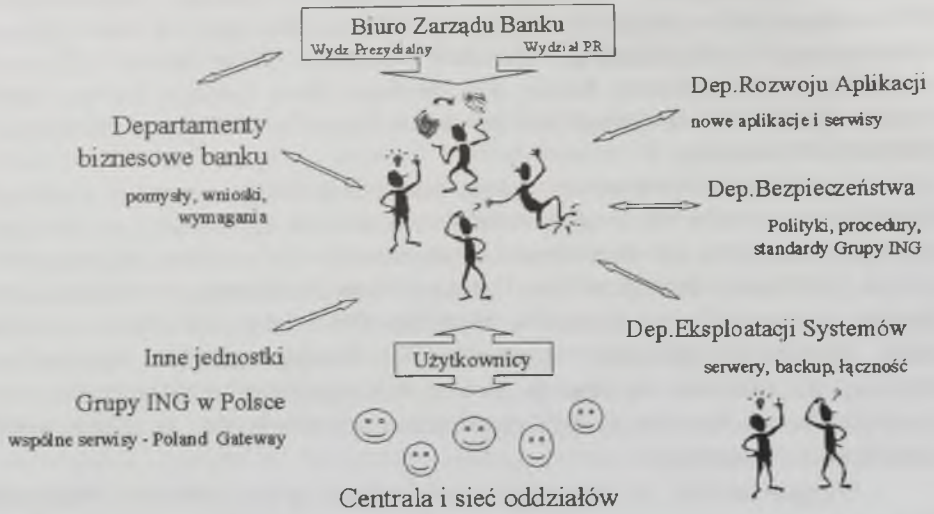
Po blisko trzech latach funkcjonowania systemu komunikacji elektronicznej w ING Banku Śląskim S.A. można stwierdzić, że ten sposób zarządzania informacją sprawdził się znakomicie. Użytkownicy mają pełne zaufanie, że w intranecie i w bazach znajdą zawsze aktualne dokumenty, które mogą uznać za obowiązujące. W wielu obszarach zrezygnowano całkowicie z dokumentów papierowych.

7. ZARZĄDZANIE SYSTEMEM KOMUNIKACJI ELEKTRONICZNEJ

Administratorzy systemu komunikacji elektronicznej pracują na styku zarzadzania informacją i informatyki. Starają się jak najlepiej spełniać potrzeby użytkowników, przekładając ich wymagania na wymagania techniczne. Tak więc z jednej strony współpracują z departamentami biznesowymi banku, a z drugiej z departamentami pionu informatycznego, który odpowiada za infrastrukturę techniczną, bezpieczeństwo i rozwój systemu.

Administratorzy Systemu Komunikacji Elektronicznej

- na styku zarządzania informacją i informatyki



Rys.5 Zarządzanie systemem komunikacji elektronicznej w banku.

Administratorzy zarządzają systemem jako całością, ekranem powitalnym użytkownika, pocztą i portalem intranetowym. Centralna architektura systemu sprawia, że wszelkie nowe aplikacje i zmiany w konfiguracji są przygotowywane przez administratorów, a użytkownicy muszą, co najwyżej odświeżyć ekran powitalny lotusa, aby ściągnąć aktualną wersję.

Blisko trzyletnia eksploatacja systemu komunikacji elektronicznej w ING Banku Śląskim wykazała jego ogromną rolę w organizacji. Dziś nikt nie wyobraża sobie funkcjonowania banku bez poczty elektronicznej, intranetu i wielu aplikacji wdrożonych w środowisku Domino. Nawet kilkuminutowa przerwa w działaniu systemu powoduje dezorganizację pracy i lawinę telefonów do administratorów. Na szczęście wybór rozwiązania informatycznego okazał się trafny i niezawodność systemu sięga prawie 100%. Na pewno jest to również zasługa wysoko kwalifikowanej kadry informatycznej banku.

Wysoka krytyczność aplikacji oraz jej stały rozwój powoduje, że duży nacisk położono w banku na odpowiedni system backupowy i plany zachowania ciągłości pracy systemu. Poza środowiskiem produkcyjnym utrzymywany jest także osobny system rozwojowy i akceptacyjny. Bank wdraża rozwiązania informatyczne zgodne pod względem bezpieczeństwa z normą ISO/IEC 17799 oraz politykami Grupy ING.

Udane wdrożenie systemu komunikacji elektronicznej w dużej, wielooddziałowej organizacji to wspólna praca wszystkich jednostek organizacyjnych, specjalistów biznesowych

i informatyków oraz grupy zarządzającej projektem, a następnie samym systemem. W przypadku ING Banku Śląskiego S.A. z powodzeniem udało się połączyć wszystkie te elementy, a efekcie uzyskano znaczną poprawę organizacji i standardów pracy, jak przystało na nowoczesną firmę w XXI wieku.

Elżbieta GOLA
Biuro Zarządu Banku,
ING Bank Śląski S.A., 40-086 Katowice,
ul. Sokolska 34 Tel.:(0-32)357-75-08
e-mail: elzbieta.gola@ingbank.pl

WSPOMAGANIE METODY *SIX SIGMA* ZA POMOCĄ OPROGRAMOWANIA *IGRAFX PROCESS 2003 FOR SIX SIGMA*

Mirosława LASEK, Marek PĘCZKOWSKI

Streszczenie: *iGrafx Process for Six Sigma* wspomaga realizację metodologii doskonalenia jakości procesów gospodarczych *Six Sigma* dzięki połączeniu wykorzystania możliwości modelowania i analizy procesów biznesowych programu *iGrafx* z możliwościami analiz statystycznych pakietu statystycznego *Minitab*. Możliwe jest zarówno przeprowadzanie analiz statystycznych danych empirycznych pochodzących z pomiarów rzeczywistych realizacji procesów biznesowych, jak i statystycznej analizy wyników symulacji wielu przebiegów procesów na danych generowanych przez użytkownika. W artykule przedstawiono podstawowe możliwości wspomagania analiz *Six Sigma* przez program *iGrafx Process 2003 for Six Sigma*. Zilustrowano je przykładami zastosowań.

1. Metoda *Six Sigma* doskonalenia jakości procesów gospodarczych

Najogólniej, metoda *Six Sigma* polega na przeprowadzaniu analizy procesów za pomocą procedur statystycznych, a następnie ich optymalizacji w oparciu o wyniki tej analizy. Pozwala połączyć metody graficznej wizualizacji, analizy i projektowania procesów gospodarczych i systemów informatycznych z rozbudowaną analizą statystyczną. Dokładniej, *Six Sigma* wykorzystuje metody gromadzenia danych i narzędzia analizy statystycznej, aby wykryć przyczyny powstawania wad procesu (braków) i znaleźć sposoby ich wyeliminowania [M. Harry, R. Schroeder 2001, s. 36]. Celem jest poprawa jakości, która przyniesie korzyści klientom i firmie. Zakłada się poprawę jakości początkowo poprzez eliminację błędów, a w dalszej kolejności poprzez doskonalenie przebiegu procesów gospodarczych, co ma prowadzić do zmniejszenia kosztów operacyjnych, podniesienia poziomu wydajności, podwyższenia marży zysku, skrócenia cyklu produkcyjnego, ograniczenia zapasów, sprawniejszego przebiegu transakcji handlowych [DOE. Warsztaty ... 2002; T.N. Goh 2002; M. Harry, R. Schroeder 2001; B.P. Lientz, K.P. Rea 2002; "Statystyka dla ..." 2002]. Występujące w nazwie metody określenie *Sigma* jest odwołaniem do odchylenia standardowego w rozkładzie normalnym, oznaczanego zwyczajowo grecką literą sigma (σ). *Six Sigma* – sześć sigma, to 3,4 wadliwych produktów na 1 milion wykonanych i pożądany poziom jakości do osiągnięcia przez firmę. Wiele zjawisk w przyrodzie i w procesach produkcji wykazuje zgodność z rozkładem normalnym. Wynika to z faktu, że jeżeli na wynik wpływa bardzo dużo różnokierunkowych i niezależnych czynników, to ich łączny efekt można wyrazić za pomocą krzywej Gaussa. Można to udowodnić matematycznie – mówi o tym centralne twierdzenie graniczne w rachunku prawdopodobieństwa. W związku z powyższym można przyjąć, że ustalony parametr procesu produkcji (np. średnica produkowanej

śrubki, długość automatycznie ciętego drutu, waga torebki z towarem) jest zmienną losową o rozkładzie normalnym. Oznacza to, że jeżeli proces technologiczny przewiduje produkowanie np. śrubek o średnicy 10 mm, to kolejne egzemplarze nie będą miały dokładnie 10 mm każdy, ale ich średnice będą odchylać się od tej liczby w górę albo w dół. Mówimy, że proces jest prawidłowo wyregulowany, jeżeli wartość oczekiwana procesu wynosi 10 mm (wtedy wartość średnia dużej liczby wyprodukowanych śrubek wyniesie z dużą dokładnością 10 mm). Średnica poszczególnych śrubek (odchyłka od normy wynoszącej 10 mm) będzie zależeć od odchylenia standardowego procesu (σ). Rozkład normalny ma tę własność, że 99,73% wartości odchylających się od wartości oczekiwanej mieści się w przedziale (-3σ; $+3\sigma$). Jeżeli ustalimy normę tak, że odchylenia mieszczące się w tym przedziale uznamy za zgodne z normą, a elementy o parametrze wykraczającym poza ten przedział uznamy za braki, to liczba braków będzie wynosić ok. 2700 na milion sztuk (2700 ppm – *parts per million*). Mówimy wtedy o jakości na poziomie 3 sigma. Możemy postąpić odwrotnie, tzn. ustalić jakąś normę (przedział, w którym powinien mieścić się parametr) i zadbać o to, żeby proces miał jakość na poziomie 3 sigma. Jeżeli nie uzyskujemy takiej jakości, to należy poprawić technologię, organizację produkcji, czy też inne czynniki. Liczba 2700 błędów na 1 milion, to jest jednak dużo błędów. Oznacza to na przykład, że 2700 przesyłek na 1 milion nadawanych na poczcie nie dochodzi do adresata. Aby zmniejszyć liczbę błędów stworzono metodologię zakładającą jakość na poziomie 6 sigma [DOE. Warsztaty... 2002; T.N. Goh 2002; M. Harry, R. Schroeder 2001; "Statystyka dla ..." 2002]. Jest to wysokie wymaganie – 6 sigma, to jest 2 razy więcej niż 3 sigma. Zgodnie z własnościami rozkładu normalnego mielibyśmy wtedy 0,002 braków na milion egzemplarzy, (czyli ponad milion razy mniej niż w jakości 3 sigma). Jest to liczba niewiarygodnie mała, np. produkując 1 sztukę na sekundę, przeciętnie trzeba by czekać na brak 15 lat. Osiągnięcie tak dobrej jakości produkcji jest w praktyce niemożliwe. W metodologii tradycyjnej (3 sigma) uczyniono założenie, że proces technologiczny jest doskonale stabilny, tzn. wartość oczekiwana parametru nie zmienia się w czasie. W rzeczywistości należy przyjąć, że jest to nierealne, i że nie tylko produkowane egzemplarze odchylają się od wzorca, ale również parametr rozkładu normalnego (wartość oczekiwana) odchyla się od wzorca. Biorąc pod uwagę doświadczenia wynikające z praktyki w produkcji (np. przemyśle elektronicznym) przyjęto, że przesunięcia wartości oczekiwanej mieszczą się w przedziale $\pm 1{,}5\sigma$. W metodologii *Six Sigma* uwzględnia się fluktuacje wartości oczekiwanej i jeżeli będziemy mieścić się w normie 6 σ (6 odchyleń standardowych) każdego z rozkładów normalnych, to uwzględniając fluktuację wartości oczekiwanej o $\pm 1{,}5\sigma$, będziemy mieścić się w przedziale $\pm 4{,}5\sigma$ dla wzorcowego rozkładu normalnego. To wskazuje na możliwość utrzymywania się w normie przy założeniu $\pm 4{,}5\sigma$, czyli 3,4 braków na milion egzemplarzy [DOE. Warsztaty... 2002; B.P. Lientz, K.P. Rea 2002; „Statystyka dla ..." 2002].

2. Oprogramowanie *iGrafx Process 2003 for Six Sigma*

Program *iGrafx Process for Six Sigma* umożliwia [„iGrafx Process ..." 2000; „Introducing ..." 2001; M. Lasek, M. Pęczkowski, B. Otmianowski 2002]: (1) znalezienie rozkładu statystycznego dopasowanego do danych empirycznych (wartości zmiennych), charakteryzujących przebieg procesu, np. czasu obsługi zlecenia klienta, czasu dostawy produktu do klienta, kosztów transportu; taki rozkład zmiennej jest traktowany jako wyrażenie (ang. *expression*), wykorzystywane w późniejszym przeprowadzaniu symulacji przebiegu procesów (opcja Fit Data programu); (2) automatyczne przeprowadzanie eksperymentów[1], przy wskazanych przez nas założeniach, dotyczących czynników decydujących o przebiegu procesu (ang. *factors*), np. sposobów realizacji zamówień klientów, stosowanych środków transportu, sposobu świadczenia usług: ekspresowego lub standardowego. W zależności od przyjętych założeń obliczane są mierniki wydajności lub jakości realizacji procesów (tzw. *responses*). Miernikami takimi mogą być przykładowo koszty i czas realizacji czynności, czy całych „ścieżek" przebiegu procesów, lub też stopień wykorzystania zasobów (ludzi i urządzeń). Mogą też być zdefiniowane i zastosowane mierniki bardziej specyficzne dla metodologii *Six Sigma*, takie jak (i) wskaźnik wydajności przejściowej *TY* (*Throughput Yield*), określający prawdopodobieństwo, że dany produkt lub usługa przejdzie przez dany etap procesu nie wykazując żadnych wad, (ii) wskaźnik wydajności mierzonej w toku *RTY* (*Rolled Throughput Yield*), przedstawiający prawdopodobieństwo, że produkt lub usługa przejdzie przez cały proces nie wykazując wad, (iii) znormalizowany wskaźnik wydajności *NY* (*Normalized Yield*), będący przeciętnym wskaźnikiem wydajności przejściowej, obliczany przy uwzględnieniu wszystkich etapów procesu, tj. średnia wydajność przejściowa przypadająca na etap procesu. Po realizacji wskazanej przez użytkownika liczby eksperymentów, wyniki mogą być poddane szczegółowej analizie statystycznej za pomocą procedur pakietu *Minitab* (opcja RapiDOE programu), gdzie *DOE* jest skrótem utworzonym z pierwszych liter słów *Design of Experiments*); (3) przekazanie („wyeksportowanie") wyników symulacji dla poszczególnych transakcji lub wyników eksperymentów do *Minitab* lub arkusza kalkulacyjnego *Excel* w celu przeprowadzenia interesujących nas w konkretnym przypadku specjalistycznych analiz (opcja Log Transactions programu); (4) przesłanie („wyeksportowanie") raportów uzyskanych za pomocą programu *iGrafx* do zewnętrznych aplikacji dla przeprowadzania dalszych analiz (opcja Export Report... programu); (5) przeglądanie i analizę przebiegu wybranej ścieżki lub ścieżek na mapie procesu w celu ich przedefiniowania lub tylko zmodyfikowania; analiza przebiegu ścieżek realizacji procesu jest dostępna w specjalnym oknie dialogowym *Process Analyzer* (opcja Analyses i View Process Analyzer programu).

---

[1] Specjaliści odpowiedzialni za zastosowanie *Six Sigma*, tzw. *black belts*, muszą znać metody planowania eksperymentu (*DOE – Design of Experiments*) [„DOE Warsztaty ..." 2002; M. Harry, R. Schroeder 2001; „Statystyka dla ..." 2002]:

3. Ustalanie rozkładów zmiennych na podstawie zebranych danych rzeczywistych

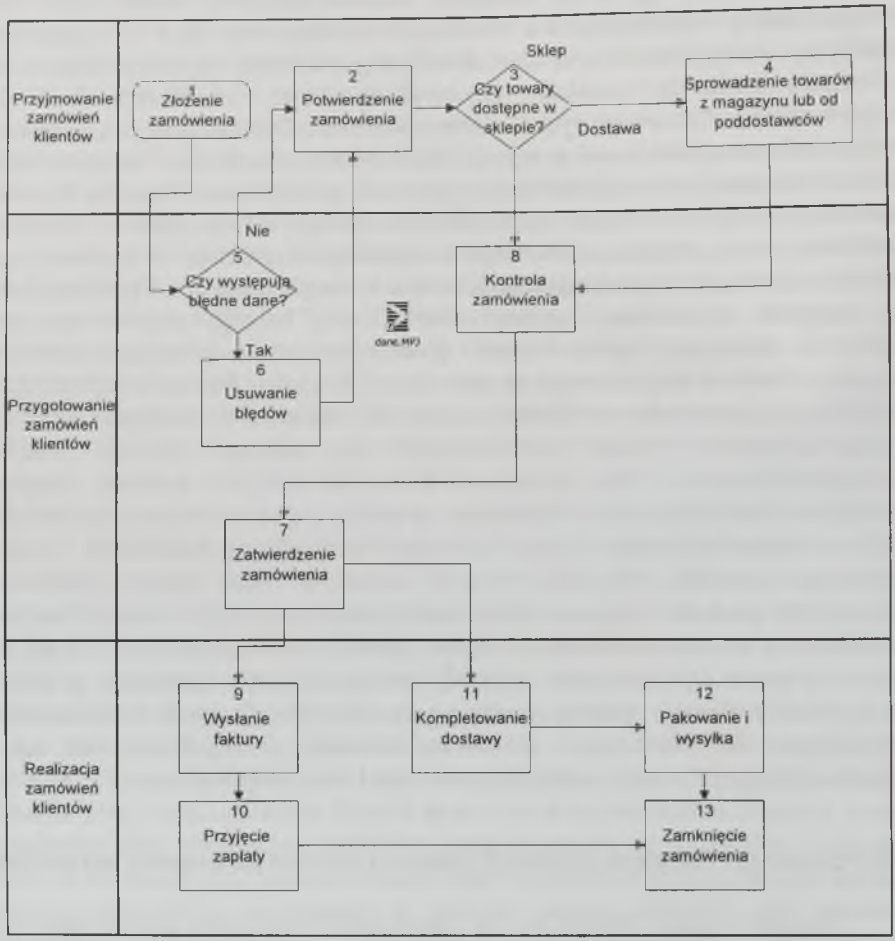
Na podstawie obserwowanych wartości danych są budowane rozkłady zmiennych. Rozkłady te są wykorzystywane dla przeprowadzania eksperymentów, dotyczących realizacji procesów. Tworzone rozkłady statystyczne na podstawie obserwowanych wartości mogą dotyczyć np. czasu przetwarzania transakcji, kosztów czynności, czy też obciążenia zasobów [M. Lasek, M. Pęczkowski, B. Otmianowski 2002; M. Pęczkowski 2002]. Do gromadzenia wartości danych rzeczywistych, konstruowania i wyboru rozkładów wykorzystywane są możliwości, jakie daje pakiet statystyczny *Minitab*. Dla zilustrowania rozpatrywanych zagadnień rozpatrzmy uproszczony schemat procesu realizacji zamówień klientów, sporządzony za pomocą programu *iGrafx Process for Six Sigma* i przedstawiony na rys. 1. Proces realizacji zamówień obejmuje trzy „jednostki organizacyjne" – sfery obsługi klientów: (1) przyjmowanie zamówień, (2) przygotowanie zamówień, (3) realizację zamówień. Celem analizy procesu może być: zbadanie wpływu sposobu sprowadzania towarów: z magazynu lub od poddostawców – w przypadku, gdy zamawiane towary nie są dostępne w sklepie – na czas i koszty realizowanych zamówień, lub zbadanie, jaki wpływ na efektywność procesu ma czas i koszty związane z wykrywaniem i usuwaniem błędów w zamówieniach składanych przez klientów, czy też rozważenie wpływu czasu i kosztów kompletowania dostawy (zamówienia klienta) na przebieg procesu. Wnioski z analizy procesu powinny pomóc w jego udoskonaleniu, tak aby zmniejszyć czas i koszty dostawy zamówionych towarów i zwiększyć zadowolenie klientów z wywiązywania się przez firmę ze zlecanych jej zamówień. Aby przeprowadzić komputerową symulację przebiegu procesu wprowadziliśmy czasy trwania poszczególnych czynności niezbędnych do realizacji zleconych firmie zamówień. Czas niezbędny dla sprowadzenia towarów z magazynu lub od poddostawców, czas usuwania błędów oraz czas kompletowania dostawy zostały wyznaczone jako realizacje rozkładów statystycznych na podstawie zebranych danych ze stu kolejnych, losowych zamówień od klientów. Dane zbierano podczas realizacji zamówień i zapisywano w arkuszu pakietu *Minitab*. Aby zapewnić sprawną współpracę programu *iGrafx Process* z poziomu opisu procesu z pakietem *Minitab*, dzięki mechanizmowi OLE, jest on traktowany jako obiekt wbudowany w proces. Na schemacie procesu (rys. 1.) widać ikonę *dane.MPJ*, kliknięcie na niej pozwala uruchomić *Minitab*. Procedury pakietu *Minitab* umożliwiają wybór teoretycznego rozkładu prawdopodobieństwa najlepiej dopasowanego do zebranych danych empirycznych – estymacja jest uruchamiana po wyborze polecenia *Six Sigma Fit Data*. Otrzymane rozkłady prawdopodobieństwa mogą być potem używane podczas symulacji jako parametry generatora transakcji [M. Lasek, M. Pęczkowski, B. Otmianowski 2002; M. Pęczkowski 2002], czas trwania czynności, koszty oraz atrybuty transakcji. Dostępnych jest kilka rozkładów teoretycznych: rozkład jednostajny (*Uniform*), rozkład normalny (*Normal*), rozkład Weibulla (*Weibull*), rozkład logarytmiczno-normalny (*LogNormal*), rozkład wykładniczy (*Exponential*). Jako kryterium wyboru rozkładu (tj. dobroci

dopasowania do danych empirycznych) proponuje się miarę *Andersona-Darlinga* (*Anderson-Darling Measure*). Wartości miary dla każdego wybranego typu rozkładu są wyświetlane w kolumnie zatytułowanej *A-D Measure*. Dopasowanie do danych jest tym lepsze im mniejszą wartość przyjmuje miara *Andersona-Darlinga*. Należy więc wybrać ten rozkład, dla którego wartość *A-D Measure* jest najmniejsza. Należy zaznaczyć, że rozkład statystyki *Andersona-Darlinga* zależy od rodzaju badanego rozkładu teoretycznego oraz od liczby obserwacji (pomiarów), na podstawie których dokonano obliczeń. Dlatego przy podejmowaniu decyzji powinniśmy traktować te wyniki jedynie jako wskazówkę i brać pod uwagę również wizualne podobieństwo funkcji gęstości na wykresie. Praktyka dowodzi, że dla dużej liczby obserwacji (rzędu kilkaset) występuje duża zgodność pomiędzy wartościami miary *Andersona-Darlinga* a wizualną interpretacją. W analizowanym przypadku dla czasów sprowadzania towarów z magazynu lub od poddostawców (*czas_magpod*) wybraliśmy rozkład *NormDist(1,73;0,60)*, czasów usuwania błędów w złożonych zamówieniach (*czas_usuw_bl*) - *NormDist(0,59;0,30)*, natomiast czasów kompletowania dostaw (*czas_kompl*) - *WeibullDist(0,84;2,02)*. Wyświetlane są parametry rozkładów oraz miary *Andersona-Darlinga*. Jeżeli czas trwania czynności został zdefiniowany za pomocą funkcji rozkładu prawdopodobieństwa – jak przedstawiono to powyżej, to podczas symulacji (eksperymentów) będzie on zdefiniowany za pomocą wyrażenia (*expression*). Inną możliwością jest założenie, że parametr jest stałą, znaną wielkością. Możemy przyjąć na przykład, że czas trwania czynności jest zawsze jednakowy (przyjmujemy parametr *constant* dla czasu trwania czynności) i wskazać ten czas. Tak zrobiliśmy dla czynności *zatwierdzenie zamówienia* – założyliśmy, że ten czas wynosi 5 minut. Możemy też założyć, że czas trwania czynności przyjmuje wartości z określonego przedziału (parametr *distributed*) zgodnie z rozkładem jednostajnym lub normalnym (*Uniform*, *Normal*). Postąpiliśmy tak np. w przypadku przyjęcia zapłaty: zakładając od 5 do 15 minut (*Uniform*).

4. Sterowanie przebiegiem symulacji i eksperymentów za pomocą parametrów

Program *iGrafx Process for Six Sigma* umożliwia różny sposób sterowania przebiegiem symulacji i eksperymentów w punktach podejmowania decyzji, takich jakimi w naszym przypadku są: *czy towary dostępne w sklepie?* oraz *czy występują błędne dane?*. Możemy przyjąć określone rozkłady wyjść, wyrażone procentowo (w programie wybieramy *Decision Statistical*). Przykładowo dla zapytania: *czy towary dostępne w sklepie?*, możemy założyć, że w 50% przypadków będzie to odpowiedź *tak*, a w 50% przypadków odpowiedź *nie*. Te wielkości procentowe odpowiedzi możemy zmieniać tak, aby otrzymać wyniki dla przebiegu procesu przy różnych założeniach. Inną możliwością sterowania przebiegiem procesów w punktach podejmowania decyzji jest zastosowanie wyrażenia (w programie przyjmujemy *Decision Expression*), którego wartość jest zależna od wielkości zdefiniowanego przez nas parametru. Zaletą takiego podejścia jest większa elastyczność. Możemy bowiem sterować rozkładem wyjść z punktu podejmowania

decyzji podczas przeprowadzania eksperymentów, tj. z okna *Design of Experiments* (*DOE*).[2] W zastosowanym wyrażeniu mogą być używane funkcje standardowe programu *iGrafx Process for Six Sigma*.



Rys. 1. Proces realizacji zamówień
Źródło: opracowanie własne przy wykorzystaniu programu *iGrafx Process for Six Sigma*

W naszym przypadku, dla punktu decyzyjnego: *czy towary dostępne w sklepie?*, zdefiniowano parametr decydujący o tym, czy będzie realizowana wersja sprzedaży bezpośrednio ze sklepu, czy też niezbędne będzie sprowadzenie towarów z magazynu lub od poddostawców. Obie sytuacje mogą być

[2] Przydatność zastosowania parametrów dobrze ilustruje przykład zamieszczony w [DOE. Warsztaty ... 2002; „Introducing ..." 2001], gdzie porównywane są dwa procesy: obecnie realizowany z proponowanym. Posłużono się parametrem, tak aby można było na przemian symulować proces obecny i proces proponowany.

przedstawione w postaci odrębnych podprocesów. Ponieważ przedstawiamy bardzo uproszczoną wersję procesu nie zostały one przedstawione jako bardziej szczegółowe, odrębne podprocesy. Gdybyśmy rozpatrzyli oba podprocesy bardziej szczegółowo, z punktu widzenia efektywności działania mogłoby okazać się istotne, czy gromadzimy więcej towarów w sklepie, aby je bezpośrednio sprzedawać klientom, czy też bardziej opłacalne jest sprowadzanie towarów z magazynu lub od poddostawców, a na miejscu zajęcie się przede wszystkim kompletowaniem zamówień i wysyłką. Zdefiniujemy tutaj dwa scenariusze (Scenario), a wprowadzony parametr będzie dotyczyć wyboru scenariusza. Nazwa typu wprowadzonego atrybutu, to *Sklep_lub_Dostawa*, którego elementami są *Sklep* (*sprzedaż wszystkiego ze sklepu*) lub *Dostawa* (*sprowadzenie z magazynu lub od poddostawców*). Są one definiowane jako atrybuty scenariusza. Wartością początkową atrybutu jest *Dostawa*. Atrybuty mogą być przypisywane takim obiektom jak: transakcja, proces, scenariusz, wydział. Istnieją określone zasady nazewnictwa atrybutów. Jeżeli obiekt jest transakcją, nazwa nie ma prefiksu atrybutu, natomiast jeżeli jest procesem, scenariuszem lub wydziałem, to przed nazwą pojawia się odpowiednio prefiks *P* (*Process*), *S* (*Scenario*), *D* (*Department*). Dla analizy naszego procesu został zdefiniowany jeszcze jeden atrybut scenariusza. Nosi on nazwę *Czy_błąd*, jest typu *Number* i ma przypisaną wartość początkową równą 20. Jest on związany z punktem decyzyjnym *Czy występują błędne dane?*

5. Zastosowanie funkcji standardowej *PercentYes*

Funkcja *PercentYes* używana w przypadku decyzji na podstawie danych statystycznych określa kierunek wyjścia transakcji. Jest to funkcja typu logicznego i przyjmuje dwie wartości „Prawda" (*Yes*) lub „Fałsz" (*No*). Funkcja *PercentYes* ma jeden argument, który określa (w procentach) częstość wybierania wartości „Prawda" (*Yes*). Parametr *Czy_błąd* określa w procentach, ile razy podczas symulacji zostanie wybrana ścieżka *Tak* (występowanie błędu). Jako wartość początkową przyjęto 20%, co można zmieniać przeprowadzając różne eksperymenty.

6. Definiowanie zasobów i przypisywanie zasobów dla realizacji czynności

W metodologii *iGrafx* przyjmujemy, że do wykonania czynności potrzebne są zasoby [M. Lasek, M. Pęczkowski, B. Otmianowski 2002; M. Pęczkowski 2002]. Trzy „bazowe", przyjęte jako standardowe w programie typy zasobów, to – *worker* (ludzie), *equipment* (maszyny i urządzenia), *other* (można zdefiniować tu dowolny rodzaj zasobów). Dla realizacji zamówień, modelowanych za pomocą opisywanego procesu, zaangażowani są pracownicy. Przewidziano zatrudnienie: w dziale „Przyjmowanie zamówień klientów": pracowników zajmujących się przyjmowaniem składanych przez klientów zamówień i potwierdzeniami zamówień (*Przyjmujących*) oraz pracowników zajmujących się dostawą towarów z magazynu lub od poddostawców (*Dostawców*), w dziale „Przygotowanie

zamówień klientów": *Kontrolerów* - zajmujących się kontrolą zamówień, wykrywaniem błędów w zamówieniach, usuwaniem błędów oraz zatwierdzaniem zamówień do realizacji, a w dziale „realizacja zamówień klientów": *Księgowego* wysyłającego faktury i przyjmującego zapłatę oraz *Sprzedawców*, kompletujących dostawy, zajmujących się pakowaniem i wysyłką oraz decydujących o *zamknięciu zamówienia*. Zgodnie z metodą postępowania systemu *iGrafx* najpierw zostały zdefiniowane zasoby (pracownicy) wraz ze stawkami godzinowymi ich wynagrodzeń, a następnie przypisano ich do realizacji określonych czynności. Sposób definiowania zasobów i przydzielania zasobów do czynności został opisany m.in. w [M. Lasek, M. Pęczkowski, B. Otmianowski 2002; M. Pęczkowski 2002] - w naszym przypadku zajmujemy się tylko zaangażowaniem pracowników, nie interesując się urządzeniami niezbędnymi do realizacji czynności.

7. Licznik błędów i monitorowanie błędów

Dla analizy czasu, kosztów, wykorzystania zasobów realizowanego procesu możemy zdefiniować własne mierniki. Jednym z najprostszych wskaźników może być zbadanie liczby błędów, jakie pojawiają się w zamówieniach, aby następnie określić wpływ ich występowania na efektywność realizacji całego procesu. Jedną z możliwości „zliczania błędów" jest wprowadzenie *licznika błędów*. Ten *licznik błędów* musi być wcześniej zdefiniowany jako atrybut, a jego wartości mogą być nadawane lub zmieniane w różnych miejscach procesu. Jednym ze sposobów jest dokonanie tego w oknie dialogowym czynności (tu: *Usuwanie błędów*) na karcie *Attributes*. Dogodnym narzędziem dla gromadzenia statystyk dotyczących przebiegu procesu (jego transakcji i atrybutów) jest zastosowanie monitorowania (śledzenia przebiegu procesu) poprzez ustawienie tzw. monitorów. Monitor jest ustawiany w miejscu realizacji określonej czynności i podczas przeprowadzania symulacji gromadzi informacje o realizowanych transakcjach i wskazanych przez nas statystykach, np. liczbie transakcji, które zakończyły się, ale podczas ich realizacji pojawił się błąd. O możliwościach wykorzystywania monitorów dla badania przebiegu procesów dzięki definiowaniu i „zliczaniu" przez nie statystyk (wskaźników) można się zapoznać m.in. w [DOE. Warsztaty ... 2002; „Introducing ..." 2001; M. Lasek, M. Pęczkowski, B. Otmianowski 2002; M. Pęczkowski 2002]. i*Grafx* ma wiele zdefiniowanych już statystyk dotyczących czasu, kosztów, wykorzystania zasobów, kolejek. Użytkownik może wybrać z nich zestaw mierników, które wyświetlają się w raportach wynikowych. Może też zdefiniować własny układ raportu i zdefiniować inne mierniki. Służy do tego opcja *Custom Statistics*. Wskaźniki metody *Six Sigma* możemy zdefiniować jako *Custom Statistics* (wybierając z menu *Report* polecenie *Custom Stats...*). Dla naszego przykładu zdefiniujemy wskaźnik wydajności mierzonej w toku (*Rolled Throughput Yield*), wyliczany wg wzoru: $RTY = e^{DPU}$, gdzie *DPU* jest liczbą błędnych danych podzieloną przez liczbę transakcji.

8. Przeprowadzanie symulacji i eksperymentów

Po wprowadzeniu danych do procesu i ustawieniu jego przebiegu za pomocą generatora możemy przeprowadzić symulację procesu. W naszym przykładzie generator wprowadza transakcje, jakimi są zamówienia klienta do punktu startowego: u nas jest nim *złożenie zamówienia*. Na temat przeprowadzania symulacji procesów za pomocą programu *iGrafx* i odczytywania wyników symulacji można się zapoznać m.in. w pracach [„iGrafx ...." 2000; M. Lasek, M. Pęczkowski, B. Otmianowski 2002; M. Pęczkowski 2002]. Program *iGrafx Process for Six Sigma* został rozbudowany w porównaniu do programu *iGrafx* o możliwość zaprojektowania i przeprowadzania eksperymentów ["iGrafx ..." 2000; „Introducing ..." 2001]. Możliwość zaprojektowania i przeprowadzania eksperymentów daje polecenie: *Six Sigma Rapid DOE* (*Design of Experiments*). Po wyborze tej opcji z menu jest wykonywany cykl eksperymentów przy narzuconych przez nas założeniach wejściowych. Są to niezależne wielkości wejściowe (*factors*) – czynniki, które kontrolujemy podczas eksperymentu W wyniku eksperymentów przedstawiane są wyniki, dotyczące interesujących nas statystyk. Są to zależne wielkości wyjściowe (*responses*), określające wyniki eksperymentów. W przedstawianym przypadku projektowania eksperymentów jako analizowane (aktywne) zostały zaznaczone wcześniej zdefiniowane podczas projektowania procesu, dwa atrybuty: *Sklep_lub_Dostawa* oraz *Czy_błąd*. Wartość pierwszego możemy zmieniać w zależności od tego, czy sprzedaż odbywa się bezpośrednio ze sklepu, czy też potrzebne jest sprowadzenie go z magazynu lub od poddostawców. W drugim przypadku ustalamy częstość pojawiania się błędów. Obliczamy 3 mierniki (statystyki): *Cost* - przeciętny koszt transakcji (u nas realizacji zamówienia), *AvgCycleTime* – średni czas przepływu jednej transakcji przez cały proces, *RTY* – wskaźnik wydajności mierzonej w toku. Replikacja (*replication*) jest krotnością, z jaką każdy eksperyment zostanie powtórzony. Replikacja równa 4 oznacza czterokrotne przeprowadzenie obliczeń przy przyjętych założeniach. W naszym przypadku mamy po 2 wartości dla każdej z rozpatrywanych zmiennych: *Sklep_lub_Dostawa* oraz *Czy_błąd*, tj. 2x2 = 4 możliwe kombinacje wartości, a uwzględniając 4 replikacje: 4x4 = 16 możliwych eksperymentów. Po wprowadzeniu ustaleń, tak jak przedstawiono to powyżej, możemy przeprowadzić eksperyment, a następnie zapisać założenia i wyniki w arkuszu programu *Minitab* w celu przeprowadzenia analizy statystycznej.

Analizę statystyczną i interpretację wyników eksperymentów ułatwiają wykresy, jakie możemy sporządzić za pomocą procedur udostępnianych przez program *Minitab*. Przykładowo można zilustrować zależność między zmienną *Sklep_lub_Dostawa* lub zmienną *Czy_błąd* a wskaźnikami: średniego kosztu realizacji zamówienia *Cost*, średniego całkowitego czasu przetwarzania zamówienia *AvgCycle Time*, wydajności mierzonej w toku *RTY*. Wykresy ilustrują jak *RTY* maleje wraz z rosnącą liczbą błędów w zamówieniach i jak zmieniają się wskaźniki kosztów w zależności od sytuacji, czy sprowadzamy towary z magazynu lub od poddostawców, czy sprzedajemy je bezpośrednio ze sklepu.

9. Śledzenie kosztów błędów w przypadku różnych założeń w realizacji procesów

Rozważmy przykład procesu dotyczącego realizacji projektów informatycznych i obejmującego cały cykl życia oprogramowania, począwszy od analizy, poprzez projektowanie, programowanie aż do etapu odbioru i użytkowej eksploatacji. Zakładamy, że proces jest realizowany w firmie produkującej systemy informatyczne na sprzedaż. Tym razem jako transakcję będziemy traktować wytwarzany system informatyczny. Celem naszych działań jest ograniczenie błędów w produkcie dostarczanym użytkownikowi oraz jak najwcześniejsze wykrywanie i usuwanie błędów, aby ograniczyć łączne koszty związane z defektami: koszty *Non-Value Added (NVA)* – koszty nie doliczane do kosztów produkcji lub usługi (nie przynoszące wartości). Koszty są związane z kontrolą na każdym etapie produkcji oprogramowania (w punktach zaznaczanych jako punkty decyzyjne) i usuwaniem wykrytych błędów (nazywanych na mapie procesu czynnościami modyfikacji). Przedstawmy proces (mapę procesu) w *iGrafx Process 2003 for Six Sigma* i „wbudujmy" arkusz *Excela*. Arkusz zawiera informacje o błędach, które mogą być nam pomocne w procesie doskonalenia produkcji programów. Zakładamy, że na poszczególnych etapach realizacji projektów zostaje usunięty tylko określony procent błędów. Przeprowadzamy symulacje zmieniając założenia, dotyczące liczby błędów wykrywanych i skutecznie usuwanych na poszczególnych etapach tworzenia systemu. Możemy sporządzić raport przedstawiający istniejącą sytuację, co do wykrywanej liczby błędów i kosztów ich usuwania, a wynikami w tym zakresie uzyskiwanymi w kolejno realizowanych symulacjach. Interesującą możliwością badania procesów jest ominięcie podczas symulacji pewnych etapów. Aby to zrobić, najprościej rozrysować proces, tak aby pominąć czynności, których nie chcemy wykonywać podczas symulacji (technika „bypass") [„"Introducing ..." 2001]. Po przeprowadzeniu symulacji możemy je ponownie „podłączyć" do procesu.

10. Analiza wyników realizacji pojedynczych transakcji

Możemy wyeksportować do *Excel'a* lub do pakietu *Minitab*, wybrane dane, dotyczące realizacji poszczególnych transakcji (np. budowy jednego systemu informatycznego). Są to dane, dotyczące pojedynczych, kolejno realizowanych transakcji, a nie – jak w raportach z symulacji – średnie wielkości z symulacji dla wielu zrealizowanych transakcji. W *Excel'u* lub *Minitab* możemy dokonywać analiz statystycznych i oceniać proces budując mierniki i sporządzając wykresy ilustrujące wyniki przebiegów poszczególnych transakcji.

11. Optymalizacja zatrudnienia

Rozpatrzmy jeszcze jeden przykład. Dotyczy on obsługi korespondencji w małej firmie. Załóżmy, że proces obsługi korespondencji przedstawiliśmy za pomocą *iGrafx Process 2003 for Six Sigma*. Proces obejmuje trzy działy - „obszary": (1) przyjmowanie korespondencji, (2) segregowanie korespondencji, (3) dostarczanie korespondencji. Celem zastosowania metodologii *Six Sigma* i programu *iGrafx Process 2003 for Six Sigma* jest analiza wykorzystania pracowników i budowa zespołu pozwalającego realizować przyjmowanie, segregowanie i dostarczanie korespondencji tak, aby osiągnąć możliwie najlepsze rozwiązanie pod względem ilości obsługiwanej korespondencji (liczby przesyłek), czasu oraz kosztów. Wykorzystujemy procedurę *RapiDOE*: *Rapid Design of Experiments*, badając jak zmiana liczby zatrudnianych pracowników na poszczególnych etapach obsługi korespondencji wpływa na liczbę obsługiwanych przesyłek (*TransCount*), koszty (*Cost* i *AvgCost*) oraz średni czas obsługi przesyłki (*AvgCycle Time*). Analizę i interpretację wyników możemy przeprowadzić posługując się wykresami sporządzonymi za pomocą procedur *Minitab*.

12. Podsumowanie

W artykule przedstawiliśmy trzy przykłady zastosowania programu *iGrafx Process 2003 for Six Sigma* do optymalizacji procesów gospodarczych: realizacji zamówień klientów, procesów produkcji oprogramowania, obsługi korespondencji. Program umożliwia optymalizację procesów pod względem czasu, kosztów i wykorzystywania zasobów. W przypadku procesu realizacji zamówień klientów oraz obsługi korespondencji do analizy statystycznej wykorzystaliśmy pakiet *Minitab*. W przykładzie, dotyczącym produkcji oprogramowania posłużyliśmy się arkuszem kalkulacyjnym *Excel*. Program *iGrafx Process 2003 for Six Sigma* okazał się bardzo wygodnym narzędziem, umożliwiającym statystyczną analizę danych i przeprowadzanie eksperymentów w oparciu o nagromadzone dane z rzeczywiście zrealizowanych procesów realizacji zamówień klientów, procesów produkcji oprogramowania, obsługi korespondencji. Program umożliwia statystyczną analizę danych zebranych w wyniku realizacji wielu (kilkudziesięciu, czy też kilkuset procesów) i w oparciu o te dane zbudowanie modelowych rozkładów prawdopodobieństwa, które możemy następnie wykorzystywać w przeprowadzanych eksperymentach (tak jak przedstawiliśmy to na przykładzie realizacji zamówień klientów). Po wykonaniu eksperymentów możemy przeprowadzić statystyczną analizę wyników – w tym bardzo ułatwiającą ich interpretację – analizą graficzną za pomocą wykresów. Zasługującą na podkreślenie zaletą programu jest możliwość analizy i porównywania przebiegu równoległych ścieżek w realizacji procesu. W ten sposób możemy zbadać wady i zalety różnych sposobów realizacji danego procesu, np. różnych sposobów

realizacji zleceń klientów. *iGrafx Process 2003 for Six Sigma* umożliwia definiowanie i stosowanie wskaźników zalecanych w metodologii *Six Sigma*, takich jak wskaźnik wydajności przejściowej, czy wskaźnik wydajności mierzonej w toku. Nasze doświadczenia z programem, nie tylko te, które opisaliśmy w artykule, wskazują, że może być on skutecznym narzędziem wspomagającym praktyczne zastosowania zdobywającej coraz większą popularność strategii doskonalenia procesów, a nawet jak można zaryzykować chyba szersze ujęcie - nie tylko strategii, ale całej nowej filozofii analizy i modelowania procesów gospodarczych, zwanej metodologią *Six Sigma*.

Literatura

1. DOE. Warsztaty: „Projektowanie eksperymentu za pomocą iGrafx System", materiały z konferencji „Six Sigma to bardzo proste"; KEMA; 18-19 października 2002; opracowanie iSolution@iSolution.com.pl; 2002.
2. Goh T.N.; "A Strategic Assessment of Six Sigma"; Quality and Reliability Engineering International (18); 403-410; 2002.
3. Harry M., Schroeder R.; „Six Sigma. Wykorzystanie programu jakości do poprawy wyników finansowych"; Dom Wydawniczy ABC – Oficyna Ekonomiczna; Kraków; 2001.
4. "iGrafx Process for Six Sigma 2000. The Intelligent Way to Achieve Six Sigma Results"; User's Guide; Micrografx Inc.; 2000.
5. "Introducing iGrafx Process 2000 for Six Sigma, training manual", Micrografx Inc.; 2001.
6. Lasek M., Pęczkowski M., Otmianowski B.; „Analiza procesów biznesowych z wykorzystaniem programów: iGrafx Process 2000 for Six Sigma / iGrafx FlowCharter 2000 Professional PL"; Wyższa Szkoła Informatyki Stosowanej i Zarządzania; Warszawa; ISBN 83-88311-50-6; 2002.
7. Lientz B.P., Rea K.P.; "Achieve Lasting Process Improvement. Reach Six Sigma Goals without the Pain"; Elsevier Academic Press; 2002.
8. Pęczkowski M.; „Modelowanie procesów gospodarczych z wykorzystaniem programu iGrafx Process 2000 / iGrafx FlowCharter 2000 Professional"; materiały dydaktyczne; Uniwersytet Warszawski; Warszawa 2002.
9. „Statystyka dla jakości produktów i usług – Six Sigma i inne strategie"; materiały z seminarium zorganizowanego przez StatSoft Polska; Warszawa 28 października 2002.

Mirosława LASEK, Marek PĘCZKOWSKI
Katedra Informatyki Gospodarczej i Analiz Ekonomicznych, Wydział Nauk Ekonomicznych, Uniwersytet Warszawski, ul. Długa 44/50, 00-241 Warszawa,
e-mail: mlasek@wne.uw.edu.pl, mpeczkowski@we.uw.edu.pl

ZINTEGROWANY SYSTEM ZARZĄDZANIA I KONTROLI IACS W POLSKIEJ RZECZYWISTOŚCI PRZED AKCESJĄ DO WSPÓLNOTY EUROPEJSKIEJ

Jarosław ŁAPETA, Tomasz LIS

Wstęp

Zintegrowany System Zarządzania i Kontroli (IACS) jest podstawowym narzędziem kontroli wydatków Sekcji Gwarancji Funduszu FEOGA UE. Znaczenie tego Funduszu rośnie – wraz z upływem czasu – jako, że stopniowo w UE coraz więcej działań jest bezpośrednio zarządzanych przez ten system. W Polsce jego wdrożenie planowane jest do czasu akcesji Polski do UE. Nic w tym nie byłoby dziwnego gdyby nie ograniczenia czasowe na wdrażanie całego systemu. W Polsce zasadniczo realizowany on jest w całym kraju od 2002 roku. Wcześniej były nieudane próby wdrożenia w kilku województwach. W między czasie podczas negocjacji w Kopenhadze zmieniono strukturę dopłat, co spowodowało również zmiany w systemie.

Brak kompletnego systemu w dniu akcesji jest równoznaczny z nieotrzymaniem dopłat do produkcji rolnej przez polskich rolników.

1. System IACS w krajach Unii Europejskiej.

System IACS składa się ze ściśle określonych zasad, reguł postępowania oraz kryteriów zgodnie, z którymi obliczane i wypłacane są dotacje dla gospodarstw i rolnictwa. System taki posiada każde państwo członkowskie UE. Rozporządzenie Rady EWG nr 3508/92, na mocy, którego wprowadzono System IACS, ustanawia 5 elementów, które są obowiązujące w państwach członkowskich:

- skomputeryzowana baza danych służąca rejestracji danych zawartych we wnioskach o pomoc dla każdego gospodarstwa rolnego,
- system identyfikacji gruntów rolnych i działek uprawnych, tj. system pozwalający na lokalizację zadeklarowanej powierzchni, tak aby można było monitorować w czasie ewentualne zmiany posiadania, powierzchni i upraw. W oparciu o ten system organizowane są kontrole krzyżowe i kontrole na miejscu wniosków o dopłaty powierzchniowe,
- system ewidencji zwierząt, który pozwala na kontrolę krzyżową i kontrolę na miejscu wniosków o dopłaty (premie) zwierzęce,
- sformalizowane procedury i tryb składania wniosków o płatności bezpośrednie w odniesieniu do powierzchni, areału upraw pastewnych i pogłowia zwierząt,
- zintegrowany system kontroli administracyjnych i kontroli na miejscu.

Skomputeryzowana baza danych – służy rejestracji wniosków, naliczeń i wypłat oraz danych uzyskiwanych z wniosków o pomoc dla każdego gospodarstwa

rolnego. Dobór systemu komputerowego pozostawia się decyzji państwom członkowskim.

System identyfikacji gruntów i działek rolnych – jest to rejestr zawierający dane w formie tekstu (dane osobowe, położenie i powierzchnię działki, rodzaj użytkowania) nazywany także bazą alfanumeryczną. Baza alfanumeryczna musi być uzupełniana bazą graficzną umożliwiającą identyfikację działek uprawnych i upraw. Dla wytworzenia bazy graficznej Unia zaleca stosowanie tzw. ortofotomapy ze zdjęć lotniczych lub satelitarnych (możliwy jest wariant mieszany). System identyfikacji i rejestracji gruntów i działek jest warunkiem niezbędnym do przeprowadzania efektywnych kontroli wniosków o pomoc powierzchniową.

System ewidencji zwierząt – jest to komputerowy rejestr tych zwierząt, na które przyznawane są płatności bezpośrednie (bydło, owce, kozy). W systemie komputerowym powinny być dostępne następujące informacje: znak zwierzęcia (kod identyfikacyjny), data urodzenia, numer identyfikacyjny gospodarstwa urodzenia i wszystkich gospodarstw, w których zwierzę przebywało, daty każdej zmiany gospodarstwa, data śmierci lub uboju. System ten musi umożliwiać jednoznaczną identyfikację zwierząt, na które składane są wnioski o płatności bezpośrednie, stanowiąc przy tym podstawę do przeprowadzania kontroli (wizytacji terenowych).

Wnioski o płatności – każdego roku rolnik jest uprawniony do złożenia jednego wniosku o pomoc powierzchniową oraz jednego lub więcej wniosków o płatności (premie) zwierzęce. Wszystkie wnioski muszą zawierać dane dotyczące tożsamości rolnika oraz pisemne zobowiązanie, że rolnik będzie przestrzegał określonych w przepisach zasad i reguł postępowania w odniesieniu do poszczególnych schematów pomocowych w zakresie płatności. Dane zawarte we wnioskach muszą umożliwiać przeprowadzenie kontroli kwalifikowanych poprzedzających naliczenie i wypłatę środków finansowych.

Zintegrowany system kontroli składa się z kontroli administracyjnych i kontroli na miejscu. Celem kontroli administracyjnych jest zapewnienie, że pomoc jest naliczona prawidłowo oraz, że żadne płatności nie są dokonywane podwójnie. Obejmują one kontrole gospodarstw wnioskujących o płatności, kontrole krzyżowe zadeklarowanych powierzchni oraz kontrole krzyżowe zadeklarowanych zwierząt. Głównym celem kontroli na miejscu jest sprawdzenie, czy wniosek rolnika odpowiada rzeczywistości (deklarowanym uprawom, co do rodzaju i powierzchni, a także liczby i rodzaju zwierząt). Kontrole na miejscu muszą obejmować, co najmniej 5% wniosków o dopłaty powierzchniowe i 10% wniosków opartych o pogłowie zwierząt.

2. Polska wersja systemu IACS.

Zintegrowany System Zarządzania i Kontroli składać się będzie ze zbioru danych o polskim rolnictwie i rolnikach oraz ściśle określonych zasad i reguł

postępowania zgodnie z którymi naliczane i wypłacane będą polskim rolnikom dopłaty bezpośrednie.

Zgodnie z Narodowym Programem Przygotowania do Członkostwa Polska, do dnia akcesji, zobowiązana jest przygotować i wdrożyć Zintegrowany System Zarządzania i Kontroli, który zapewni administrowanie i kontrolę płatności bezpośrednich. Zadanie wykonania i prowadzenia tego Systemu – na mocy ustawy o krajowym systemie ewidencji gospodarstw rolnych i zwierząt gospodarskich – powierzono Agencji Restrukturyzacji i Modernizacji Rolnictwa.

W celu ustanowienia podstaw prawnych dla funkcjonowania w Polsce systemu ds. administrowania dopłatami bezpośrednimi dla rolnictwa po wejściu Polski do Unii Europejskiej, Sejm uchwalił ustawę z dnia 25 lipca 2001 r. o krajowym systemie ewidencji gospodarstw rolnych i zwierząt gospodarskich oraz o zmianie niektórych ustaw. Ustawa w sposób niewystarczający stanowi o systemie kontroli, który jest zasadniczym elementem prawodawstwa europejskiego dotyczącego systemu IACS. W rezultacie powołany ustawą z dnia 25 lipca 2001 r. system w niewielkim zakresie nabiera cech systemu operacyjnego opartego na kontroli wniosków, naliczeń i wypłat, co stanowi o istocie Zintegrowanego Systemu Zarządzania i Kontroli w państwach UE.

Przepisy cytowanej ustawy stanowią, że system ten składał się będzie z następujących elementów:

- ewidencji gospodarstw rolnych,
- ewidencji zwierząt gospodarskich,
- rejestru płatności bezpośrednich,
- dokumentacji związanej z ewidencją gospodarstw rolnych oraz przyznawaniem i wypłatą płatności bezpośrednich,
- dokumentacji związanej z przeprowadzanymi przez ARiMR kontrolami przestrzegania przepisów dotyczących płatności bezpośrednich.

W skład systemu wchodzą także rejestry producentów bydła, owiec oraz kóz, którzy są uprawnieni do otrzymywania limitów określających maksymalną liczbę zwierząt, za które mogą być przyznane płatności bezpośrednie na podstawie odrębnych przepisów.

Agencja Restrukturyzacji i Modernizacji Rolnictwa powierzyła początkowo całokształt zadania związanego z budową systemu wybranemu w drodze przetargu wykonawcy. Wykonanie tego Systemu – w części informatycznej i nieinformatycznej, ale bez agencji płatniczej – określono na dzień 31 grudnia 2002r.

Brak postanowień dotyczących budowy agencji płatniczej, a więc elementu warunkującego wypłatę środków rolnikom, stanowił zasadniczą wadę umowy zawartej z Wykonawcą systemu. Wątpliwości budziło także powierzenie Wykonawcy budowy tzw. części nieinformatycznej systemu IACS. Zdecydowano zatem, że część nieinformatyczną Agencja wykona samodzielnie. Podpisane w dniu 12 marca br. Porozumienie – aneks z Hewlett-Packard Polska usunęło wymienione powyżej zasadnicze wady umowy konieczne do poprawy funkcjonalności i akredytacji systemu IACS. Powstała jednak konieczność przyspieszenia tempa prac w zakresie dostosowania organizacyjnego Agencji do

nowych zadań związanych m.in. z przejęciem – w wyniku renegocjacji umowy – całej części tzw. nieinformatycznej. W ramach realizacji zadań związanych z częścią nieinformatyczną Systemu – ARiMR jest zobowiązana przygotować strukturę organizacyjną (od centrali do powiatów), przygotować, pozyskać i wyposażyć lokale pod siedziby IACS (16 województw, 315 powiatów), a także przeprowadzić:

- rekrutację i szkolenie personelu do obsługi IACS,
- kampanię informacyjną na temat Systemu,
- szkolenie rolników,
- szkolenie pracowników ARiMR oraz
- zapewnić środki transportu dla IACS.

Zadaniem niezwykle ważnym, bo warunkującym poprawność działania i akredytację Systemu jest przygotowanie struktury organizacyjnej i funkcjonalnej przyszłej agencji płatniczej. Zadanie to w części dotyczącej zaprojektowania odpowiedniej architektury systemu informatycznego ciąży – po renegocjacji kontraktu – na firmie Hewlett Packard. Obejmować ona musi obsługę 3 podstawowych funkcji przyszłej agencji: autoryzację płatności, polecenie dokonywania płatności dla rolników (we współpracy z bankami) oraz księgowanie. Agencja zobowiązana jest do przygotowania wszystkich procedur dotyczących agencji płatniczej, a także procedur dla wszystkich czynności związanych procesem obsługi wniosków, kontroli administracyjnych i kontroli na miejscu oraz naliczeń płatności, które obsługuje System IACS. Obecnie trwają intensywne prace, których rezultatem będzie przygotowanie Systemu i Agencji do wykonywania funkcji agencji płatniczej oraz akredytacja Systemu nie później niż do daty akcesji.

3. Informatyczna strefa IACS.

Zgodnie z umową na budowę Zintegrowanego Systemu Zarządzania i Kontroli (IACS) zawartą w dniu 28 czerwca 2001 roku z firmą Hewlett-Packard Polska Sp. z o.o. z Agencją Restrukturyzacji i Modernizacji Rolnictwa. Zawarcie Umowy nastąpiło w wyniku postępowania o udzielenie zamówienia publicznego zgodnie z Ustawą o Zamówieniach Publicznych. Przetarg nieograniczony został ogłoszony przez ARiMR 16 listopada 2000 roku. Zgodnie z procedurami przygotowano Specyfikację Istotnych Warunków Zamówienia, którą odebrały 62 firmy. Do przetargu zgłosiło się czterech oferentów. Agencja rozstrzygnęła przetarg w kwietniu 2001 roku, powierzając w wyniku przeprowadzonej procedury przetargowej realizację powyższego kontraktu firmie Hewlett-Packard Polska sp.zo.o. Dnia 28 czerwca 2001 roku ARiMR oraz HP Polska podpisały umowę na opracowanie i wdrożenie Zintegrowanego Systemu Zarządzania i Kontroli IACS.

Łączna wartość wynagrodzenia HP z tytułu wykonania zadania przewidzianego umową wynosi równowartość 67.335.902 Euro w złotych polskich. Płatności następują zgodnie z harmonogramem na podstawie faktur, po akceptacji kolejnych etapów projektu przez komitet sterujący i koordynatora ze

strony ARiMR. Realizacja kontraktu w całości finansowana jest ze środków budżetowych i nie ma w nim udziału finansowego Unii Europejskiej.

Jeśli chodzi o zaawansowanie prac nad informatyczną stroną IACS nie ma żadnych informacji oficjalnych o fazie prac, oraz modelu systemu. Czyżby była to kolejna „czarna skrzynka" polskiej informatyki?

4. Podsumowanie.

Prace nad zbudowaniem IACS oraz nad zorganizowaniem Agencji Płatniczej znajdują się obecnie, w większości zadań, na ścieżce kryzysowej. Agencja Restrukturyzacji i Modernizacji Rolnictwa odpowiada za przygotowanie systemu IACS i będzie pełnić funkcję Agencji Płatniczej. Stwierdzenie, że realizacja większości zadań znajduje się obecnie na ścieżce kryzysowej, oznacza, że teraz jakiekolwiek opóźnienia w realizacji programu IACS grożą już jego nieprzygotowaniem do momentu akcesji, a to oznaczałoby, że polscy rolnicy nie mogliby z budżetu Unii Europejskiej otrzymać wynegocjowanych w Kopenhadze płatności bezpośrednich.

Można długo myśleć nad i zastanawiać się, dlaczego tak ważna struktura organizacyjna realizująca korzyści dla polskiego rolnictwa z integracji do UE jest tak słabo przygotowana. Na leży jednak podkreślić, że nie wynika to z wolnych działań ARiMR. Opisana publikacja oparta jest na obserwacjach w śląskim oddziale agencji w Częstochowie. Łatwo tam zauważyć, że pracownicy dają z siebie wszystko. Mimo to jest poważne zagrożenie powodowane głównie brakiem czasu i płynności finansowej. Obecnie w województwie śląskim w zaawansowanym stadium znajdują się prace na infrastrukturą techniczną tzn. budynki, komputery, okablowania łącza zewnętrzne. Jednak najbardziej niebezpieczne jest wdrożenie systemy informatycznego, który budowany od podstaw przy uwzględnianiu wielu zmian może stać się wielkim fiaskiem. Obserwując niepowodzenia polskiej informatyki (ZUS, Wybory 2002) można mieć mieszane uczucia, co do sukcesu wdrożeniowego tego przedsięwzięcia.

Ocenę całokształtu wdrożenia można potraktować jako temat wyjściowy na kolejną konferencję naukową.

Literatura

1. www.minrol.gov.pl
2. www.arimr.gov.pl
3. www.ppr.pl
4. Informacje i obserwacje wykonane w śląskim oddziale ARiMR.

Mgr inż. Jarosław Łapeta
Politechnika Częstochowska;
Wydział Zarządzania
Instytut Ekonometrii i Informatyki
Ul. Armii Krajowej 19B
42-200 Częstochowa
e-mail: jlapeta@zim.pcz.czest.pl

mgr inż. Tomasz Lis
Politechnika Częstochowska;
Wydział Zarządzania
Instytut Ekonometrii i Informatyki
Ul. Armii Krajowej 19B
42-200 Częstochowa
e-mail: tomlis1@poczta.wp.pl

PRAKTYCZNE ASPEKTY WDRAŻANIA ALTERNATYWNEJ TECHNOLOGII DOSTĘPU DO INTERNETU PRZEZ SIEĆ ENERGETYCZNĄ- POWER LINE COMMUNICATIONS (PLC)

Rafik NAFKHA, Michał KARDAŚ, Paweł PIOTROWSKI

Streszczenie: Zmiany dokonujące się na krajowym rynku energetycznym wpływają na poszerzenie oferty usług spółek dystrybucyjnych. Ciekawą i perspektywiczną usługą wydaje się dostęp do Internetu i usług mutlimedialnych przy wykorzystaniu przewodów sieci energetycznej. Ta forma dostępu jest testowana w wielu krajowych spółkach dystrybucyjnych, w tym również w STOEN S.A.[13]. Z praktycznego wdrażania tej technologii wynikają zarówno wady jak i zalety. Tekst omawia zarówno praktyczne aspekty wdrażania jak i opis samej technologii, która ma wiele odmian i jest w różny sposób wykorzystywana w poszczególnych krajach europejskich oraz w Stanach Zjednoczonych.

1. Rys historyczny

W roku około 1920 rozpoczęto prace nad techniką modulacji częstotliwości nośnej i wykorzystaniem linii energetycznych wysokiego napięcia do transmisji danych [1]. W trakcie upowszechniania elektryczności i rozbudowy sieci energetycznej WN, istotne stało się zapewnienie wysokiej niezawodności w dostawach energii do odbiorcy. Dostawcy energii potrzebowali informacji o aktualnym stanie sieci, stacji transformatorowych, pozwalających na bezzwłoczną lub nawet zautomatyzowaną reakcję w wyniku np. zwarcia lub zerwania linii. Stworzenie oddzielnej sieci diagnostycznej było kosztowne. Zdecydowano się wykorzystać istniejącą infrastrukturę sieciową WN do przesyłu danych. Już wtedy rozważano ich wykorzystanie do przesyłu głosu. Od tego momentu technologia przesyłu danych sieciami energetycznymi zaczęła ewoluować, stając się coraz bardziej niezawodną techniką. W roku 1929 w Szwajcarii zaczął pracować system, który potrafił przesyłać sygnały typu włącz/wyłącz [2]. Polecenia te służyły do sterowania m.in. oświetleniem ulicznym, podgrzewaczami wody i światłami wystaw sklepowych. Opracowany przez firmę Siemens i wdrożony w Poczdamie w 1930 roku system o nazwie Telenerg posiadał wiele częstotliwości nośnych rozłożonych w zakresie częstotliwości od 280 do 600 Hz. W 1935 roku AEG uruchomił w Magdeburgu i Stuttgarcie system o nazwie Transkommando, wykorzystujący modulację amplitudy napięcia sieciowego [1]. Po roku 1935 rozwijane były techniki multipleksowania z jedną częstotliwością nośną, przy czym najczęściej stosowano modulację częstotliwości zbliżonej do częstotliwości napięcia sieciowego. Dzięki temu sygnały bez problemu przenoszone były przez transformatory i nie były tłumione w stacjach rozdzielczych. Na takiej właśnie koncepcji opiera się stosowana współcześnie technika TRT. Prace nad przesyłem

głosu w sieci energetycznej prowadziła również firma Bell [2]. Badania rozpoczęto w roku 1936. Po dziesięciu latach opracowano system, który mógł przenosić głos na odległość do 32 km. W latach 50-tych przewody energetyczne były dość powszechnie wykorzystywane jako medium transportujące dane kontrolne, diagnostyczne oraz głos. Powstawały jednak zakłócenia generowane przez transmisję, sieć podatna była na złą pogodę oraz inne szumy.

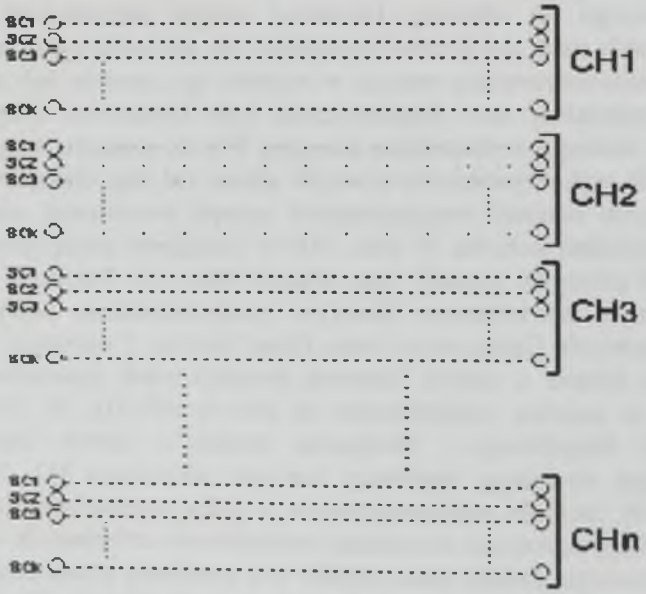
2. Realizacja sieci PLC

Stosowane są dwie metody modulacji i multipleksacji dla sieci PLC [3]:

- CDM – Code Division Multiplexing
- OFDM – Orthogonal Frequency Division Multiplexing

CDM zostało scharakteryzowane jako niewrażliwe na wąskopasmowe zakłócenia i wybiórcze tłumienie, działa przy niskiej mocy sygnału, który jest istotny z powodu zjawiska promieniowania. OFDM zostało zaprojektowane jako najlepszy kandydat do zastosowania w systemach transmisji PLC z większymi prędkościami danych, ponieważ jego wydajność szerokości pasma jest wysoka. OFDM transmituje dane za pomocą licznych podnośnych, które stwarzają możliwość odchylenia od krytycznych częstotliwości używanych przez inne systemy komunikacji).

Systemy transmisji OFDM używają licznych podnośnych (SC) rozłożonych w spektrum częstotliwości (rysunek 1). Każda podnośna posiada przepustowość transmisyjną i możliwe jest stworzenie grupy podnośnych by powiększyć kanał transmisyjny (CH) z wyższą przepustowością.



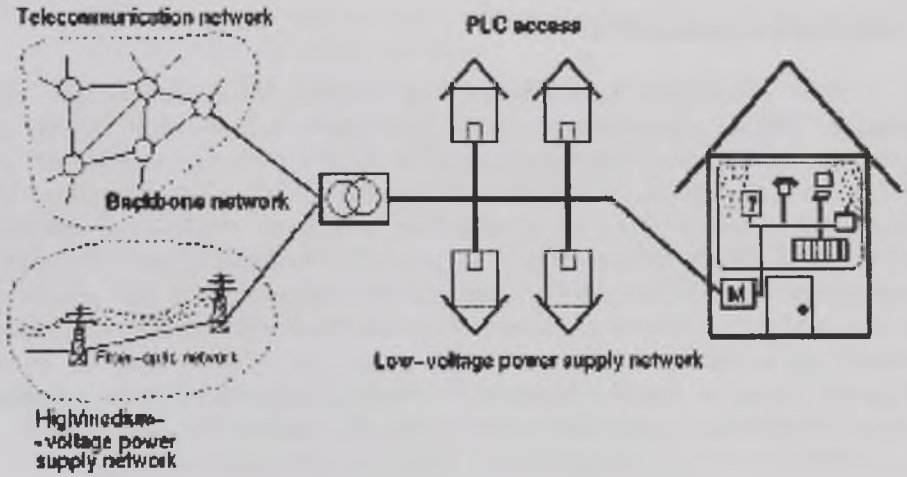
Rys. 1. Struktura kanału OFDM, , źródło: [4]

Istnieje możliwość zarządzania podnośnymi OFDM by uniknąć krytycznych widm częstotliwości.

Z powodu bardzo gęstego pokrycia spektrum częstotliwości przez usługi radiowe, moc sygnału w systemach PLC musi pozostać w niskim zakresie [4].

Niski poziom mocy sygnału w systemach PLC przysparza na dłuższych dystansach problemów w transmisji. Może to być rozwiązane przez zastosowanie repeaterów w sieci PLC. Repeatery wzmacniają moc sygnału do poziomu, który jest wystarczający do utrzymania możliwej transmisji na odcinku sieci. Z drugiej strony, długości odcinków sieci powinny być tak dobrane, aby transmisja była możliwa z taką mocą sygnału, która jest niższa od wartości granicznej. Jednakże, zastosowanie repeaterów w sieciach PLC zwiększa ich koszty.

Zwykle sieć dostępowa PLC jest podłączona do swojej sieci szkieletowej przez stację bazową (rysunek 2) [5]. Oznacza to, że cała komunikacja między użytkownikami sieci PLC a światem jest przeprowadzana przez stację bazową. Zakładamy również, że wewnętrzna komunikacja między użytkownikami sieci PLC jest dokonana przez stację bazową.



Rys. 2. Sieć dostępowa PLC, źródło: [5]

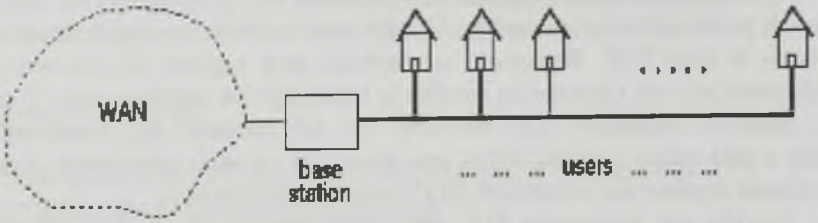
Istnieją dwa kierunki transmisji w sieci PLC:

- „Downlink"/„downstream" od stacji bazowej do użytkowników
- Uplink"/„upstream" od użytkowników do stacji bazowej

Jeśli założymy, że stacja bazowa jest umiejscowiona w stacji transformatorowej (to również obowiązuje jeśli stacja bazowa jest usytuowana w każdej innej stacji w sieci) dostrzegamy następujące rodzaje transmisji:

- Informacja wysłana przez stację bazową w kierunku „downlink" jest transmitowana do wszystkich sieciowych podsekcji i jest odbierana przez wszystkich użytkowników w sieci

- W kierunku „uplink", informacja wysłana przez użytkownika jest transmitowana nie tylko do stacji bazowej ale również do wszystkich użytkowników w sieci

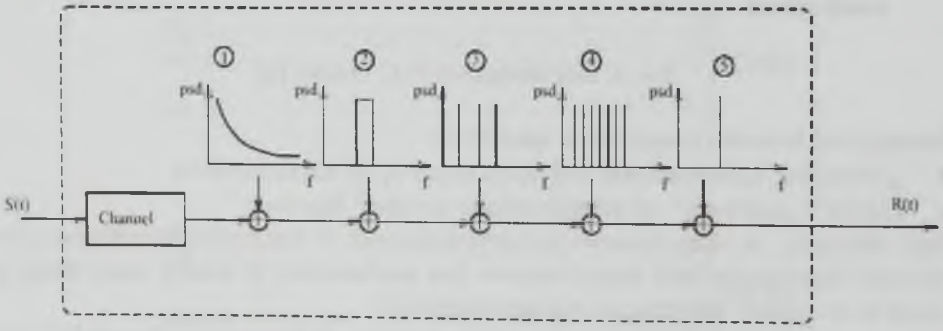


Rys. 3. Model sieci PLC, źródło: [5]

Oznacza to, że transmisyjny nośnik/kabel PLC posiada strukturę magistrali pomimo faktu, że niskonapięciowe sieci zasilające mają topologię drzewa.

3. Zakłócenia w sieciach PLC

Aby efektywnie komunikować się wykorzystując rozdzielcze sieci zasilające, trzeba przezwyciężyć wiele problemów technicznych, takich jak niestabilne charakterystyki transmisji, bardzo niska impedancja kanału, itd. [6]. Wśród tych technicznych spraw, jedną z najbardziej istotnych jest projektowanie systemu komunikacyjnego z uwzględnieniem unikalnych właściwości zakłóceń. Zakłócenia na liniach energetycznych są głównie powodowane przez elektryczne urządzenia podłączone do tych linii. Statystyczne zachowanie się tego szumu jest całkowicie różne od ustalonego białego szumu Gaussa a jego charakterystyki mogą zmieniać się w bardzo krótkich okresach czasu [6]. Według różnych źródeł, addytywny szum w szerokopasmowych kanałach komunikacyjnych powerline może być uważany za sumę pięciu typów szumu, jak pokazano to na rysunku 4.



Rys. 4. Addytywny szum w środowisku Powerline, źródło: [7]

Opisy poszczególnych typów zakłóceń:

- Typ 1 – barwny szum tła, głównie powodowany przez sumowanie licznych źródeł zakłóceń o małej mocy. Jego poziom zmienia się nieznacznie w przeciągu np. godzin.
- Typ 2 – szum wąskopasmowy: jego nazwa oznacza, że szum ten jest ograniczony do wąskopasmowej części pasma częstotliwości, przy której poziom jest w przybliżeniu stały. Ten typ szumu jest głównie powodowany przez wnikanie stacji radiowych do zakresów transmisji fal średnich i krótkich. Poziom ten ogólnie zmienia się w zależności od pory dnia (wysoki w porze wieczornej, znacznie niższy podczas godzin dziennych).
- Typ 3 – okresowy szum impulsowy, (asynchroniczny wobec częstotliwości sieci elektrycznej), bardzo często ten typ szumu jest powodowany przez przełączenia urządzeń zasilających. Te impulsy mają w większości przypadków częstotliwość powtarzania między 50 kHz do 200 kHz, co daje w rezultacie widmo dyskretnych linii, których odstęp częstotliwości jest dyktowany przez częstotliwość powtarzania.
- Typ 4 – okresowy szum impulsowy, (synchroniczny wobec częstotliwości sieci elektrycznej): te impulsy mają częstotliwość powtarzania 50 Hz lub 100 Hz i są synchroniczne wobec okresu sieci zasilającej. Mają one krótkie czasy trwania (kilka mikrosekund)
- Typ 5 – asynchroniczny szum impulsowy: ten typ szumu impulsowego jest powodowany przez stany przejściowe w sieci. Te impulsy mają czas trwania od mikrosekund do kilku milisekund i są losowe

W celu przeprowadzenia bardziej wyraźnych badań zakłóceń, ogólna klasyfikacja tych zakłóceń może być dokonana na podstawie ich zachowania w czasie. Pomiary pokazują, że te pięć typów może być sklasyfikowanych w dwie kategorie: szum tła włączając w to pierwsze trzy typy i szum impulsowy, który reprezentuje dwa ostanie typy.

4. Zagrożenia ze strony sieci PLC

Najszerzej rozpowszechnione są miedziane sieci przewodowe rozprowadzające energię elektryczną i brane są coraz częściej pod uwagę jako środek dla telekomunikacji cyfrowej. Sygnały cyfrowe były nakładane na przewody zasilające 50 Hz od dawna, na przykład do celów zdalnego odczytu liczników energii u odbiorców. Sygnały te mają małą częstotliwość i pojawiają się w bardzo ograniczonym czasie. Od pewnego czasu, te same sieci zasilające są używane do przesyłania szerokopasmowych (0,5 MHz – 30 MHz) sygnałów cyfrowych w technologii PLC. Technologia PLC ma również wady, które mogą znacznie ograniczyć dalszy rozwój tej usługi. Chodzi tu przede wszystkim o możliwe niepożądane promieniowanie, towarzyszące przesyłowi danych. Sieć energetyczna została zaprojektowana do transmisji na poziomie 50 Hz, podczas gdy PLC wykorzystuje wartości wielokrotnie większe, rzędu 30 MHz. Może to spowodować, że sieć zacznie działać jak antena promieniująca.

Jeśli sygnały wielkiej częstotliwości są doprowadzone do przewodów elektrycznych, występuje zjawisko promieniowania elektromagnetycznego [8]. Jednakże promieniowanie to może zostać ograniczone przez:

- ekranowanie przewodów (przewody współosiowe)
- stosowanie linii zrównoważonych, z zasady swej nie promieniujących (linie antenowe, skręcone przewody telefoniczne)

Z drugiej strony, istnieje szereg czynników sprzyjających promieniowaniu, jak:

- niedostateczne ekranowanie
- niedopasowanie impedancji linii i jej zakończenia
- nie zrównoważenie linii otwartych

Najgorszym przypadkiem promieniowania jest jednak technologia PLC. Przewody energetyczne ani nie są ekranowane, ani nie są zrównoważone. Co więcej, jako zakończenia sieci stosowane są przez użytkowników końcowych najróżniejsze urządzenia. Wynikiem tego są wysokie poziomy promieniowania wielkiej częstotliwości. Systemy PLC oraz tyrystorowe sterowanie układów napędowych generują harmoniczne wchodzące w zakres fal długich [10]. Ponieważ transmisja odbywa się po przewodach, nie podlega ona bezpośrednio przepisom o emisjach radiowych (fal elektromagnetycznych), a jedynie przepisom o zakłóceniach wywoływanych przez urządzenia elektryczne takie jak silniki lub wyłączniki. Nie uwzględniają one istnienia promieniowania elektromagnetycznego z nie ekranowanych linii przesyłowych. Jak długo odbywało to się na m.cz., nie budziło większych protestów, z chwilą wejścia w zakres w.cz., wywołało to liczne, słuszne protesty. Sygnały PLC mogą powodować zakłócenia w działaniu urządzeń domowych takich jak odbiorniki radiowe, sprzęt wideo itp. [11]. Innym problemem jest zakłócanie częstotliwości służb radia krótkofalowego i fal krótkich [9]. Istnieje również problem bezpieczeństwa przesyłanych danych – nie ma na razie pewności, czy informacje przekazywane tą drogą nie będą narażone na podsłuch lub przechwycenie. W chwili obecnej trwają dyskusje pomiędzy aktualnymi użytkownikami radiokomunikacji krótkofalowej i użytkownikami PLC zamierzającymi wykorzystywać istniejące sieci kablowe do transmisji danych w paśmie HF. Wiadomo już, że przeciwko wprowadzeniu technologii PLC zdecydowanie protestują profesjonalne i amatorskie służby radiowe. Nie są temu również chętne siły zbrojne, dla których radiokomunikacja krótkofalowa jest podstawowym środkiem łączności podczas operacji militarnych.

5. Zainteresowanie technologią PLC ze strony spółek dystrybucyjnych

Próby wprowadzenia technologii PLC w Wielkiej Brytanii podjęto już w roku 2000. Testy przeprowadzone w sieciach energetycznych oraz pomiary poziomu zakłóceń wykonane przez właściwe służby państwowe (przy współpracy związku krótkofalowców) wykazały znaczny wzrost zakłóceń i szumów elektromagnetycznych (nawet o 20dB). Odpowiednik polskiego URTiP zakazał wprowadzenia tej techniki do użytku.

W Niemczech koncern RWE miał przygotowany dostęp do Internetu za pośrednictwem linii elektroenergetycznych. W zamierzeniach tej firmy miała to być konkurencja dla oferowanego przez Deutsche Telekom ADSL. W Niemczech drogę dla PLC otworzyła złagodzona norma NB30, która dopuściła większy poziom interferencji z linii energetycznych. Po kilkunastu miesiącach koncern RWE (pionier instalacji PLC w Europie) ogłosił wycofanie się z dalszych badań i rozpowszechniania technologii. Decyzję uzasadniono trudnościami technicznymi i szybkim rozwojem innych technik telekomunikacji.

Stosunek Ameryki Północnej i Kanady do PLC szerokopasmowej, można wnioskować z postawy Nortela, który nie zakwestionował samej technologii, lecz jej nikłe efekty ekonomiczne. Natomiast dostęp xDSL liczy się tam w milionach użytkowników.

Nieco inaczej wygląda sytuacja we Francji. Energetyczny monopolista EDF (Electricite de France) oficjalnie stwierdza, że dla niego jest nieco za wcześnie na oferowanie takich usług jak PLC szerokopasmowa, czyli Internet w kablu elektroenergetycznym. Wprawdzie jego statut uniemożliwia mu świadczenie innych usług niż energetyczne. Przeprowadzono już pewne testy we francuskich liceach, zbliżone do eksperymentów Norweba i Nortela w Anglii (Manchester). Rynek francuski jest stosunkowo duży i należy się spodziewać, że energetyczny gigant pewnie nie zrezygnuje bez walki o zyski, jakie może ewentualnie przynieść PLC. Dość szybko technologia PLC zdobyła duży rozgłos również w naszym kraju [9]. Na polskim rynku działają obecnie dwaj dostawcy technologii PLC – szwajcarski Ascom oraz polsko-amerykański Pattern Communications, który dostarcza sprzęt izraelskiej firmy Main.net.

W Polsce testowanie PLC szerokopasmowej miało miejsce m.in. w Wilanowie przez warszawski STOEN przy użyciu sprzętu firmy ASCOM, w Nowej Hucie przez ZE Kraków przy użyciu sprzętu Main.net oraz na jednym z osiedli w Płocku przez ZE Płock przy użyciu sprzętu firm ASCOM i Main.net.

Przepływności oferowane przez dostępne systemy DPL/PLC na tle innych rozwiązań (tablica 1) wyglądają obiecująco, pod warunkiem, iż uwzględni się brak konieczności tworzenia całkowicie nowej infrastruktury [12]. Jeżeli pominie się tą ważną zaletę proponowanych rozwiązań, to atrakcyjność PLC maleje. Oferowane przepływności stanowią jedynie konkurencję dla dostępu za pośrednictwem publicznej sieci telefonicznej.

Na podstawie wprowadzonych w Polsce systemów PLC można ocenić najważniejszy dla użytkownika wskaźnik, tzn. cenę 1 MB. W zależności od wykupionego pakietu usług zróżnicowana jest ilość danych, które możemy ściągnąć w ramach abonamentu. W pakiecie najtańszym oferowanym przez Pattern Communications wynoszącym 85,60 zł jest to 250MB, za każdy następny MB zapłacimy 0,214 PLN. Dlatego szacunki względem 1 MB nie są miarodajne.

Tablica 1 Porównanie przepływności różnych technik dostępu do Internetu, źródło: [12]

Parametr	ISDN	PSTN	CATV	ADSL	DPL 1000 firmy NorWeb	E-PLC SD V1 firmy Ascom	PLNet 2001 Inst. Łączności
Przepływność binarna systemu DPL/PLC [kbit/s]	128	56	38000	8192	1280	1500 lub 4500	1,024
Przepływność przydzielana abonentowi [kbit/s]	128	56	64-2048	64-8192	32-1280	15-1500/45-4500	64-960

Cenniki przedstawione w poniższej tablicy pochodzą z drugiego kwartału 2003 roku i uwzględniają należny podatek VAT. Ceny w PLN.

Tablica 2. Porównanie cen dostępu do Internetu poprzez PLC oraz komutowane łącze TP S.A., źródło: [12,13]

	Pattern Comunications			TPS.A.	STOEN Info Sp. z o.o.
	Pakiet srebrny	Pakiet złoty	Pakiet platynowy		
Instalacja i aktywacja	244*)	244*)	244*)	nd	1220**)
Kaucja zwrotna za modem	120,78*)	120,78*)	120,78*)	nd	nd
Cena abonamentu	85,60*)	128,40*)	212,93*)	64,20	74,90**)
Ilość danych dostępna do ściągnięcia w abonamencie	250	750	nd	0	nd
Cena dodatkowych MB [zł]	0,214	0,171	nd	???	nd
Dla 250 MB	85,6	128.4	nd	38,5	nd
Dla 750 MB	192,6	128,4	nd	74,9	nd
Dla 1 GB	246,1	171,2	nd	93,6	nd
Dla 1.5 GB	353,1	256,8	nd	131,1	nd
Dla 3 GB	674,1	513,6	nd	242,3	nd

Legenda: *nd – nie dotyczy, bd – brak danych, dla TP S.A. wybrano pakiet „30 godzin"*, *) przy minimalnej liczbie 10 chętnych użytkowników, **) przy minimalnej liczbie 30 chętnych użytkowników

Należy pamiętać również o czasie pobierania, który dla 1,5 GB wynosi prawie 60 godzin dla dostępu konwencjonalnego i 13 godzin dla PLC (przy gwarantowanych 256 kbit/s). Poza tym w trakcie operacji konwencjonalnej telefon pozostaje cały czas zajęty, może dojść do zerwania połączenia, założono także najbardziej optymistyczny przypadek, gdy transmisja stale sięga 56 kbit/s.

6. Wnioski

Usługą PLC są zainteresowani przede wszystkim dostawcy usług energetycznych z dużych miast [9]. Tam bowiem PLC może być dochodowe. Według specjalistów z ZE Warszawa-Teren SA ze stacji bazowej musi korzystać minimum 30 użytkowników – inaczej inwestycja będzie nieopłacalna. Jak zapewnić taką liczbę chętnych w małych miejscowościach? Energetycy z zakładów energetycznych w Gorzowie i Jeleniej Górze stwierdzili wręcz, że wprowadzenie technologii PLC na tamtejszym obszarze jest niecelowe – „rozrzedzona" infrastruktura i brak dużych jednostek transformatorowych bardzo podniosłyby koszty wdrożenia nowej technologii, co odbiłoby się negatywnie na późniejszym cenniku.

Generalnie z problemami natury organizacyjno-prawnej stykają się też te zakłady, które wdrażają PLC. Wg Stefana Wieczorka z Energetyki Poznańskiej SA przeszkodą są na przykład uregulowania dotyczące montażu urządzeń w stacji transformatorowej. Zgodnie z przepisami mogą to uczynić tylko upoważnione osoby z zakładu energetycznego, w związku z czym nie za bardzo wiadomo, jak umożliwić dostęp do tych pomieszczeń pracownikom firm wdrażających technologię PLC. Michał Pączko z Lubzelu SA informuje, że nie wiadomo jeszcze, jaką metodą (radiową czy za pomocą kabla) dostarczony zostanie sygnał do stacji transformatorowej. Przedstawiciele wszystkich uczestniczących w projekcie zakładów energetycznych podkreślają jednak, że zarówno Ascom, jak i Pattern Communications zapewniają kompleksową obsługę i pomoc przy wdrażaniu PLC w danym okręgu.

Dla samych zakładów atrakcyjniejsze jest skupienie się na dostawie energii elektrycznej i wykorzystanie technologii do np. zdalnych odczytów liczników [12]. Mniej interesuje ich zapewnianie swoim abonentom dodatkowej usługi, jaką jest transmisja danych. Poza aspektami technicznymi pod wielkim znakiem zapytania stoi opłacalność tej technologii. Może okazać się, że pieniądze pochodzące z abonamentu nie będą w stanie zamortyzować poniesionych kosztów instalacji, ponieważ nikt nie jest w stanie przewidzieć liczby użytkowników.

Można być pewnym, że obecnie nie nastąpi gwałtowny rozwój technik dostępowych do Internetu dla użytkowników indywidualnych poprzez PLC.

Aczkolwiek z drugiej strony według np. ekspertów ABI (Allied Buisnes Intelligences) w najbliższym pięcioleciu globalny rynek urządzeń PLC osiągnie wielkość 2,5 mld $ przy rocznej dynamice wzrostu 48%.

Literatura

1. Marcin Pawlak; „Sieci pod napięciem"; Chip-CD 10/2001
2. Marcin Nowak; „Niech moc będzie z wami"; http://www.chip.pl/archiwum/article_9290html
3. Halid Hrasnica, Ralf Lehnert; "Powerline Communications in Telecommunication Access Area"; http://www.ifn.et.tu-dresden.de/~hrasnica/research/plc1.pdf
4. Halid Hrasnica, Ralf Lehnert; "Extended ALOHA and Hybrid-Polling Reservation MAC Protocols for Broadband Powerline Communications Access Networks"; http://www.ifn.et.tu-dresden.de/~hrasnica/research/plc18.pdf
5. Halid Hrasnica, Ralf Lehnert; "Powerline communications for access networks – Performance Study of the MAC Layer"; http://www.ifn.et.tu-dresden.de/~hrasnica/research/plc2.pdf
6. David Hines, John Dickinson, Peter Nicholson, Ralf Lehnert, Halid Hrasnica, Abdelfatteh Haidine, Mariana Stantcheva, Marco Langhof, Uwe Leicht; "PALAS – Powerline as an Alternative Local AccesS IST-1999-11379; D2: PLC Technology Inventory and Development Roadmap; June 2000"; http://palas.regiocom.net/intern/upload/hines/D2.doc
7. Halid Hrasnica, Abdelfatteh Haidine; "Modeling MAC Layer for Powerline Communications Networks"; http://www.ifn.et.tu-dresden.de/~hrasnica/research/plc10.pdf
8. Gaston Bertels; EUROCOM Newsletter, 02.02.2003, „Power Line Technology : HF bands under fire"; tłumaczenie: Vy 73! Piotr SP2JMR, Prezes PZK; „PLC w ofensywie"; http://www.pzk.org.pl/modules.php?op=modload&name=News&file=article&sid=186
9. Marek Staniewicz; „Sieć z elektrowni"; http://www.chip.pl/arts/n/article_23320.html
10. Zdzisław Bieńkowski; „Co to jest PLC i jakie niesie zagrożenia"; http://www.pzk.org.pl/komunika/2002/PLCoto.htm
11. „Krytyczna ocena studiów dotyczących konsekwencji transmisji/komunikacji z wykorzystaniem energetycznych linii zasilających (Power Line Transmission/Communication – PLT/PLC)"; Dokument przygotowany przez TNO-FEL na posiedzenie SE35 w grudniu 2001 w Helsinkach, na zlecenie użytkowników widma częstotliwości KF w Holandii (NARFA-NL, Radio Netherlands, NOZEMA, VERON, VRZA); tłumaczenie: Krzysztof Słomczyński MON-WBZC
12. Przemysław Pawełczak, Marcin Suszkiewicz; „Przyszłość PLC w Polsce"; http://zstux.ita.pwr.wroc.pl/iref-2002/17.pdf
13. Kardaś Michał, „Technologia PLC w praktyce na przykładzie usługi dostępnej w spółce dystrybucyjnej STOEN S.A.", praca dyplomowa magisterska, Politechnika Warszawska, Wydział Elektryczny, Warszawa 2003

dr inż. Rafik Nafkha
STOEN S.A.
ul. Wybrzeże Kościuszkowskie 41,
00-347 Warszawa
e-mail: Rafik.Nafkha@stoen.pl

dr inż. Paweł Piotrowski
Politechnika Warszawska, Instytut Elektroenergetyki
ul. Koszykowa 75,
00-662 Warszawa
e-mail: pawel.piotrowski@ien.pw.edu.pl

mgr inż. Michał Kardaś
Absolwent Wydziału Elektrycznego Politechniki Warszawskiej
ul. Nicejska 1 m.67
02-763 Warszawa
e-mail: mkardas@ee.pw.edu.pl

WDRAŻANIE SYSTEMÓW ERP – PORADNIK PREZESA I NIE TYLKO CZYLI GRZECHY GŁÓWNE I PRZYKAZANIA

Andrzej SAKOWSKI

Istnieje co najmniej 100 powodów by twierdzić, że wdrażanie pełnego systemu ERP jest przeprawą przez nieznane wody dla przedsiębiorstwa w którym następuje wdrożenie i poważną chorobą dla niedoświadczonych konsultantów. Często są ofiary po obu stronach.

O ile dla project managera i konsultantów z firmy wspomagającej wdrożenie mieści się to w ramach ryzyka zawodowego i przeważnie zdają sobie sprawę z zagrożeń, to firma w której będzie miało miejsce wdrożenie zwykle nie wie co ją czeka.

Ci, u których wdrożenie zakończyło się sukcesem (i mogą po roku od zakończenia wdrożenia twierdzić, że osiągnęli cele i jeszcze rozszerzyli zakres), są słusznie ogromnie dumni z tego osiągnięcia i odważnie patrzą w przyszłość. Jeśli przedsiębiorstwo utrzyma ten sam zespół ludzi, którzy współdziałając we wdrażaniu stał się prawdziwym team'em, nie straszne mu jest żadne nowe wyzwanie. Przedsiębiorstwo zyskało dużo więcej niż funkcjonujący system, a każdy z tych ludzi, którzy działając aktywnie doczekali sukcesu jest wart 2 razy więcej niż przed rozpoczęciem tego projektu (i to na każdym rynku !).

Ta wspomniana przeprawa jest zarządzalna i warto zadbać by była możliwie bezpieczna i krótka. Trzeba więc skorzystać z istniejących doświadczeń, przygotować ją i wykonać zgodnie z zasadami i uniknąć błędów. Statystyka mówi, że dużo więcej błędów popełnia firma w której następuje wdrożenie co jest zrozumiałe, ponieważ robi to po raz pierwszy. Poza tym to właśnie ta firma ponosi prawie całe ryzyko niepowodzenia przedsięwzięcia. Dlatego główną uwagę skupimy na tej stronie, chociaż ten wóz ciągną 2 zaprzęgi.

Nazwijmy firmę u której następuje wdrożenie ERP – BENEFICJENTem, a zewnętrzną firmę wspomagającą wdrożenie konsultantami –KONSULTANTem.

Rozpocznijmy od zasad prawidłowego zarządzania firmą z poziomu zarządzania strategicznego. Wrażanie ERP z całą pewnością jest decyzją strategiczną. Na razie nie wiemy jeszcze czy system ERP będzie nam potrzebny i czy już w najbliższym czasie i czego od tego systemu będziemy oczekiwać. Winno to wynikać z cyklicznego działania Zarządu, przedstawionego poniżej (co najmniej raz w roku).

Model działań:

1. Analiza wyników działania firmy, weryfikacja prawidłowości przyjętej misji i strategii firmy BENEFICIENTa. Wytyczenie celów do osiągnięcia.
2. Opracowanie/weryfikacja koncepcji rynkowej i organizacji funkcjonowania firmy BENEFICIENTa, systemu pomiaru oraz programu poprawy i utrzymania efektywności firmy, lista procesów w firmie BENEFICIENTa,

zdefiniowanie metod organizacyjnych, zdefiniowanie podstawowych procesów

3. Plan wdrożenia całego systemu poprawy/utrzymania efektywności firmy. (Dopiero z tego (pkt. 2 i 3) wynikają wymogi co systemu i kryteria doboru systemu wspomagającego zarządzanie (ERP), kolejność i terminarz działań wdrożeniowych (podział na podprojekty, terminy itp.), wybór KONSULTANTa.)
4. Proces wdrożeniowy (ERP + ewentualnie inne przedsięwzięcia)
5. Ocena rezultatów i plan dalszych działań usprawniających działanie firmy
6. Realizacja innych elementów planu poprawy i utrzymania efektywności

Jeśli z w/w działań Zarząd zdecydował wstępnie o wdrożeniu ERP to do wykonania są niżej wymienione fazy prac, w trakcie których okaże się czy decyzja wstępna będzie podtrzymana czy nie (okaże się to w fazie „A" –omówionej dalej). W każdym razie proponowana kolejność buduje solidne uzasadnienie decyzji zarówno na TAK jak i NIE oznaczającej przerwanie dalszych prac.

Fazy wdrażania:

A. Przygotowanie
B. Realizacja procesu wdrożeniowego
C. Ustanowienie mechanizmu dalszego rozwoju usprawniania procesów

O ile faza A i B będą określone w czasie, to faza C jest działaniem ciągłym, które nie kończy się nigdy. W ramach fazy C realizowane będą cyklicznie mini-fazy A i B.

Konsultanci wdrożeniowi będą niezbędni przy pierwszym cyklu ABC, a w następnych A1,B1 będą odgrywać dużo mniejszą rolę, w kolejnych jeszcze mniejszą, choć prawdopodobnie nie da się z nich całkiem zrezygnować. Jest to więc partnerstwo długofalowe (z czego trzeba sobie zdawać sprawę i dlatego trzeba dokonać mądrego wyboru partnera).
Omówmy teraz faza po fazie najistotniejsze grzechy i przykazania.

Faza A – Przygotowanie

Obejmuje ona okres od powstania idei zastosowania ERP do momentu rozpoczęcia procesu wdrożeniowego. Jest to faza w której BENEFICJENT i KONSULTANT popełniają najpoważniejsze grzechy, które wręcz decydują o powodzeniu lub klęsce dalszych działań.
Obie strony nie doceniają jej znaczenia –a jest to faza NAJWAŻNIEJSZA.

Grzechy:

- decyzja o wdrożeniu ERP podjęta z braku koncepcji innych działań -nie wpisuje się w strategię BENEFICJENTA a czasem ma zastąpić strategię
- BENEFICJENT nie zdaje sobie sprawy z rozległości, znaczenia projektu wdrożenia ERP, czasu trwania i kosztów.
- nie określono celów biznesowych wdrożenia
- nie rozpoczęto od analizy i poprawy procesów
- nie określono zakresu i pracochłonności prac przygotowawczych i towarzyszących
- nie stworzono projektu przygotowującego do wdrożenia (zasoby, środki, nadzór)
    - nie zapewniono zasobów realizacyjnych
    - nie zbudowano realistycznego ramowego planu całego przedsięwzięcia
- nie przeprowadzono analizy opłacalności przedsięwzięcia
- źle wybrano system ERP i firmę KONSULTANTA, metodę wdrożenia
- nie uzgodniono harmonogramu wdrożenia i podziału zadań i odpowiedzialności stron przed kontraktem
- przyjęto niejasne sformułowania kontraktowe (w tym zakres dokumentacji, szkoleń, forma, miejsce, koszty dojazdu i pobytu)
- nie stworzono i nie przygotowano zespołu wdrożeniowego BENEFICJENTa
- nie określono zasad sterowania projektem wdrożeniowym przez obie strony
- nie zapewniono czynnego udziału zarządu BENEFICJENTA w nadzorowaniu projektu i podejmowaniu skutecznych i szybkich interwencji (duże uprawnienia)
- nie zakończono prac przygotowawczych przed decyzją rozpoczęcia projektu i nie sprawdzono tego
- grzechy KONSULTANTA polegają to brak cierpliwości i nadmiernym naciskaniu na potencjalnego BENEFICJENTa by jak najszybciej podpisać kontrakt (martwić się będziemy później). Z tego wynikają i nieprecyzyjne sformułowania w kontrakcie i niedopowiedzenia, czasem zdarzają się obiecanki bez pokrycia. Nacisk ten jest zrozumiały przy obecnej konkurencji na rynku polskim i braku ogłoszenia przejrzystych reguł postępowania selekcyjnego ze strony BENEFICJENTów. BENEFICIENT liczy na to, że „przy okazji" wdrożenia ERP niejako za darmo dostanie poprawę procesów, bezpłatny program usprawnień. KONSULTANT twierdzi, że będą efekty z wdrożenia ERP więc BENEFICIENT liczy na to, że samo wdrożenie już wystarczająco usprawni jego działalność. Nie czuje presji na inne dodatkowe działania. Kontrakt więc zawierany jest często z niedopowiedzeniami. Strony różnie rozumieją zakres, ale mimo braku jednoznaczności, uzgadniają jakąś cenę licząc na swoją siłę przekonywania. W trakcie wdrożenia dochodzi do sporów, wyjaśniania, przekraczania kosztów i terminów i często obie strony są rozczarowane.

Przykazania:

- Negocjuj, wyjaśniaj PRZED, nigdy W TRAKCIE projektu, a narzekanie PO jest już jawnym przyznaniem się do porażki. Wykorzystaj czas w fazie „A" by wymienić poglądy, wydyskutować cele i kierunki rozwiązań, ustalić z partnerami wszelkie warunki współpracy, podzielić się wątpliwościami i je wyjaśnić. W trakcie projektu jest to zbyt kosztowne i psuje współpracę.
- podejmując decyzję o wdrożeniu ERP musisz zdawać sobie sprawę z rodzaju tego przedsięwzięcia. Musi to wynikać i być częścią długofalowej strategii firmy. Zakres systemu ERP obejmuje wspomożenie najistotniejszych dla firmy procesów gospodarczych decydujących o jej bycie. Ponieważ wdrożenie będzie trwało jakiś czas, zamierzamy systemem wspierać procesy docelowe (określone strategią), a nie te wg obecnego kształtu. Jeśli te nowe procesy zadecydują o bycie firmy to należy do ich opracowywania przywiązać najwyższą wagę i dać najlepszy personel, a zarząd winien je zatwierdzić i w ich opracowywaniu współdziałać.
- ustal cele biznesowe stawiane przed projektem, ustal kryteria sukcesu, miary
- przygotuj proces wdrożeniowy by nie było sytuacji „co kartofel to decyzja". Im dłużej proces wdrożeniowy trwa tym więcej kosztuje. Uprzednie przygotowanie personelu BENEFICIENTA i danych w firmie, kształtu procesów umożliwi możliwy płynny i szybki przebieg projektu. Ponieważ Murphy nigdy nie śpi to i tak coś „wyskoczy" ale im lepsze przygotowanie tym mniej wyskoków. Najgorsze jest prowadzenie w trakcie wdrożenia ERP rozległych projektów organizacyjnych w rodzaju zmiana indeksacji... Dlatego zaleca się wykonanie możliwie wcześnie „Implementation study", które określi wstępnie jakie korzyści BENEFICIENT może zyskać z wdrożenia ERP i określi listę niezbędnych prac przygotowawczych. Oszacuj pracochłonność i czas trwania tych prac. Teraz jest moment na pierwszą decyzję o kontynuacji lub zaniechaniu przygotowań. Dalszym krokiem będzie zdefiniowanie i przeprowadzenie projektu realizacji prac przygotowawczych, do czego wymagane już będzie zaangażowanie środków i prace potrwają jakiś czas. Korzyścią jest to, że mogą to zrobić ludzie BENEFICIENTA z niewielką pomocą doradcy zewnętrznego (stosunkowo nieduże koszty, mamy możliwość swobodnego czasu trwania) i nie jesteśmy pod presją KONSULTANTa. Udokumentowanie wszelkich ustaleń w pracach przygotowawczych będzie bezcenne w projekcie wdrożeniowym.
- nigdy nie powierzaj zarządzania projektem wdrożeniowym informatykom, ponieważ zawodowe zainteresowania ukierunkują wysiłki na stronę techniczną, a nie biznesową –więc koniecznie wciągnij zarząd

- przed projektem wdrożeniowym zrób analizę opłacalności np. ROI, a po projekcie zmierzysz rezultaty
- przygotuj kasę –wdrożenie ERP jest to kosztowne przedsięwzięcie - kosztuje i komputer i oprogramowanie i konsultanci i odrywasz twoich ludzi i (jeśli nie –przygotowałeś tego)wyskoczy wiele pilnych podprojektów organizacyjnych, porządkujących –bądź na to przygotowany.
- przed kontraktem ustal schemat wdrażania (strategię wdrażania, zakres, etapy, rezultaty..). Najważniejsze jest uzgodnienie strategii wdrożeniowej tzn czy chcemy wdrażać 2-stopniowo czy 1-stopniowo. W ramach 2-stopniowej strategii, pierwszy projekt jest zorientowany na użycie podstawowej funkcjonalności systemu ERP przy czym BENEFICIENT upraszcza swoje procesy dopasowując się do możliwości systemu, a użytkownicy ucząc się szybko dość standardowych funkcji systemu stają się dobrym partnerem do drugiego stopnia wdrożenia. Pierwszy stopień trwa stosunkowo krótko 3-4 miesiące, może kończyć się startem eksploatacji lub można poprzestać na ograniczonej próbce danych rzeczywistych. Drugi stopień polega na pogłębieniu wdrożenia i odwzorowaniu w pełni docelowych procesów w systemie, co jest już łatwiejsze ponieważ personel BENEFICIENTa rozumiejąc filozofię systemu już świadomie współdziała w rozszerzaniu możliwości poprzedniego rozwiązania. Model 2-stopniowy zwiększa czas wdrożenia (w porównaniu z 1-stopniowym) ale zmniejsza się ryzyko niepowodzenia i poprawia jakość rozwiązania końcowego. Drugim pod względem istotności elementem strategii wdrożeniowej jest decyzja czy zaczynamy od startu eksploatacji w całym przedsiębiorstwie jednocześnie, czy też obejmiemy najpierw jeden fragment, zaplanujemy stopniowe rozszerzenie, czy podejdziemy do wdrożenia funkcjami, stopniowo zwiększając ich ilość. Jest to również czynnik zmniejszania ryzyka projektu. Że wpływa to na plan wdrożenia w kontrakcie i koszty, nie trzeba nikomu tłumaczyć. Ale dlatego otwarcie trzeba to przedyskutować i wpisać w kontrakt.
- nie rozpraszaj środków i zasobów na inne działania równoległe (zwykle występuje problem krótkiej ławki –w firmie często jest tylko mała ilość osób zdolnych do samodzielnego, twórczego myślenia). Wytypuj dodatkowo młode, obiecujące osoby -wschodzące gwiazdy- i daj im możliwości działania. Część się sprawdzi, w ten sposób wydłużysz ławkę i będą osoby zdolne do kontynuacji w fazie „C".
- dobrze wybierz system ERP i firmę wspomagającą wdrożenie. Ważka i trudna decyzja, ale rozsądnie przeprowadzony proces wyłaniania partnera (a nie dostawcy!) przy współpracy doradcy winien dać wystarczające przesłanki wyboru, przy zachowaniu właściwej kolejności działań. Nie należy się przy tym spodziewać, że to doradca wybierze.
- nie wstydź się swojej niewiedzy –weź niezależnego doświadczonego doradcę. Zwykle podejmujesz taki projekt po raz pierwszy, więc nie masz doświadczeń. Koszt jego pracy będzie nikły w porównaniu z kosztami

błędów, których dzięki niemu unikniesz. Zwróć uwagę na małe firmy doradcze - one nie oferują systemów ERP, ani ich wdrażania - więc nie będzie ukrywanego konfliktu interesów i będzie się liczył tylko interes BENEFICIENTa.

- przed decyzją o podpisaniu kontraktu, kiedy znasz już wszystkie uwarunkowania wróć raz jeszcze do strategii firmy i upewnij się:
  - czy planowane i uszczegółowione (w wyniku fazy A-Przygotowanie) efekty wdrożenia i zaplanowany proces wdrożeniowy służy realizacji strategii ?
  - czy masz zaplanowane wystarczające zasoby realizacyjne ?
  - czy harmonogram uzyskiwania efektów współgra z innymi elementami strategii i czy są przewidziane rezerwy na zdarzenia nie przewidziane ?
  - jakie są zagrożenia dla powodzenia projektu i jak jeszcze można temu przeciwdziałać?
  - jakie byłyby skutki nie osiągnięcia sukcesu ?

- Reasumując: zastosuj poniższą sekwencję działań przygotowawczych fazy „A", która gwarantuje poprawność i skuteczność przygotowań ponieważ poprzednia czynność przygotowuje materiał niezbędny do czynności następnych:

1. znajdź doradcę, opracujcie razem schemat prac przygotowawczych
2. zleć wykonanie „Implementation study"
3. zleć doradcy przeprowadzenie warsztatów (dla Zarządu i grupy zajmującej się strategią) „ustalenie celów biznesowych projektu i kształtu procesów strategicznych"
4. dokonaj uszczegółowienia kształtu procesów do poziomu czynności z pełną listą wariantów
5. wykonaj definicję projektu przygotowania organizacyjnego firmy do wdrożenia ERP (w tym oszacuj pracochłonność, potrzebne zasoby, możliwe terminy realizacji i koszty- uwzględnić szkolenia w nowoczesnych technikach zarządzania)
6. oszacuj wstępnie z doradcą koszty i wynikający ze strategii terminarz wdrożenia ERP
7. przeprowadź analizę opłacalności inwestycji (jakąkolwiek wybraną metodą) biorąc pod uwagę dotychczasowe szacunki
8. podejmij decyzję czy kontynuować przygotowania
9. wytypuj osoby biorące udział w dalszych pracach i szkoleniach, określ ich zadania i zrealizuj szkolenia przygotowujące ich do ról
10. opracuj materiały dla oferentów (w tym strategią wdrażania, zakres, procesy i pożądany terminarz działań, szereg pytań ankietowych, oczekiwana zawartość oferty, kryteria oceny) i zaproś ich współpracy
11. dokonaj wstępnej selekcji oferentów (zawężenie do 3-4) przez wytypowany zespół

12. sformułuj jednakowy przykład do pokazów dla pozostałych i ramowy plan działań w którym nacisk położony jest na zapewnienie porównywalności ofert, ale pozostawiona jest część (max 1/5) dla oferenta na dowolne uzupełnienia
13. przeprowadź prezentacje działania systemu na przykładzie z ewentualnie dodatkowymi wyjaśnieniami i dokonaj z zespołem wyboru 2 oferentów do negocjacji końcowych
14. Dokonaj końcowych dodefiniowań i ustaleń, co do planu wdrożenia, obowiązków stron, zobowiązań, sposobu zarządzania projektem i zadbaj o jasność sformułowań i zgodność z planem prac przygotowawczych.
15. Po analizie dokonaj porównania rezultatów uzgodnień ze strategią firmy i ostatecznie oceń celowość i kierunki działań i biorąc pod uwagę fakt, że de facto wybierasz partnera (a nie dostawcę) na długie lata podpisz kontrakt lepiej dopasowany do tej strategii.

Po takim przygotowaniu, koszty, tempo i efekty wdrożenia będą zdecydowanie korzystniejsze dla BENEFICIENTA. BENEFICIENT ma szansę wybrać system tańszy stosowny do potrzeb, nie kupuje niepotrzebnych modułów, nie wdraża niepotrzebnej funkcjonalności systemu. KONSULTANT nie zabezpiecza się wyższą ceną przed nieznanym zakresem, wydłużaniem wdrożenia, brakiem decyzji. Dla KONSULTANTA utrzymania tempa procesu wdrożeniowego, skoncentrowanie w czasie również procesu wyboru i jasność oczekiwań jest długofalowo korzystna. Koszty więc mogą być zdecydowanie niższe i znacznie bardziej kontrolowalne

Faza B – Realizacja procesu wdrożeniowego

Obejmuje okres od rozpoczęcia realizacji kontraktu do protokołu zamykającego wdrożenie (niekoniecznie od podpisania kontraktu, bo data startu projektu może być późniejsza).

Jest to faza w której najistotniejsza jest współpraca między ludźmi z 2 różnych firm o odmiennej kulturze. Ponieważ metoda wdrożeniowa i plan wdrożenia został już uzgodniony pozostaje jego realizacja we współpracy z partnerem. Otwarte są tylko sprawy zarządzania zespołami ,poszczególnymi osobowościami i wzajemnej komunikacji.

Grzechy:

- wycofywanie się z ustaleń lub próba ich zmiany drogą jednostronnej interpretacji przez przedstawicieli obu stron
- wyszukiwanie drobnych potknięć drugiej strony by zdobyć „przewagę"
- przywiązywanie takiej samej wagi do mało znaczących drobiazgów jak i spraw zasadniczych

- wyolbrzymianie specyficznych jednostkowych przypadków prowadzące do koncentracji na wyjątkach zamiast głównych strumieniach transakcji
- brak rzeczywistej współpracy wywodzący się z rozumowania: "za tę cenę to sami musicie mi zrobić wszystko", "nie wyobrażam sobie byście tego nie zrobili" i stała postawa roszczeniowa wynikająca z niewłaściwego rozumienia interesu BENEFICJENTa
- generowanie żądań przez końcowych użytkowników z ich punktu widzenia, który nie pokrywa się z interesem firmy BENEFICJENTa.
- nie wytworzenie przez kierownictwo projektu i obu firm myślenia w kategoriach : celem obu stron jest sukces tego projektu
- nie zapewnienie dostatecznych zasobów do projektu .Często pracownicy BENEFICJENTa nie mają czasu na projekt wdrożeniowy obok swych normalnych obowiązków, w których nikt nie zadbał by im ulżyć.
- brak bieżącego zarządzenia problemami
- brak silnego wsparcia projektu przez zarząd BENEFICJENTa
- brak bieżącej komunikacji i wymiany informacji wśród uczestników projektu
- lekceważenie terminów, nie trzymanie się harmonogramu projektu i traktowanie prac projektowych jako mniej ważnych od codziennych obowiązków –do wykonania kiedy będzie wolny czas
- szukanie pretekstu by nie zamknąć ostatecznie projektu i mnożenie żądań dodatkowych

Przykazania:

- buduj atmosferę współpracy i życzliwości od samego początku i powstrzymuj pierwsze negatywne emocje (niestety często dopiero pod koniec projektu tworzy się ta atmosfera)
- nie oczekuj, że zewnętrzni konsultanci podejmą za ciebie decyzje jak masz działać, zażądaj omówienia/listy możliwości, przedyskutuj skutki, musisz decydować na bieżąco
- nie walcz o wdrożenie wszystkiego, co jest w systemie a tylko o to co naprawdę ci potrzebne, ogranicz ambicje automatyzacji wszystkiego
- robisz wdrożenie po raz pierwszy, zaufaj wybranemu PM -nie wywracaj kolejności prac, a jeśli już to uzgodnij z wyprzedzeniem, unikniesz rozdęcia budżetu i dezorganizacji
- w trakcie projektu organizuj, koordynuj, jest to fragment przebudowy kultury firmy
- współpracuj z konsultantami, stwórz im życzliwą atmosferę, a dadzą z siebie więcej
- obserwuj swoich ludzi-nie wszyscy się sprawdzą, reaguj, zarządzaj ludźmi, buduj team
- koncentruj się na celach i rezultatach, a nie na problemach
- zadbaj o informowanie w projekcie i sprawne komunikowanie

- wyjaśniaj jak najszybciej swoje wątpliwości, nie ma głupich pytań merytorycznych
- nie walcz z ludźmi, a z problemami (problem nie zarządzony rośnie i tyka, kiedyś wybuchnie)
- stosuj uzgodnione standardy i trzymaj się uzgodnionej metody wdrożeniowej
- informuj zawczasu partnera o swoich trudnościach lub błędach

Faza C - Ustanowienie mechanizmu dalszego usprawniania procesów

Celem tej fazy jest wypracowanie stałej procedury i mianowanie osób odpowiedzialnych za stałe dostosowywanie procesów i wspomagających narzędzi ERP do zmieniających się potrzeb przedsiębiorstwa. A wiadomo, że one się będą zmieniać.

Faza rozpoczyna się z chwilą opracowania pod koniec fazy B „listy dalszych usprawnień", a w zasadzie nie kończy się nigdy. Jest to mechanizm cyklicznych mini-projektów usprawnienia rozpoczynanych po uzbieraniu i zatwierdzeniu nowej listy. Jedynie pierwszy obrót A1,B1 wymaga ustanowienia mechanizmu tzn pewnych zasad, które będą obowiązywać w tym zakresie.

Grzechy:

- niechęć do specyfikacji potrzeb i uzgadniania ich z innymi procesami
- skłonność do stosowania zastępczych procedur „ręcznych" co prowadzi do sytuacji, że istniał będzie system nieformalnych zapisów, ewidencji obok sytemu komputerowego. Osoba czerpiąca dane z systemu komputerowego dostanie niepełne dane i spadnie zaufanie do systemu, co prowadzi do zamierania jego użycia, a erozji ulega w ten sposób cały informacyjny przepływ w firmie i decyzje oparte na niepełnych danych będą złe, wyniki firmy będą gorsze itd.
- Niechęć do zebrania kilku zmian i zastosowania projektowej, zorganizowanej formy wprowadzenia ich w życie (która sprawdziła się w trakcie pierwszego projektu wdrożeniowego)

Przykazania:

- natychmiast po zamknięciu pierwszego projektu wdrożeniowego należy powołać stały komitet złożony z osób odpowiedzialne za poszczególne procesy w firmie, spośród uczestniczących w projekcie wdrożeniowym (doświadczenia) by rejestrowały wnioski i opracowywały propozycje zmian w procesach i po uzgodnieniu(też z KONSULTANTEM) przedstawiały propozycje do akceptacji Zarządu. Ten komitet musi mieć wyznaczony budżet na swoją stałą działalność.

- projekty zmian należy akceptować co najmniej raz w roku, inicjując tym samym nowy kolejny projekt z jego budżetem, organizacją, wyznaczeniem ról i mechanizmami kontroli (jak to przećwiczono praktycznie przy pierwszym wdrożeniu)
- z poziomu Zarządu należy żądać stałych półrocznych sprawozdań z prac komitetu

Podsumowanie

Powyżej w dużym skrócie przedstawiono propozycję „właściwego porządku rzeczy" czyli takiej kolejności działań by uzyskać w sposób kontrolowany z góry uzgodnione cele w przewidzianym czasie i przy racjonalnym poziomie kosztów. W tej sekwencji każdy z podmiotów współdziałających ma wszelkie potrzebne informacje i gwarancje tego czego może się spodziewać. Nie powinno więc marnowania wysiłku ludzi, czasu i pieniędzy na prace przedwczesne, nieprzygotowane czy niepotrzebne. Przedstawmy w skrócie zalety proponowanego podejścia, które oparte jest na obserwacji wielu projektów ERP.

Moment i decyzja wdrażania -

Mamy pełne podstawy do racjonalnego wyboru systemu (czy na pewno ERP), wiemy co dla nas jest ważne, o co pytać KONSULTANTA, znamy uwarunkowania czasowe i budżetowe. Mamy pełne podstawy by zaplanować i przeprowadzić przygotowanie organizacyjne firmy do wdrożenia (szkolenia w nowych metodach, uporządkowanie danych wspólnych, rejestrów).

Kształt rozwiązania i przebieg wdrażania –

Rozwiązanie w całej naszej organizacji jest zdefiniowane i spójne, (lokalnie uzgodnienia na poziomie użytkownik-konsultant nie będą w stanie zdegenerować rozwiązania). Wybrane algorytmy działania dobrane są pod kątem dobra całej firmy a nie lokalnego użytkownika.

Przy okazji ujednolicamy sposób funkcjonowania wszystkich oddziałów firmy. KONSULTANT ma szansę przygotować za pierwszym razem dobrą konfigurację systemu.

Tempo wdrażania już nie wpływa na jakość rozwiązania i może być znacznie szybsze. Czyli dostaniemy szybciej i mniejszym kosztem lepszy rezultat w procesie wdrożeniowym.

Efekty - Jeśli mamy zdefiniowany cel wdrożenia, znamy metody zarządzania firmą, których chcemy użyć to możemy PRZED zawarciem kontraktu wdrożeniowego dość dokładnie oszacować efektywność inwestycji (np.ROI), a na końcu zmierzyć osiągnięte rezultaty. Koszty zaś procesu wdrożeniowego będą zdecydowanie niższe jeśli ten proces przygotujemy i przygotujemy organizacyjnie firmę. Ci, którzy już ten proces przeszli wiedzą, że nie są to koszty bagatelne.

Uważny czytelnik zauważy oczywiście, że zajmują czas i kosztują czynności poprzedzające sam proces wdrożeniowy. To prawda.

Ale jaki mają one charakter ? Moim zdaniem stały, cykliczny. Zarząd firmy musi reagować na zmieniającą się sytuację rynkową, warunki gospodarowania,

poczynania konkurencji. Ponieważ tempo zmian otoczenia jest coraz większe, nieuniknione jest częstsze wykonywanie tych czynności. Proszę zwrócić uwagę, że naszym celem jest poprawić i UTRZYMAĆ produktywność, a więc reagować i ciągle ulepszać, poprawiać. Kaizen. Czy to wiązać zawsze z wdrażaniem ERP ?.
Oczywiście, że ze względu na inną częstotliwość tych procesów winny to być rozdzielne ciągi czynności. Nie każda zmiana metody i reorganizacja pociąga za sobą zmiany parametrów w systemie ERP (chociażby przez sterowanie danymi) i odwrotnie. Rezultatem prac może być rozpoczęcia innych działań nie związanych z ERP. Tylko więc w niektórych przypadkach konieczne będzie rozpoczynanie procesu wdrażania / dostosowania systemu ERP.

Są to więc czynności stałe, winny być rozpatrywane jako stały koszt zarządzania firmą w tych dynamicznych czasach. Personel przywoływany do nich jest zazwyczaj inny niż biorący udział w typowym wdrażaniu ERP.
Tempo tych prac może być regulowane przez firmę BENEFICIENTA dowolnie i wg potrzeb.

Ta opisana „metoda" nie jest żadnym rewelacyjnym wynalazkiem.
Czy jest to najlepsze i jedynie słuszne spojrzenie? Być może nie, ale jest w tym logika prowadząca do eliminacji kosztów nie dodających wartości i uniknięcia konfliktów psujących współpracę.

Celowym wydaje się nazwanie rzeczy po imieniu i przedstawienie racji i problemów każdej ze stron by przywrócić zainteresowanie usprawnianiem firm (w tym wdrażaniem rozwiązań zintegrowanych). Do tego w Polsce mamy wszystko:

- firmy BENEFICIENTÓW potrzebujące wsparcia (w tym też systemów ERP) –choć często niezbyt zasobne, nieprzygotowane organizacyjnie, bojące się partnerstwa i generujące niepotrzebne koszty u dostawców
- systemy ERP, które nie przestały być potrzebne, a które mogłyby być bardziej przyjazne użytkownikom, lepiej dobrane do potrzeb BENEFICIENTA i tańsze
- firmy KONSULTANTów, które muszą przygotować bardziej racjonalne-tańsze oferty dla BENEFICIENTa, choć niekoniecznie mniej opłacalne dla siebie
- doświadczonych Kierowników Projektów (Project Managerów) –również jako kontraktowych uczestników procesu wdrożeniowego. Jest to nowe ale bardzo interesujące zjawisko na polskim rynku, bo stwarza BENEFICIENTOM zaangażowania kompetentnej osoby (nie związanej układami wewnętrznymi) tylko na czas wdrożenia
- rosnącą ilość konsultantów biznesowych-doradców (nie mylić z konsultantami- wdrożeniowcami), którzy mogą pomóc BENEFICIENTOM w przygotowaniu strategii, doborze metod zarządzania i poprawy konkurencyjności. Dla przypomnienia to właśnie zastosowanie nowych metod zaowocowało Amerykanom. Nie są to duże firmy, często pojedyńcze osoby, ale znacznie tańsze niż renomowane światowe firmy konsultingowe. Wydaje się, że nadchodzi czas realizacji procesów naprawczych w dużej skali by

poprawić konkurencyjność polskich firm szczególnie w perspektywie wejścia do Unii Europejskiej.

Dlatego też mam nadzieję na namówienie BENEFICIENTÓW do świadomego rozpoczęcia budowania własnej drogi rozwoju/naprawy, a wszystkich innych wymienionych do przygotowania możliwie taniej oferty dla BENEFICIENTÓW, by doprowadzić do WSPÓŁPRACY z poszanowaniem wzajemnych interesów. Już najwyższy czas na konkretne działania, ale jeśli potrzebna jest jeszcze dyskusja rozpocznijmy ją czym prędzej.

Zachęcam.

Proponuję jednak unikać obwiniania drugiej strony, a wskazywać na możliwości współdziałania. WSPÓŁPRACY. Podobnie jak w zarządzaniu łańcuchem dostaw –nie chodzi o przerzucenie kosztów na partnera, a o zbudowanie mechanizmów eliminacji niekoniecznych kosztów z całego procesu.

Andrzej Sakowski
Niezależny konsultant ERP.
ERP-EKSPERT
tel: 0-502 768 163
andrzej@saratoga.neostrada.pl

ZASTOSOWANIE SYSTEMÓW INFORMATYCZNYCH W GEOLOGII

Krzysztof SZAMAŁEK, Waldemar GOGOŁEK, Bogusław KAZIMIERSKI, Stanisław PRZENIOSŁO, Stefan MŁYNARSKI

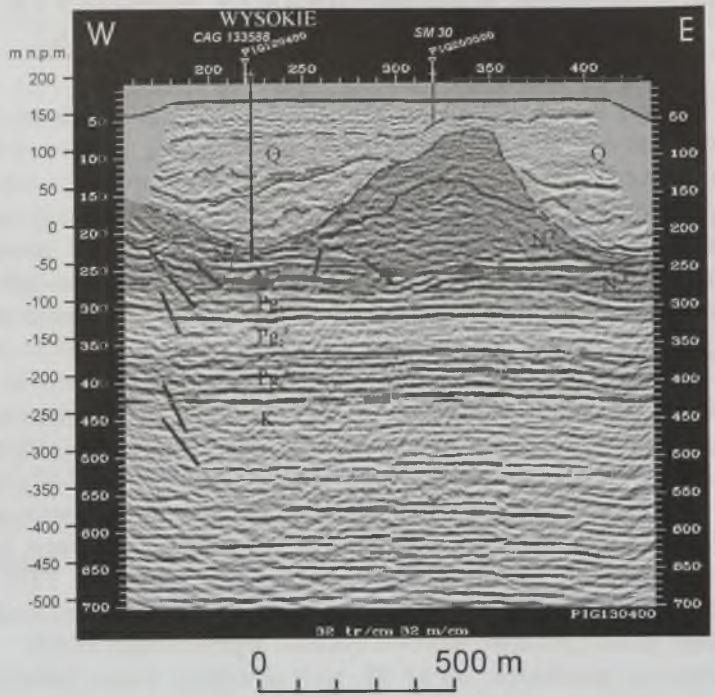
Geologia współczesna to wiele specjalistycznych zagadnień, dla których prowadzone badania w szerokim zakresie wykorzystują systemy informatyczne. Omówienie wszystkich dziedzin geologii, w których informatyka odgrywa istotną rolę jest zbyt obszerne, aby można było omówić w wystąpieniu o ograniczonych ramach czasowych. W związku z powyższym przedstawione zostaną wybrane zagadnienia, gdzie informatyka odgrywa istotną rolę i nie można sobie wyobrazić prowadzenia współczesnych badań bez udziału programów komputerowych.

Geofizyka jest jedną z dziedzin geologii, w której szeroko stosowane są systemy informatyczne. Już w końcu lat 60-tych ubiegłego wieku w Polsce w badaniach sejsmicznych wprowadzono do rejestracji amerykańskie i francuski aparatury z zapisem magnetycznym i opracowanie sekcji czasowych na centralach analogowych. Wykorzystywano do tego celu komputery „ODRA". Następnie już w latach 70-tych wprowadzono w sejsmice technikę cyfrową, zarówno w zakresie rejestracji, jak i przetwarzania danych. Szerszy zakres cyfrowego przetwarzania danych sejsmicznych był możliwy dzięki wprowadzeniu w 1976 roku przez Przedsiębiorstwo Badań Geofizycznych systemu WARS, będącego adaptacją stosowanego przez górnictwo naftowe systemu sejsmicznego SYSIS.

Obecnie, poczynając od lat 90- tych ubiegłego wieku badania sejsmiczne prowadzone są o szeroko wykorzystywane systemy informatyczne. Do przetwarzania danych sejsmicznych stosowany był przykładowo w Państwowym Instytucie Geologicznym system FOCUS firmy CogniSeis Devleopment Inc, a do interpretacji i prezentacji, zarówno w wersji czasowej, jak i głębokościowej, oraz dla konstrukcji przekrojów i map sejsmicznych, oprogramowanie firmy LANDMARK. Stworzyło to szerokie możliwości prezentacji uzyskiwanych wyników. Na przykładowym przekroju sejsmicznym w wersji czasowej ze skalą głębokościową widać urozmaiconą interpretację danych sejsmicznych.

Na szczególne podkreślenie zasługuje wykorzystanie informatyki w geofizycznych metodach pól potencjalnych, a w szczególności w badaniach magnetycznych i grawimetrycznych. Stworzony w latach 90 –tych ubiegłego wieku komputerowy bank danych grawimetrycznych był podstawą do wydania seryjnej mapy grawimetrycznej Polski w skali 1 : 200 000, a w roku 1995 publikacji (Cz. Królikowski i Z. Petecki) Atlasu grawimetrycznego Polski. Niemniej dopiero zaawansowane oprogramowania informatyczne pozwoliły w ostatnim czasie na prezentacje obrazów geofizycznych o szerokiej gamie uwypuklania kierunków obserwowanych anomalii. Pozwala to na interpretację wybranych elementów budowy geologicznej, zarówno w opcji globalnej, w powiązaniu z danymi z Europy, jak również w przypadkach analizy lokalnej.

Przykładowo dla szczegółowej lokalizacji stref dyslokacyjnych na mapie geologicznej Polski.

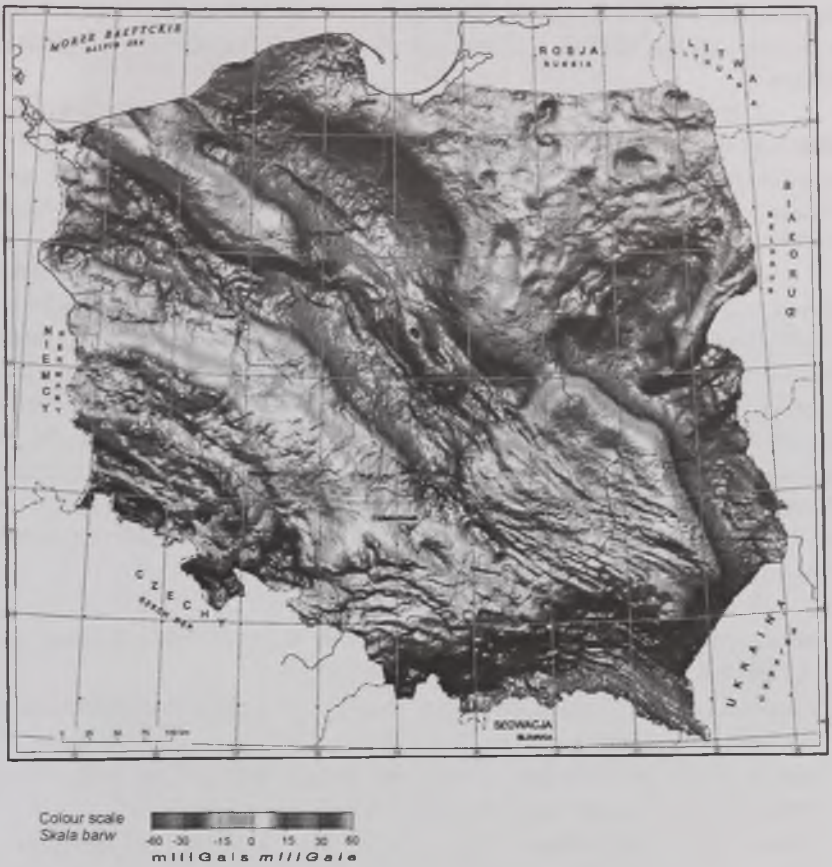


Rys. 1. Przekrój sejsmiczny w wersji czasowej ze skalą głębokościową

Przykładem możliwości prezentacji danych jest załączony jeden z wielu publikowanych w pracach S. Wybrańca (1999 - Transformations and visualization of potential field data, PGI Special Papers, v.1.) obraz grawimetryczny Polski dający dobre podstawy do na wnikliwą interpretacji geologicznej danych grawimetrycznych.

Dzięki coraz szybszemu rozwojowi funkcjonalności oprogramowania i lawinowemu przyrostowi mocy sprzętu komputerowego w ostatnich latach było możliwe wprowadzenie nowych technologii w **kartografii geologicznej**.

Realizacja zadań kartograficznych przy wykorzystaniu techniki komputerowej była możliwa dzięki zleceniom Ministra Środowiska na ich wykonywanie i finansowaniu z dwóch źródeł: budżetu w 3% i Narodowego Funduszu Ochrony Środowiska i Gospodarki Wodnej w 97%.



Rys. 2. Mapa grawimetryczna Polski jako przykład wizualizacji pól potencjalnych

Do najważniejszych opracowań realizowanych z wykorzystaniem metod i narzędzi informatycznych należą obecnie trzy seryjne mapy w skali 1:50 000, atlasy geochemiczne, mapy dna Bałtyku i mapy geologiczno-inżynierskie. Wymienione mapy i atlasy były i są wykonywane przy zaangażowaniu znacznych środków finansowych oraz dużych zespołów specjalistów. Ich głównym wykonawcą i koordynatorem jest Państwowy Instytut Geologiczny

Szczegółowa mapa geologiczna Polski w skali 1:50 000 (SMGP) jest podstawową mapą geologiczną kraju, realizowaną od 1953 r. i składającą się z 1069 arkuszy. Jej autorskie opracowanie zbliża się ku końcowi. Trwającą już 50 lat edycję mapy będzie można zakończyć w ciągu kilku lat, również dzięki zastosowaniu systemów informatycznych. Mapa jest opracowaniem kompleksowym, powstającym na podstawie zatwierdzonego projektu prac

geologicznych, w wyniku szczegółowego zdjęcia geologicznego w skali 1:25 000, które obejmuje prace polowe, wiercenia, pomiary geofizyczne, badania laboratoryjne i prace kameralne. Obecnie typowy arkusz powstaje w ciągu trzech lat. Każdy arkusz SMGP jest monograficzną dokumentacją geologiczną danego obszaru.

Z inicjatywy Departamentu Geologii MŚ w 1994 r. rozpoczęto program komputerowego opracowania mapy. Do jej opracowania od początku używano zaawansowanego sprzętu: pierwotnie wyłącznie stacji roboczych i graficznych pracujących w oparciu o system Unix-Irix, obecnie pracujących w systemie MS Windows NT/2000.

Podstawowym oprogramowaniem używanym do cyfrowego opracowania mapy jest system GIS firmy ESRI. Początkowo było to oprogramowanie Arc/Info w wersji 6.0, obecnie oprogramowanie to jest tylko częścią, choć podstawową, znacznie rozszerzonego funkcjonalnie systemu ArcGIS w wersji 8.3. Baza danych powstaje przy wykorzystaniu systemu relacyjnych baz danych Oracle. Podstawą opracowania są jednak autorskie aplikacje przygotowane specjalnie dla wspierania realizacji konkretnych zadań związanych z cyfrowym opracowaniem mapy: aplikacja ArcTeren dla autorów arkuszy, a aplikacje ArcSMGP i Koda dla operatorów tworzących bazę danych i redaktorów technicznych przygotowujących mapę do druku.

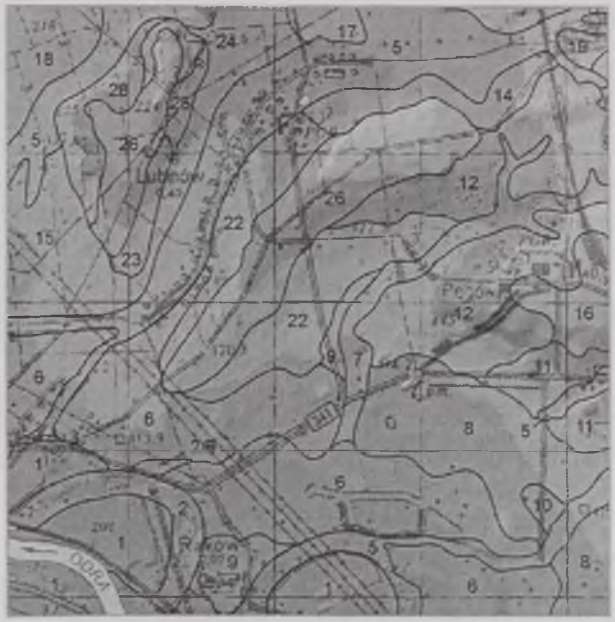
W wyniku realizacji programu opracowano:

- nową metodykę i *Instrukcję* wykonania mapy uwzględniające potrzeby i możliwości cyfrowego opracowania mapy;
- zaprojektowano i wdrożono bazę danych, w której zgromadzono do chwili obecnej 43% arkuszy mapy;
- opracowano technologię cyfrowego przygotowania arkuszy do druku z wykorzystaniem zasobów bazy danych SMGP;
- opracowano technologię komputerowego opracowania arkuszy mapy i objaśnień w celu ich udostępniania w formie cyfrowej i druku na zamówienie.

Przez 9 lat realizacji programu cyfrowego opracowania SMGP oprogramowanie i stosowane rozwiązania metodyczne były stale udoskonalane. Najbardziej znaczącymi momentami w tych latach było:

- wydrukowanie w 1996 r., po raz pierwszy, przygotowanych cyfrowo do druku arkuszy SMGP, 10 offsetowo i 10 na ploterze;
- przejście w 2000 r. na udostępnianie mapy tylko na zamówienie i zastąpienie druku offsetowego ploterowym;
- przygotowanie narzędzi i procedur usprawniających udostępnianie przy wykorzystaniu Internetu (ten sposób udostępniania danych dotyczy trzech map seryjnych SMGP, MHP i MGGP).

Rys. 3. Fragment arkusza Szczegółowej mapy geologicznej Polski w skali 1:50 000 przygotowany do druku ploterowego na zamówienie



Rys. 4. Dane z bazy Szczegółowej mapy geologicznej Polski w skali 1:50 000 zestawione z cyfrowym modelem terenu (fragment opracowania związanego z analizą problemów związanych z powodziami)

Podstawowe korzyści z realizacji SMGP przy wykorzystaniu systemów informatycznych to:

- usprawnienie i przyspieszenie edycji mapy – do 1995 r. drukowano po kilka arkuszy rocznie, obecnie 40;
- utworzenie bazy danych SMGP, z której dane służą przygotowaniu analiz i map tematycznych dla różnych dziedzin gospodarki, geologii, administracji różnego szczebla;
- ułatwienie dostępu do podstawowej wiedzy o budowie geologicznej kraju, możliwej do uzyskania w formie tradycyjnej, czyli wydruków lub nowoczesnej formie cyfrowej w formatach rastrowych i wektorowych, które pozwalają na skomplikowane analizy przestrzenne.

Mapa hydrogeologiczna Polski w skali 1:50 000 (MHP) jest opracowywana od 1998 r przy wykorzystaniu metod cyfrowych w systemie GIS. MHP jest mapą seryjną w cięciu arkuszowym, która powstaje jako baza zawierająca dane przestrzenne związane z występowaniem wód podziemnych, pozyskiwane z pomiarów terenowych, archiwów i innych baz danych. Celem mapy jest odwzorowanie warunków hydrogeologicznych oraz tych elementów gospodarczych i sozologicznych, które wiążą się z zagrożeniem i ochroną wód podziemnych. Baza danych MHP współpracuje z centralnym bankiem danych hydrogeologicznych HYDRO oraz systemami obserwacji i monitoringu SOH i MONBADA. Baza może być wykorzystana w typowych pracach geologicznych oraz do zadań specjalnych:

- waloryzacji wód podziemnych w celu ich wykorzystania i ochrony,
- lokalizacji ujęć wód podziemnych,
- sporządzania bilansów wodno-gospodarczych,
- planowania lokalnych i regionalnych badań hydrogeologicznych,
- sporządzania i korekty planów zagospodarowania przestrzennego.

Mapa geologiczno-gospodarcza Polski w skali 1:50 000 (MGGP) jest kartograficznym opracowaniem, które w sposób syntetyczny przedstawia zagadnienia środowiskowe na tle zjawisk geologicznych. MGGP jest mapą seryjną w cięciu arkuszowym wykonywanym na podstawie inwentaryzacji i reinterpretacji materiałów archiwalnych z pewnym udziałem prac terenowych. Pojedynczy arkusz jest wykonywany w ciągu jednego roku.

Zawartość mapy i jej bazy dzieli się na kilka bloków tworzących jednocześnie grupy warstw informacyjnych w systemie GIS. Są to:

- kopaliny,
- górnictwo i przetwórstwo kopalin,
- wody powierzchniowe i podziemne,
- warunki podłoża budowlanego,
- elementy ochrony przyrody, krajobrazu i zabytków kultury.

Baza danych MGGP współpracuje z bankiem złóż surowców mineralnych MIDAS.

MGGP ma służyć do wspomagania regionalnych i lokalnych działań gospodarczych, prac studialnych nad uwarunkowaniami i kierunkami zagospodarowania przestrzennego.

Do tworzenia i edycji MHP i MGGP stosowane jest oprogramowanie MGE, firmy Intergraph. Silnikiem graficznym aplikacji MGE jest MicroStation popularny system CAD. W plikach MicroStation jest przechowywana graficzna prezentacja obiektów obu map, informacje opisowe o obiektach są przechowywane w relacyjnych bazach danych opartych o system Oracle. Wydruki poszczególnych arkuszy map są wykonywane przy wykorzystaniu rodziny programów InterPlot IPLOT.

Dzięki zastosowaniu systemów informatycznych realizacja MHP i MGGP jest niezwykle wydajna co pozwoli na ukończenie ich edycji odpowiednio w 2004 i 2006 roku.

Przy wykorzystaniu metod cyfrowych opracowano liczne atlasy i mapy nieseryjne. Warto tu wymienić ważniejsze z nich.

Atlasy i mapy geochemiczne wykonywane w różnych skalach dla wybranych regionów i aglomeracji. Zawierają one opisy metodyki prac, geologii, omówienia wyników badań próbek gleb, wód powierzchniowych i podziemnych oraz tablice z mapami występowania różnych pierwiastków.

Mapy i atlasy geologiczno-inżynierskie zawierają bogaty i usystematyzowany materiał archiwalny, również w postaci baz danych, z których dane pozwalają na wykonywanie aplikacji analitycznych.

Mapy geologiczne Bałtyku w skali 1:200 000 i 1:500 000 są kompleksowymi opracowaniami zawierającymi wyniki wieloletnich badań dna morskiego. Mapy zawierają różne tematyczne plansze i przekroje.

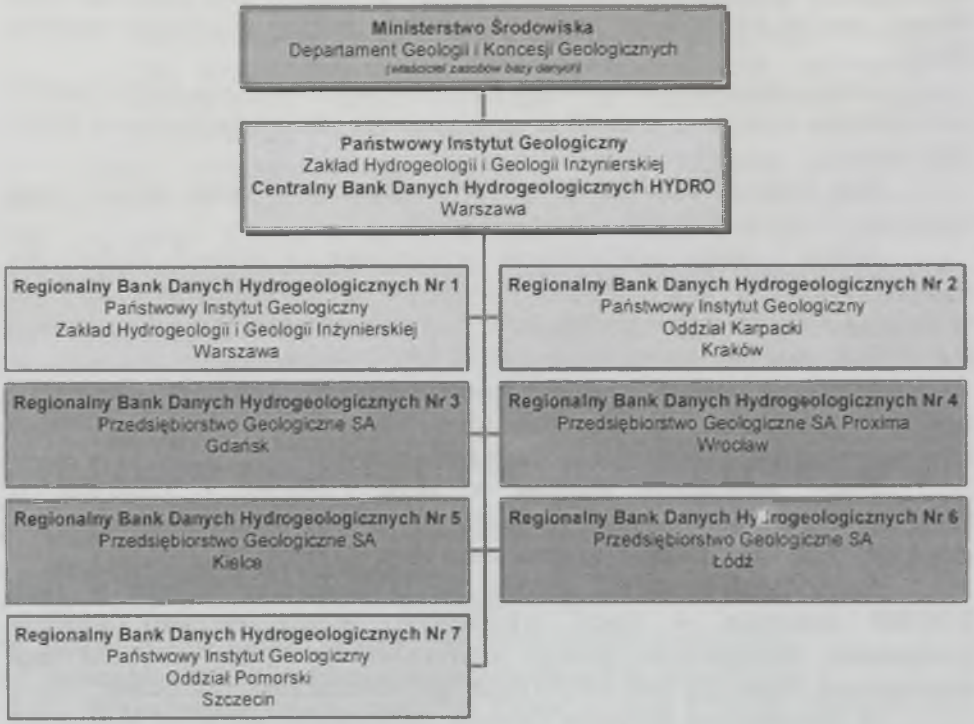
Mapa geodynamiczna polskiej strefy brzegowej Bałtyku w skali 1:10 000 wykonana w cięciu arkuszowym zawiera elementy geologii, geodynamiki, hydrogeologii, geologii inżynierskiej i surowcowej, waloryzacji sozologicznej. Mapa i jej baza danych mają ogromne znaczenie praktyczne.

W Państwowym Instytucie Geologicznym eksploatowane są następujące bazy danych hydrogeologicznych: Bank HYDRO, System Obserwacji Hydrogeologicznych SOH, Monitoringowa Baza Danych MONBADA.

Baza danych hydrogeologicznych Bank HYDRO zawiera dane dokumentacyjne o odwiertach, ujęciach i źródłach wód podziemnych zwykłych (pitnych), mineralnych i termalnych Polski. Pierwszy projekt bazy powstał w latach siedemdziesiątych ubiegłego wieku. Wielokrotnie dokonywano gruntownej modernizacji bazy, założenia projektowe aktualnie eksploatowanej bazy opracował, z udziałem firmy Intergraph (Europe) Polska, zespół w składzie: L. Skrzypczyk; B. Kazimierski, J. Mikuszewska, K. Kupiszak - PIG Warszawa, G. Piłat, Z. Bartoś, T. Głowacki - Intergraph), a w roku 2000 system został wdrożony również w bankach regionalnych.

Bank HYDRO zawiera dane dokumentacyjne o ok. 122 tysiącach obiektów hydrogeologicznych (odwierty, ujęcia i źródła zwykłych wód podziemnych, wód mineralnych i termalnych) występujących na obszarze Polski. Dane z obszaru kraju

gromadzone są w Centralnym Banku Danych Hydrogeologicznych HYDRO (CBDH) Państwowego Instytutu Geologicznego w Warszawie. Aktualną strukturę organizacyjną sieci Banków HYDRO tworzą: Centralny Bank Danych Hydrogeologicznych HYDRO oraz siedem Regionalnych Banków Danych Hydrogeologicznych HYDRO.



Rys. 5. Struktura organizacyjna sieci Banków HYDRO

Podstawowe zadania realizowane przez wykonawców projektu Bank HYDRO:

- Gromadzenie i aktualizacja danych o odwiertach, ujęciach i źródłach zwykłych wód podziemnych, wód mineralnych i termalnych z obszaru Polski
- Udostępnianie danych hydrogeologicznych Banku HYDRO użytkownikom według zasad i norm określonych przez Ministerstwo Środowiska
- Prowadzenie szkoleń dla użytkowników w zakresie obsługi bazy danych i systemu Banku HYDRO
- Pomoc w wdrażaniu systemu Banku HYDRO w państwową administrację geologiczną, uczelnie wyższe, firmy geologiczne, firmy i instytucje ochrony środowiska

- Prowadzenie serwisu informacyjnego Banku HYDRO (Internet, publikacje)
- Gromadzenie, opracowanie i aktualizacja cyfrowej kartografii dokumentacyjnej Banku HYDRO

W pracach modernizacyjnych systemu Banku HYDRO w 1998 r opracowano nową relacyjną bazę danych HYDRO w systemie Oracle oraz wykonano nowe oprogramowanie komputerowe bazy danych – nowe narzędzia do obsługi bazy danych to aplikacje HYDRO2000 i GeoHydro/GeoMedia GIS.

Ogólnie system Banku HYDRO pozwala użytkownikowi na szybkie uzyskanie informacji na temat:

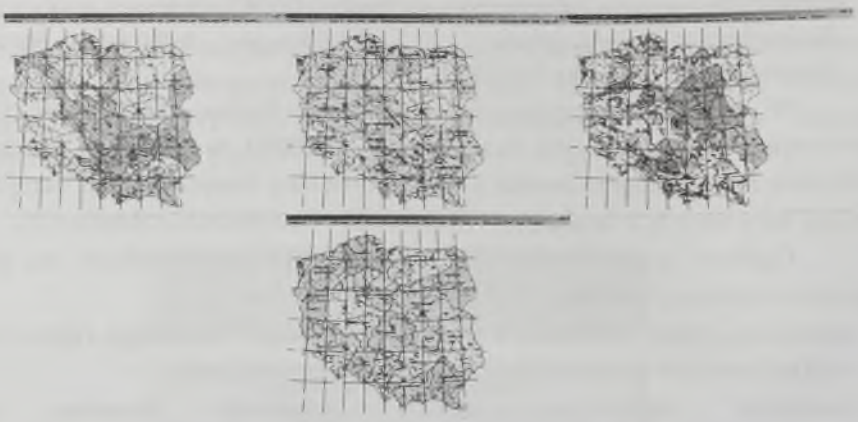
- aktualnego stanu rozpoznania hydrogeologicznego wybranego rejonu (ilość i rozkład otworów hydrogeologicznych w zadanym rejonie);
- tworzenia wydruków, szkiców lokalizacji otworów, profili litostratygraficznych, kart otworów, wyników analiz fizyczno-chemicznych, danych o ujęciach wód podziemnych itd.;
- tworzenie dowolnych zestawień danych hydrogeologicznych według zadanego przez użytkownika klucza w formie wydruku lub zbioru komputerowego z danymi Banku HYDRO z przeznaczeniem do dalszego przetwarzania danych na dowolnym arkuszu kalkulacyjnym, oprogramowaniu typu GIS itd.

Więcej informacji o banku Hydro można uzyskać na stronie PIG, pod adresem www.pgi.waw.pl/hydro.

System obserwacji hydrogeologicznych SOH – baza danych. funkcjonuje lat siedemdziesiątych ubiegłego wieku, a w roku 1996 został gruntownie zmodernizowany. Aktualnie funkcjonuje w wersji PC. Dane aktualizowane są na bieżąco, w miarę ich napływu od obserwatorów, w formacie access'a. Pod adresem: http://www.pgi.waw.pl/soh/ istnieje internetowa wersja bazy danych z ograniczonym dostępem do danych archiwalnych. Prowadzone są prace związane z prezentacją graficzną SOH w sieci wewnętrznej PIG (intranet) z wykorzystaniem GeoMedia Web Map.

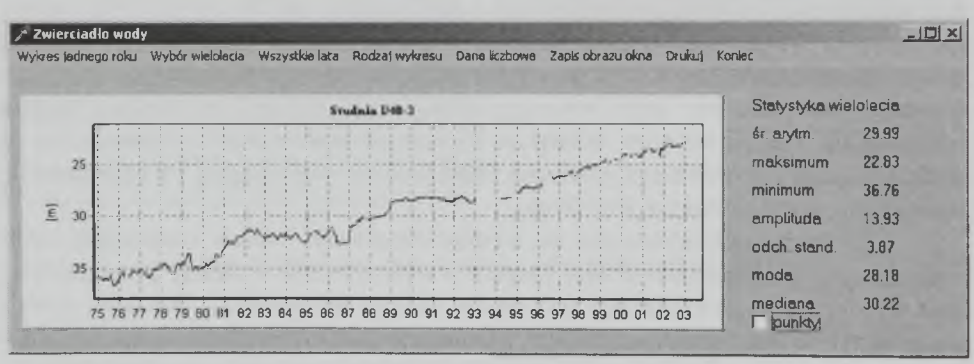
W bazie archiwizowane są wyniki obserwacji wahań zwierciadła wód podziemnych i wydajności źródeł (pomiary wykonywane w każdy poniedziałek) oraz wyniki analiz chemicznych (raz na rok lub raz na dwa lata). Początek okresu obserwacji w wielu przypadkach sięga połowy lat 70-tych. Baza obejmuje wyniki z obserwacji zarówno z punktów aktualnie działających jak i archiwalnych. Aktualnie baza jest modernizowana, prowadzone są również prace związane z prezentacja graficzną SOH w sieci wewnętrznej PIG (intranet).

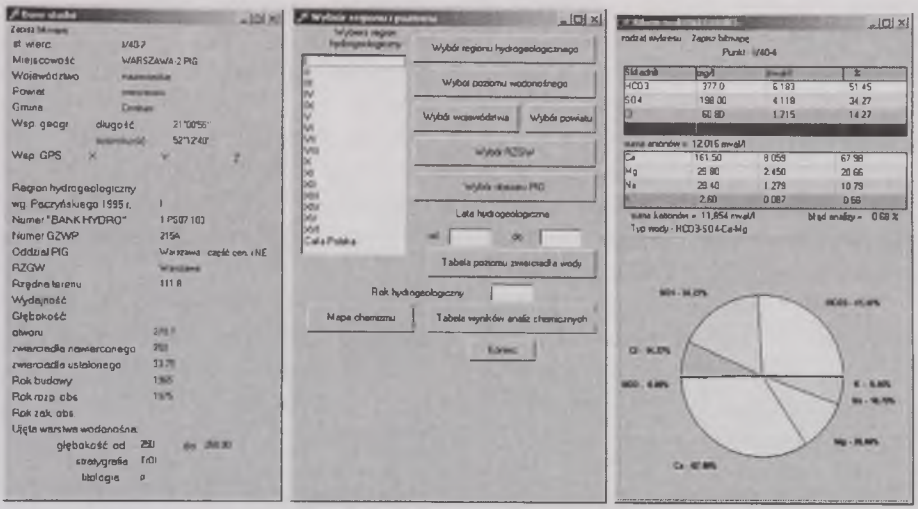
Punktem wyjścia przy korzystaniu z bazy SOH jest wybór jednej z map – tła z lokalizacją punktów: zasięgi głównych użytkowych poziomów wodonośnych, jednostki hydrogeologiczne (regiony, subregiony, rejony), Główne Zbiorniki Wód Podziemnych (GZWP), zasięgi działania Regionalnych Zarządów Gospodarki Wodnej (RZGW), obszary działania Oddziałów Państwowego Instytutu Geologicznego, województwa, zlewnie.



Rys. 6. Przykład hydrogeologicznych map tematycznych, stanowiących tło dla wyboru punktów obserwacyjnych wód podziemnych.

Istnieje możliwość dostępu do danych przez wybór punktu obserwacyjnego na mapie, wybór numeru punktu lub jego nazwy. Dane można przeglądać w postaci zestawień tabelarycznych i prezentacji graficznej. Można wykorzystywać programy selekcji i przetwarzania danych, co jest niezwykle pomocne w przygotowywaniu wyników obserwacji do udostępniania czy sporządzania raportów o stanie wód. Przykład prezentacji graficznej przedstawiono poniżej:





Rys. 7. Przykład prezentacji graficznej wyników obserwacji oraz informacji o punkcie obserwacyjnym.

Monitoringowa baza danych MONBADA zawiera komplet analiz wykonanych w ramach Państwowego Monitoringu Środowiska począwszy od 1991 roku. W chwili obecnej, w związku ze zmianą oprogramowania, następuje przenoszenie bazy danych w nową strukturę. Specjalistyczne oprogramowanie GIS umożliwia połączenie informacji geograficznej z danymi atrybutowymi, takimi jak informacje o punkcie, wyniki poszczególnych analiz chemicznych, itp. Coroczne sprawozdania z badań monitoringowych (statystyka, zestawienia tabelaryczne, wykresy, mapy w skali przeglądowej) przedkładane do GIOŚ przez Zakład Hydrogeologii i Geologii Inżynierskiej PIG generowane są w oparciu o narzędzia GIS. Zastosowane narzędzia umożliwiają bezpośredni dostęp i analizę danych pochodzących z różnych źródeł, którymi mogą być projekty GIS w różnych formatach.

Dają możliwość pozyskiwania, weryfikacji topologicznej, zarządzania i aktualizacji danych geograficznych. Umożliwiają wyświetlanie danych w różnych układach współrzędnych, niezależnych od odwzorowania danych źródłowych. Istnieje możliwość zmiany odwzorowania podczas pracy z projektem, a wszystkie dane z podłączonych projektów są przeliczane do nowego układu odniesienia w czasie rzeczywistym. Posiadają wiele funkcji, które pozwalają na wykonywanie skomplikowanych analiz przestrzennych, a także na filtrowanie danych względem atrybutów i wybranych operatorów przestrzennych.

Wykorzystując funkcje analiz przestrzennych, wyniki analiz chemicznych z laboratorium przechodzą przez odpowiedni proces obróbki, dzięki czemu można je później zaprezentować na mapach. W raporcie rocznym zobrazowania wyników przedstawiono na mapach w skali 1 : 2 500 000.

Wprowadzenie nowoczesnego oprogramowania z zakresu GIS spowodowało istotne skrócenie czasu obróbki danych, spowodowało rozszerzenie

możliwości analizy danych oraz istotnie usprawniło prace nad projektami GIS. Poprzez wykorzystywanie systemów informacji geograficznej gromadzone i przetwarzane są duże ilości niezmiernie wartościowych danych. Stanowią one istotny wkład zarówno w odniesieniu lokalnym jak i globalnym. Istnieje możliwość integracji tych danych z bazami danych z innych krajów (np. Unii Europejskiej). Dzięki temu można stworzyć projekty o skali globalnej. Jeden z takich projektów, dotyczący wdrażania Ramowej Dyrektywy Wodnej, przygotowywany jest przez Ministerstwo Środowiska we współpracy z Państwowym Instytutem Geologicznym.

Zasoby hydrogeologicznych baz danych wykorzystywane są do celów projektowych i dokumentacyjnych w dziedzinie ochrony środowiska, w administracji rządowej, samorządowej oraz służbach państwowych jako system wspomagania decyzji w zarządzaniu środowiskiem oraz coraz częściej dla potrzeb sporządzania raportów dla Instytucji Unii Europejskiej z zakresu oceny stanu środowiska i gospodarki wodnej. Dla państwowej administracji geologicznej i gospodarki wodnej zasoby tych baz danych stanowią główne źródło informacji o stanie rozpoznania i zagospodarowania wód podziemnych regionu. Do głównych ich użytkowników można zaliczyć: Ministerstwo Środowiska, administrację publiczną wszystkich szczebli, przedsiębiorstwa i firmy geologiczne, fundacje i organizacje ochrony środowiska, uczelnie i instytucje naukowe, Wojsko Polskie, Regionalne Zarządy Gospodarki Wodnej, Parki Narodowe i Dyrekcje Lasów państwowych, Inspekcję Ochrony Środowiska.

Problematyka geologii gospodarczej i zapotrzebowanie na informacje w tym zakresie, szczególnie w okresie przemian gospodarczych Polski, spowodowało utworzenie w Państwowym Instytucie Geologicznym kilku systemów informatycznych. Należą do nich: system Gospodarki i Ochrony Bogactw Mineralnych „MIDAS" wraz z włączonym do niego Rejestrem Obszarów Górniczych, „PRICESMIN" oraz INFOGEOSKARB.

System „MIDAS" wykonany w 1988 roku, pomyślany był jako baza danych o złożach kopalin Polski wraz z oprogramowaniem pozwalającym na wszechstronne zarządzanie ta bazą. Ten początkowy cel został znacznie rozszerzony o potrzeby wynikające z konieczności wykonywania analiz gospodarki poszczególnymi surowcami lub złożami, szczególnie w okresach wieloletnich.

System ten pozwala na gromadzenie danych o zasobach, jakości kopaliny, a w złożach wielo surowcowych o wszystkich kopalinach złoża (głównej i towarzyszących), o warunkach geologiczno-górniczych, poziomach wodonośnych itp. Gromadzenie informacji o ruchu zasobów wykonywane jest w okresach rocznych, a do innych zbiorów następuje w sposób ciągły. Obecnie w bazach systemu znajdują się dane od 1986 r. Pozwala to na prześledzenie gospodarki złożem w dłuższym okresie i uzyskanie danych o wielkości bazy zasobowej, przyrostach zasobów w wyniku pozytywnych badań geologiczno-poszukiwawczych, a także o ubytkach z tytułu wydobycia i rzeczywistych stratach zasobów czy stopniu wykorzystania złoża.

Drugim ważnym zadaniem dla systemu ochrony złóż jest racjonalne wykorzystanie surowców mineralnych, w tym odpadów pogórniczych i przeróbczych. Z jednej strony, odpady takie mogą być odpadowym surowcem mineralnym możliwym do wielokierunkowego wykorzystania. Z drugiej strony ich utylizacja ogranicza zanieczyszczenie środowiska. Aby sprostać tym zadaniom, system gromadzi informacje o składowiskach, na których deponowane odpady pogórnicze i przeróbcze, z rejestracja ilościową tych odpadów, a także danych jakościowych. W składowiskach administrowanych przez zakłady górnicze bilansowane są przychody i ubytki odpadów. Na składowiskach komunalnych i tzw. centralnych rejestruje się jedynie przyrosty odpadów pogórniczych. Do swoistych odpadów pogórniczych należą wody kopalniane. W systemie rejestrowane są wody kopalniane wraz z ich głównym składem chemicznym i ilością wód wykorzystanych lub zrzucanych do rzek.

Innym blokiem informacji w systemie MIDAS są bazy danych i system grafiki pozwalającej na prezentacje konturów złóż, obszarów górniczych, obszarów chronionych, obszarów koncesji poszukiwawczych itp. na mapach topograficznych. Wbudowane algorytmy pozwalają na transformacje współrzędnych w różnych układach państwowych jak "1965", "1942" a nawet z układów lokalnych.

Bazy systemu MIDAS zawierają ponad 600 000 rekordów. Rocznie przybywa ponad 40 tys. rekordów nowych lub modyfikujących zmiany. Do największych podbaz należą „punkty konturowe" (ponad 180 tys.) i „zasoby" (ponad 140 tys.).

Bazy danych **Rejestru Obszarów Górniczych** prowadzonego w PIG od 1995 r. zostały włączone jako podzbiory omówionego powyżej systemu MIDAS. Rejestrują one dane o nazwie obszaru i terenu górniczego, powierzchni, dacie ustanowienia, terminie ważności, adresie podmiotu gospodarczego wydobywającego kopalinę itp. Połączenie ze szczegółową informacją o złożu kopaliny pozwala na rozszerzenie informacji i jej jednoznaczne przypisanie do konkretnego złoża. Wykorzystanie kartometrycznych baz danych i grafiki umożliwia szczegółowe prezentacje konturów obszarów górniczych w stosunku do topografii i granic złoża.

Rejestr obszarów górniczych jest publicznie dostępny i stanowi podstawową informację o lokalizacji eksploatowanych złóż i zakładów górniczych.

Zbiory danych o produkcji, eksporcie i imporcie surowców mineralnych na świecie oraz cen surowców energetycznych, metalicznych, chemicznych i skalnych będących przedmiotem obrotu międzynarodowego gromadzone są w infosystemie **PRICESMIN** (**P**rodukcja **I** **CE**ny **S**urowców **MIN**eralnych). W bazach zestawiane są informacje pochodzące z notowań na giełdach światowych: LME, LBM, NYMEX, COMEX, a także - od 1996 r. - ceny z giełdy rosyjskiej obejmujące wyroby z żelaza, tytan, miedź i jej stopy, nikiel, aluminium, cynk, cyna. Rejestrowano są ceny z różnych giełd surowcowych oraz ceny na wolnym rynku i u producentów. Zastosowane algorytmy przeliczeniowe umożliwiają automatyczne transformacje cen w różnych walutach a przypisane deflatory zezwalają na

przedstawianie trendów w tzw. cenach stałych surowców mineralnych w stosunku do dowolnie wybranego roku.

W bazie gromadzone są również informacje dotyczące polskiego eksportu i importu surowców mineralnych z i do różnych krajów świata, uwzględniające zarówno kraje pochodzenia i przeznaczenia, jak również kraje zakupu i sprzedaży. surowców. Umożliwia to opracowywanie trendów w zakresie gospodarki surowcami mineralnymi i wykorzystywanie danych w analizie kierunków rozwoju przemysłu surowcowego oraz ocenie konkurencyjności gospodarki krajowej na tle rynku światowego. System dostarcza również informacji o możliwościach zaopatrzeniowych w zakresie surowcowym i umożliwia przeprowadzanie analiz wskazujących na potencjalne możliwości i kierunki rozwojowe gałęzi przemysłu opartych na krajowym wydobyciu. Opracowywane dane dotyczące eksportu i importu stanowią pożyteczne źródło informacji dla potencjalnych inwestorów krajowych i zagranicznych i są zestawiane w rozdziale p.t. Eksport i import surowców mineralnych w Polsce", publikowanym corocznie w „Bilansie zasobów kopalin i wód podziemnych w Polsce".

Zadaniem systemu **INFOGEOSKARB** jest gromadzenie danych o informacji geologicznej o złożach kopalin ze szczególnym uwzględnieniem wiadomości o prawach własności do informacji geologicznej.

Katalogowanie informacji geologicznej i danych o złożach jest niezbędne ze względu na zróżnicowanie praw własności zarówno do złóż jak i opisujących je dokumentacji geologicznych czyli do informacji geologicznej zarówno właścicieli dokumentacji, podstawy ich tytułu własności, jak również różnego rodzaju obciążenia praw do informacji geologicznych ciążących na właścicielach. System ma służyć stopniowemu porządkowaniu zagadnień związanych z prawem własności do informacji geologicznej, uwzględniający aktualny status prawny przedsiębiorstw zajmujących się poszukiwaniem, rozpoznawaniem i wydobywaniem surowców mineralnych.

INFOGEOSKARB działa w powiązaniu z istniejącym w PIG systemem MIDAS oraz z Centralną Bazą Danych Geologicznych. Z pierwszego z systemie otrzymuje informacje o złożach ich lokalizacji, kopalinach, koncesjach, itp., a z CBDG, gdzie gromadzone są m. in. informacje o dokumentacjach geologicznych archiwizowanych w Centralnym Archiwum Geologicznym, dane dotyczące dokumentacji jak tytuły opracowań, autorzy, data wykonania itp. Zespół tych informacji jest uzupełniany informacjami dotyczącymi praw własności.

Oddzielną sprawą jest baza danych geologicznych dostarczająca informacji do interpretacji geologicznej. **Centralna Baza Danych Geologicznych** jest zlokalizowana w Państwowym Instytucie Geologicznym. Podstawowe informacje o zawartości bazy są udostępnione w serwisie internetowym www.pgi.waw.pl.

W wyniku dotychczasowych prac analitycznych, projektowych, programistycznych i wdrożeniowych Centralna Baza Danych Geologicznych obejmuje obecnie następujące podsystemy:

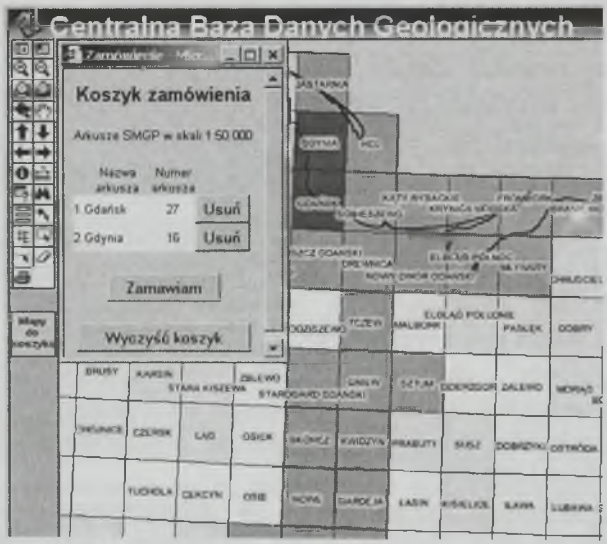
- Podsystem "Dokumenty", obsługujący dane katalogowe o archiwalnych opracowaniach geologicznych (w tym — kartach otworów wiertniczych i

dokumentacjach otworów głębokich), mapach publikowanych zgromadzonych w zbiorach Archiwum Map CAG, mapach cyfrowych, zdjęciach lotniczych i satelitarnych; podsystem używany jest również do rejestracji wypożyczeń tych dokumentów;

- Podsystem „Otwory", obsługujący informacje otworowe: dane lokalizacyjne i inne informacje podstawowe („metryki otworów"), profile stratygraficzne i litologiczne, informacje o historii prac na otworze, dane o uzysku rdzenia; podsystem umożliwia gromadzenie i udostępnianie szczegółowych informacji dotyczących litologii (w tym: dane o strukturach sedymentacyjnych i tektonicznych, położeniu warstw, dane paleontologiczne), występowania wód podziemnych i bituminów, opróbowania otworu, a także informacji typu technicznego (dane o metodach wiercenia, zarurowaniu i zafiltrowaniu otworu); podsystem „Otwory" został ostatnio rozszerzony o blok informacyjny dotyczący zasobów rdzeni wiertniczych w magazynach rdzeni Centralnego Archiwum Geologicznego;
- Podsystem „Punkty badawcze", ściśle związany z podsystemem „Otwory", przeznaczony do obsługi informacji pochodzących z powierzchniowych punktów obserwacyjnych, jak odsłonięcia, wkopy itp.; podsystem umożliwia gromadzenie zarówno danych lokalizacyjnych o takich punktach, jak i szczegółowych informacji geologicznych (litologia, stratygrafia);
- Podsystem „Geofizyka wiertnicza" przeznaczony do gromadzenia i udostępniania danych geofizyki wiertniczej, w tym — graficznej wizualizacji krzywych karotażowych;
- Podsystem „Analizy", obsługujący szeroki zakres danych analitycznych i pomiarowych będących wynikami badań wykonanych w terenie (w otworach wiertniczych lub nieotworowych punktach badawczych) lub wynikami późniejszych badań pobranych próbek;
- Podsystem „Złoża", obsługujący katalog złóż systemu MIDAS wraz z informacjami o kopalinach złóż, ich budowie geologicznej i zasobach, a także o pierwiastkach lub związkach chemicznych uwzględnionych w krajowym bilansie zasobów kopalin;
- Podsystem „Kolekcje geologiczne" (w przygotowaniu), umożliwiający katalogowanie w CBDG zbiorów próbek, preparatów i okazów geologicznych;
- Podsystem przestrzenny (GIS) wraz z przeglądarką geograficzną CBDG, umożliwiającą wizualizację lokalizacji obiektów (otworów, złóż itp.) opisanych w CBDG, ich wyszukiwanie i wybieranie na mapach wyświetlanych na ekranie;
- Przeglądarkę profilową CBDG, umożliwiającą graficzną prezentację i przeglądanie profilów otworów, przy czym litologia, stratygrafia oraz inne dane geologiczne i techniczne przedstawiane są za pomocą symboli graficznych; przeglądarka profilowa obsługuje także wizualizację krzywych karotażowych, gromadzonych w podsystemie CBDG „Geofizyka wiertnicza";
- Przeglądarkę tabelaryczną CBDG, przeznaczoną do prezentacji danych, wyszukanych w CBDG, w formie tabel, z możliwością dodatkowego

przeszukiwania, filtrowania, sortowania, zmiany szerokości i kolejności kolumn, wydruku;

- Aplikację „Administracja", przeznaczoną do obsługi słowników bazy przez administratorów CBDG;
- Aplikację internetową, obejmującą zarówno dane tekstowe i materiał ilustracyjny o Centralnej Bazie Danych Geologicznych, jak i mechanizm wyszukiwania danych według atrybutów oraz według lokalizacji obiektów na ekranie, wizualizowanych w prostej przeglądarce GIS. Aplikacja internetowa wyposażona jest także w formularze zapytań, umożliwiające wyszukiwanie danych według wartości ich atrybutów. Witryna internetowa CBDG zawiera także roboczą wersję słownika jednostek litostratygraficznych Polski (z możliwością przeszukiwania przez użytkowników), a ponadto jest połączona z serwisem systemu INFOGEOSKARB, opartego na CBDG i na bazie MIDAS. Obecnie aplikacja internetowa CBDG jest rozszerzana o warstwy skorowidzowe map SMGP, MHP i MGGP, a także o opcje umożliwiające zamawianie wybranych przez użytkownika arkuszy tych map przez Internet.



Rys. 8. Fragment przeglądarki geograficznej CBDG z funkcją zamawiania seryjnych map geologicznych z wykorzystaniem Internetu.

Podsumowanie

Z przedstawionych przykładów wynikają następujące wnioski:

- Wykorzystanie systemów informatycznych w geologii jest szeroko stosowane.
- W geofizyce informatyka jest nieodzowna dla prac interpretacji i powiązania uzyskanych wyników z geologią

- W kartografii geologicznej, dzięki informatyce uzyskujemy łatwy dostęp do map, a co istotniejsze powstały możliwości łatwego dokonywania bieżących aktualizacji.
- Hydrogeologiczna baza danych jako jedna z pierwszych baz danych geologicznych udowodniła potrzebę stosowania metod informatycznych.
- Nieodzowne jest wykorzystanie baz danych dotyczących problematyki geologii gospodarczej i rejestru obszarów górniczych.
- Centralna Baza Danych Geologicznych spełnia rolę zbioru informacji, wymaga dalszego wypełniania i rozwoju.
- Perspektywicznie należy rozszerzać stosowanie systemów informatycznych w geologii we wszystkich dziedzinach aktualizując dane i wprowadzając nowe oprogramowania.

Dr hab. Krzysztof Szamałek – Ministerstwo Środowiska
Sekretarz Stanu – Gł. Geolog Kraju
00-922 Warszawa, ul. Wawelska 52/54 tel. 0-22-57 92 337

Mgr Waldemar Gogołek –
Dr Bogusław Kazimierski –
Doc. Dr Stanisław Przeniosło –
Doc. Dr Stefan Młynarski –
Państwowy Instytut Geologiczny
00-975 Warszawa, ul. Rakowiecka 4
Tel. 0-22- 849 5351

KOMPUTEROWY SYSTEM BIBLIOTECZNY APIS-ZB BIBLIOTEKI GŁÓWNEJ POLITECHNIKI CZĘSTOCHOWSKIEJ

Lidia SZCZYGŁOWSKA

Świat zmierza do stworzenia społeczeństwa informacyjnego, nazywanego niekiedy społeczeństwem globalnej informacji. Ważnym wyróżnikiem społeczeństwa informacyjnego są metody komputerowego wspomagania nauczania i tu niebagatelną rolę odgrywają biblioteki.

W ostatnich latach obserwujemy duże zmiany w polskich bibliotekach akademickich, związane między innymi z rozwojem techniki komputerowej i sieci komputerowych. Sieci stały się doskonałym medium komunikacyjnym, pozwalając na szybkie porozumiewanie się oraz szeroki dostęp do informacji naukowej gromadzonej w bazach danych. Jedne z największych baz powstających w bibliotekach naukowych to katalogi biblioteczne. Do ich tworzenia wykorzystywane są zintegrowane systemy biblioteczne.

Decyzja Biblioteki Głównej Politechniki Częstochowskiej o zakupie bibliotecznego systemu komputerowego APIS-ZB zapadła po długich konsultacjach, gdzie nie bez znaczenia były możliwości finansowe. Za przykładem Biblioteki Głównej Politechniki Warszawskiej jako Pierwszej Centralnej Biblioteki Technicznej zdecydowaliśmy się na zakup systemu komputerowego autorstwa informatyków zatrudnionych w Bibliotece Głównej Politechniki Gdańskiej. Ścisłe kontakty pomiędzy ośrodkami w których proces automatyzacji bibliotek akademickich jest daleko zaawansowany, a takimi w których się dopiero rozpoczyna, pozwala zarówno na sprawne wprowadzenie nowych programów, jak też rozszerzenie programów już wdrożonych. Efektem takiego podejścia jest uniknięcie błędów merytorycznych oraz minimalizacja nakładów czasu i pracy.

1. Krótka charakterystyka systemu APIS-ZB

System APIS-ZB komputeryzuje w pełnym zakresie działalność biblioteki dotyczącą obsługi czytelników, począwszy od katalogowania książek, a kończąc na ich wypożyczeniu. System pozwala też na usprawnienie wielu czynności związanych z funkcjonowaniem służb bibliotecznych. Ponadto umożliwia bieżące śledzenie pracy całej biblioteki pod kątem realizacji zamówień złożonych przez czytelników na wypożyczenie, dostarczając informacji niezależnie od typowej statystyki bibliotecznej.

Niewielkim nakładem finansowym udało nam się stworzyć sieć komputerową Novell NetWare w której APIS-ZB praktycznie funkcjonuje do dziś. Rozwój systemu

jest permanentny i prowadzony w kierunku rozwoju systemu oraz rozwoju usług sieciowych.
System APIS-ZB składa się z następujących podsystemów:

- KANAB
  podsystem realizuje funkcje związane z katalogowaniem książek. Opisy katalogowe tworzone są zgodnie z międzynarodowym standardem USMARC.
- KATALOG KSIĄŻEK
  podsystem udostępnia czytelnikowi komplet funkcji pozwalający na samodzielny dostęp do katalogu komputerowego i rezerwację książek do wypożyczenia, z możliwością sprawdzenia własnego konta czytelniczego.
- WYPOŻYCZENIA
  podsystem realizuje wszystkie funkcje związane z obsługą czytelników. Rejestruje dane o czytelniku, dane o wypożyczeniach, a także pozwala na administrowanie systemem.
- MAGAZYN
  Realizuje funkcje związane z pracą magazynu
- CZASOPISMA
  podsystem realizuje wszystkie funkcje akcesji i tworzenia katalogu komputerowego czasopism i wydawnictw ciągłych zgodnie ze standardem USMARC.
- KATALOG CZASOPISM
  moduł umożliwiający wyszukiwanie w katalogu czasopism i wydawnictw ciągłych

Komputerowy system APIS-ZB umożliwia szybkie wyszukiwanie książki oraz jej rezerwację bez konieczności wypisywania rewersu. Czytelnicy zarejestrowani są w bazie komputerowej i mają samodzielny dostęp do katalogu i swojego konta czytelniczego, którego stan mogą w każdej chwili sprawdzić. Dostęp do kont czytelniczych zabezpieczony jest kodem bezpieczeństwa, nadawanym przez system w momencie rejestracji czytelnika i jego zmiana może być dokonana tylko przez bibliotekarza na życzenie czytelnika.

2. Korzystanie z systemu

System APIS –ZB udostępnia menu z różnymi funkcjami. Informacje wyświetlane u dołu ekranu wskazują zakres funkcji możliwych do zrealizowania.

Wyszukiwanie możliwe jest poprzez INDEKSY: autorski, tytułów, haseł przedmiotowych, UKD, serii, nazw konferencji, roku wydania, itd. Jest to najprostszy sposób wyszukiwania, dostępny w funkcji WYDSZUKIWANIE PROSTE. Tryb wyszukiwania poprzez PYTANIE odbywa się wg warunków określonych przez czytelnika. Stopień złożoności pytania może być, zależnie od potrzeb, dowolnie wysoki z zastosowaniem operatorów, logicznych (alternatywy, koniunkcji,

wyłączenia), opcji fleksji (różnych końcówek nazw), czy opcji maski (wyboru nazw zawierających ciąg określonych znaków). Proces formułowania PYTANIA ułatwiony jest poprzez możliwość podglądu i wyboru haseł z INDEKSÓW. Ten zaawansowany sposób odszukiwania książek dostępny jest w funkcji WYSZUKIWANIE ZŁOŻONE.

Wyszukane opisy użytkownik może zobaczyć na ekranie z pełną informacją dotyczącą „statusu zajętości" poszukiwanej książki tzn. czy książka jest wypożyczona, czy jest dostępna w Czytelni Głównej lub Czytelni Wydziałowej. Jeśli książka jest dostępna czytelnik dokonuje rezerwacji i automatycznie rewers drukowany jest w Magazynie, jeśli książka nie jest dostępna system zaproponuje ustawienie w kolejce do wypożyczenia z informacją o miejscu zajmowanym w kolejce. Ponadto czytelnik otrzyma informację o stanie realizacji zamówień, a więc czy zamówienie jest dopiero w Magazynie, czy też książki oczekują już w Wypożyczalni.

APIS-ZB jest systemem łatwym w obsłudze i stosowaniu i jak wynika z doświadczeń współpracujących z nami bibliotek oraz naszych obserwacji nauka posługiwania się systemem przez czytelników trwa bardzo krótko.

3. Zarządzanie pracą systemu

Eksploatacja systemu APIS-ZB sterowana jest poprzez zbiór parametrów definiowany w momencie rozpoczynania prac wdrożeniowych w bibliotece i może być zmieniana w trakcie eksploatacji systemu zależnie od potrzeb biblioteki. Zmiany mogą być wprowadzane globalnie, a także indywidualnie dla określonego czytelnika. Jest to szczególnie przydatne w przypadku zmian w regulaminie biblioteki, gdyż mogą one być natychmiast wprowadzane dla wszystkich czytelników. System ułatwia pracę personelowi biblioteki kontrolując kompletność danych i ich wzajemne relacje, a także w każdej chwili dostarcza informacje dotyczące realizacji zamówień złożonych przez użytkowników, co pozwala na bieżące śledzenie pracy biblioteki

4. Tworzenie baz danych

Prace z systemem APIS-ZB rozpoczęliśmy dwutorowo, tworząc

- katalogową bazę danych w podsystemie KANAB
- bazę kont czytelniczych wraz ze stanem wypożyczeń.

Jedną z pierwszych decyzji towarzyszących wyborowi zintegrowanego systemu bibliotecznego była decyzja dotycząca wyboru formatu USMARC. W dobie połączeń sieciowych stosowanie przez biblioteki tego samego formatu katalogowania ma istotne znaczenie. Przyjęcie jednolitych rozwiązań odnośnie formatu opisu bibliograficznego pozwala na uzyskanie pełnej wymienności danych.

Katalogowe bazy danych zwykle rozpoczyna się budować od jakiejś ustalonej daty, na ogół przyjmuje się datę wydania dokumentu albo datę wpływu do biblioteki,

bądź do oddziału opracowania. My wybraliśmy za podstawę datę wpływu do opracowania. Ze względu na strategię wdrażania modułu udostępniania za niezbędną uznaliśmy pełną retrokonwersję księgozbioru studenckiego. Użytkownicy naszej biblioteki najczęściej korzystają z pozycji tego katalogu, a także z pozycji najnowszych.

5. Wdrażanie systemu

Prace wdrożeniowe przebiegały w dwóch etapach:

- próbna eksploatacja systemu
- pełne wdrożenie

Przed przystąpieniem do próbnej eksploatacji systemu wymagane jest stworzenie dwóch baz danych:

- bazy kartoteki czytelników zawierającej dane personalne wraz ze stanem kont
- bazy katalogowej wydawnictw zwartych

Eksploatacja próbna polegała na wykonywaniu wszystkich funkcji systemu równolegle z obsługą tradycyjną użytkowników. Pozwoliło to na wyeliminowanie ewentualnych błędów w zgromadzonych danych, ujawnieniu niedociągnięć organizacyjnych, a także dało możliwość przyswojenia przez pracowników biblioteki wszystkich funkcji systemu.

System w momencie pełnego wdrożenia powinien cechować się wysoką niezawodnością, kompletnością i wiarygodnością danych. Wszelkie powiązania między bazą katalogową, oraz bazą kont czytelniczych powinny być jednoznaczne i kompletne.

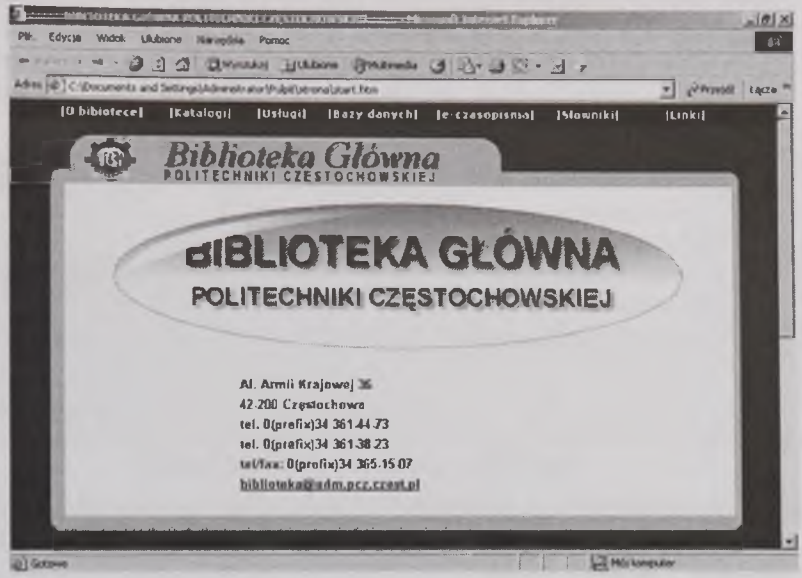
6. Stan aktualny komputeryzacji

W dniu dzisiejszym użytkownicy Biblioteki Głównej maja dostęp w sieci lokalnej do KATALOGU książek od 1994 roku oraz wszystkich podręczników gromadzonych w Bibliotece. Czas oczekiwania na realizacje zamówienia został w dużym stopniu przyspieszony poprzez wyeliminowanie rewersów tradycyjnych i możliwości sprawdzenia stanu zajętości książki. Zmieniła się także jakość obsługi czytelnika, co umożliwiło wprowadzenie oznaczenia książek sygnaturami z kodami paskowymi oraz szybkiej identyfikacji czytelnika dzięki legitymacji z kodem paskowym.

Czytelnik ma także dostęp do katalogu CZASOPISM, w którym wyszukiwanie jest oparte na podobnych zasadach jak katalogu druków zwartych.

Kolejnym ułatwieniem jest internetowy dostępu do katalogów komputerowych książek i czasopism naszej biblioteki z Witryna Biblioteki Głównej Politechniki Częstochowskiej **www.bg.pcz.czest.pl**. Poprzez stworzenie

własnej strony WWW Biblioteka i jej zasoby stały się integralną częścią globalnej sieci, a także źródłem bieżącej i wiarygodnej informacji.



Rys. 1. Strona internetowa Biblioteki głównej Politechniki Częstochowskiej

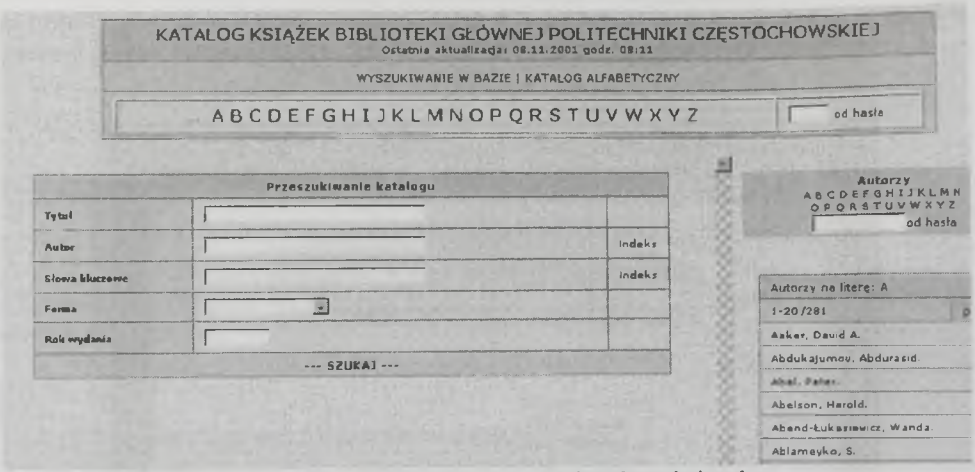
Interfejs katalogu APIS-ZB pod WWW wymagał stworzenia narzędzi importu danych z Novella z oprogramowania sieci lokalnej APIS-ZB, stworzenia bazy danych na nowym systemie, następnie stworzenie witryny internetowej dla bazy danych. System pracuje na serwerze Windows 2000, gdzie zainstalowano serwer Apache charakteryzujący się dużą odpornością na wirusy.

KATALOG KSIĄŻĘK oraz KATALOG CZASOPISM zawiera opisy dostępnych w Bibliotece woluminów i czasopism bazując na opisach znajdujących się w bazie danych APIS-ZB (Novell). Aktualność danych (ilość opisów, stan egzemplarzy w magazynie) uzależniona jest od przeprowadzanych aktualizacji - stan prezentowany w opisach rekordów datowany jest na czas ostatniej aktualizacji.

System umożliwia przeszukiwanie bazy danych książek w dwojaki sposób:

- z wykorzystaniem formularza, w którym podajemy dane dotyczące poszukiwanej książki (autor, tytuł, fragment tytułu, słowo kluczowe itp.) – WYSZUKIWANIE W BAZIE

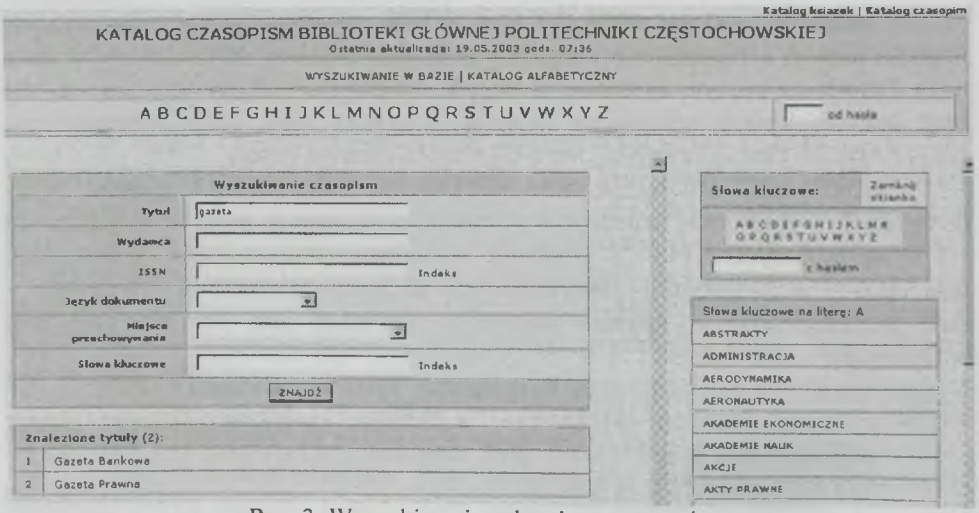
- poprzez KATALOG ALFABETYCZNY oraz przeglądanie tytułów od hasła



Rys. 2. Wyszukiwanie w katalogu książek

W przypadku KATALOGU CZASOPISM, wyszukiwanie odbywa się także:

- poprzez formularz podając dane dotyczące czasopisma takie jak (tytuł, czy fragment tytułu, słowo kluczowe ewentualnie wydawcę) – WYSZUKIWANIE W BAZIE
- poprzez KATALOG ALFABETYCZNY oraz przeglądanie tytułów od hasła



Rys. 3. Wyszukiwanie w katalogu czasopism

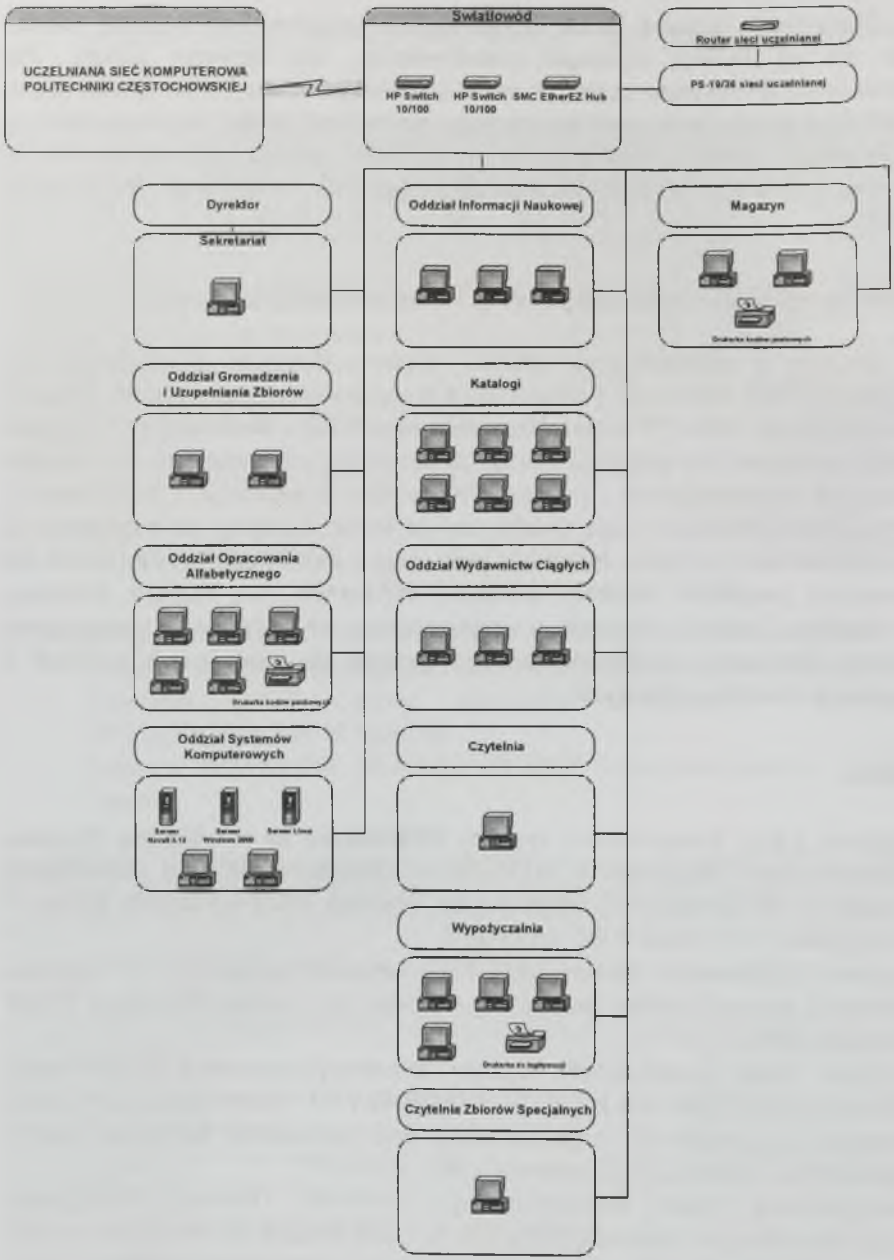
Zamawianie książek przez użytkowników jest możliwe poprzez telnet. Interfejs do tej funkcji wymagał posadowienia na serwerze Linux.. Po załogowaniu się do systemu czytelnik może dokonać rezerwacji wyszukanej przez siebie pozycji po podaniu numeru swojego konta oraz kodu bezpieczeństwa i zaakceptowaniu wyboru. Funkcja ta umożliwia zdalne przeszukiwanie i ewentualną rezerwacje wybranych pozycji dostępnych w katalogu Novllowym APIS-ZB.

7. Biblioteka wirtualna jako kolejny etap komputeryzacji biblioteki

Decyzja o udostępnianiu pełnych tekstów skryptów uczelnianych na stronie internetowej Biblioteki Politechniki Częstochowskiej już zapadła. Była to wspólna inicjatywa Władz Uczelni, Biblioteki oraz Działu Wydawnictw. Obecnie projekt jest na etapie prac przygotowawczych obejmujących konkretne rozwiązania organizacyjne, technologiczne i prawne. Prowadzimy konsultacje z bibliotekami, które mają doświadczenie i dużą wiedzę na ten temat. Liczymy na współpracę z innymi bibliotekami, a także bierzemy pod uwagę możliwość przystąpienia do realizowanych projektów budowy bibliotek cyfrowych. Na bieżąco śledzimy proces tworzenia kolekcji e-książek, a użytkownikom odwiedzającym naszą stronę internetową oferujemy możliwość łatwego dostępu do istniejących polskich i zagranicznych zbiorów cyfrowych.

Literatura

1. Zięborak Lech: Komputerowe systemy biblioteczne na przykładzie Systemu Udostępniania Wydawnictw APIS-ZB w Bibliotece Głównej Politechniki Gdańskiej. W: Działalność informacyjna bibliotek szkół wyższych. Kielce – Ameliówka, 25-27 maja 1994, s.123-130
2. Ligman J., Zięborak l.: System APIS-ZB – perspektywy rozwoju. W: Zadania biblioteki wyższej uczelni technicznej wczoraj, dziś i jutro. Warszawa, 07-08 listopada 1995, s. 75-78
3. Zięborak Lech: Komputerowy system informacji naukowej w Bibliotece Głównej Politechniki Gdańskiej. W: INFOBAZY' 97. Bazy danych dla nauki. Materiały z konferencji organizowanej pod patronatem Komitetu Badań Naukowych. Gdańsk, 23-25 czerwca 1993, s. 284-292
4. Szczygłowska Lidia: Komputeryzacja Biblioteki Głównej Politechniki Częstochowskiej w systemie APIS-ZB. W: INFOBAZY'99. Bazy danych dla nauki. Materiały z konferencji organizowanej pod patronatem Komitetu Badań Naukowych Gdańsk, 30 sierpnia–01 września 1999, s. 193-197



SCHEMAT SIECI KOMPUTEROWEJ BIBLIOTEKI GŁÓWNEJ POLITECHNIKI CZĘSTOCHOWSKIEJ

mgr inż. Lidia Szczygłowska
Kustosz, Z-ca Dyrektora
Dyrektora Biblioteki Głównej
ds. Komputeryzacji
Politechnika Częstochowska

ROZDZIAŁ 5

METODY MATEMATYCZNE

ROZMYTA OPTYMALIZACJA DZIAŁALNOŚCI DYSTRYBUTORA

Marek DOLATA, Ludmiła DYMOWA, Janusz GRABARA,

Streszczenie. W zagadnieniu optymalizacji działalności dystrybutora uwzględniono nie tylko koszty transportu ale ograniczenia związane z możliwością spełnienia kontraktu zawartego pomiędzy dystrybutorem i odbiorcami oraz między dystrybutorem i dostawcami. W odróżnieniu do podejść klasycznych do formalizacji istniejących niepewności za pomocą metod probabilistycznych użyte zostały elementy teorii zbiorów rozmytych pozwalające na uwzględnienie nie tylko obiektywnych informacji otrzymanych za pomocą statystycznej obróbki danych, ale wiedzy i intuicji specjalistów z dziedziny oraz decydentów.
Dla rozwiązania problemu użyto rozmytego uogólnienia tradycyjnej metody programowania liniowego simplex za pomocą programowania obiektowego. Przy tym w realizacji operacji arytmetycznych na liczbach rozmytych użyto procedury ich przedstawienia w postaci sieci α-przekrojów. Istotnym problemem w realizacji tego podejścia jest porównywanie liczb rozmytych, dlatego używano oryginalnej procedury opartej na probabilistycznej interpretacji przedziałów ostrych i rozmytych. Rezultaty testów porównywano z analogicznymi otrzymanymi za pomocą procedury losowania Monte-Carlo. Ciekawym wynikiem zastosowania podejścia Monte-Carlo jest niejednoznaczność rezultatów co sprawia problemy interpretacji wyników. Udowodniono, że jednoznaczna interpretacja tego wyniku może być uzyskana za pomocą podejścia rozmyto-przedziałowego.

Kluczowe słowa: Problem dystrybutora; programowanie liniowe rozmyte; problem transportowy romyty; metoda Monte-Carlo.

1. Wprowadzenie

Zagadnienie optymalizacji działalności dystrybutora może być sformułowane jako uogólnienie klasycznego problemu transportowego. Konwencjonalny problem transportowy jest specjalnym typem programowania liniowego gdzie sam problem jak i ograniczenia są opisane szczególną matematyczną strukturą. Źródłem dostaw może być producent, magazyn itp. dla którego mamy przypisane odpowiednie parametry, podobnie jak dla celu dostaw; dodatkowo znane są koszty transportu na danych trasach. W klasycznym podejściu chodzi o minimalizację kosztów poniesionych przez pośrednika podczas transportowania towaru od M producentów do N konsumentów, w opisywanym podejściu postanowiono dokonać maksymalizacji zysku dystrybutora przy tych samych warunkach.

W 1979, Isermann [1] przedstawił algorytm, dla rozwiązywania problemu, przy pomocy którego został wyliczony komplet wszystkich skutecznych

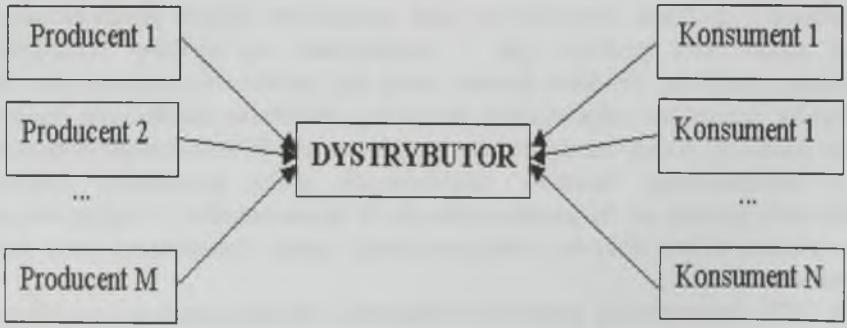
rozwiązań. Ringuest i Rinks [2] proponowali dwa algorytmy iteracyjne dla rozwiązywania liniowego wielokryterialnego problemu transportowego. Podobne rozwiązanie zaproponowane w [3].

Różne efektywne algorytmy zostały opracowane dla tego problemu transportowego ale z uwzględnieniem stałych parametrów zadania opisanych w postaci liczb rzeczywistych. Jednakże takie warunki są spełnione rzadko albo niemalże nigdy ze względu na wahania parametrów; przykładowo ciężko ustalić stały koszt dla określonej trasy. W pracy [4] taki problem był rozwiązany w warunkach przedziałowej niepewności kosztów transportowych.
W pracach S.Chanasa i D. Kuchty [5, 6] rozwinięto podejście oparte na rozmyto-przedziałowym przedstawieniu niepewnych parametrów modelu. Rozwój tego podejścia przedstawiony jest w pracy [7].

Ogólną charakterystyką omówionych prac jest wprowadzenie ograniczeń dotyczących formy funkcji przynależności. To pozwala autorom przekształcić pierwotny problem rozmytego programowania liniowego w sieć zwykłych zadań programowania liniowego za pomocą procedur analitycznych. W praktyce jednak funkcje przynależności opisujące parametry niepewne używanych modeli mogą mieć dość skomplikowane formy, oprócz tego istotnym momentem opracowania algorytmów programowania rozmytego jest niezbędność porównywania liczb rozmytych. Istnieje wiele podejść do tego ale będziemy używali podejścia probabilistycznego [8, 9] pozwalającego za pomocą tylko jednego zupełnie naturalnego założenia stwierdzającego, że przedział jest przedziałem liczby losowej ze stałą gęstością prawdopodobieństwa, otrzymać cały zbiór operacji porównywania przedziałów ostrych, przedziałów rozmytych (z użyciem α-przekrojów) oraz przedziałów i liczb rzeczywistych. Proponowane podejście pozwala na bezpośrednie rozmyte rozszerzenie klasycznego algorytmu simplex z implementacją metody za pomocą programowania obiektowego.

2. Postać matematyczna problemu

Problem dystrybutora można zdefiniować przyjmując, że pośrednik zaopatruje się u M producentów i dostarcza towar do N konsumentów Rys1.



Rys. 1. Graficzna prezentacja problemu dystrybutora

Założono, że wiadome są maksymalne możliwości producentów dotyczące ilości wyprodukowanego surowca wynoszące a_i ,(i=1,2,..., M) i maksymalne zdolności odbioru towarów przez konsumentów b_j, (j=1, 2,...,N). Dystrybutor posiada informację o cenach za jednostkę towaru który kupuje u każdego producenta i o cenach sprzedaży dla każdego konsumenta. Wiadomo, że straty na dostarczenie jednostki towaru od *i-tego* producenta do *j*-tego konsumenta są równe c_{ij}, (i=1,2,..., M; j=1, 2,...,N). Zgodnie z zawartymi umowami dystrybutor jest zobowiązany kupować u *i*-tego producenta minimum p_i jednostek towaru po cenie t_i za jednostkę oraz gwarantować dostarczenie *j*-temu konsumentowi minimum q_j jednostek tego towaru po cenie s_j za jednostkę. Całą ilość towaru powyżej omówionej w kontrakcie wartości p_i, dystrybutor kupuje po cenie promocyjnej k_i za jednostkę. Z kolei konsument kupuje całą ilość towaru powyżej q_j także po cenie promocyjnej r_j za jednostkę. Rozwiązaniem problemu są optymalizowane ilości towaru kupowanego u każdego *i*-tego producenta oraz dostarczonego i sprzedanego *j*-temu konsumentowi x_{ij} (i=1,2,..., M; j=1, 2,...,N) w warunkach ograniczeń związanych z podpisanymi umowami z producentami i konsumentami dotyczącymi ilości kupna i sprzedaży.

W wyniku dochód dystrybutora ***D*** może być przedstawiony przez wyrażenie

$$D=\sum_{j=1}^{N}\left(q_j * s_j\right)-\sum_{i=1}^{M}\left(p_i * t_i\right)-\sum_{i=1}^{M}\sum_{j=1}^{N}\left(c_{ij} * x_{ij}\right)+$$
$$+\sum_{j=1}^{N}\left(\sum_{i=1}^{M} x_{ij}-q_j\right) * r_j-\sum_{i=1}^{M}\left(\sum_{j=1}^{N} x_{ij}-p_i\right) * k_i \qquad (1)$$

W rezultacie zadanie redukuje się do znalezienia wszystkich x_{ij} maksymalizujących dochód ***D*** przy ograniczeniach:

- dotyczących górnych granic popytu i podaży

$$\sum_{j=1}^{N} x_{ij} \le a_i (i=1..M); \qquad \sum_{i=1}^{M} x_{ij} \le b_j (j=1..N);$$

(2)

- dotyczących dolnych granic popytu i podaży

$$\sum_{j=1}^{N} x_{ij} \ge p_i (i=1..M); \qquad \sum_{i=1}^{M} x_{ij} \ge q_j (j=1..N); \qquad (3)$$

Sformułowany problem można rozwiązać jako zadanie programowania liniowego przy wykorzystaniu algorytmów programowania liniowego np. simplex.

Uwzględniając, że wszystkie parametry w (1)-(3) są danymi niepewnymi będziemy przedstawiali je za pomocą liczb rozmytych o trapezoidalnej formie. Wtedy

zagadnienie optymalizacji działalności dystrybutora może być przedstawione w formie:

$$\hat{D} = \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\hat{z}_{ij} * \hat{x}_{ij} \right) \rightarrow \max \qquad (4)$$

- ograniczenia dotyczące górnych granic popytu i podaży

$$\sum_{j=1}^{N} x_{ij} \leq \hat{a}_i (i = 1..M); \qquad \sum_{i=1}^{M} x_{ij} \leq \hat{b}_j (j = 1..N); \qquad (5)$$

- ograniczenia dotyczących dolnych granic popytu i podaży

$$\sum_{j=1}^{N} x_{ij} \geq \hat{p}_i (i = 1..M); \qquad \sum_{i=1}^{M} x_{ij} \geq \hat{q}_j (j = 1..N); \qquad (6)$$

Gdzie $\hat{z}_{ij} = \hat{r}_j - \hat{k}_i - \hat{c}_{ij}$ dla każdego i=1..M, j=1..N.

W wyrażeniach (4)-(6) $\hat{D}$, $\hat{z}_{ij}$, $\hat{a}$, $\hat{b}$, $\hat{q}$, $\hat{p}$ są liczbami rozmytymi.

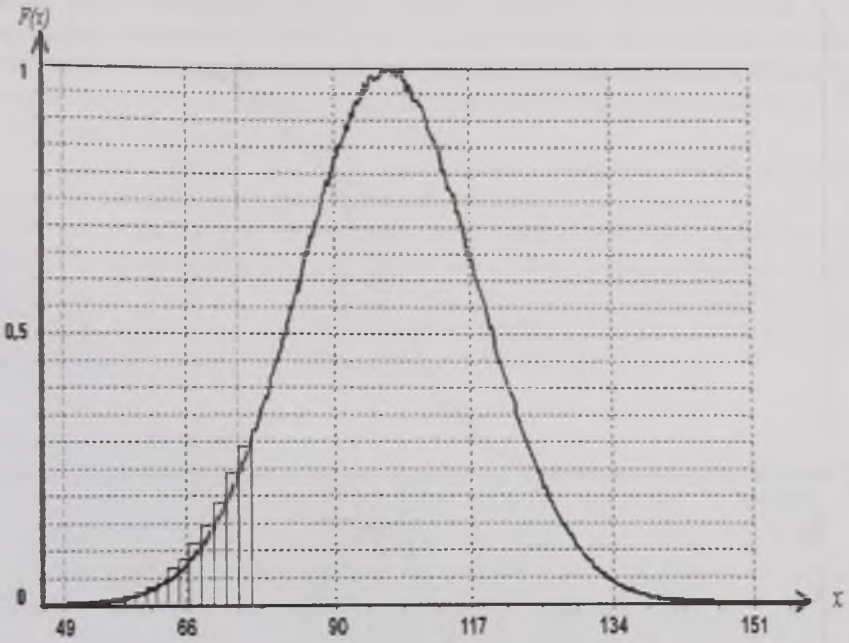
W rezultacie otrzymamy zagadnienie maksymalizacji rozmytego dochodu (4) w warunkach ograniczeń (5) i (6).

W praktyce często mamy problem związany z różnymi dokładnościami przedstawienia danych niepewnych np. część opisanych powyżej parametrów może być przedstawiona w postaci trapezoidalnych liczb rozmytych na podstawie opinii ekspertów. Druga część może mieć postać np. histogramu lub gęstości prawdopodobieństwa w dość skomplikowanej formie otrzymaną w wyniku badań statystycznych. W tych wypadkach zasady ogólno metodologiczne sugerują przekształcenie wszystkich danych do formy o najmniejszym poziomie dokładności. Z tego tez względu istnieje potrzeba transformacji danych przedstawionych w postaci rozkładu prawdopodobieństwa lub histogramu do funkcji przynależności do liczby rozmyto-przedziałowej.

Aby uzyskać dane wejściowe w postaci liczb rozmyto-przedziałowych należy najpierw zastosować algorytm budujący funkcję przynależności na podstawie gęstości zmiennych losowych, jeżeli takie istnieją lub bezpośrednio na podstawie histogramu. Najczęstszą spotykaną w praktyce formą przedstawiania niepewności związanej z jakimś parametrem decyzyjnym jest podanie wartości średniej i odchylenia standardowego σ, które to parametry łatwo przedstawić w postaci rozkładu prawdopodobieństwa.

W opisywanym przypadku zastosowano algorytm, który pozwala na przejście od wartości podanej w postaci dowolnego histogram lub rozkładu gęstości zmiennej losowej do liczby rozmyto-przedziałowej. Otrzymana w wyniku liczba jest przedstawiona w postaci czwórki liczb reprezentujących liczbę rozmyto-przedziałową w formie trapezu.

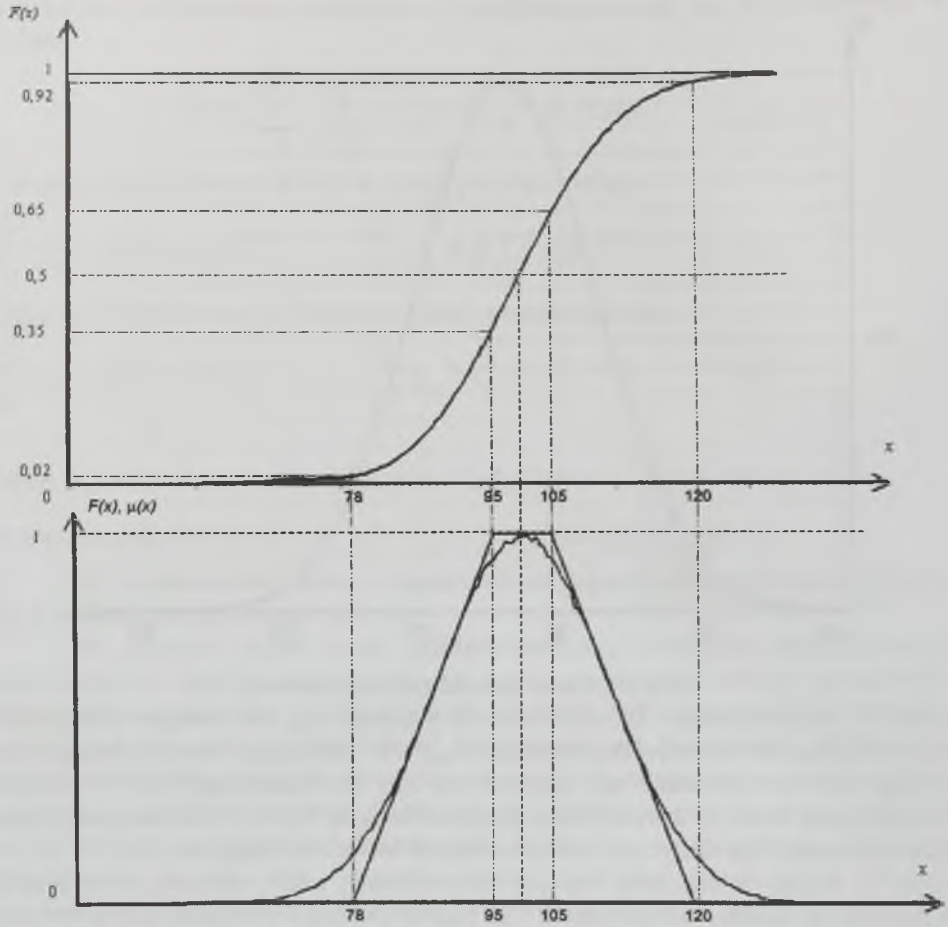
Dla rozkładu prawdopodobieństwa zmiennej losowej przedstawionego na rysunku Rys. 2 wykonujemy następujące kroki algorytmu:



Rys. 2 Rozkład gęstości zmiennej losowej.

Krok 1. Rozpoczynając od najmniejszej wartości x_{min} (w naszym przykładzie x_{min}=50) dążąc do wartości maksymalnej x_{max}=151 obliczamy wartość funkcji $F(x_i)$ równą polu powierzchni pod krzywą od x_{min} do aktualnego x_i ; w wyniku otrzymujemy funkcję skumulowaną przedstawioną na Rys. 3. Wiadomo, że funkcja skumulowana $F(x_i)$ faktycznie jest prawdopodobieństwem tego, że $x<x_i$.

Krok 2. Mając wyznaczoną funkcję skumulowaną należy wyznaczyć w sposób subiektywny zależny od decydenta i od rozpatrywanego problemu cztery wartości $F(x_i)$ dla $i=0,..3$, które określają po zrzutowaniu na oś odciętych dolne i górne przedziały ufności dla liczby rozmytej. Dla przedstawianego przypadku dobrano wartości zgodnie z rysunkiem Rys. 4. W rezultacie na Rys. 4 przedział [95, 105] odpowiada prawdopodobieństwu 30%; przedział [78, 120] prawdopodobieństwu 90%



Rys 3. Realizacja przejścia od funkcji skumulowanej do liczby rozmytej.

Krok 3. Wyznaczone po zrzutowaniu wartości z osi odciętych zapisujemy jako liczbę rozmyto-przedziałową w postaci uporządkowanej czwórki liczb. Dla przykładu:

$\hat{a}$=[78, 95, 105, 120]

Z rysunku Rys. 3 dodatkowo można ocenić poprawność odwzorowania rozkładu gęstości zmiennej losowej na liczbę rozmyto-przedziałową. Można stwierdzić, że dokładność tego odwzorowania zależy od doboru odpowiednich wartości dla funkcji skumulowanej $F(x_i)$ dla $i=0,..3$ określających przedziały ufności liczby rozmyto-przedziałowej.

Trzeba także zauważyć, że ze względu na to, iż teoretycznie krzywa reprezentująca rozkład gęstości zmiennej losowej zawiera się od $+\infty$ do $-\infty$ to należy wybrać punkty, względem których ograniczamy dolny przedział ufności liczby rozmyto-przedziałowe.

Wartości $F(x_i)$ określające górny i dolny przedział ufności zostały wybrane symetrycznie względem środka krzywej skumulowanej. Ich rozpiętość jest zależna od decydenta, dobrana jednak tak aby jak najdokładniej odwzorować funkcję przedstawiającą rozkład gęstości zmiennej losowej i ustalić odpowiednią szerokość dla przedziału o największym stopniu ufności.

Opisana metoda pozwala przedstawić wszystkie dane niepewne w jednolitej formie trapezoidalnych liczb rozmytych.

Proponowana metoda rozwiązania zagadnienia programowania rozmytego (4)-(6) realizowana za pomocą przedstawienia wszystkich liczb rozmytych w postaci zbiorów odpowiednich α-przekrojów, faktycznie redukuje zagadnienie rozmyte w sieć zagadnień programowania ostro-przedziałowego z wykorzystaniem probabilistycznej metody porównywania przedziałów [8, 9].

3. Przykład numeryczny.

Dla uproszczenia i przejrzystości rezultatów przypuśćmy że mamy tylko trzech producentów i trzech producentów M=3; N=3.

W celu umożliwienia porównywania wyników programowania rozmytego z rezultatami otrzymanymi za pomocą tradycyjnej procedury Monte-Carlo będziemy używali parametrów niepewnych w pierwotnej postaci gęstości prawdopodobieństwa Gaussa z wartościami oczekiwanymi przedstawionymi w abstrakcyjnych jednostkach miary, j.m., w tablicy 1. Przy tym w celu uproszczenia odchylenie standardowe σ dla wszystkich gęstości przyjęto równe 10 j.m.

Tablica 1 Wartości oczekiwane rozkładów Gaussa niepewnych parametrów zagadnienia.

a_1=460	b_1=410	p_1=440	q_1=390	t_1=600	s_1=1000	k_1=590	r_1=990
a_2= 460	b_2=510	p_2=440	q_2=490	t_2=491	s_2=1130	k_2=480	r_2=1100
a_3= 610	b_3=610	p_3=590	q_3=590	t_3=581	s_3=1197	k_3=570	r_3=1180
c_{11}=100	c_{12}=30	c_{13}=100					
c_{21}=110	c_{22}=36	c_{23}=405					
c_{31}=120	c_{32}=148	c_{33}=11					

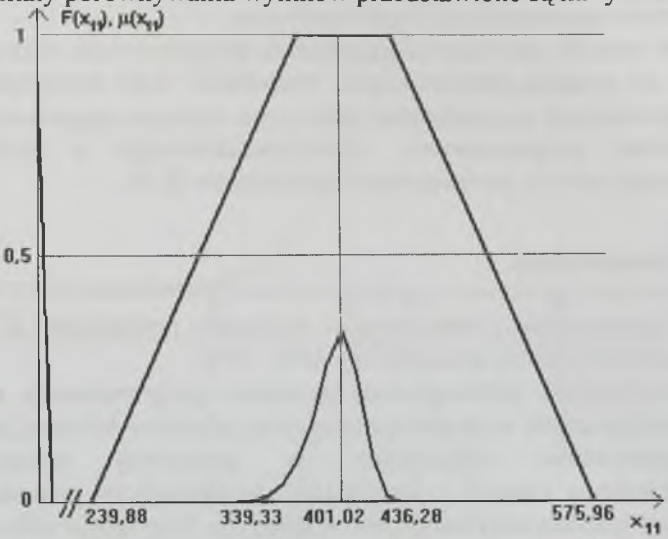
Za pomocą procedury opisanej w rozdziale 2 gęstości prawdopodobieństwa przekształcone zostały w trapezoidalne przedziały rozmyte, które przedstawione są w formie cztero punktowej w tablicy 2

Tablica 2 Postać rozmyto-przedziałowa parametrów zagadnienia (4)-(6)

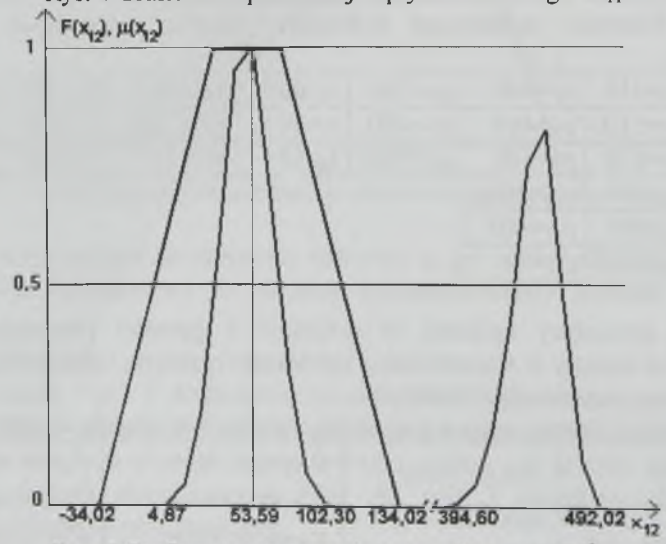
$\hat{a}_1$=[437, 455, 464, 479], $\hat{a}_2$=[437, 455, 464, 479], $\hat{a}_3$=[587, 605, 614, 629],	$\hat{b}_1$=[387, 405, 414, 429], $\hat{b}_2$=[487, 505, 514, 529], $\hat{b}_3$=[587, 605, 614, 629],	$\hat{c}_{11}$=[277, 295, 304, 319], $\hat{c}_{12}$=[457, 475, 484, 499], $\hat{c}_{13}$=[467, 485, 494, 509],	$\hat{c}_{21}$=[377, 395, 404, 419], $\hat{c}_{22}$=[561, 579, 588, 603], $\hat{c}_{23}$=[272, 290, 299, 314],
$\hat{p}_1$=[417, 435, 444, 459], $\hat{p}_2$=[417, 435, 444, 459], $\hat{p}_3$=[567, 585, 594, 609],	$\hat{q}_1$=[367, 385, 394, 409], $\hat{q}_2$=[467, 485, 494, 509], $\hat{q}_3$=[567, 585, 594, 609],	$\hat{c}_{31}$=[277, 295, 304, 319], $\hat{c}_{32}$=[359, 377, 386, 401], $\hat{c}_{33}$=[576, 594, 603, 618],	

Oprócz zagadnienia rozmyto-przedziałowego (4)-(6) zastosowano także tradycyjną procedurę Monte-Carlo dla wartości parametrów zagadnienia pierwotnego (1)-(3) wylosowanych w zgodności z rozkładami prawdopodobieństwa Gaussa z wartościami oczekiwanymi przedstawionymi w tablicy 1 i odchyleniami standardowymi $\sigma = 10$.

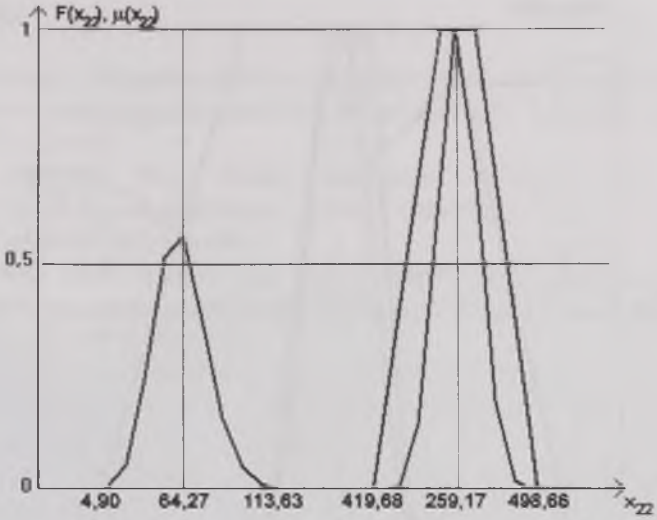
Niektóre rezultaty porównywania wyników przedstawione są na rysunkach 4-8



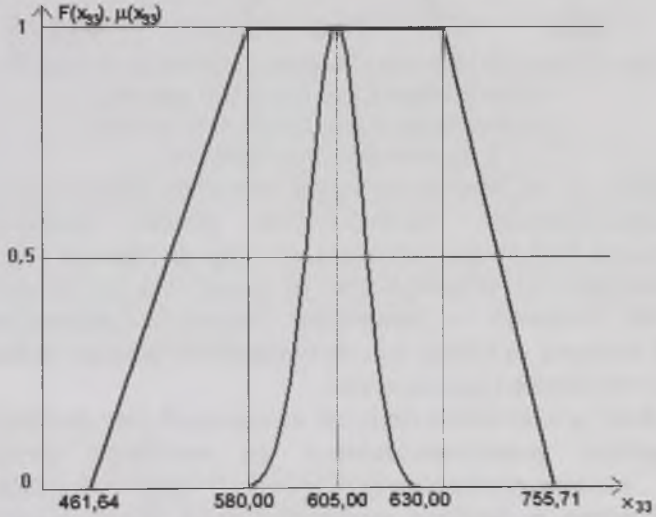
Rys. 4 Graficzna reprezentacja optymalizowanego $\hat{x}_{11}$



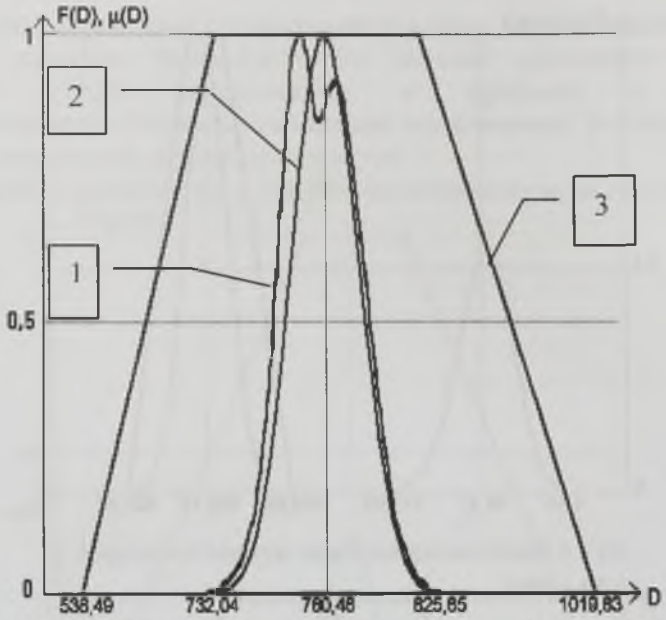
Rys. 5 Graficzna reprezentacja optymalizowanego $\hat{x}_{12}$



Rys. 6 Graficzna reprezentacja optymalizowanego $\hat{x}_{22}$



Rys. 7 Graficzna reprezentacja optymalizowanego $\hat{x}_{33}$



Rys. 8 Graficzna reprezentacja optymalizowanego dochodu D:
1-metoda Monte-Carlo dla 10 000 losowań;
2- metoda Monte-Carlo dla 1000000 losowań;
3- metoda rozmytoprzedziałowa

Widzimy, że w wyniku używania procedury Monte-Carlo otrzymamy rezultaty niejednoznaczne: optymalizowane gęstości prawdopodobieństwa optymalizowanych ilość towaru kupowana u i – tego producenta i sprzedawana j – temu konsumentowi. x_{ij} przedstawione są przez funkcję dwuekstremalne co sprawia pewne trudności w interpretacji rezultatów optymalizacji. Wyniki optymalizacji rozmytej są bliskie tym otrzymanym za pomocą procedury Monte-Carlo ale dość zrozumiałe i jednoznaczne.

Szerokość przedziałów rozmytych wynikowych jest większa niż wizualna szerokość gęstości prawdopodobieństwo jest rezultatem uwzględnienia w procedurze rozmyto-przedziałowej nawet tych wartości, których prawdopodobieństwo w zwykłym sensie jest prawie równe zero. Z rysunku 8 wynika, że dla otrzymania dość gładkich funkcji prawdopodobieństwa gęstości dochodu potrzeba zbyt wielu losowań co faktycznie przekreśla używanie metody Monte-Carlo w praktyce.

Wszystko to świadczy o skuteczności i wystarczającej dokładności proponowanego podejścia do rozmyto-przedziałowego problemu optymalizacji działalności dystrybutora.

LITERATURA

1. H. Isermann, *The enumeration of all efficient solution for a linear multiple-objective transportation problem*, Naval Research Logistics Quarterly 26 (1979) 123-139;
2. J.L. Ringuest, D.B. Rinks, *Interactive solutions for the linear multiobjective transportation problem*, European Journal of Operational Research 32 (1987) 96-106.
3. A.K. Bit, M.P. Biswal, S.S. Alam, *Fuzzy programming approach to multicriteria decision making transportation problem*, Fuzzy Sets and Systems 50 (1992) 135-142.
4. S.K. Das, A. Goswami, S.S. Alam , *Multiobjective transportation problem with interval cost, source and destination parameters*, European Journal of Operational Research 117 (1999) 100-112
5. S. Chanas, M. Delgado, J.L Verdegay and M.A. Vila, *Interval and fuzzy extensions of classical transportation problems*, Transportation Planning Technol. 17(1993) 203-218.
6. S. Chanas, D. Kuchta, *Fuzzy integer transportation problem*, Fuzzy Sets and Systems 98 (1998) 291-298
7. Waiel F. Abd El-Wahed, *A multi-objective transportation problem under fuzziness*, Fuzzy Sets and Systems 117 (2001) 27-33
8. P. Sewastianow, P. Róg, K. Karczewski, *A Probabilistic Method for Ordering Group of Intervals*, Informatyka teoretyczna i stosowana/Computer Science. Politechnika Częstochowska, Rocznik 2, 2 (2002), 45-53
9. Sewastianow, P. Róg, *A Probability Approach to Fuzzy and Crisp Intervals Ordering*, Task Quarterly 7 No 1 (2003), 147-156, Politechnika Częstochowska

Ludmiła Dymowa,
Politechnika Częstochowska
Instytut Informatyki Teoretycznej i Stosowanej
ul. Dąbrowskiego 73, 42-200 Częstochowa

Janusz Grabara
Politechnika Częstochowska
Instytut Ekonometrii i Informatyki
Ul. Armii Krajowej 19b, 42-200 Częstochowa
e-mail: grabara@zim.pcz.czest.pl

Marek Dolata
Instytut Informatyki Teoretycznej i Stosowanej
ul. Dąbrowskiego 73, 42-200 Częstochowa
e-mail: marek@zapr.com.pl

WYKORZYSTANIE MODELOWANIA PROBABLISTYCZNEGO W SYSTEMIE EKSPERTOWYM WSPOMAGAJĄCYM ZARZĄDZANIE RYZYKIEM OPERACYJNYM

Dorota DŻEGA

Streszczenie: W niniejszym artykule podjęta zostanie próba implementacji modelowania probabilistycznego w obszarze zarządzania ryzykiem operacyjnym. Pod pojęciem ryzyka operacyjnego rozumiane są wszystkie defekty systemów informacyjnych lub wewnętrznych regulacji prowadzące do niespodziewanych strat, a także do utraty reputacji. Rozwijając powyższą definicję stwierdza się, że powstaje ono najczęściej w wyniku zaniedbań lub nieodpowiedniego dopasowania procesów zachodzących w niniejszych instytucjach. Dotyczy również błędów personelu, jakości dokumentacji, awarii systemów, oraz zdarzeń losowych.

Pomiar ryzyka operacyjnego obejmuje procesy analizy, monitoringu, oraz kontroli, prowadząc do wypracowania odpowiednich zasad i procedur wspomagających zarządzanie nim. Mając na uwadze zarówno ilościowy, jak i jakościowy charakter czynników wpływających na ten rodzaj ryzyka podjęta w pracy próba pomiaru ryzyka operacyjnego opierać się będzie na wykorzystaniu metod probabilistycznych. Metody te bazując na teorii prawdopodobieństwa, wykorzystują zarówno logiczną, jak i statystyczną interpretację prawdopodobieństwa. Umożliwiają tym samym ocenę prawdopodobieństw danego zdarzenia na podstawie faktów oraz prawdopodobieństwa faktów na podstawie statystycznych informacji.

1. Ryzyko operacyjnym – zagadnienia wstępne

Ryzyko operacyjne jest pojęciem stosunkowo nowym, brakuje uogólnionej definicji. Jedna z definicji dostępnych z literaturze przedmiotu proponuje rozpatrywać je w dwóch aspektach. Propozycja pierwsza to ujęcie pośrednie, pod pojęciem, którego zawarte są wszystkie ryzyka występujące w bankowości za wyjątkiem ryzyka rynkowego, kredytowego oraz płynności finansowej. Natomiast, aspekt bezpośredni mianem ryzyka operacyjnego określa wszystkie straty wynikające z nieadekwatności procesów wewnętrznych, personelu, systemów lub zdarzeń zewnętrznych[1]. Straty powstałe wskutek oddziaływania czynników generujących ryzyko operacyjnego to nie tylko straty finansowe, ale i zagrożenie utraty reputacji. Dlatego też, należy uwidocznić znaczenie powyższych zagrożeń dla systemu bankowego i dążyć do wypracowania odpowiednich standardów zarządzania tym rodzajem ryzyka w integracji z pozostałymi rodzajami ryzyka, to znaczy ryzykiem kredytowym, płynności oraz z ryzykiem rynkowym.

---

[1] Geiger H.:Regulating and Supervising Operational Risk for Bank. Conference „Future of Financial Regulation: Global Regulatory Reform and Implications for Japan". Tokyo 2000

Rys. 1 Systematyka ryzyka operacyjnego
Źródło: Opracowanie własne na podstawie: Lewandowski D.: Ryzyko operacyjne w działalności banków – nowe wyzwania, pilna potrzeba zarządzani. Bank i Kredyt, maj 2001r.

Ryzyko administracyjne związane jest z zagrożeniami powstającymi wewnątrz instytucji finansowej, dotyczy funkcjonowania banku jako przedsiębiorstwa. Obejmuje przede wszystkim procesy planowania działalności na poziomie operacyjnym, taktycznym oraz strategicznym. Związane jest ponadto z przepływem informacji. Obok czynników wewnętrznych duże znaczenie odgrywają czynniki zewnętrzne, a przede wszystkim zmiany w otoczeniu, przejawiające się najczęściej poprzez wzrost konkurencyjności innych podmiotów działających na rynku finansowym, postęp techniczny, a co z tym idzie konieczność ciągłego śledzenia nowości technologicznych.

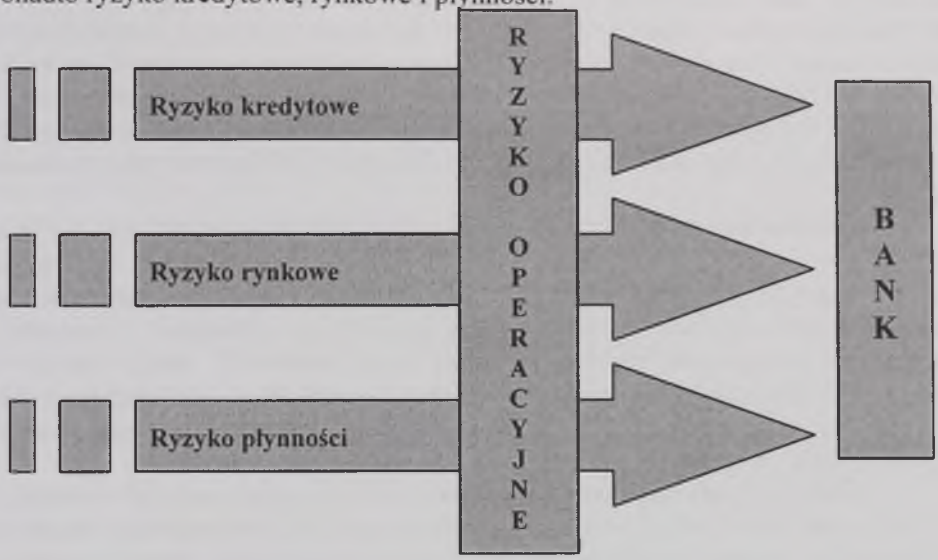
Ryzyko prawne to nieustanna niepewność wywołana zmianami przepisów prawnych. Pociąga za sobą konieczność odpowiedniego dostosowywania się instytucji do obowiązujących regulacji.

Ryzyko transakcyjne związane z zakłóceniami powstającymi w rozliczeniach i systemach płatności, uzależnione jest od problemów natury technicznej, a tym samym od ryzyka technologicznego.

Ryzyko technologiczne potęgują awarie oraz przerwy w funkcjonowaniu systemów teleinformatycznych. Może wynikać z błędów pracowników, celowych nadużyć, czy też z niewłaściwego systemu informowania kierownictwa. Powiązane jest z ryzykiem kadrowym oraz administracji.

Ryzyko kadrowe związane jest bezpośrednio z umiejętnościami pracowników oraz z ich zachowaniami. Nasila się w wyniku braku profesjonalizmu, w przypadkach nadużyć, czy też w wyniku łamania zasad bezpieczeństwa i niewłaściwego korzystania z dostępnych w miejscu pracy technologii informatycznych.

Proces zarządzania ryzykiem operacyjnym powinien zachowywać zgodność z praktyką stosowaną w banku, z przepisami prawnymi obowiązującymi w Polsce i na świecie oraz z regulacjami Komitetu Bazylejskiego. Przed bankami stoi zadanie przeprowadzenia poprawnej identyfikacji czynników generujących ryzyko operacyjne, powiązanie komórek audytu wewnętrznego, stworzenie zasad i procedur oraz wypracowanie mechanizmów kontroli wewnętrznej. Powyższa problematyka wymaga wypracowania odpowiednich metod uwzględniających ponadto ryzyko kredytowe, rynkowe i płynności.



Źródło: Opracowanie własne

Umiejętność zarządzania ryzykiem operacyjnym uzależniona jest w dużej mierze od właściwego przepływu informacji między wszystkimi pracownikami banku. Dlatego też, znaczenia nabierają metody analizy i szacowania poziomu ryzyka operacyjnego w obszarze systemów wspomagania decyzji, między innymi systemów ekspertowych, informowania kierownictwa, projektowych, analitycznych oraz systemów księgowych.

2. Systemy ekspertowe jako narzędzia wspomagające procesy zarządzania ryzykiem operacyjnym

Pod pojęciem **systemów ekspertowych (SE)** rozumiane są programy komputerowe, zaprojektowane w celu wykonywania złożonych zadań, a w

szczególności wyciągania wniosków i podejmowania decyzji na podstawie zgromadzonej wiedzy w sposób zbliżony do rozumowania człowieka. Systemy ekspertowe stworzono celem zastąpienia wysokiej klasy specjalistów. Zasadniczy podział systemów ekspertowych rozróżnia **systemy doradcze**, prezentujące rozwiązania dla użytkownika, oceniającego ich jakość, systemy **podejmujące decyzje bez kontroli człowieka** oraz **systemy krytykujące** przedstawiające **problem**, a następnie jego rozwiązanie[2].

Jednym z wyróżników systemów ekspertowych powinien być wysoki poziom wydawanych ekspertyz, co oznacza rozwiązywanie zadań, dysponowanie strategiami pozwalającymi na imitację wiedzy oraz intuicji specjalistów. Jakość ekspertyz uzależniona jest przede wszystkim od stopnia komplikacji systemu ekspertowego, a ten z kolei od zdefiniowania odpowiednio dużej ilości reguł opisujących zakres heurystyk z danej dziedziny. Dodatkowo, praca z systemem powinna zapewniać uzasadnienie przyjętych rozwiązań na każdym etapie procesu wnioskowania. Podstawą tworzenia systemu ekspertowego opartego na bazie wiedzy jest zakodowanie informacji dotyczących zagadnień związanych z daną dziedziną, pochodzących od specjalistów, będących wynikiem często wieloletnich doświadczeń. Zabieg taki pozwala na wielokrotne i efektywne dysponowanie informacją.

Fundament systemu ekspertowego stanowi **baza wiedzy** definiowana jako szczegółowy, logicznie powiązany zbiór danych, wraz z regułami logicznymi umożliwiającymi jej zastosowanie. Zbiór ten tworzą informacje pozyskane od ekspertów dziedzinowych, są nimi między innymi opisy obiektów i zależności, sposoby rozwiązywania problemów, fakty oraz heurestyki[3]. Baza wiedzy w systemach ekspertowych opiera się przede wszystkim na zagadnieniach reprezentacji, pozyskiwania i użycia wiedzy, ponadto na procedurach objaśniających i uczenia się.

Widoczne jest oddzielenie bazy wiedzy od pozostałych elementów systemu ekspertowego, czyli pamięci roboczej, mechanizmu wnioskowania oraz interfejsu użytkownika. **Pamięć robocza** zwana również przestrzenią roboczą podlega ciągłemu przeglądowi i uaktualnianiu poprzez wykorzystanie wiedzy zawartej w bazie wiedzy. Podstawowym zastosowaniem pamięci roboczej jest krótkotrwałe przechowywanie danych związanych z aktualną problematyką. Proces przeglądania i uaktualniania niniejszej przestrzeni odbywa się poprzez wykorzystanie **mechanizmu wnioskowania**, czyli zbioru reguł reprezentujących dane fragmenty wiedzy. Mechanizm wnioskowania jako element kontrolujący system ekspertowy uruchamiany jest przez użytkownika systemu w momencie inicjacji konsultacji. Niniejszy mechanizm odpowiada za wybór i kolejność reguł wnioskowania, wykorzystując w tym celu logikę formalną. W systemach klasy SE stosowane są

---

[2] Domański Cz.: Statystyczne systemy ekspertowe. Wydawnictwo Uniwersytetu Łódzkiego; Łódź 1998 r.

[3] Kisielnicki J., Sroka H.: Systemy Informacyjne Biznesu. Agencja Wydawnicza „Placet", Warszawa 1999 r.

dwa rodzaje wnioskowania, a mianowicie **wnioskowanie w przód** i **wnioskowanie wstecz**, celem sprawdzenia czy system rozpoczyna pracę od przesłanek, czy od konkluzji. Spotykane jest połączenie wnioskowania wprzód z wnioskowaniem wstecz celem zwiększenia efektywności systemu. Niezależność reguł wnioskowania umożliwia dodawanie nowych reguł do bazy wiedzy, a przez to wpływa na rozszerzenie całego systemu ekspertowego. Aby zapobiec trudnościom powstającym przy przeszukiwaniu rozbudowanych baz wiedzy możliwa jest segmentacja wiedzy w zbiory reguł.

3. Modelowanie probablistyczne

Metody probabilistyczne pozwalają zarówno na logiczną, jaki statystyczną interpretację prawdopodobieństwa. Statystyczna interpretacja prawdopodobieństwa bazuje na zasadzie, zgodnie, z którą nieznane prawdopodobieństwa pewnych zdarzeń można szacować na podstawie informacji, określając częstość ich występowania. Logiczna interpretacja prawdopodobieństwa nie opiera się na subiektywnych przekonaniach, lecz dąży do uprawdopodobnienia pewnych hipotez przez fakty[4]. Zaprezentowana interpretacja prawdopodobieństwa pozwala na ocenę prawdopodobieństw hipotez na podstawie faktów oraz prawdopodobieństwa faktów na podstawie statystycznych obserwacji. W praktyce na podstawie informacji trenującej szacowane są pewne prawdopodobieństwa dotyczące właściwości danej dziedziny, następnie na ich podstawie z wykorzystaniem probabilistycznych mechanizmów wnioskowania, prawdopodobieństwa przypisywane są hipotezom. Może zdarzyć się tak, że zbiory prawdopodobieństw są hipotezami na podstawie, których wnioskuje się o prawdopodobieństwach obserwacji, co oznacza, że dokonywana jest indukcyjna predykcja. Przedstawione mechanizmy wnioskowania w większości przypadków bazują na twierdzeniu Bayesa i określane są mianem **metod probabilistycznych**. Do najpopularniejszych z nich zalicza się metody **klasyfikacji bayesowskiej, sieci bayesowskie** oraz **metody oparte na minimalnej długości kodu.**

Twierdzenie Bayesa umożliwia określenie prawdopodobieństwa *a posteriori* hipotezy *h* jako prawdopodobieństwa warunkowego *Pr(h/D)* [5]

$$\Pr(h/D) = \frac{\Pr(h)\Pr(D/h)}{\Pr(D)}$$

gdzie:
- *D* jest pewnym zbiorem danych potencjalnie wpływających na ocenę prawdopodobieństwa hipotez,
- *Pr(h)* prawdopodobieństwo hipotezy h, w przypadku, gdy dane D nie wpływają na hipotezę h,

---

[4] Cichosz P: Systemy uczące się. Wydawnictwo Naukowo - Techniczne Warszawa 2000 r.
[5] Cichosz P: Systemy uczące się.../ op.cit/.

- *Pr(D)* prawdopodobieństwo wystąpienia danych D, bez uprzedniego zakładania poprawności hipotezy h,
- *Pr (D/h)* prawdopodobieństwo wystąpienia danych D, przy założeniu poprawności hipotezy h.

Twierdzenie Bayesa wraz z regułą łańcuchową opartą na prawdopodobieństwach warunkowych[6]:

$$P(h_1, h_2, ..., h_n) = P(h_1 / h_2, ..., h_n) P(h_2 / h_3, ..., h_n) P(h_{n-1} / h_n) P(h_n)$$

oraz z innymi prawami rachunku prawdopodobieństwa tworzy podstawy funkcjonowania sieci bayesowskich ujmowanych w bazie modeli stochastycznych. **Baza modeli stochastycznych** tworzy architekturę symulacyjną systemu ekspertowego opartego na modelach numerycznych danej dziedziny oraz na zależnościach funkcyjnych pozwalających na zdefiniowanie reguł w postaci **sieci przyczynowych powiązań** lub **probablistycznych sieci bayesowskich**.

Sieci bayesowskie zwane również jako **BBN** (ang. *Bayesian Belief Networks*) tworzą usystematyzowany zapis wiedzy ekspertów o danym zjawisku, czynnikach je modelujących oraz o wzajemnych interakcjach zachodzących między nimi. Istotną zaletę sieci BBN stanowi elastyczność wnioskowania, która predysponuje je jako bazę wiedzy Systemu Ekspertowego. Jako grupa graficznych modeli decyzyjno – analitycznych, posiadających zdolność do kodowania subiektywnej wiedzy ekspertów oraz statystycznej analizy danych ilościowych służą do szacowania prawdopodobieństw rozbudowanych hipotez. Przyjmują postać struktury grafowej, gdzie zmiennymi opisującymi, czyli parametrami są węzły modelu, a jednokierunkowe połączenia pomiędzy nimi to zapisy wykrytych bądź założonych zależności statystycznych o charakterze przyczynowo – skutkowym. Sieć jako całość stanowi odzwierciedlenie struktury problemu, natomiast lokalne połączenia pomiędzy parametrami są kwantyfikowane w postaci tablic prawdopodobieństw warunkowych[7]. Wnioskowanie o stanie określonych zdarzeń może być przeprowadzane zarówno wprzód, jak i wstecz.

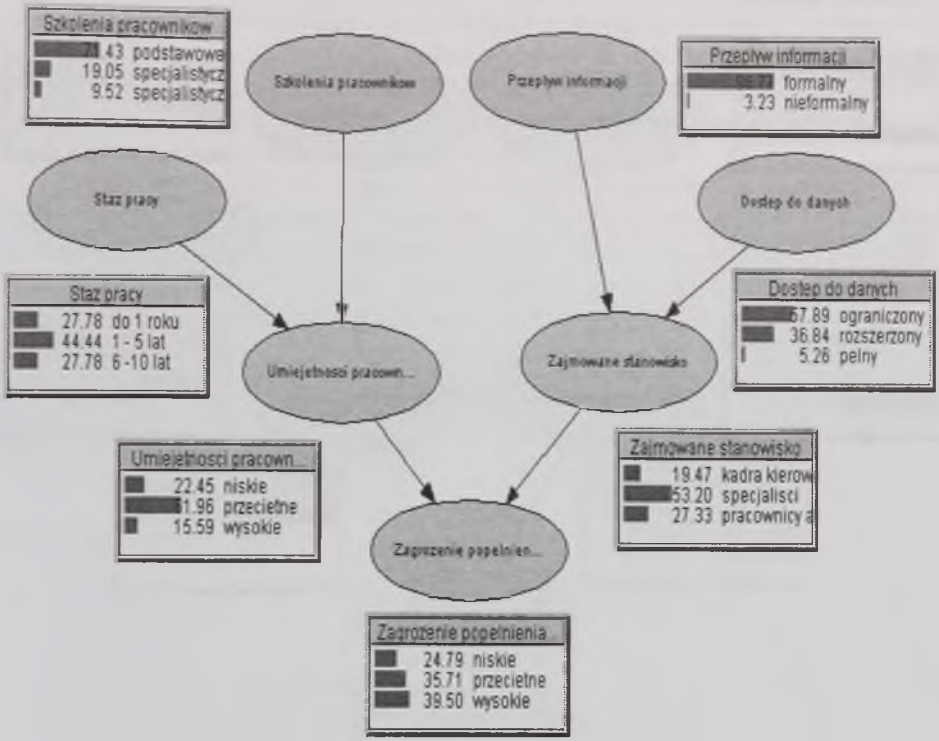
Pierwszym etapem budowy struktury sieciowej jest zdefiniowanie węzłów, a następnie wykrycie pomiędzy nimi zależności o charakterze przyczynowo-skutkowym. W ten sposób generowana jest optymalna struktura bazująca na czynnikach opisowych. Kolejny krok to konstrukcja tablic prawdopodobieństw warunkowych, stanowiących ilościową specyfikację modelu. Utworzenie sieci oraz wypełnienie tablic prawdopodobieństwa umożliwia przeprowadzanie wnioskowania – propagacji zdarzeń, polegającej na uaktualnieniu rozkładów prawdopodobieństw na podstawie nowych, dostępnych obserwacji. Złożoność niektórych problemów wymaga zastosowania algorytmów bazujących na dekompozycji sieci, przez co problem może zostać rozłożony na szereg

---

[6] Fenton N: Probablity Theory and Bayesian Belief Nets. Computer Science Department of Informatics and Mathematical Sciences. Queen Mary and Westfield College.

[7] Fenton N.: Probablity Theory and Bayesian Belief Nets.../op. cit./

mniejszych, prostszych do rozwiązania. Podział na podsieci dokonywany jest w oparciu o rozdział w węzłach zawierających zdarzenia zaobserwowane.



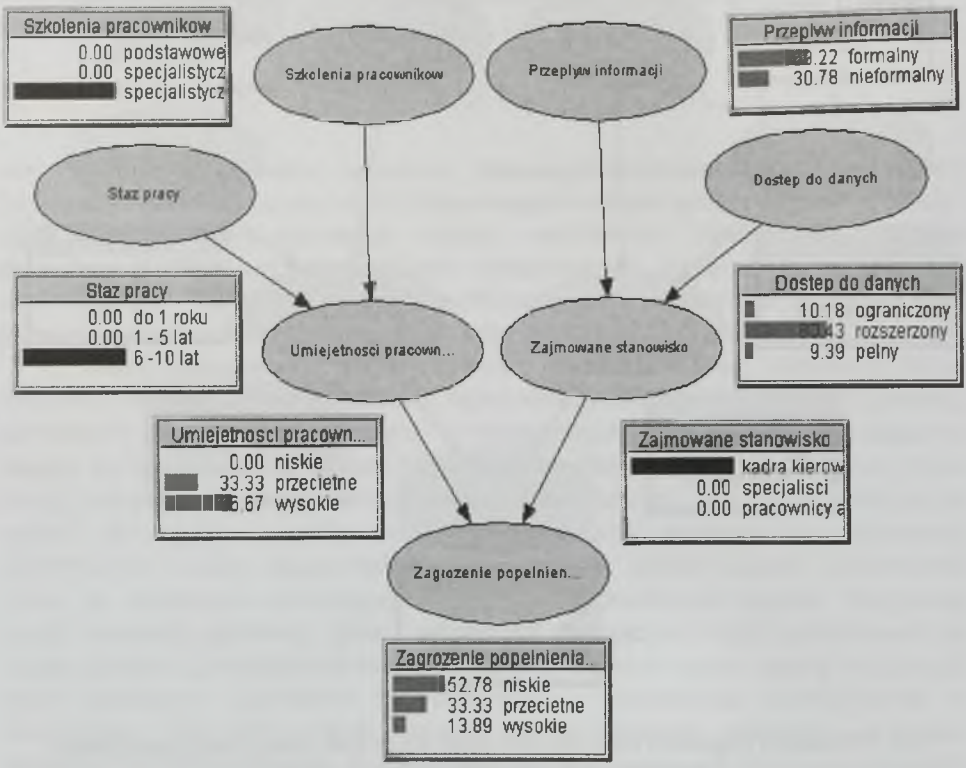
Rys. 3 Struktura fragmentu sieci bayesowskiej służącej do identyfikacji możliwości popełnienia błędów przez pracowników
Źródło: Opracowanie własne

Zaprezentowana powyżej struktura służąca do oceny możliwości popełnienia błędów przez pracowników stanowi fragment sieci bayesowskiej identyfikującej ryzyko operacyjne w obszarze ryzyka kadrowego. Przedstawiona na rysunku 1 sieć została zainicjowana, to znaczy wprowadzono do niej dane wejściowe. Rozważania stanowią model teoretyczny i opierają się na subiektywnej wiedzy autora. Przykładowe wnioskowanie przeprowadzone zostało w oparciu o algorytm propagacji **Lauritzen & Spiegelhalter**[8].

Wprowadzenie dowolnych zmian w węzłach umożliwia przeprowadzenie propagacji, odczytanie nowych rozkładów prawdopodobieństwa i co najważniejsze

[8] Jensen F.: Implementation Aspects of Various Propagation Algorithms in Hugin. Department of Mathematics and Computer Science Aalborg University, Denmark 1994

ułatwia podjęcie decyzji. Do przeprowadzenia propagacji nie jest wymagana znajomość obserwacji we wszystkich węzłach. Elastyczność sieci bayesowskich pozwala na przeprowadzenie wnioskowania na podstawie informacji zawartych w wybranych węzłach.



Rys. 4 Struktura fragmentu sieci bayesowskiej służącej do identyfikacji możliwości popełnienia błędów przez pracowników przy znanych obserwacjach wybranych węzłów
Źródło: Opracowanie własne

Zdolność do wnioskowania i diagnozowania w warunkach niepełności informacji, a także niezwykła elastyczność sieci bayesowskich, przejawiająca się między innymi podatnością na ulepszanie struktury sieci poprzez wprowadzanie nowych węzłów oraz wykrywanie nowych związków o charakterze przyczynowo-skutkowym przemawia za implementacją sieci bayesowskich jako bazy wiedzy w Systemie Ekspertowym. Umiejętność dekompozycji sieci stanowi dodatkowy atut w przypadku problemów złożonych wymagających przeprowadzenia segmentacji bazy wiedzy.

Podsumowanie

Postęp technologiczny, a wraz z nim wzrost konkurencja w sektorze usług bankowych spotęgowało zagrożenie powstania strat finansowych oraz możliwości utraty reputacji. Wzrost znaczenia czynników generujących ryzyko operacyjnego, w tym czynników o charakterze subiektywnym dodatkowo potęguje złożoność problemu, jaki stanowią kwestie związane z procesem zarządzania ryzykiem operacyjnym. Dlatego też, tak istotne jest opracowanie odpowiednich narzędzi służących do usprawnienia niniejszego procesu.

Zaprezentowana w niniejszym artykule zdolność do wnioskowania i diagnozowania w warunkach niepełności informacji, podatność na ulepszanie struktury sieci poprzez wprowadzanie nowych węzłów oraz wykrywanie nowych związków o charakterze przyczynowo-skutkowym przemawia za zasadnością implemenatacji sieci bayesowskich jako bazy wiedzy w Systemie Ekspertowym wspomagającym zarządzanie ryzykiem operacyjnym. Uwzględniając zarówno ilościowy, jak i jakościowy charakter czynników wpływających na ten rodzaj ryzyka można twierdzić, że wykorzystanie metod probablistycznych w tym przypadku jest jak najbardziej trafne.

Literatura

1. Cichosz P: Systemy uczące się. Wydawnictwo Naukowo - Techniczne Warszawa 2000
2. Domański Cz.: Statystyczne systemy ekspertowe. Wydawnictwo Uniwersytetu Lódzkiego; Łódź 1998
3. Fenton N: Probablity Theory and Bayesian Belief Nets. Computer Science Department of Informatics and Mathematical Sciences. Queen Mary and Westfield College.
4. Geiger H.:Regulating and Supervising Operational Risk for Bank. Conference „Future of Financial Regulation: Global Regulatory Reform and Implications for Japan". Tokyo 2000
5. Jensen F.: Implementation Aspects of Various Propagation Algorithms in Hugin. Department of Mathematics and Computer Science Aalborg University, Denmark 1994
6. Kisielnicki J., Sroka H.: Systemy Informacyjne Biznesu. Agencja Wydawnicza „Placet", Warszawa 1999
7. Lewandowski D.: Ryzyko operacyjne w działalności banków – nowe wyzwania, pilna potrzeba zarządzani. Bank i Kredyt, maj 2001
8. Mulawka J.: Systemy ekspertowe. Wydawnictwo Naukowo – Techniczne, Warszawa 1996
9. http://www.wib.org.pl/1_1_3.html

mgr Dorota Dżega
Wydział Informatyki Politechniki Szczecińskiej
ul. Żołnierska 48 Tel.: (0-91) 449-56-69
e-mail: ddzega@wi.ps.pl

ROZWIĄZYWANIE ZAGADNIEŃ EKONOMICZNYCH Z WYKORZYSTANIEM ARKUSZY KALKULACYJNYCH

Iwona ISKIERKA

Streszczenie: W pracy przedstawiono sposoby uproszczenia i skrócenia czasu procesu planowania. Wykorzystano możliwość zastosowania macierzowej postaci modelu bilansu produkcji, obliczenia przeprowadzono w arkuszu kalkulacyjnym.

Wstęp

W roku 1973 amerykański ekonomista Wassily Leontief otrzymał nagrodę Nobla w dziedzinie ekonomii za stworzenie podstaw metody analizy nakładów i wyników produkcji. Zgodnie z zasadami ekonomii podstawowa zasada planowania opiera się na tzw. bilansie gospodarki narodowej. W. Leontief zaproponował ułożenie bilansu gospodarki narodowej w taki sposób, aby znane były ilości produkcji i jej przydziały do poszczególnych działów gospodarki.

1. Statyczny model przepływów międzygałęziowych

Założono, że cała gospodarka narodowa jest podzielona na n działów. Planowaną produkcję każdego działu, która jest wytworzona w jednostce czasu oznaczamy przez $Q_1,\dots Q_n$. Ustalone są także jednostki miary produkcji. Część produkcji każdego działu przechodzi do innego działu, część zostaje zużyta w tym samym dziale, część jest przeznaczona na cele nie związane z bieżącą produkcją - jest to produkt końcowy każdego działu (q_i). Symbol q_{ij} $(i,j=1,\dots,n)$ oznacza ilość produkcji i-tego działu, która zużywana jest na cele produkcyjne w dziale j [1].

Okazuje się, że w gospodarce zamkniętej, w której nie ma wymiany z zagranicą produkty końcowe muszą być określone warunkiem $q_i \geq 0$, gdzie $i=1,\dots,n$

Bilans gospodarki narodowej można, więc zapisać macierzą blokową (macierzą bilansu produkcji) następującej postaci:

$$\left[\begin{array}{c|ccc|c} Q_1 & q_{11} & \cdots & q_{1n} & q_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ Q_n & q_{n1} & \cdots & q_{nn} & q_n \end{array}\right]$$

Wielkości $a_{ij} = \frac{q_{ij}}{Q_j}$, gdzie $i, j = 1,...,n$ nazywa się technicznymi współczynnikami produkcji.

Wszystkie współczynniki techniczne produkcji tworzą macierz stopnia n: macierz techniki wytwarzania

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

Przyjmując oznaczenia

$$\begin{bmatrix} Q_1 \\ Q_2 \\ \dots \\ Q_n \end{bmatrix} = Q, \qquad \begin{bmatrix} q_1 \\ q_2 \\ \dots \\ q_n \end{bmatrix} = q$$

i wykonując odpowiednie przekształcenia otrzymujemy:

$$Q = (I - A)^{-1} q$$

Powyższe równanie umożliwia szybkie opracowanie różnych wariantów planu produkcji materialnej, w zależności od wariantów założonego produktu końcowego [1].

Do określenia wielkości przepływów międzygałęziowych można zaproponować wykorzystanie arkusza kalkulacyjnego, w którym wykonano obliczenia zgodnie z układem widocznym poniżej (oczywiście liczba działów może ulegać zmianie).

W konkretnym przypadku elementy macierzy $(I - A)^{-1}$ określają ilościowe związki między produkcją każdej gałęzi (działu) a produktami końcowymi wszystkich działów.

Zastosowanie algebry macierzy w ekonomii

Gałąź produkcji materialnej	Wartość produkcji globalnej poszczególnych gałęzi	Przepływy międzygałęziowe do gałęzi			Produkt końcowy	
1		0	0	0		0
2		0	0	0		0
3		0	0	0		0

1	0	0
0	1	0
0	0	1

Macierz jednostkowa

Macierz technicznych współczynników

1	0	0
0	1	0
0	0	1

I-A

1,000	0,000	0,000	
0,000	1,000	0,000	
0,000	0,000	1,000	

$[I-A]^{-1}$

Wartość produkcji globalnej poszczególnych gałęzi

Zastosowanie algebry macierzy w ekonomii

Gałąź produkcji materialnej	Wartość produkcji globalnej poszczególnych gałęzi	Przepływy międzygałęziowe do gałęzi			Produkt końcowy	
1	635,6	0,0	231,0	304,6	100	635,6
2	770,1	317,8	0,0	152,3	300	770,1
3	761,4	254,2	231,0	76,1	200	761,4

1	0	0
0	1	0
0	0	1

Macierz jednostkowa

0	0,3	0,4
0,5	0	0,2
0,4	0,3	0,1

Macierz technicznych współczynników

1	-0,3	-0,4
-0,5	1	-0,2
-0,4	-0,3	0,9

I-A

1,822	0,846	0,998	635,6
1,150	1,605	0,868	770,1
1,193	0,911	1,844	761,4

$[I-A]^{-1}$

Wartość produkcji globalnej poszczególnych gałęzi

W przypadku kilku wariantów zapotrzebowania na produkt końcowy stosowanie powyższej metody obliczeń pozwala studentom szybko określić wielkość przepływów międzygałęziowych w zależności od określonego wariantu.

Wykonywanie żmudnych rachunków zastąpiono opracowanymi formułami, co wpłynęło na czas wykonania zadań. Widoczny jest sens ekonomiczny poszczególnych elementów, które można odpowiednio symulować.

2. Określanie całkowitych nakładów pracy

Obliczanie całkowitych nakładów pracy sprowadza się do obliczenia nakładów pracy uprzedmiotowionej oraz nakładów pracy żywej.
Przy założeniu produkcji n produktów i wprowadzeniu następujących oznaczeń

t_j - nakład pracy żywej przy wytwarzaniu j - tego produktu

T_j - całkowity nakład pracy żywej i uprzedmiotowionej przy wytwarzaniu j - tego produktu

a_{ij} - zużycie i - tego produktu niezbędne do wytworzenia jednostki j - tego produktu

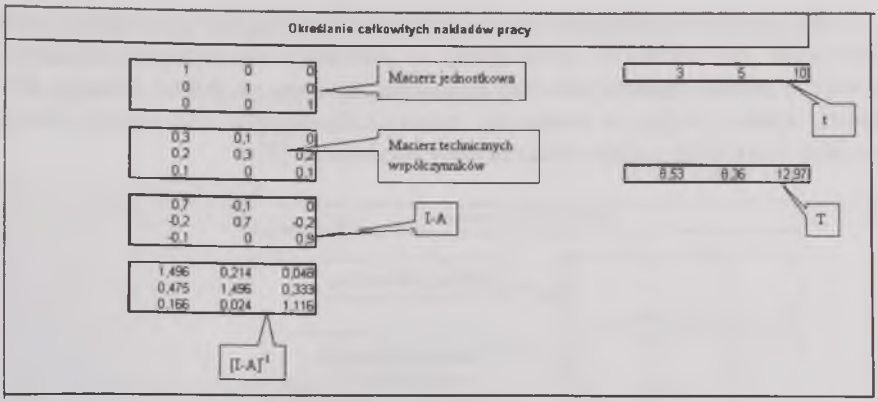
$$T = [T_1 \quad T_2 \quad \dots \quad T_n], \quad t = [t_1 \quad t_2 \quad \dots \quad t_n]$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

można otrzymać równanie:

$$T = t(I - A)^{-1}$$

Korzystając z możliwości skonstruowania odpowiednich formuł w arkuszu kalkulacyjnym obliczenie całkowitych nakładów pracy sprowadza się w konkretnym przypadku do obliczenia iloczynu macierzy nakładów pracy żywej przez macierz $(I - A)^{-1}$ [3].



3. Określanie cen produktów

Przy założeniu produkcji n produktów w ilościach odpowiednio $Q_1, Q_2, \ldots Q_n$ i wprowadzeniu następujących oznaczeń:

a_i - amortyzacja na jednostkę produkcji i - tego działu

d_i - produkcja czysta przypadająca na jednostkę produkcji i - tego działu

p_i - cena jednostki i - tego produktu (cena jednolita)

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}$$

A^T - macierz transponowana macierzy techniki wytwarzania

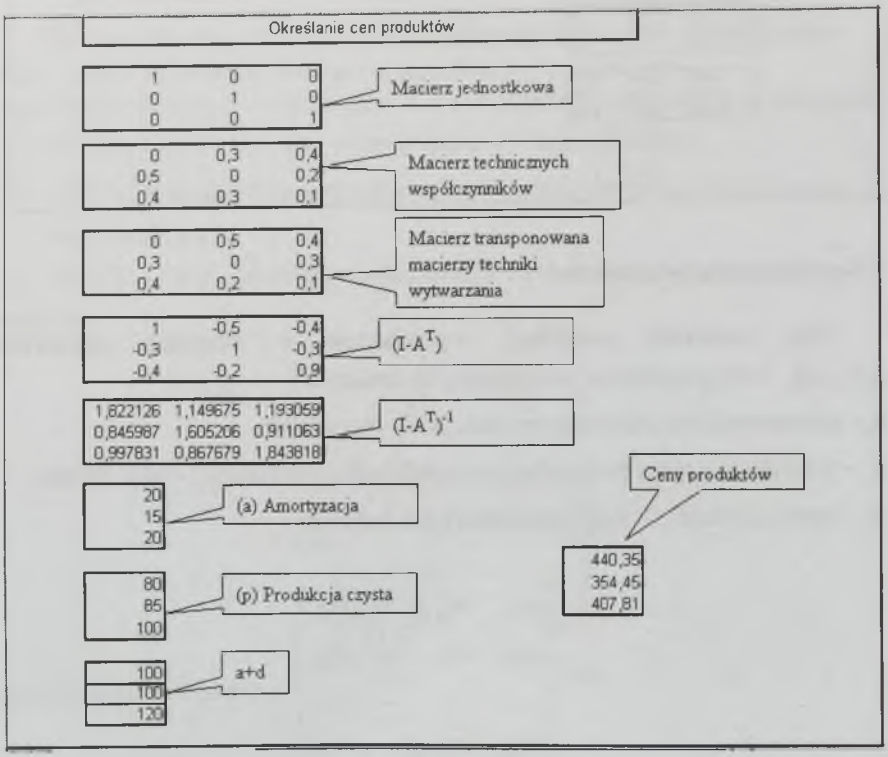
otrzymujemy:

$$\begin{bmatrix} p_1 \\ p_2 \\ \ldots \\ p_n \end{bmatrix} = p \quad , \quad \begin{bmatrix} a_1 \\ a_2 \\ \ldots \\ a_n \end{bmatrix} = a \, , \quad \begin{bmatrix} d_1 \\ d_2 \\ \ldots \\ d_n \end{bmatrix} = d$$

a następnie następujące równanie:

$$p = \left(I - A^T\right)^{-1} (a + d)$$

Pamiętając o zależności cen odpowiednich produktów przy danej technice wytwarzania od wielkości amortyzacji i produkcji czystej przypadającej na jednostkę produktu można określić odpowiednią cenę produktu każdego działu. Studenci, wykorzystując te zależności, arkusz kalkulacyjny oraz algebrę macierzy są w stanie w krótkim czasie obliczyć ceny produktów [3].



Proponowane rozwiązywanie problemu może być zastosowane w czasie wielu zajęć ze studentami, tam gdzie liczy się wykorzystanie umiejętności twórczego, samodzielnego rozwiązywania za pomocą komputera zarówno problemów typowych jak i problemów niestandardowych. Zaletą tego typu obliczeń jest skupienie uwagi użytkownika na analizie problemu z pominięciem nużących rachunków. Współczesność wymaga od użytkownika komputera wykazania się umiejętnościami posługiwania się technologiami informatycznymi [2]. Dlatego też użytkownik powinien umieć samodzielnie określić takie sytuacje, w których można rozwiązać problem posługując się komputerem i odpowiednim oprogramowaniem.

Literatura

1. Piszczała J.: Matematyka i jej zastosowanie w naukach ekonomicznych. Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań 2000.

2. Andrzej Maria Wilk.: Polska w drodze do Społeczeństwa Informacyjnego. Warszawa 2002.
3. Kryński H E.: Matematyka dla ekonomistów. PWN, Warszawa

Dr inż. Iwona ISKIERKA
Instytut Informatyki i Ekonometrii
Politechnika Częstochowska
ul. H. Dąbrowskiego 69,
42-200 Częstochowa
tel.: (0-34) 3 250 242
e-mail: iskierka@zim.pcz.czest.pl

GRY SYMULACYJNE DYDAKTYCZNE ZARZĄDZANIA I ICH PODSTAWY OBIEKTOWE

Aleksander KATKOW, Agnieszka ULFIK

Wstęp

Nikogo nie dziwi fakt, że w procesie szkolenia pilotów stosowane są symulatory. Praca osoby siedzącej za sterami samolotu jest niezwykle odpowiedzialna i wymaga solidnego przygotowania. Osoba która podejmując decyzje, kieruje losami firmy ma równie odpowiedzialne zadanie. Niewłaściwe decyzje mogą doprowadzić przedsiębiorstwo do bankructwa. Z tego właśnie powodu zaczęto stosować symulatory w postaci gier zwanych symulacyjnymi, aby przygotować przyszłych menadżerów do pracy w rzeczywistych warunkach rynkowych.

Historia gier symulacyjnych jest dość długa choć były one znane pod innymi nazwami. Pierwsze gry symulacyjne stosowano już w starożytności, a dotyczyły one działań wojennych. Szczególnie dużą popularność w tym aspekcie gry uzyskały w siedemnastym wieku, kiedy to przed zbudowaniem fortecy najpierw przeprowadzano symulacje mające na celu ustalenie jej zdolności obronnych. W późniejszych latach zaczęto szeroko stosować tak zwane gry militarne. Miały one na celu wyuczenie reakcji na zdarzenia jakie mogły by zaistnieć na polu walki. Gry takie stosowano często w szkołach wojskowych, a wyniki uzyskane przez studentów pozwalały ocenić ich zdolności podejmowania strategicznych decyzji. W aspekcie zarządzania zaczęto stosować gry symulacyjne dopiero w drugiej połowie XX wieku [4].

1. Gry i edukacja

Gra symulacyjna to zazwyczaj zastosowanie pewnego modelu przedsiębiorstwa usytuowanego w pewnym wirtualnym otoczeniu rynkowym. Należy pamiętać o tym, że każdy model jest jedynie pewnym przybliżeniem rzeczywistości. W zależności od postaci tego modelu może on bardziej lub mniej dokładnie odzwierciedlać świat realny [5]. W grze zawsze mamy do czynienia z pewnym uproszczeniem rzeczywistości. Każda gra powinna w sposób satysfakcjonujący przedstawiać główne aspekty rzeczywistości [6].

Istnieje wiele rodzajów tego typu gier oraz wiele dziedzin w których są one stosowane. Gry takie mogą mieć za zadanie symulowanie pracy przedsiębiorstwa jako całości, zachowań wybranego wycinka rynku, czy pracy wybranego obszaru w przedsiębiorstwie [3, 7].

Dobrodziejstwem, jakie płynie z zastosowania gier symulacyjnych jest fakt edukacji na wirtualnym modelu, a nie rzeczywistym przedsiębiorstwie. Kolejnym pozytywnym aspektem jest fakt przyspieszenia upływu czasu [6]. Dzięki temu

skutki naszych decyzji są odczuwalne niemal natychmiast, podczas gdy w świecie realnym, procesy takie trwają miesiącami, a nawet latami.

Gry są obecnie szeroko stosowane w procesie kształcenia na wielu kierunkach studiów (głównie ekonomicznych i zarządczych). Takie gry według dr Jana Korbuta [10] umożliwiają:

- integrację wiedzy pochodzącej z przedmiotów wykładanych na studiach, kursach czy szkoleniach,
- praktyczne zapoznanie uczestników gry z zasadami zarządzania firmą w warunkach konkurencji,
- kształcenie umiejętności zarządzania czynnikiem ludzkim w warunkach gospodarki rynkowej,
- przekazanie metodyki sprawnego i atrakcyjnego prowadzenia treningu menedżerskiego,
- opanowanie podstawowych zasad opracowania skutecznego biznes planu firmy oraz wykorzystania uzyskanych wyników finansowych do oceny stopnia jego realizacji,
- wykształcenie umiejętności pracy zespołowej.

W trakcie gry jej uczestnicy muszą zazwyczaj podejmować decyzje dotyczące rozmiarów produkcji, ceny wyrobu, jego jakości, polityki kadrowej, badań marketingowych, nakładów na reklamę, polityki płacowej itd. Dzięki temu w trakcie symulacji dochodzi do [2, 6]:

- wykorzystania adrenaliny w procesie nauczania,
- zaangażowania wszystkich uczestników dzięki mechanizmowi rywalizacji,
- nabycia doświadczenia bez ponoszenia realnych strat,
- dostarczenia informacji zwrotnej,
- kształtowania pożądanych postaw,
- integracji zespołów,
- a przede wszystkim do szybkiego i skutecznego uczenia, co jest spowodowane atrakcyjnością samej metody.

Gry symulacyjne wykorzystywane są nie tylko jako przedmiot w programie studiów [1, 11]. Wiele firm wykorzystuje tego typu symulację aby zwiększyć wydajność pracowników, czy poprawić kontakt z klientami [3]. Organizowane są również konkursy takie jak np. Euromanager czy L'Oreal E-start Challenge. Gry takie zazwyczaj rozgrywane są za pośrednictwem Internetu i trwają kilka tygodni. Zespoły (często z całego świata) rywalizując ze sobą, tworzą „wirtualną" konkurencję. Najlepsi wygrywają a wszyscy uczestnicy otrzymują doświadczenie, którego nie da się niczym zastąpić.

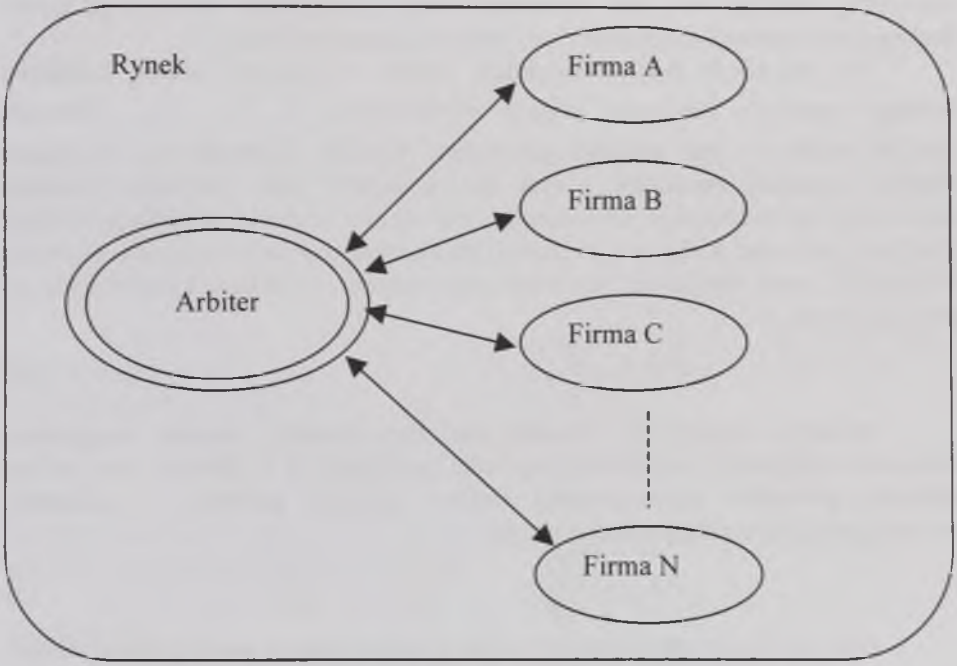
2. Gra produkcyjno – decyzyjna

Proponowana gra symulacyjna oparta jest na ekonometrycznym modelu przedsiębiorstwa. W modelu tym zastosowano analizę produkcji, kosztów i popytu.

Gra polega na komunikowaniu się ze sobą kilku firm produkcyjnych. Firmy te produkują podobne, substytucyjne produkty. Każda firma to zespół kilku

osób. Osoby te w każdej rundzie muszą podjąć pewne decyzje, które mają na celu zmaksymalizowanie zysku oraz rozwój przedsiębiorstwa.

Gra składa się z kolejnych rund symbolizujących pewne okresy czasu (np. miesiąc lub kwartał). Wszystkie firmy (zespoły graczy) rozpoczynają grę mając takie same warunki startowe. Po każdej rundzie następuje wymiana informacji pomiędzy firmami tzn. wszystkie decyzje podjęte przez zespoły trafiają do arbitra gry, który symbolizuje rynek i jako wyniki finansowe trafiają w postaci informacji zwrotnej do zespołów graczy. Firmy nie komunikują się zatem między sobą bezpośrednio, a jedynie poprzez arbitra (rys. 1). Po otrzymaniu informacji zwrotnej rozpoczyna się kolejna runda w której zespoły znów podejmując decyzje, kierują losami swojego wirtualnego przedsiębiorstwa. Firmy można zatem potraktować jako obiekty tej samej klasy, natomiast Arbitra możemy traktować jako obiekt zaprzyjaźnionej klasy.



Rys. 1. Schemat relacji w prezentowanej grze decyzyjnej

3. Model przedsiębiorstwa

Model przedsiębiorstwa opiera się na ekonometrycznych założeniach [9] i jak każdy model jest tylko pewnego rodzaju przybliżeniem rzeczywistości. W modelu tym zostały ujęte trzy zagadnienia:

- produkcja,
- koszty,
- popyt.

3.1. Analiza produkcji

Funkcja produkcji to następujący model ekonometryczny:

$$Q = f(X_1, X_2, ..., X_m) \tag{1}$$

gdzie: Q - wielkość produkcji (produkt całkowity) w określonym przedziale czasu,

$X_1, X_2, ..., X_m$ - wielkości nakładów i czynników produkcji niezbędnych do wytworzenia tej produkcji

Funkcja ta najczęściej przybiera jedną z dwóch postaci:

- liniową: $$Q = a_0 + a_1 X_1 + a_2 X_2 + ... + a_m X_m \tag{2}$$
- potęgową: $$Q = b_0 X_1^{b_1} X_2^{b_2} ... X_m^{b_m} \tag{3}$$

Zazwyczaj funkcja ta jest zdeterminowana warunkami technologicznymi. Doskonalenie technologii prowadzi do zmiany funkcji produkcji.

Na podstawie funkcji produkcji można wyznaczyć szereg wielkości charakteryzujących zależności między wielkościami $X_1, X_2, ..., X_m$. Pierwszą z takich wielkości jest produkt przeciętny (produkt jednostkowy) względem nakładu czynnika produkcji i jest on rozumiany jako stosunek produktu całkowitego wyznaczonego ze wzoru (1) do danego czynnika produkcji. Produkt przeciętny jest więc wielkością produkcji jaka można uzyskać z jednostki czynnika produkcji X, przy ustalonym poziomie pozostałych czynników i przedstawia go poniższy wzór:

$$PP_X = \frac{Q}{X} \tag{4}$$

Kolejna wielkość to produkt krańcowy (inaczej produkt marginalny, krańcowa wydajność) względem czynnika produkcji X i określa ona zmianę wielkości produkcji spowodowanej zmiana czynnika produkcji o jednostkę, a wyznaczana jest z następującego wzoru:

$$PK_X = \frac{\partial Q}{\partial X} \tag{5}$$

Inna miarą jest elastyczność produkcji wyznaczana z następującego wzoru:

$$EP_X = \frac{\Delta Q}{Q} : \frac{\Delta X}{X} = \frac{PK_X}{PP_X} \tag{6}$$

Elastyczność produkcji względem nakładów czynnika produkcji X wskazuje, jakie relatywne zmiany rozmiarów produkcji odpowiadają relatywnej zmianie tego czynnika produkcji. Inaczej można powiedzieć, że elastyczność produkcji względem nakładu czynnika produkcji X informuje, o ile procent zmieni się produkcja, jeśli nakłady czynnika X zwiększymy o 1%.

3.2. Analiza kosztów

Funkcja kosztów to badanie zależności rozmiaru kosztów od skali produkcji. Funkcję kosztów globalnych dobrze przybliża funkcja, która jest wielomianem stopnia trzeciego, rosnącym, nie posiadającym ekstremów:

$$K = a_0 + a_1 Q + a_2 Q^2 + a_3 Q^3 \tag{7}$$

gdzie: K - koszty globalne w określonym przedziale czasu,

a_0, a_1, a_2, a_3 - oszacowane parametry funkcji kosztów, spełniające następujące warunki:

$$a_0, a_1, a_3 \rangle 0,$$

$$a_2 \langle 0,$$

$$a_2 \langle 3a_1 a_3.$$

W pewnych przedziałach funkcja kosztów globalnych może być przybliżona do linii prostej a wówczas:

$$K = a_0 + a_1 Q \tag{8}$$

Niezależnie od postaci funkcji kosztów globalnych – wzór (7) lub (8) – można wyznaczyć koszty jednostkowe będące ilorazem kosztów globalnych przez rozmiar produkcji:

$$k = \frac{K}{Q} \tag{9}$$

Można także wyznaczyć koszty krańcowe będące pochodną funkcji kosztów globalnych i mającą jedną z następujących postaci (w zależności od tego czy analizujemy koszty globalne wg wzoru (7) czy (8)):

$$K' = a_1 + 2a_2 Q + 3a_3 Q^2 \tag{10}$$

$$K' = a_1 \tag{11}$$

Koszty krańcowe określają zmianę wielkości kosztów globalnych spowodowanych zmianą wielkości produkcji o jednostkę.

Z analizą kosztów związany jest także zysk rozumiany jako przychód minus koszty:

$$Z = P - K \tag{12}$$

gdzie: Z – zysk,

P – przychód rozumiany jako wielkość produkcji razy cena ($P=c*Q$),

K – koszty globalne.

Jak widać, na zysk bezpośredni wpływ ma cena produktu. Zysk jednostkowy możemy zatem policzyć z następującego wzoru:

$$z = c - k \tag{13}$$

gdzie: z – zysk jednostkowy,

c – cena jednostkowa,

k – koszt jednostkowy.

3.3. Analiza popytu

Zrozumienie ekonomicznej koncepcji popytu jest podstawą efektywnego podejmowania krótko- i długoterminowych decyzji produkcyjnych w firmie. Popyt jest definiowany jako liczba jednostek pewnego produktu lub usługi, którą konsumenci są skłonni nabyć w ciągu pewnego czasu w określonych warunkach. Punktem wyjścia w analizie popytu w naszej grze jest następująca funkcja:

$$P = f(C, R, D, C_{K1}, R_{K1}, C_{K2}, R_{K2}, ..., C_{KN}, R_{KN}) \tag{14}$$

gdzie: C – cena naszego produktu,

R – wielkość nakładów na reklamę przedsiębiorstwa,

D – średnia wielkość dochodów potencjalnych klientów,

$C_{K1}, C_{K2}, ..., C_{KN}$ - ceny produktów produkowanych przez pozostałe zespoły graczy,

$R_{K1}, R_{K2}, ..., R_{KN}$ - wielkość wydatków na reklamę zainwestowane przez pozostałe zespoły graczy.

W praktyce spotykane są dwa rodzaje funkcji popytu:

- liniowa,
- potęgowa.

Dodatkowe wielkości stosowane w analizie popytu to elastyczności względem ceny, wydatków na reklamę, względem dochodów, a także elastyczności krzyżowe względem ceny i wydatków na reklamę produktów pozostających w związku z rozpatrywanym produktem. Wszystkie te elastyczności liczymy wg następującego wzoru:

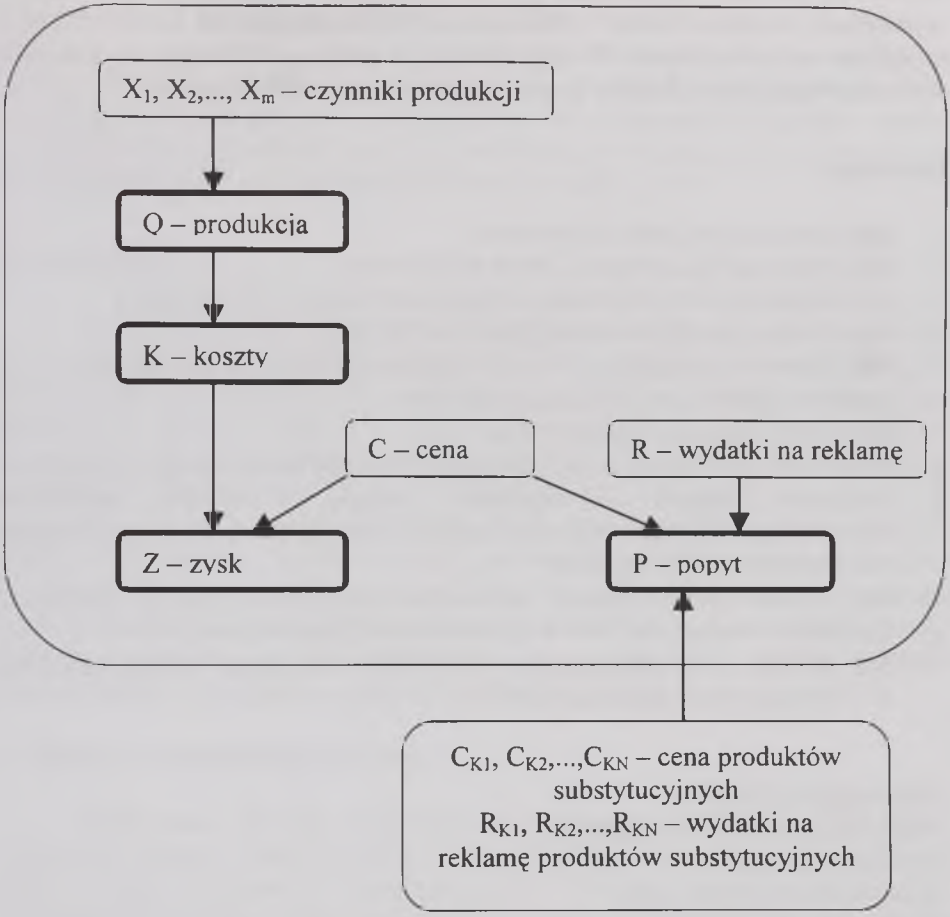
$$E_X = \frac{\partial P}{\partial X} \cdot \frac{X}{P} \tag{15}$$

a interpretujemy jako relatywną zmianę wielkości popytu jaka odpowiada relatywnej zmianie badanej wielkości.

4. Algorytm gry

Proponowany model przedsiębiorstwa, zastosowany w grze symulacyjnej mającej na celu symulację zachowania zespołów przedsiębiorstw zajmujących się produkcją substytucyjnych wyrobów przedstawiono na rysunku 2. Zgodnie z założeniami gracze w poszczególnych przedsiębiorstwach ustalają $X_1, X_2, ..., X_m$, czyli wielkości nakładów i czynników produkcji niezbędnych do wytwarzania. W rzeczywistym świecie mogą to być np. ilości maszyn, nakłady poszczególnych surowców czy ilości zatrudnionych pracowników. W zależności od tego jakie wielkości zostaną ustalone – taka będzie wielkość produkcji obliczona według wzoru (1). Mając obliczoną wielkość produkcji Q, gracze mogą obliczyć produkty przeciętne (wzór (4)), produkty krańcowe (wzór (5)) oraz elastyczności produkcji względem wybranych czynników produkcji (wzór (6)).

Następnym krokiem jest liczenie kosztów (wzór (7) lub (8)), a także kosztów jednostkowych (wzór (9)) i kosztów krańcowych (wzór (10) lub (11)). Na podstawie obliczonych kosztów gracze mogą ustalić planowane zyski poprzez ustalanie kolejnej wielkości: ceny.



Rys. 2. Graficzne przedstawienie modelu przedsiębiorstwa

Po ustaleniu ceny należy ustalić wydatki na reklamę. Przy ustalonej wysokości ceny i wydatków na reklamę można przystąpić do analizowania wielkości popytu (wzór (14)). Popyt jest również uzależniony od cen i wydatków na reklamę ustalonych przez pozostałe zespoły graczy. Te dane zbiera Arbiter gry i ostateczny wynik dotyczący wielkości popytu przekazuje każdemu zespołowi graczy.

Po otrzymaniu informacji na temat popytu gracze przechodzą do kolejnej rundy i znów modyfikują kolejne wielkości takie jak: czynniki produkcji, cenę produktu i wydatki na reklamę.

5. Wnioski

Proponowany model przedsiębiorstwa na pewno jest jedynie pewnym przybliżeniem rzeczywistości. Model ten jednak pozwala uczestnikom gry na obserwowanie pewnego wirtualnego rynku oraz jego reakcji na ich zachowania. Model taki pozwala uczestnikom gry na „bezbolesne" zapoznanie się z realiami rządzącymi „światem biznesu" ponieważ nawet jeśli poniosą oni klęskę, to będzie to jedynie wirtualna klęska. Wiedza zdobyta na polu tego typu gry zaowocuje na pewno mniejszą ilością błędów popełnianych w rzeczywistym świecie.

Literatura

1. http://antenor.pol.lublin.pl/users/wit/
2. http://praca.korba.pl/trendy/1,8598,469,art.html
3. http://strategie.info.pl/biblioteka/artykul.htm?artykul_id=4629&s=1
4. http://www.cire.pl/praca/kwalifikacje/art_002.doc
5. http://www.ceo.org.pl/epio/metody_nauczania/metody_symulacje.htm
6. http://www.door.com.pl/wersja_pl/gry.htm
7. http://www.isosmg.com.pl/gry.html
8. http://www.progresja.com.pl/index.php?DOC=/szkolenia/gry/gra_sym.htm
9. Krzysztof Jajuga, „Ekonometria, metody i analiza problemów ekonomicznych", wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław 1999
10. Jan Korbut, „Wprowadzenie teoretyczne, gry symulacyjne w dydaktyce", dokument wewnętrzny WSP w Częstochowie, Częstochowa 2001
11. Jan Korbut, „Gra symulacyjna GRAKOR", dokument wewnętrzny WSP w Częstochowie, Częstochowa 2001

Aleksander KATKOW
Instytut Ekonometrii i Informatyki
Wydział Zarządzania
Politechnika Częstochowska
Ul. Armii Krajowej 19b
42-200 Częstochowa

Agnieszka ULFIK
Instytut Ekonometrii i Informatyki
Wydział Zarządzania
Politechnika Częstochowska
Ul. Armii Krajowej 19b
42-200 Częstochowa

NOWE NURTY W ROZWOJU NARZĘDZI DO SYMULACJI KOMPUTEROWEJ

Małgorzata ŁATUSZYŃSKA

Streszczenie: W artykule przedstawiono ewolucję narzędzi wykorzystywanych w symulacji komputerowej: od języków programowania ogólnego przeznaczenia poprzez języki i pakiety symulacyjne do systemów symulacyjnych (SS). Zarysowano również aktualne trendy rozwojowe polegające na łączeniu technik symulacyjnych z technikami sztucznej inteligencji prowadzące do pojawienia się nowej kategorii SS – inteligentnych systemów symulacyjnych.

Wprowadzenie

Symulacja komputerowa jest od dawna podstawową metodą badania systemów szczególnie złożonych. Podobnie jak inne metody musi się rozwijać i dostrajać tak, aby utrzymać swą dotychczasową pozycję wśród innych metod badawczych. Rozwój symulacji odbywa się poprzez wyszukiwanie punktów stycznych z innymi metodami badawczymi, rozwijającymi się niezależnie, przykładowo wywodzącymi się z nurtu badań nad sztuczną inteligencją (rozwój na zewnątrz), a także poprzez doskonalenie stosowanych technik (rozwój wewnętrzny). Obydwa kierunki rozwoju powodują konieczność tworzenia narzędzi implementacyjnych dostosowanych do nowej jakościowo symulacji komputerowej. Ewolucja instrumentarium symulacyjnego jest możliwa dzięki postępowi w technologii komputerowej, w tym głównie w zakresie języków programowania i systemów wspomagania decyzji.

1. Istota symulacji komputerowej

Najprościej mówiąc, symulacja komputerowa to technika służąca do imitowania działania całego systemu[1] lub też tylko naśladowania pewnej sytuacji poprzez użycie programów komputerowych[2]. Głównym celem symulacji jest dostarczenie informacji o badanym realnym systemie lub takim, który ma dopiero powstać. Informacje te mogą być wykorzystane w procesie podejmowania decyzji dotyczących danego systemu. Bardzo często stosuje się symulację również w celach dydaktycznych.

---

[1] Fizycznego, ekonomicznego, transportowego itd. W niniejszym artykule skupiono się na zastosowaniach symulacji komputerowej do badania systemów ekonomicznych.

[2] Według Wielkiej Internetowej Encyklopedii Multimedialnej symulacja komputerowa to metoda wnioskowania o zachowaniu się obiektów rzeczywistych na podstawie obserwacji programów komputerowych symulujących to zachowanie, Stosowana gdy bezpośrednie obserwowanie zachowania się obiektu jest trudne lub niemożliwe: http://wiem.onet.pl.

Obiektem, którym manipuluje się w trakcie badań symulacyjnych jest model. Jest to pojęcie bardzo ogólne, oznacza reprezentację danego systemu w postaci innej niż w rzeczywistości. Pragmatyczna definicja modelu mówi, że: „...jest to narzędzie, za pomocą którego można opisać system i jego zachowanie się w różnych warunkach zewnętrznych"[3]. W badaniach systemów stosuje się różnego rodzaju modele[4]. W symulacji największe znaczenie rzecz jasna mają modele matematyczne, które opisują rzeczywistość za pomocą równań, nierówności matematycznych, związków logicznych itp.

Symulację komputerową bardzo często myli się z metodami Monte Carlo, grami symulacyjnymi, modelami analitycznymi rozwiązywanymi na pomocą maszyny cyfrowej oraz sztuczną inteligencją[5].

Metody Monte Carlo służą do dokonywania za pomocą komputera eksperymentów losowania. W eksperymentach tego typu czas nie odgrywa żadnej roli w przeciwieństwie do symulacji komputerowej jako techniki badania systemu w czasie. Metody Monte Carlo mogą być wykorzystywane w trakcie badań symulacyjnych w celu ułatwienia manipulowania dużymi zbiorami danych i generowania liczb losowych.

Zgodnie z literaturą, gry symulacyjne służą przede wszystkim do wtórnego odzwierciedlania procesów decyzyjnych w przedsiębiorstwie. Polegają w skrócie na tym, iż pewna liczba osób (graczy) wykonuje pewne role decyzyjne według sztywno przypisanych reguł. W grze symulacyjnej można wykorzystywać modele symulacyjne w celu ustalenia oddziaływania decyzji na efekt końcowy. W przypadku gdy model ten jest do dyspozycji w postaci programu komputerowego, mówi się o symulacji człowiek-komputer. W symulacji komputerowej nie występują gracze - jest to symulacja maszynowa w czystej postaci czyli symulacja typu komputer-komputer.

Symulacja komputerowa różni się także od modeli analitycznych rozwiązywanych za pomocą komputera, gdyż modele te nie ujmują upływu czasu. W budowie modelu symulacyjnego można jednak posłużyć się do rozwiązania problemu cząstkowego modelami analitycznymi. Symulacja komputerowa służy głównie do rozwiązywania takich problemów, które są zbyt złożone dla klasycznych metod analitycznych[6].

Mniej wyraźne różnice występują pomiędzy symulacją komputerową a sztuczną inteligencją, która zajmuje się odzwierciedlaniem procesów myślowych człowieka. W modelach symulacyjnych również odzwierciedla się procesy

---

[3] Zob. (Findeisen W. i Gutenbaum J., 1985, s. 303).

[4] Liczne przekroje klasyfikacyjne prezentuje literatura, przykładowo: (Gordon G., 1974, s. 24-30), (Fishman G.S., 1981, s. 24-25), (Findeisen W. i Gutenbaum J., 1985, s. 303-324)

[5] Zob. (Biniek Z., 2002, s. 96) oraz (Fishman G.S., 1981, s. 33).

[6] Problem nieadekwatności modeli analitycznych do badania systemów szczególnie złożonych, które są podstawowym przedmiotem zainteresowania symulacji komputerowej jako techniki badawczej, był wielokrotnie dyskutowany w literaturze, por.: (Forrester J.W., 1971), (Simon H., 1982), (Niemeyer G., 1977).

myślowe, w inny jednak sposób i z tą różnicą, że procesy te są fragmentem realnej rzeczywistości.

Symulacja komputerowa ma wiele zalet, które czynią ją atrakcyjnym narzędziem analizy[7]. Po pierwsze może ona "skondensować" w takim stopniu czas, że istnieje możliwość „wysymulowania" kilku lat działalności systemu w ciągu minuty, lub w niektórych przypadkach w ciągu kilku sekund - zależy to od stopnia skomplikowania problemu i możliwości komputera. Zdolność ta pozwala badaczowi rozważyć różnorodne plany, które chciałby zbadać w bardzo krótkim czasie, podczas gdy przeprowadzenie próby na każdym, realnym systemie (jeśli w ogóle byłoby to możliwe) trwałoby bardzo długo. Symulacja komputerowa pozwala również na rozszerzenie czasu działania systemu, gdyż można za jej pomocą zbadać szczegółową strukturę zmian, których nie można byłoby zaobserwować w czasie rzeczywistym. Każdy eksperyment komputerowy, w odróżnieniu od eksperymentów laboratoryjnych, można powtórzyć w tych samych warunkach. Wyniki eksperymentów można bardzo łatwo przechowywać i porównywać. Wymienione wyżej zalety symulacji komputerowej jako metody badania powodują, że sięga się po nią przy rozpatrywaniu różnorodnych problemów[8].

Badacz wybierający symulację komputerową, powinien jednak wiedzieć, że opracowanie i uruchomienie programu symulacyjnego jest zajęciem trudnym i nie istnieją gwarancje, że wysiłek włożony w przygotowanie modelu zwróci się w postaci wartościowych wyników. Zawsze istnieje niebezpieczeństwo, iż model, jako uproszczony zapis rzeczywistości, nie obejmie wszystkich istotnych z punktu widzenia celu badania, zależności istniejących w realnym systemie[9].

2. Techniki symulacji komputerowej

Klasycznie, techniki symulacji komputerowej można podzielić na ciągłe i dyskretne[10]. Symulacja ciągła polega na modelowaniu systemu za pomocą ciągłych równań, opisujących zmiany atrybutów badanego systemu w czasie. Najprostsze modele tego typu mają postać układów liniowych równań różniczkowych o stałych współczynnikach. Symulacja ciągła jest możliwa do przeprowadzenia za pomocą komputerów, wówczas gdy stosuje się do całkowania równań małych przyrostów

---

7 Por. (Fishman G.S., 1981, s. 30-33).

8 Nie ma praktycznie dziedziny, w której nie próbowano by rozwiązać pewnych problemów za pośrednictwem symulacji komputerowej. Znane są symulacyjne modele transportowe, ekonomiczne, ekologiczne, społeczno-ekonomiczne, energetyczne.

9 Dla zwiększenia wiarygodności komputerowego modelu symulacyjnego wykonuje się różnego rodzaju testy jakościowe i ilościowe. Na ten temat obszernie w (Naylor T.H, 1975).

10 W literaturze wymienia się również symulację mieszaną, w której wykorzystuje się techniki ciągłe i dyskretne jednocześnie w sposób szeregowy lub równoległy. Tego typu symulacja służy głównie do modelowania układów elektronicznych, pozostających poza zainteresowaniem niniejszego artykułu.

czasowych. Taka symulacja nazywa się symulacją metodą stałego kroku[11]. Polega ona na powiększaniu czasu symulacyjnego o stałą, stosunkowo małą wielkość i przetwarzaniu wszystkich zdarzeń, które są do przetworzenia w określonym czasie. W symulacji dyskretnej decydującym elementem są zdarzenia. Opis systemu składa się tu głównie z równań logicznych, określających warunki występowania poszczególnych zdarzeń. Symulację dyskretną tego typu realizuje się na komputerach zarówno metodą stałego kroku jak i tzw. metodą kolejnych zdarzeń[12]. Polega ona na tym, że po przetworzeniu każdego zdarzenia czas symulacyjny jest ustawiany na chwilę wystąpienia kolejnego zdarzenia, a sekwencja zdarzeń jest określana w procesie symulacji. Techniki symulacji dyskretnej są szeroko stosowane w analizie systemów masowej obsługi, zarządzaniu zapasami oraz systemach planowania produkcji.

3. Ewolucja narzędzi symulacyjnych do badania systemów ekonomicznych

Podstawowym warunkiem zastosowania symulacji komputerowej jako techniki badawczej jest efektywne oprogramowanie modelu. Początkowo do tego celu służyły języki ogólnego zastosowania, następnie zaczęto opracowywać specjalistyczne języki symulacyjne[13] - wielokrotnie za pomocą języków ogólnego przeznaczenia takich jak przykładowo Fortran (CSL, FORSIM IV, GASP II, SIMSCRIPT), Algol (ESP, SIMON, SIMULA, SOL), PL/I (SIMPL), C (DYNAMO)[14].

Kolejny etap w ewolucji narzędzi do symulacji komputerowej to powstanie tzw. pakietów symulacyjnych[15], obejmujących z reguły następujące elementy[16]: *język symulacyjny*; *bibliotekę programów*, zawierającą zbiór makrorozkazów (oraz procedur np. matematyczno-statystycznych) o standaryzowanej zawartości; *program organizujący* (sterujący), za pomocą którego opracowuje się w danym języku symulacyjnym program realizujący zamierzony proces symulacji na określonym komputerze; oraz *kompilator*[17]. Wiele współczesnych pakietów symulacyjnych jest wyposażonych dodatkowo w moduł graficzny, umożliwiający modelującemu opisywanie struktury modelowanego systemu w postaci sformalizowanego schematu (np. VENSIM, STELLA). Na podstawie takiego schematu generowana jest częściowo procedura symulacyjna dla danego modelu.

---

[11] W literaturze określa się taką symulację również mianami: symulacji ze stałym przyrostem czasu, z zegarem lub o stałym skoku (Zajchowska-Lipiec M., 1988, s. 25).

[12] Inne nazwy to symulacja bez zegara, o zmiennym skoku, poprzez zdarzenia (tamże).

[13] O zastosowaniu języków ogólnego przeznaczenia w symulacji komputerowej oraz przyczynach opracowania i własnościach języków symulacyjnych w: (Naylor T.H., 1975, s. 536-544).

[14] Opisy różnych języków symulacyjnych między innymi: (tamże, s. 580-632); (Tyszer J., 1990, s. 101-144); (Fishman G.S., 1981, s.113-152).

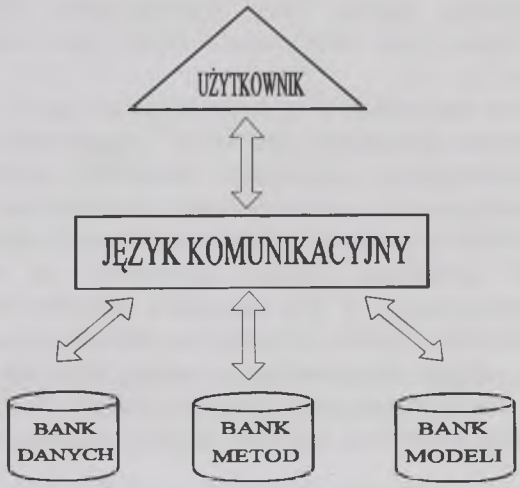
[15] Por. (Evans G.W. i in., 1973).

[16] (Lipiec-Zajchowska M., 1988, s. 28-29).

[17] Typowym pakietem symulacyjnym był przykładowo PDPlus opracowany specjalnie dla potrzeb jednej z technik symulacji ciągłej – Dynamiki Systemów.

Od przedstawionej koncepcji czteroelementowego, lub raczej pięcioelementowego (moduł graficzny), pakietu symulacyjnego dzieli nas już tylko krok do następnej kategorii narzędzi symulacyjnych, tzw. systemu symulacyjnego (SS)[18]. W literaturze przedmiotu na temat SS nie ma zbyt wielu opracowań[19]. Najczęściej określa się system symulacyjny jako taki pakiet komputerowy, który umożliwia wielokrotne wprowadzanie danych i modeli oraz wykonywanie obliczeń symulacyjnych za pomocą właściwych algorytmów i zapamiętywanie tych obliczeń na trwałych nośnikach danych[20]. Jest to jednak zbyt wąska definicja, która utożsamia SS z pakietem symulacyjnym. Nie jest również właściwym uznawanie za SS pakietu do modelowania graficznego[21].

Za system symulacyjny można uznać takie narzędzie, które umożliwia pełną integrację najróżniejszych metod, modeli i danych w procesie modelowania[22]. W nowoczesnych SS użytkownik ma do dyspozycji indywidualne, przystosowane do własnych potrzeb bazy metod, danych i modeli dostępne w trybie interaktywnym, pozwalającym na wielokrotne używanie tych samych elementów do tworzenia i rozwiązywania nowych modeli symulacyjnych[23]. Komunikację pomiędzy składnikami systemu symulacyjnego, a użytkownikiem zapewnia język komunikacyjny - aktualnie najczęściej w postaci interfejsu graficznego (rys.1).



Rys. 1. Struktura systemu symulacyjnego
Źródło: opracowanie własne.

---

18 (Zwickert E., 1981, s. 522).
19 (Biniek Z., 2002, s. 200).
20 (Sauerbier T., 1999, s.20).
21 Jak to ma miejsce w opracowaniu (Witte T. (red.), 1999).
22 Pierwsze systemy symulacyjne to między innymi TROLL oraz COMOS (Zwickert E., 1981, s. 568-569).
23 Por. (Biniek Z. 2002, s. 200).

Bank danych w systemie symulacyjnym zawiera dane empiryczne w postaci np. szeregów czasowych, dotyczące modelowanego systemu. Dane te są wykorzystywane do określania warunków początkowych symulacji, parametrów i związków funkcjonalnych pomiędzy zmiennymi modeli. Bank metod obejmuje dostępne procedury i funkcje matematyczno-statystyczne, które są niezbędne do obróbki danych empirycznych, w przypadku szacowania struktury i parametrów modelu, jak również walidacji modeli symulacyjnych. W banku modeli natomiast przechowuje się gotowe modele.

Rozwój obiektowych języków programowania otworzył nowe możliwości przed twórcami systemów symulacyjnych. Stało się możliwym wbudowane do SS mechanizmów obsługi wszystkich technik symulacji komputerowej[24] oraz implementacja technik sztucznej inteligencji[25].

4. Zastosowanie technik inteligentnych w modelowaniu symulacyjnym

W procesie modelowania symulacyjnego można wyróżnić trzy podstawowe fazy[26]: tworzenie modelu symulacyjnego, rozwiązanie modelu oraz eksperymentowanie na modelu. Przy założeniu, że komputerowa implementacja modelu symulacyjnego nastąpi przy wykorzystaniu obiektowego języka programowania, każda faza modelowania może być wspierana technikami inteligentnymi (rys. 2).

Przykładowo zagnieżdżenie systemu ekspertowego w strukturze systemu symulacyjnego stwarza sprzyjające warunki do[27]: sugerowania sposobu realizacji eksperymentu symulacyjnego, objaśniania wyników uzyskanych w trakcie eksperymentu symulacyjnego, automatycznego konstruowania nowych równań modelowych na podstawie analizy dotychczas wykonanych eksperymentów.

Algorytmy genetyczne można zastosować na etapie planowania eksperymentu symulacyjnego. W tym przypadku algorytm staje się brakującym ogniwem łączącym zbiór wyników symulacji ze zbiorem zmiennych decyzyjnych. Poprzez operacje selekcji, krzyżowania i mutacji algorytm identyfikuje takie wartości danych wejściowych, przy których wartość funkcji przystosowania mierzona jako funkcja addytywna wyników eksperymentu jest najwyższa[28].

---

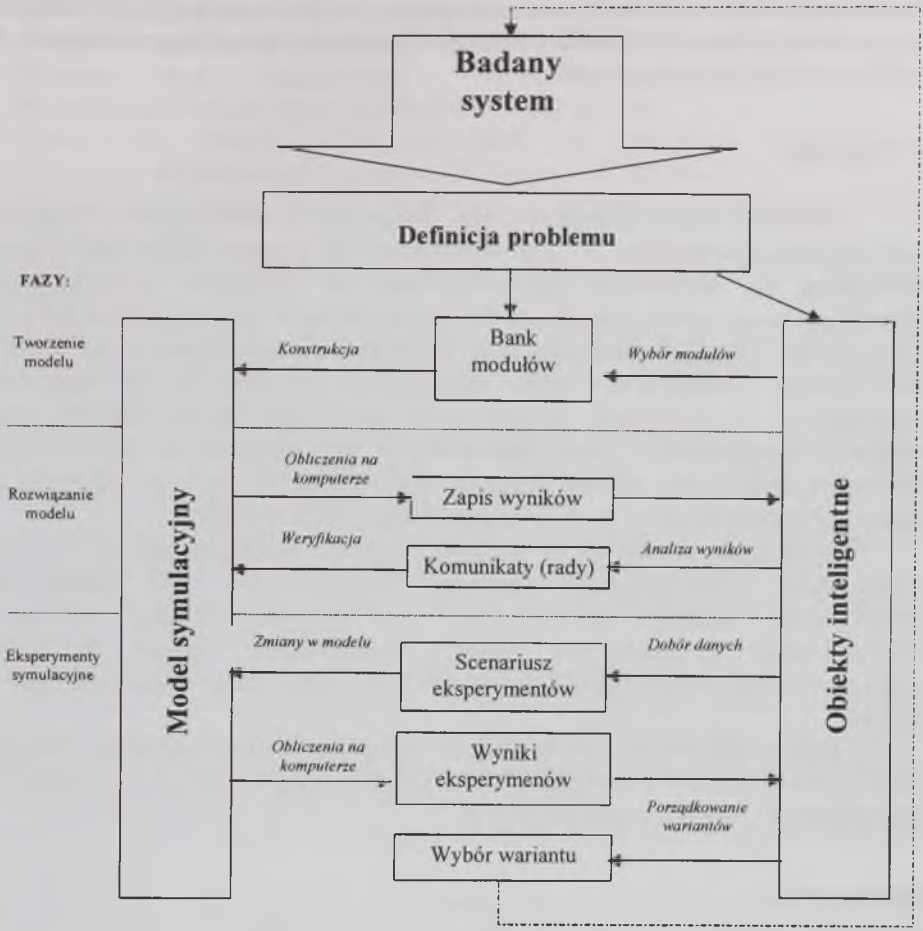
[24] Przykładami SS umożliwiających jednoczesne korzystanie z techniki symulacji ciągłej i dyskretnej jest przykładowo MOOSE (Cubert R. M. i Fishwick P. A., 1997) oraz FUNSYS (Biniek Z., 2002).

[25] Por. (Bolte J.P. i in., 1993) i (Fishwick P.A., 1994) .

[26] (Gordon G., 1974, s. 37-39)

[27] (Radosiński E., 1998).

[28] Tego typu zastosowanie algorytmu genetycznego opisano w: (Jankowski M. i Łatuszyńska M. 2001, s. 5-16).



Rys. 3. Obiekty inteligentne w procesie modelowania symulacyjnego
źródło: (Łatuszyńska M., 2002, s. 34)

Sieci neuronowe są kolejną propozycją metodologiczną, która może być użyta w środowisku symulacyjnym w trakcie konstruowania modelu, na przykład przy doborze elementów modelu czy odkrywaniu istotnych powiązań między nimi. Dzięki możliwości filtracji danych, sieci neuronowe nadają się również do stosowania na etapie opracowywania danych wejściowych do modelu symulacyjnego. Przydatność sieci neuronowych do rozwiązywania zadań optymalizacyjnych pozwala także na zastosowanie ich do poszukiwania najlepszego rozwiązania wśród uzyskanych w wyniku eksperymentowania na modelu[29].

W symulacji komputerowej można wreszcie zastosować teorię zbiorów rozmytych. W literaturze wykazuje się skuteczność tej techniki w sytuacji braku

[29] (Fishwick P., 1989, s. 702-710).

danych potrzebnych do przeprowadzenia eksperymentów symulacyjnych. Przydaje się ona szczególnie w przypadku niemożności zebrania danych statystycznych dla określonych zmiennych modelu[30].

5. Wnioski

Wszelkie układy hybrydowe, tzn. integrujące w sobie różne podejścia do rozwiązywania problemów, w tym wywodzące się z nurtu badań nad sztuczną inteligencją są konstrukcją przyszłościową dla systemów symulacyjnych. Uzupełniające się cechy różnych technik inteligentnych pozwalają tworzyć nowe rozwiązania o możliwościach znacznie większych niż oferowane przez każde z nich osobno. Przykładowo brak umiejętności uczenia się w warunkach niepewności w systemach ekspertowych może być przezwyciężony przez włączenie mechanizmów sieci neuronowych czy algorytmów genetycznych. Algorytmy genetyczne można również wykorzystać w procesie identyfikacji struktury bądź parametrów modelu neuronowego czy rozmytego.

Koncepcja zastosowania różnych technik inteligentnych w procesie modelowania symulacyjnego wymaga dopracowania na płaszczyźnie narzędziowej. Aktualnie realizowany jest projekt budowy obiektowo zorientowanego symulacyjnego systemu wspomagania decyzji, w którym zastosowano ideę algorytmów genetycznych do planowania eksperymentów na modelu[31].

Reasumując można stwierdzić, że komplementarne właściwości różnych technik powodują, że ich łączenie prowadzi do tworzenia następnej generacji SS – inteligentnych systemów symulacyjnych.

Literatura

1. Biniek Z.: *System symulacyjny jako system wspomagania decyzji* [w:] Problemy informatyki stosowanej. Roczniki informatyki stosowanej WIPS Nr 3, Wydawnictwo PS, Szczecin 2000.
2. Biniek Z.: *Elementy teorii systemów, modelowania i symulacji*. INFOPLAN 2002, wydanie internetowe: http://www.finus.com.pl.
3. Bolte J.P., Fisher J.A., Ernst D.H., *An Object-Oriented, Message-Based Environment for Integrating Continuous, Event-Driven and Knowledge-Based Simulation*, [w:] materiały na konferencję „Application of Advanced Information technologies: Effective Management of Natural Resources", ASAE, Czerwiec 1993, Spokane, WA.
4. Bonarini A., Bontempi G., *A Qualitative Simulation Approach for Fuzzy Dynamical Models*, ACM Transactions on Modelling and Computer Simulation, Vol. 4, October 1994, s. 285-313.

[30] (Bonarini A. i Bontempi G., 1994, s. 285-313).

[31] (Łatuszyńska M. i Jankowski M., 2001, s. 301-316).

5. Cubert, R. M., Fishwick, P. A.: *Moose: An object-oriented multimodeling and simulation application framework*. Departament of Computer & Information Science and Engineering, University of Florida 1997. ftp://ftp.cis.ufl.edu/pub/tech-reports/tr97/tr97-012.ps.gz.
6. Evans G.W., Wallace G.F.II, Sutherland G.L.: *Symulacja na maszynach cyfrowych*, Wydawnictwa Naukowo-Techniczne, Warszawa 1973.
7. Findeisen W (red.): *Analiza systemowa – podstawy i metodologia*. PWN, Warszawa. 1985.
8. Fishman G.S.: *Symulacja komputerowa. Pojęcia i metody.* PWE, Warszawa 1981.
9. Fishwick P., *Neural Network Models in Simulation: A Comparison with Traditional Modelling Approaches*, Winter Simulation Conference, Washington 1989, s. 702-710
10. Fishwick P.A., *Computer Simulation: Growth through Extension*, European Simulation Multiconference, Barcelona, Spain 1994, http://www.cis.ufl.edu/~fishwick/paper.htm
11. Forrester J.W.:*Planung unter dem Einfluss Komplexer Sozialer Systeme*. [w:] Politische Planung in Theorie und Praxis. Piper Verlag, Munchen 1971.
12. Gogg T.J., Mott J.R.A.: *Introduction to Simulation*. [w:] Winter Simulation Conference Proceedings 1993.
13. Gordon G.: *Symulacja systemów*. WNT, Warszawa 1974.
14. Jankowski M., Łatuszyńska M., *Dobór wartości parametrów do modelu symulacyjnego przedsiębiorstwa transportu samochodowego za pomocą algorytmu genetycznego*, Problemy Ekonomiki Transportu nr 2, Ośrodek Badawczy Ekonomiki Transportu, Warszawa 2001.
15. Law A.M, W. Kelton D.: *Simulation Modeling & Analysis*. McGraw-Hill, New York 1982.
16. Łatuszyńska M., Jankowski M.:*Symulacyjny system wspomagania decyzji z inteligentnym modułem planowania eksperymentów*. [w:] Badania Systemowe tom 28, Warszawa, 2001, s. 301-316.
17. Naylor T.H.: *Modelowanie cyfrowe systemów ekonomicznych*, PWN, Warszawa 1975.
18. Niemeyer G.: *Kybernetische System und Modelltheorie System Dynamics*, Vahlen Verlag, Munchen 1977.
19. Radosiński E., *Techniki inteligentne w analizie ekonomicznej firmy*, Szkoła Symulacji Systemów Gospodarczych, Zakopane-Antałówka 1998, http://chimera.ae.krakow.pl/~ketrii/sssg/sssgdan/a98/
20. Sauerbier T.: *Theorie un Praxis von Simulationssystemen*, Vieweg Studium Technik, Wiesbaden 1999.
21. Simon H.: *Podejmowanie decyzji kierowniczych. Nowe nurty*. PWE, Warszawa. 1982.
22. Tyszer J.: *Symulacja cyfrowa*, WNT, Warszawa. 1990.
23. Witte T. (red.): *Simulation als betriebliche Entscheidungshilfe*, Physica-Verlag, Heidelberg 1999.

24. Zajchowska-Lipiec M.: *Metody symulacji komputerowej w prognozowaniu makroekonomicznym*. PWN, Warszawa 1988.
25. Zwickert E.: *Simulation und Analyse dynamischer Systeme*, Walter de Gruyter, Berlin-New York 1981.

A NEW TRENDS IN COMPUTER SIMULATION TOOLS DEVELOPMENT

Abstract: The article presents an evolution of computer simulation tools: from general purpose programming languages, through simulation languages and packages to simulation systems (SS). It is also sketched a current development trends based on combining simulation techniques with artificial intelligence resulting in emerging a new qualitatively intelligent SS.

dr Małgorzata Łatuszyńska

Uniwersytet Szczeciński
Instytut Informatyki w Zarządzaniu
e-mail: malgorzata.latuszynska@uoo.univ.szczecin.pl

ZADANIA IDENTYFIKACJI MODELI MATEMATYCZNYCH W TWORZENIU SYSTEMU WSPOMAGANIA DECYZJI

Małgorzata PELCZAR

Streszczenie: Wdrażanie systemów informacyjnych w różnych organizacjach, wymaga umieszczenia w nich systemu wspomagania decyzji (SWD), który będzie zawierać różne modele matematyczne (optymalizacyjne i symulacyjne), pozwalające rozwiązywać problemy pojawiające się w tych organizacjach.

Zarządzanie organizacją realizuje się na trzech poziomach: strategicznym, taktycznym oraz operacyjnym. Dotychczas największy nacisk położony został na zastosowanie systemów wspomagania decyzji na poziomie zarządzania strategicznego.

W artykule przedstawione zostało rozwiązanie zadań identyfikacji modeli matematycznych w systemie wspomagania decyzji na poziomie zarządzania taktycznego, gdzie również pojawiają się słabo ustrukturalizowane problemy, na przykładzie systemu wspomagania decyzji dla Oddziału Funduszu Zdrowia i organizacji związanych z ochroną zdrowia.

1. System informacyjny Oddziału Funduszu Zdrowia

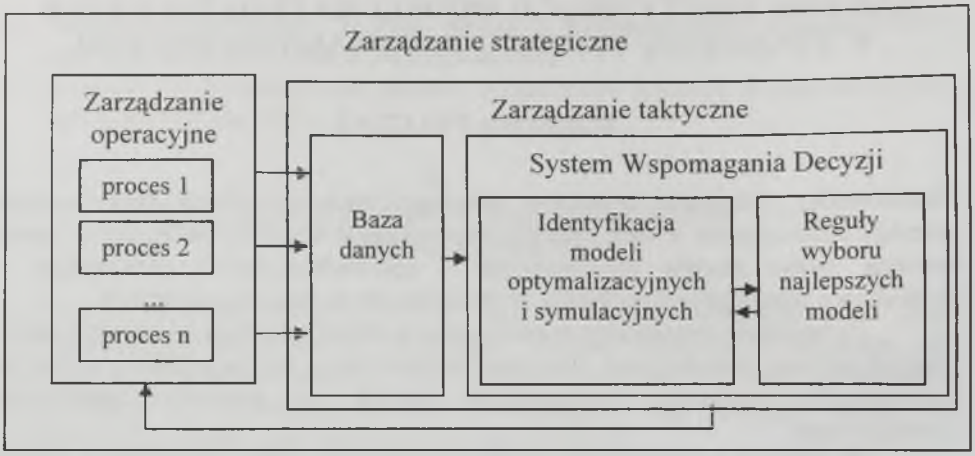
Istnieje wiele definicji systemu informacyjnego organizacji w zależności od źródła ich pochodzenia, dla potrzeb niniejszego artykułu zacytuję jedną z nich.

System informacyjny jest to wielopoziomowa struktura pozwalająca użytkownikowi na przekształcanie określonych informacji wejścia na pożądane informacje wyjścia za pomocą odpowiednich procedur i modeli.

Poniższy schemat (Rys.1.) przedstawia ogólną strukturę systemu informacyjnego organizacji ze szczególnym wyróżnieniem systemu wspomagania decyzji na poziomie zarządzania taktycznego, zawierającego proponowane moduły identyfikacji modeli matematycznych oraz reguł wyboru najlepszych modeli.

W systemie wspomagania decyzji powinny być umieszczone różne modele matematyczne tzn. z różnymi zmiennymi objaśniającymi oraz uzyskane w wyniku różnych metod identyfikacji. Następnie na podstawie wyników identyfikacji oraz doświadczeń decydenta można w danej chwili wybierać odpowiedni model, który będzie przydatny do podjęcia decyzji w określonej sytuacji.

Wdrażanie reformy ochrony zdrowia wiąże się z wdrażaniem systemów informacyjnych do organizacji związanych ze służbą zdrowia. Przykładem takiego systemu jest system informacyjny dawniej Regionalnych Kas Chorych, a od kwietnia 2003 roku Oddziałów Funduszu Zdrowia.



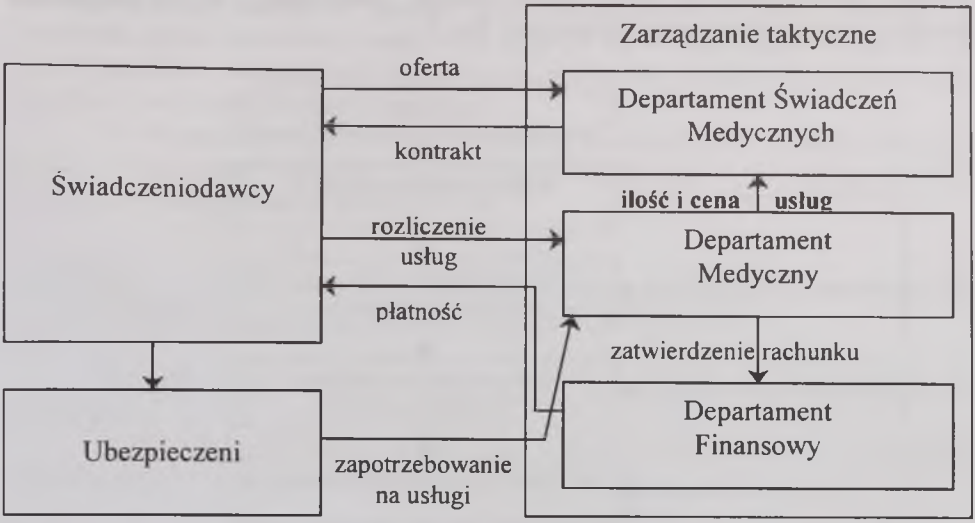
Rys. 1. Struktura systemu informacyjnego organizacji z uwzględnieniem proponowanych modułów systemu wspomagania decyzji. Źródło: opracowanie własne

Głównymi zadaniami Oddziału Funduszu Zdrowia są:

- ustalanie zapotrzebowania na usługi medyczne;
- zawieranie umów o udzielanie świadczeń na rzecz ubezpieczonych;
- ustalanie planu finansowego;
- finansowanie świadczeń realizowanych na rzecz ubezpieczonych.

Zadania te realizowane są przez odpowiednie departamenty. I tak, Departament Medyczny ustala ilość i wartość usług, które są kontraktowane przez Departament Świadczeń Medycznych ze świadczeniodawcami w oparciu o ich oferty. Następnie po realizacji usługi następuje jej rozliczenie, które wpływa do Departamentu Medycznego i tam jest zatwierdzane pod względem merytorycznym, a następnie Departament Finansowy realizuje płatność.

Z zadań tych i sposobu ich realizacji wynika ogólna struktura systemu informacyjnego Oddziału Funduszu Zdrowia, którą przedstawia poniższy schemat (Rys. 2.).



Rys. 2. Struktura systemu informacyjnego Oddziału Funduszu Zdrowia na poziomie zarządzania taktycznego. Źródło: opracowanie własne

W związku z tym, podstawowym zagadnieniem decyzyjnym pojawiającym się w zarządzaniu opieką zdrowotną jest ustalanie ilości i wartości zakupywanych dla ubezpieczonych świadczeń medycznych. Trudności polegają nie tylko na ustaleniu cen usług, ale również na tym od jakich elementów uzależnić koszty usług medycznych. Widać to najwyraźniej przy ustalaniu kosztów leczenia szpitalnego, które pochłania najwięcej środków przeznaczonych na świadczenia zdrowotne (około 40%), gdzie obecnie stosuje się modele finansowania za hospitalizację lub za szczegółowo opisaną procedurę medyczną.

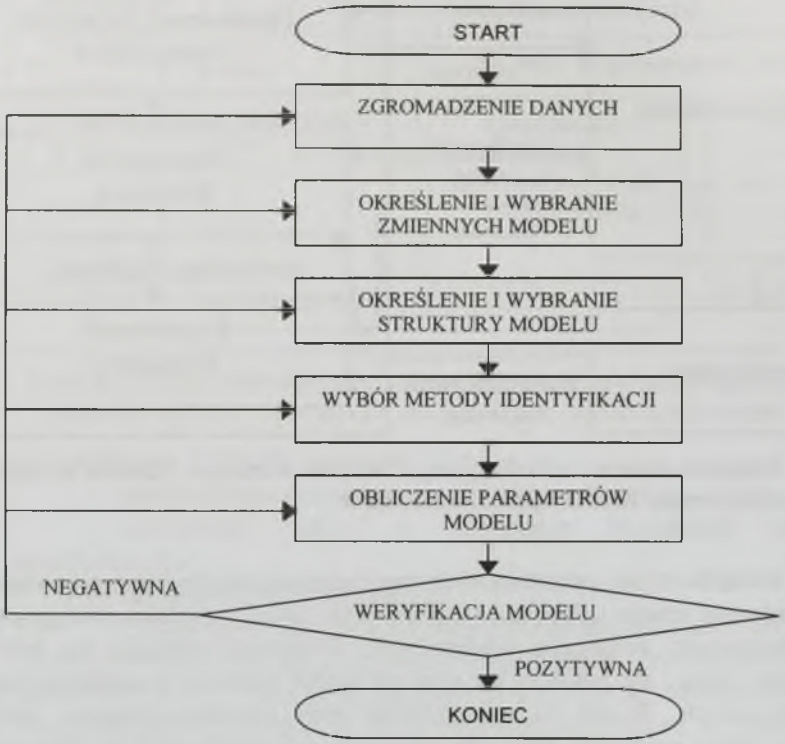
Konieczne jest zatem rozwiązanie zadania identyfikacji modeli kosztów usług medycznych oraz opracowanie reguł wyboru najlepszych modeli we wdrażanym systemie wspomagania decyzji dla Oddziału Funduszu Zdrowia.

2. Identyfikacja modeli matematycznych

Identyfikacja jest to określenie struktury i parametrów modelu matematycznego na podstawie danych empirycznych, wynika stąd niejednoznaczność rozwiązania zadania identyfikacji tak w sensie struktury jak i parametrów modelu.

Zadanie identyfikacji modeli matematycznych rozumiane w szerokim znaczeniu, czyli od ustalenia zmiennych wejściowych i wyjściowych oraz postaci

modelu matematycznego przez wybór metody identyfikacji, aż do weryfikacji modelu, przedstawia poniższy algorytm (Rys.3.).



Rys. 3. Schemat zadania identyfikacji.

W związku z tym, zadanie to może być rozwiązane w oparciu o różne metody, zarówno statystyczne jak i na przykład wykorzystujące teorię zbiorów przybliżonych. Omówię tu szerzej przesłanki zastosowania teorii zbiorów przybliżonych w systemie wspomagania decyzji.

Teoria zbiorów przybliżonych stoi u podstaw syntezy zaawansowanych i efektywnych metod analizy i redukcji zbiorów danych. W szczególności metody te mogą być stosowane w eksploracji danych, rozwiązywaniu złożonych zadań klasyfikacji i komputerowego wspomagania różnorodnych procesów decyzyjnych [6].

Zbiory przybliżone mają zastosowanie w systemach wspomagania decyzji między innymi z dwóch powodów.

Po pierwsze informacja w systemie wykorzystującym zbiory przybliżone jest zawarta w tablicy, gdzie kolumny zawierają atrybuty przykładów i podjęte na ich podstawie decyzje, natomiast wiersze odpowiadają poszczególnym przykładom. Zapis

ten reprezentuje wiedzę w postaci reguł decyzyjnych IF-THEN zebraną na podstawie opinii ekspertów, mogą więc pojawić się w nim reguły niepewne lub nawet sprzeczne. Systemy wspomagania decyzji oparte na zbiorach przybliżonych dają możliwości przetwarzania tych reguł.

Po drugie natomiast systemy te operują na wartościach dyskretnych, które często występują w systemie informacyjnym organizacji i nie można do nich zastosować znanych metod statystycznych.

3. Modele matematyczne kosztów usług medycznych w systemie wspomagania decyzji dla Oddziałów Funduszu Zdrowia

Ogólna struktura modeli kosztów usług medycznych przedstawia się następująco:

Zakładając, że koszt usługi zdrowotnej zależy od n - zmiennych objaśniających można je przedstawić w postaci funkcji:

$$K_i = F(x_1, x_2, ..., x_n, w, \varepsilon),$$

Gdzie

K_i – koszty ustalonej usługi zdrowotnej, x_j – zmienne objaśniające, w – wskaźnik wzrostu cen, ε - czynnik losowy.

Na koszt usług kontraktowanych przez Fundusz Zdrowia składa się suma kosztów poszczególnych usług zdrowotnych świadczonych przez danego świadczeniodawcę.

Biorąc pod uwagę odcinkowość modelu, a także zależności pomiędzy zmiennymi decyzyjnymi a zmiennymi objaśniającymi można przyjąć jego liniową postać:

$$K_i = c_1*x_1+c_2*x_2+...+c_n*x_n+\varepsilon,$$

gdzie

c_i – parametry modelu.

Na podstawie danych zebranych ze szpitalnych oddziałów za lata 1996, 1997, 1998 dotyczących kosztów rocznych, liczby leczonych pacjentów, liczby osobodni, liczby łóżek na oddziałach zidentyfikowane zostały modele kosztów świadczonych usług z wykorzystaniem metody regresji wielorakiej. Poniżej przedstawione są przykładowo wyniki identyfikacji modeli kosztów rocznych dla oddziałów internistycznych wraz z interpretacją ich współczynników.

Model 1. Funkcja kosztów rocznych dla szpitalnych oddziałów chorób wewnętrznych ma postać:

$$K= 652,7*n_1+42,3*n_2+6509,3*n_3-159949,3*n_4$$

gdzie n_i są to zmienne objaśniające:

n_1 – liczba leczonych pacjentów,

n_2 – liczba osobodni,

n_3 – liczba miejsc na oddziale,

n_4 – poziom referencyjny szpitala.

Model ten uwzględniający wszystkie zmienne objaśniające pokazuje, że poziom referencyjny szpitala jest uwzględniony ze znakiem ujemnym, wynika to ze słabej korelacji pomiędzy kosztami a poziomem referencyjnym dla szpitalnych oddziałów chorób wewnętrznych, należy więc pominąć go przy szacowaniu kosztów.

Model 2. Jeżeli nie uwzględniać poziomu referencyjnego szpitala to funkcja kosztów rocznych dla szpitalnych oddziałów chorób wewnętrznych ma postać:

$$K=558,4*n_1+51,8*n_2+2940,3*n_3$$

gdzie n_i są to zmienne objaśniające:

n_1 – liczba leczonych pacjentów,

n_2 – liczba osobodni,

n_3 – liczba miejsc na oddziale.

Model drugi bez uwzględniania poziomu referencyjnego szpitala pokazuje, że:

- koszty stałe zależą od liczby miejsc na oddziale i wynoszą rocznie 2940,3 zł. za każde tzw. łóżko;
- koszt leczenia 1 pacjenta wynosi 558,4 zł. plus 51,8 zł. za każdy dzień jego pobytu na oddziale.

Model 3. Jeżeli nie uwzględniać w modelu liczby miejsc na oddziale oraz poziomu referencyjnego szpitala to funkcja kosztów rocznych dla szpitalnych oddziałów chorób wewnętrznych ma postać:

$$K=605,9*n_1+57*n_2$$

gdzie

n_i są to zmienne objaśniające:

n_1 – liczba leczonych pacjentów,

n_2 – liczba osobodni.

Model bez uwzględniania poziomu referencyjnego szpitala oraz liczby miejsc na oddziale wskazuje, że koszt leczenia 1 pacjenta wynosi 605,9 zł. plus 57 zł. za każdy dzień jego pobytu na oddziale.

Model 4. Jeżeli nie uwzględniać w modelu liczby osobodni oraz poziomu referencyjnego szpitala to funkcja kosztów rocznych dla szpitalnych oddziałów chorób wewnętrznych ma postać:

$$K=1012,9*n_1+5136*n_3$$

gdzie

n_i są to zmienne objaśniające:

n_1 – liczba leczonych pacjentów,

n_3 – liczba miejsc na oddziale.

Model bez uwzględniania liczby osobodni oraz poziomu referencyjnego szpitala pokazuje, że:

- koszty stałe zależą od liczby miejsc na oddziale i wynoszą rocznie 5136 zł. za każde tzw. łóżko;

- koszt leczenia 1 pacjenta wynosi 1012,9 zł.

Model 5. Jeżeli uzależnić koszty usług tylko od liczby leczonych pacjentów to funkcja kosztów rocznych dla szpitalnych oddziałów chorób wewnętrznych ma postać:

$$K=1065,6*n_l+202796,7$$

gdzie:

n_l – *liczba leczonych pacjentów.*

Model ten najprostszy ze względu na ilość zmiennych objaśniających ustala koszt leczenia 1 pacjenta na poziomie 1065,6 zł , natomiast koszty stałe oddziału na poziomie 203 tys. zł. rocznie bez względu na poziom referencyjny czy liczbę miejsc na oddziale. Ten koszt stały jest osiągany w modelu czwartym przy 40 miejscach na oddziale.

Modele od drugiego do czwartego pozwalają zróżnicować koszty stałe w zależności od szpitala. Wskaźniki dopasowania tych modeli wynoszą 0,95 , czyli odzwierciedlają 95% obserwacji. Wskaźnik dopasowania modelu piątego wynosi 0,68.

Analiza wyników badania pojemności nośników informacji metodą Hellwiga wskazuje modele trzeci i piąty jako optymalne.

Wyniki identyfikacji parametrów takich modeli wskazują od jakich czynników uzależniać koszty leczenia szpitalnego, a także pozwalają oddzielić koszty stałe, które powinien zabezpieczyć powiat od kosztów zmiennych, które są pokrywane przez Fundusz Zdrowia.

Ponadto opracowana została metoda ustalania ceny hospitalizacji na oddziałach zachowawczych (np. pediatrycznych) oparta na teorii zbiorów przybliżonych.

Za atrybuty przykładów zostały przyjęte następujące elementy:

q1 – rodzaj schorzenia (lekkie – A, średnie – B, ciężkie – C)

q2 – długość hospitalizacji (do 24 godzin – dobowa D, od 2 do 3 dni – krótka K, powyżej 3 dni – standardowa S)

q3 – poziom referencyjny szpitala (podstawowy – 1, wyższy – 2, kliniki – 3)

Możliwymi do podjęcia decyzjami są koszty leczenia: minimalny, średni, maksymalny, które są ustalane w zależności od rodzaju oddziału szpitalnego.

Na podstawie dotychczasowych kosztów hospitalizacji na oddziałach zachowawczych oraz opinii ekspertów stworzona została informacyjna baza danych, a następnie na podstawie teorii zbiorów przybliżonych opracowane zostały reguły decyzyjne pozwalające zróżnicować koszty usług w zależności od wymienionych wyżej atrybutów [7].

Tak opracowane modele umożliwiają przeprowadzenie symulacji kosztów usług medycznych w zależności od przyjętego modelu i na podstawie tej symulacji oszacować, który model jest najlepszy w sensie największej efektywności i użyteczności.

4. Podsumowanie

W artykule przedstawione zostało zastosowanie różnych metod identyfikacji modeli matematycznych wykorzystujące zarówno tradycyjną teorię regresji wielorakiej, jak również teorię zbiorów przybliżonych w projektowaniu systemu wspomagania decyzji dla Oddziału Funduszu Zdrowia. Jest to przykład wskazujący, że w trakcie tworzenia systemów wspomagania decyzji dla organizacji można przeprowadzić identyfikację różnych modeli zarówno liniowych jak i nieliniowych, deterministycznych i stochastycznych, statycznych i dynamicznych, korzystając z różnych metod identyfikacji w zależności od rodzaju problemów decyzyjnych.

Przedstawiony przykład wskazuje również, że umieszczenie w systemie wspomagania decyzji na poziomie zarządzania taktycznego modułu modeli matematycznych oraz reguł wyboru najlepszych modeli pozwala na efektywniejsze zarządzanie organizacją.

Literatura

1. Beynon-Davies P., Inżynieria systemów informacyjnych, WNT, Warszawa 1999.
2. Drelichowski L., Elementy teorii i praktyki zarządzania z technikami informacyjnymi w przedsiębiorstwie, Wydawnictwa Uczelniane ATR w Bydgoszczy, Bydgoszcz 2000.
3. Gutenbaum J., Modelowanie matematyczne systemów, Omnitech, Warszawa 1998.
4. Kiełbasiński A., Schwetlick H., Numeryczna algebra liniowa, WNT, Warszawa 1992.
5. Kisielnicki J., Sroka H., Systemy informacyjne biznesu. Informatyka dla zarządzania. Metody projektowania i wdrażania systemów, Agencja Wydawnicza Placet, Warszawa 1999.
6. Mrózek A, Płonka L., Analiza danych metodą zbiorów przybliżonych, Akademicka Oficyna Wydawnicza PLJ, Warszawa 1999.
7. Pelczar M., Rough Sets Using for Medical Services Costs Modelling in Decision Support System, Materiały Konferencji Advanced Computer Sytems ACS 2002, Informa, Szczecin 2002.
8. Radosiński E., Systemy informatyczne w dynamicznej analizie decyzyjnej, Wydawnictwo Naukowe PWN, Warszawa – Wrocław 2001. Soderstrom T., Stoica P., Identyfikacja systemów, Wydawnictwo Naukowe PWN, Warszawa1997.
9. Soderstrom T., Stoica P., Identyfikacja systemów, Wydawnictwo Naukowe PWN, Warszawa1997.
10. Zawadzka L., Metody ilościowe w organizacji i zarządzaniu, Politechnika Gdańska, Gdańsk 1996.

mgr Małgorzata Pelczar
Wydział Informatyki Politechniki Szczecińskiej
Instytut Sztucznej Inteligencji i Metod Matematycznych
Zakład Metod Matematycznych
ul. Żołnierska 49, 71-210 Szczecin
tel. (+4891) 449 55 83
e-mail: mpelczar@wi.ps.pl

ROZMYTA OPTYMALIZACJA PORTFELA PAPIERÓW WARTOŚCIOWYCH.

Paweł SEWASTIANOW, Monika JOŃCZYK

Wstęp

Problem optymalizacji portfela papierów wartościowych jest jednym z ważniejszych zagadnień w teorii i praktyce działalności finansowych. W pionierskiej pracy Markowitza [3] zagadnienie portfelowe rozwiązano z uwzględnieniem niepewności informacji wstępnych (ceny akcji oraz ich dochodowości) za pomocą metod teorii prawdopodobieństwa, przy tym parametry niepewne przedstawione zostały w postaci gęstości zmiennych losowych. Poważnym uproszczeniem realnej sytuacji używanym w ramach klasycznego podejścia jest rozpatrywanie tylko jednego kryterium. Zakładano, że inwestor ma na celu minimalizację np. ryzyka przy gwarantowanym poziomie dochodu lub przeciwnie maksymalizację dochodu przy ograniczonym poziomie ryzyka. Jednak w praktyce, możliwości tego bardzo atrakcyjnego z teoretycznego punktu widzenia podejścia są ograniczone właśnie naturą rynku finansowego. Jak udowodnili czołowi amerykańscy specjaliści w tej dziedzinie [18], główne założenie teorii Markowitza o występowaniu normalnych rozkładów prawdopodobieństwa opisujących parametry finansowe, najczęściej nie zdarza się w praktyce. Oprócz tego rozkładu prawdopodobieństw nie są stałe, z czego wynika ze trudno przewidywać ich formy, jeżeli chodzi o planowanie przyszłego portfela. Wymienione problemy zastały w dużej mierze rozwiązane wraz z pojawieniem się teorii zbiorów rozmytych [19]. Dziś można przytoczyć cały szereg prac w tej dziedzinie [1]-[5], [8], [11]-[13]. Główną zaletą podejścia rozmytego jest możliwość zmiany gęstości prawdopodobieństwa parametrów finansowych przez odpowiednie dobranie funkcji przynależności zbiorów rozmytych, co pozwala na uwzględnienie dowolnej struktury danych otrzymanych w trakcie badań statystycznych rynku papierów wartościowych. Hipoteza normalności rozkładów prawdopodobieństw nie jest już potrzebna, dodatkową zaletą jest także możliwość wykorzystania wiedzy, doświadczenia i intuicji ekspertów do opisania funkcji przynależności. Prawie wszystkie wymienione powyżej prace, redukują sformułowane w sposób rozmyty zagadnienie do szeregu związanych między sobą problemów programowania liniowego. Przy tym używane są różnego rodzaju aproksymacje liniowe funkcji przynależności opisujących parametry finansowe, co można rozpatrywać jako poważne ograniczenie, ze względu na to, że rzeczywiste funkcje przynależności otrzymane w trakcie badań statystycznych rynku finansowego mogą mieć bardzo skomplikowane formy. Tylko w [13] problem portfelowy sformułowany został jako zagadnienie programowania nieliniowego. Warto zaznaczyć, że wszystkie wymienione powyżej podejścia oparte na metodach teorii zbiorów rozmytych, naśladują w pewnym sensie ideologię klasycznego podejścia Markowitza, ponieważ wyodrębniają tylko jedno kryterium , które trzeba

maksymalizować lub minimalizować w zależności od sytuacji. Wynika z tego, że istniejące podejścia do problemu portfelowego w ramach teorii zbiorów rozmytych są w gruncie rzeczy jednokryterialne.

Jednak wielokryterialność problemu portfelowego odnotowana została jeszcze na poziome merytorycznym przez specjalistów z dziedziny finansów [7], [8], [9], [10].

W przedstawianym artykule zagadnienie optymalizacji portfela sformułowane zostało w formie zagadnienia wielokryterialnego, nieliniowego, rozmytego programowania. Uwzględniono rywalizujące lokalne kryteria maksymalizacji zysku i minimalizacji ryzyka. W trakcie realizacji algorytmu numerycznego realizującego proponowaną metodę używamy oryginalnego probabilistycznego podejścia do porównywania liczb rozmytych.

1. Narzędzia matematyczne.

Ponieważ w ramach proponowanego podejścia używamy bezpośrednio operacji na liczbach rozmytych, poniżej przedstawione zostały podstawowe pojęcia z teorii zbiorów rozmytych.

Zbiór rozmyty A możemy zdefiniować jako:

$$A = \{(x, \mu_A(x)) : x \in X, \quad \mu_A(x) \in [0,1]\},$$

gdzie $\mu_A : X \to [0,1]$ jest funkcją przynależności elementów X do zbioru A.

Funkcja przynależności stanowi uogólnienie funkcji charakterystycznej, zbiór rozmyty zaś – uogólnienie zbioru zwykłego.

Najbardziej konstruktywnym podejściem do implementacji arytmetyki rozmytych przedziałów jest podejście oparte na reprezentacji przedziału rozmytego za pomocą α- przekrojów .

$$A = \bigcup_{\alpha} \alpha A_{\alpha},$$

gdzie A_α jest przedziałem $\{x: \mu_A(x) \geq \alpha\}$, αA_α jest przedziałem rozmytym $\{(x, \alpha): x \in A_\alpha\}$.

Niech A i B będą rozmytymi przedziałami. Zostało udowodnione, że wszystkie operacje na tych rozmytych przedziałach mogą zostać sprowadzone do operacji na przedziałach ostrych odpowiadających α- przekrojom:

$$(A @ B)_\alpha = A_\alpha @ B_\alpha$$

Z tego powodu problemy arytmetyki przedziałów rozmytych można zredukować do problemu arytmetyki ostrych przedziałów.

W przypadku prezentacji liczby rozmytej w postaci α- przekrojów, która opiera się o liczby przedziałowe, należy przedstawić podstawowe pojęcia dotyczące arytmetyki przedziałowej.

Istnieje kilka definicji arytmetyki przedziałowej, ale udowodniono, że przy zastosowaniu ich w praktycznych aplikacjach tak zwana forma „naiwna" jest formą najlepszą. Według tego, jeśli $A = [a_1, a_2]$ i $B = [b_1, b_2]$ są przedziałami, wtedy

$Z = A@B=\{ z=x@y, \ \forall x \in A, \forall y \in B \}$.

Bezpośrednią konsekwencją powyżej definicji są następujące wyrażenia:

$A+B=[a_1+b_1, b_2+b_2]$, $A-B=[a_1-b_2, a_2-b_1]$,

$A \cdot B=[min(a_1 \cdot b_1, a_2 \cdot b_2, a_1 \cdot b_2, a_2 \cdot b_1), max(a_1 \cdot b_1, a_2 \cdot b_2, a_1 \cdot b_2, a_2 \cdot b_1)]$,

$A/B=[a_1, a_2] \cdot [1/b_2, 1/b_1]$

Oczywiście istnieje wiele wewnętrznych problemów stosowanej analizy przedziałowej, np. dzielenie przez przedział zawierający zero, ale generalnie może być ona rozważana jako dobre narzędzie matematyczne dla modelowania i podejmowania decyzji w warunkach niepewności.

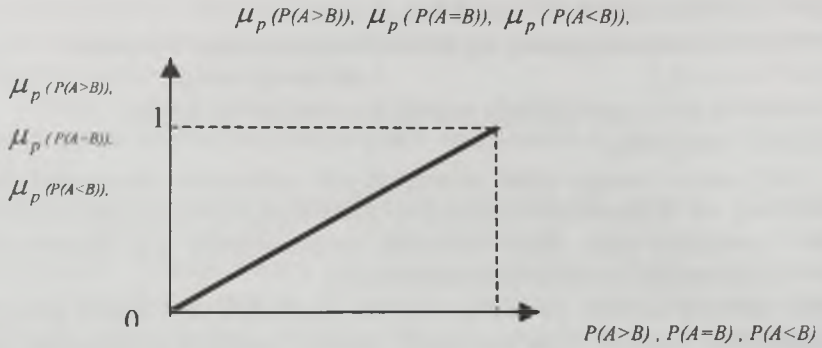
Ponieważ założone zostało, że liczba rozmyta reprezentowana będzie przez zbiór jej α-przekrojów, które są w gruncie rzeczy, zborem zwykłych przedziałów, można więc użyć arytmetyki przedziałowej, by rozwiązać problem porządkowania przedziałów rozmytych.

Kluczowym problemem w zagadnieniach optymalizacji rozmyto - przedziałowych jest porównywanie liczb lub przedziałów rozmytych. Oczywistym jest, że w ramach opisanego powyżej podejścia problem porównywania przedziałów rozmytych redukuje się do porównywania przedziałów ostrych. Należy zaznaczyć, że porównywanie przedziałów jest problemem bardzo kontrowersyjnym. Istnieją dziesiątki podejść do formułowania tej procedury, ale będziemy używali podejścia probabilistycznego, proponowanego w pracach [16] i [17]. Zaletą tego podejścia jest używanie tylko jednego podstawowego założenia, że przedział jest przedziałem stałej gęstości zmiennej losowej . Na podstawie tego, bez dodatkowych przypuszczeń, w pracach [16] i [17] udało się otrzymać kompletny zbiór wzorów dla prawdopodobieństwa nierówności P(A<B) i równości P(A=B) przedziałów ostrych i rozmytych oraz przedziałów i liczb rzeczywistych. Jednak zaproponowana w [16] i [17] metoda pozwala tylko na ocenę prawdopodobieństwa np. że B>A przy tym nie uwzględnia ona w sposób jawny stosunków szerokości porównywanych przedziałów, które mają dla nas poważne znaczenie. Rzeczywiście jeżeli chcemy maksymalizować rozmyto - przedziałowy dochód sumaryczny portfela inwestycji, jednocześnie chcielibyśmy zmniejszyć ryzyko finansowe przedstawione bezpośrednio po przez szerokości przedziałów w punkcie optimum tzn. że w trakcie realizacji algorytmu numerycznego na każdym jego kroku chcielibyśmy zwiększyć prawdopodobieństwo wzrostu otrzymanej przedziałowej/rozmyto przedziałowej wartości dochodu przy jednoczesnym zmniejszeniu szerokości reprezentującego ją przedziału.

Dlatego proponujemy dwu-kryterialne podejście dla porównywania przedziałów ostrych i rozmytych w zagadnieniach optymalizacji w warunkach niepewności (ryzyka).

Pierwsze kryterium lokalne może zostać wyrażone przez funkcję przydatności (dla problemu wyboru większego przedziału) przedstawioną na rys. 1. Zgodnie z tym kryterium, im większe jest prawdopodobieństwo A>B, tym większa wartość funkcji kryterialnej $\mu_p(P(A>B))$. Analogicznie dla przedziału B.

W ogólnym przypadku, można zapisać te kryteria jako:



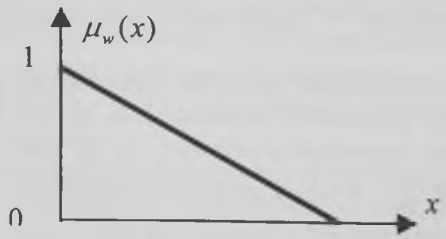
Rys. 1. Kryterium użyteczności dla prawdopodobieństwa P(A>B) i P(A<B) i P(A=B)

Drugie kryterium, związane z oceną ryzyka, może zostać przedstawione przez stosunek szerokości porównywanych przedziałów jak na Rys. 2.
W przypadku ogólnym, można kryterium to zapisać w następującej postaci:

$$\mu_w(x_A) = 1 - x_A \quad \text{, gdzie} \quad x_A = \frac{W_A}{\max(W_A, W_B)}$$

$$\mu_w(x_B) = 1 - x_B \quad \text{, gdzie} \quad x_B = \frac{W_B}{\max(W_A, W_B)}$$

Jasnym jest, że wartość kryterium jest tym większa, im mniejsza szerokość ocenianego przedziału, względem szerokości przedziału, z którym go porównujmy. W praktyce, dla przedziału szerszego wartość kryterium wynosi 0. Jeżeli np. porównujemy przedziały, A i B, i okazuje się, że przedział B jest szerszy, wówczas w x_B w mianowniku i liczniku występuje ta sama wartość, co daje wynik równy 1, w ten sposób kryterium $\mu_w(x_B) = 1 - 1 = 0$, natomiast w przeciwnym wypadku kryterium $\mu_w(x_A)$ =0. Dla przedziału o mniejszej szerokości, wartość kryterium jest tym większa im większa jest różnica szerokości pomiędzy przedziałami. Jeśli przedział węższy ma szerokość równą 0, wartość kryterium dla tego przedziału wynosi 1. W przypadku, kiedy oba przedziały mają szerokość równą 0, czyli są zdegenerowane, rozsądnym wydaje się przyjęcie dla obu przedziałów wartości tego kryterium równej 0, co wyeliminuje z oceny kryterium ryzyka, które traci w tym momencie swój sens.



Rys. 2. Kryterium użyteczności dla względnej szerokości przedziałów

Mamy dwa kryteria lokalne, które w praktyce zawsze zachowują się kontrowersyjnie, ponieważ otrzymanie większego zysku, bardziej prawdopodobne jest przy odpowiednim zwiększeniu ryzyka. Dlatego też wprowadzamy współczynniki względnej ważności kryteriów r_p, r_w charakteryzujące odpowiednio preferencje dycydenta, co do rentowności i ryzyka. Współczynniki te muszą spełniać ogólnie przyjęte ograniczenie
$(r_p + r_w)/2=1$.

Mając określone dla problemu optymalnej selekcji portfela lokalne kryteria prawdopodobieństwa i ryzyka oraz współczynniki ich względnej ważności można skonstruować globalne kryteria umożliwiające porównanie przedziałów, za pomocą najczęściej używanego addytywnego sposobu agregowania:

$$D_{A<B}(A,B) = \frac{1}{2}\cdot(r_p\mu_p(P(A<B)) + r_w\mu_w(x_A))$$

(1),

$$D_{A>B}(A,B) = \frac{1}{2}\cdot(r_p\mu_p(P(A>B)) + r_w\mu_w(x_B))$$

(2),

$$D_{A=B}(A,B) = \max(D'_{A=B}(A,B), D''_{A=B}(A,B))$$

(3)

gdzie:

$$D'_{A=B}(A,B) = \frac{1}{2}\cdot(r_p\mu_p(P(A=B)) + r_w\mu_w(x_A))$$

$$D''_{A=B}(A,B) = \frac{1}{2}\cdot(r_p\mu_p(P(A=B)) + r_w\mu_w(x_B))$$

Przedstawione powyżej kryteria wymagają kilku słów komentarza.
Pierwsze kryterium globalne $D_{A<B}(A,B)$ dotyczy sytuacji w której, sprawdzamy spełnienie warunku $A < B$. Występują tu dwa kryteria lokalne μ_p-odpowiadające za zysk, określające prawdopodobieństwo przypadku że B>A, oraz μ_w - odpowiadające za ryzyko, czyli dotyczące szerokości przedziałów A i B. Drugie kryterium globalne $D_{A>B}(A,B)$ dotyczy sytuacji dokładnie przeciwnej do opisanej powyżej, a więc przypadku gdy A > B. Trzecie kryterium globalne dotyczy sytuacji gdy A=B. Kryterium to składa się z dwóch podkryteriów globalnych: $D'_{A=B}(A,B)$ i $D''_{A=B}(A,B)$, równe jest temu podkryterium, które jest większe. Podkryteria $D'_{A=B}(A,B)$ i $D''_{A=B}(A,B)$ skonstruowane są na tej samej zasadzie jak kryterium globalne pierwsze i drugie, czyli z dwóch kryteriów lokalnych.

2. Metoda rozmyta optymalizacji portfela papierów wartościowych

Sformułujemy zadanie optymalizacji portfela jako uogólnienie klasycznego podejścia Rozmyty dochód sumaryczny portfela:

$$\hat{F} = \sum_{j=1}^{n} x_j \cdot \hat{C}_j \qquad (4)$$

gdzie: $\hat{F}$ – rozmyto- przedziałowy dochód portfela ; x_j – udział akcji j w portfelu (liczba rzeczywista);

$\hat{C}_j$ – rozmyto- przedziałowa stopa zysku z j akcji .

Jednocześnie będziemy używać standardowego ograniczenia :

$$\sum_{j=1}^{n} x_j = 1 \qquad (5)$$

Oczywiście problem polega na maksymalizacji dochodu $\hat{F}$ w sensie opisanego w rozdziale 2 podejścia probabilistycznego przy jednoczesnym wypełnieniu kryterium minimalizacji ryzyka , czyli minimalizacji szerokości $\hat{F}$. Dlatego w zgodzie z wynikami z rozdziału 2 metoda numeryczna dwu kryterialnej maksymalizacji $\hat{F}$ z uwzględnieniem wzorów (1)-(3) może być przedstawiona jako stopniowe (krok po kroku) maksymalizowanie globalnego kryterium mającego postać :

$$D_{\hat{F}_k < \hat{F}_{k+1}}(\hat{F}_k, \hat{F}_{k+1}) = \frac{1}{2} \cdot (r_p \mu_p(P(\hat{F}_k < \hat{F}_{k+1})) + r_w \mu_w(x_{\hat{F}_k})) \qquad (6)$$

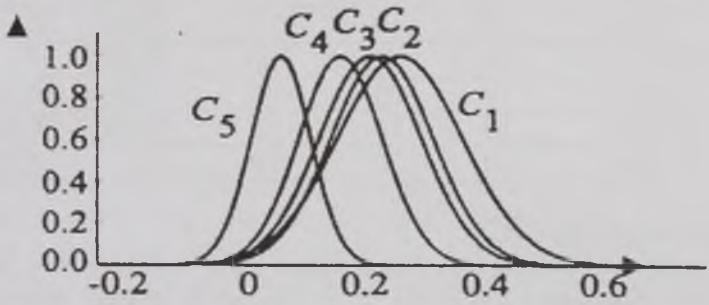
gdzie: k- numer poszczególnych kroków przejść algorytmu, $\hat{F}_k$- dochód rozmyto-przedziałowy (patrz wyrażenie (4)) w kroku k, $\mu_w(x_{\hat{F}_k}) = 1 - x_{\hat{F}_k}$ - kryterium lokalne odpowiadające za ryzyko, $x_{\hat{F}_k} = \dfrac{W_{\hat{F}_k}}{\max(W_{\hat{F}_k}, W_{\hat{F}_{k+1}})}$, $W_{\hat{F}_k}$ -szerokość przedziału funkcji F w kroku k, $W_{\hat{F}_{k+1}}$ -szerokość przedziału funkcji F w kroku $k+1$, $\mu_p(P(\hat{F}_k < \hat{F}_{k+1}))$ - kryterium lokalne odpowiadające za zysk.

W celu realizacji algorytmu numerycznego opracowane zostało rozmyto - przedziałowe rozszerzenie algorytmu bezpośredniego poszukiwania losowego [15]. Do tego algorytmu została wprowadzona dodatkowa metoda losowego wyboru wektora $x=(x_1,x_2,\ldots,x_n)$ umożliwiająca spełnienie ograniczenia (5), z określoną dokładnością ε, na każdym kroku algorytmu. Wszystkie operacje rozmyto

przedziałowe w tym operacja porównywania przedziałów rozmytych opracowane zostały za pomocą programowania obiektowego.

3.Przykład dwu-kryterialnej, rozmytej optymalizacji portfela pięciu akcji.

W celu porównywania rezultatów otrzymanych za pomocą proponowanego podejścia z wynikami innych autorów, użyliśmy przykładu dokładnie przedstawionego w [4]. W tej pracy rozpatrywano przykład portfela składającego się z pięciu rodzajów papierów wartościowych, przewidywane zyski opisane są funkcjami przynależności podobnymi do normalnych rozkładów Gaussa (patrz Rys. 3)



Rys. 3 Rozmyto-przedziałowe wartości oczekiwanego zysku z akcji C_1,C_2,C_3,C_4,C_5, [4]

Zgodnie z metodologią opisaną w rozdziale 2, przedstawione powyżej funkcje przynależności zostały przekształcone w odpowiednie zbiory α-przekrojów po czym zastosowano procedurę optymalizacji opisaną w rozdziale 3. W tabeli 1, przedstawione zostały wyniki otrzymane dla różnych współczynników względnej ważności kryteriów maksymalizacji zysku i minimalizacji ryzyka . Należy zaznaczyć, że w trakcie zmiany wartości tych współczynników otrzymano (jako przypadki szczegółowe) wszystkie rezultaty przedstawione w pracy [4], wynikające z najbardziej renomowanych jednokryterialnych metod : Markowitza, Kataoki, minimalnego ryzyka, minimalizacji rozpiętości, fraktalnych, „Modality model", „Minimax regret model".

Tablica1.Uzyskane wyniki

Współczynnik względnej ważności zysku r_p	Współczynnik względnej ważności ryzyka r_w	x_1	x_2	x_3	x_4	x_5
2	0	1	0	0	0	0
0,5971	1,4029	0,70	0,30	0	0	0
0,5970	1,4030	0,6	0,4	0	0	0
0,5969	1,4031	0,44	0,56	0	0	0
0,5968	1,4032	0,23	0,77	0	0	0
0,5960	1,4040	0	1	0	0	0
0,4000	1,6000	0	0,99	0	0	0
0,2620	1,7380	0	0,9224	0	0,0685	0
0,2600	1,7400	0	0,35	0	0,64	0
0,2500	1,7500	0	0	0	0,99	0
0,2420	1,7580	0	0	0	0,76	0,23
0,2400	1,7600	0	0	0	0,18	0,82
0,2300	1,7700	0	0	0	0	0,99

Przeanalizujmy otrzymane wyniki.

W kolumnie pierwszej znajdują się zmniejszające się współczynniki względnej ważności zysku, począwszy od 2 do 0. W kolumnie drugiej odpowiadające współczynnikom względnej ważności zysku, współczynniki względnej ważności ryzyka, spełniające zasadę $(r_p + r_w)/2=1$. Zmniejszenie współczynnika względnej ważności zysku oznacza zwiększenie współczynnika ryzyka. W kolumnie 3- 7 umieszczone są wartości udziału poszczególnych rodzajów akcji w portfelu X_1-X_5, są to konkretne wartości liczbowe. Przy czym udział x_i dotyczy akcji C_i , i=1,...,5. Rodzaje poszczególnych akcji umieszczone są (patrząc od lewej do prawej) od tych z największym oczekiwanym średnim zyskiem (i największym ryzykiem) do tych z najmniejszym zyskiem (i najmniejszym ryzykiem). Analizując powyższe wyniki wyraźnie widać, że wraz ze zwiększeniem współczynnika ważności ryzyka do portfela akcji zostają wybierane akcje z mniejszym oczekiwanym średnim zyskiem.

Przyjrzyjmy się uważniej uzyskanym wynikom. Dla współczynnika względnej ważności zysku równego 2 do portfela akcji wybierany jest pierwszy rodzaj akcji C_1.

Dla współczynnika względnej ważności zysku równego 0,5960 do portfela akcji wybrany został tylko C_2- rodzaj akcji. Sytuacja ta odpowiada wynikowi uzyskanemu przy zastosowaniu modelu fraktalnego (patrz [4]).

Przeanalizujmy wyniki uzyskane dla współczynników względnej ważności zysku zmieniających się od 0,5971 do 0,5968 wówczas do portfela akcji wybierane są akcje typu C_1 i C_2. Wraz ze zmianą układów rang zmieniają się proporcje wziętych do portfela akcji. Wraz ze zmniejszaniem współczynnika względnej ważności zysku wybierana jest większa ilość akcji typu C_2. Jest to jak najbardziej zrozumiałe i oczywiste. Akcje typu C_2 charakteryzują się mniejszym oczekiwanym średnim

zyskiem, a więc przy zmniejszaniu współczynnika względnej ważności zysku wybierana jest większa ilość tych właśnie akcji.

Podobną sytuację obserwujemy dla współczynników względnej ważności zysku zmieniających się od 0,262 do 0,26. W tym wypadku do portfela akcji wybrane zostały dwa typy akcji: akcje typu C_2 i akcje typu C_4. Tak jak poprzednio wraz ze zmianą układów współczynników względnych ważności kryteriów zmieniają się proporcje wybranych do portfela akcji. Co ciekawe wynik uzyskany dla współczynnika ważności zysku równego 0,26, odpowiada rezultatowi uzyskanemu przy zastosowaniu modelu minimalizacji rozpiętości (patrz [4]).

Obserwując wyniki uzyskane przy dwukryterialnym podejściu rozmytym, a zgromadzone w tabeli 1, można zauważyć, że akcje typu C_3 nigdy nie są brane do portfela akcji. Dzieje się tak ze względu na to że przegrywają one w rywalizacji z akcjami typu C_2- które to charakteryzują się większym średnim oczekiwanym zyskiem, a przy tym charakteryzują się taka samą niepewnością co akcje typu C_3 .

4. Uwagi końcowe

Proponowane podejście do wielokryterialnej, rozmytej, nieliniowej optymalizacji portfela papierów wartościowych zrealizowane zostało za pomocą oryginalnej metody numerycznej opartej na dwu - kryterialnym podejściu do porównywania przedziałów ostrych i rozmytych. Wprowadzone operacje matematyczne realizowane są za pomocą programowania obiektowego. Opracowana metoda sprawdzona została na przykładzie optymalizacji portfela papierów wartościowych składającego się z pięciu akcji. Udowodniono ze podejście wielokryterialne do optymalizacji portfela jest uogólnieniem wszystkich najbardziej renomowanych metod optymalizacji portfela jednokryterialnego , przy tym rezultaty otrzymane w oparciu o przedstawione podejście są szczególnymi przypadkami realizacji zagadnienia wielokryterialnego dla konkretnie wybranych współczynników względnej ważności kryteriów lokalnych , maksymalizacji zysku i minimalizacji ryzyka. Podejście wielokryterialne nie tylko uogólnia podejście klasyczne ale wyraźnie lepiej odzwierciedla naturę problemu optymalizacji portfela papierów wartościowych .

Literatura

1. Endre Pap, Zita Bosnjak, Sasa Bosnjak, *Application of fuzzy sets with diferent t-norms in the interpretation of portfolio matrices in strategic management*, Fuzzy Sets and Systems 114 (2000) 123.
2. Hideo Tanaka, Peijun Guo, I. Burhan Turksen, *Portfolio selection based on fuzzy probabilities and possibility distributions,* Fuzzy Sets and Systems 111 (2000) 387.
3. H.M. Markowitz, *Portfolio Selection: Efcient Diversıfcation of Investments,* Wiley, New York, 1959.

4. Masahiro Inuiguchi, Jaroslav Ramik, *Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem,* Fuzzy Sets and Systems 111 (2000) 3
5. Masahiro Inuiguchi , Tetsuzo Tanino, *Portfolio selection under independent possibilistic information,* Fuzzy Sets and Systems 115 (2000) 83
6. C. Zopounidis, *Multicriteria decision aid in financial management,* European Journal of Operational Research 119 (1999) 404-415
7. Jacquillat B., *Les modèles dÕévaluation et de sélection des valeurs mobilières: Panorama des recherches américaines,* Analyse Financière (1972), 11, 4e trim, p. 68-88.
8. Colson, G., Zeleny, M., 1979, *Uncertain prospects ranking and portfolio analysis under the condition of partial information,* Mathematical Systems in Economics 44, Verlag Anton Hain, Maisenheim.
9. Zeleny, M., 1977, *Multidimensional measure of risk: The prospect ranking vector. in: Zionts, S. (Ed.), Multiple Criteria Problem Solving,* Springer, Heidelberg, pp. 529-548.
10. Zeleny, M., 1982, *Multiple Criteria Decision Making,* Mc-Graw-Hill, New York.
11. T. Lorenzana, N. Márquez, S. Sardà, *An aproach to the problem of portfolio selection,* Fuzzy Economic Review Number 1, Volume I. May 1996 62-74
12. Peijun Guo, Hideo Tanaka, *Possibilistic data analysis and its aplication to portfolio selection problems,* Fuzzy Economic Review Number 2, Volume III. November 1998 1-11
13. Francesc J. Ortí, José Sáez, Antonio Terceño, *On the treatment of uncertainty in portfolio selection,* Fuzzy Economic Review Number 2, Volume VII. November 2002 22-31
14. Kaufmann A., Gupta M, *Introduction to fuzzy arithmetic-theory and applications*, New York: Van Nostrand Reinhold, 1985 – 349 p.
15. Fortemps, M. Roubens, *Ranking and defuzzification methods based on area compensation,* Fuzzy Sets and Systems (1996) 319-330.
16. P. Sewastianow, P. Róg, K. Karczewski, *A Probabilistic Method for Ordering Group of Intervals*, Informatyka teoretyczna i stosowana/Computer Science. P.Cz., Rocznik 2, 2 (2002), s. 45-53
17. Paweł Sewastianow, Paweł Róg, *A Probability Approach to Fuzzy and Crisp Intervals Ordering,* Task Quarterly 7 No 1 (2003), 147-156, Politechnika Częstochowska
18. Linsmeier T. J., Pearson N. D., *Risk measurement : an introduction to value at risk,* Champaign, IL: University of Illinois, 1996
19. L.A. Zadech, Fuzzy sets, Inform. And Control 8 (1965) 338-353

Paweł Sewastianow,
Politechnika Częstochowska
Instytut Informatyki Teoretycznej i Stosowanej,
Ul. Dąbrowskiego 73, 42-200 Częstochowa
e-mail sevast@k2.pcz.czest.pl

Monika Jończyk
Instytut Informatyki Teoretycznej i Stosowanej, Politechnika Częstochowska
Ul. Dąbrowskiego 73, 42-200 Częstochowa
e-mail: monikjon@wp.pl

WYKORZYSTANIE SYMULACJI MONTE CARLO W SZACOWANIU WARTOŚCI NARAŻONEJ NA RYZYKO (VAR) – METODOLOGIA

Wioletta SKRODZKA

Streszczenie: Jedną z podstawowych kwestii dla ilościowych miar ryzyka jest występowanie w wielu wypadkach różnic pomiędzy empirycznymi rozkładami zmian stóp procentowych, a rozkładem normalnym. Niestety większość metod pomiaru bazuje na założeniu, że względne zmiany czynników ryzyka mają rozkład normalny. Konsekwencją tego założenia jest niedoszacowanie wartości narażonej na ryzyko. Straty przekraczające jej wartość będą zdarzały się częściej, niż wynika to z teorii. Metoda Monte Carlo pozwala na uchylenie założeń stawianych przy wykorzystaniu metody parametrycznej, pozwala na nieliniowe oraz liniowe relacje między czynnikami ryzyka a kryterium pomiarowym. Stosując metodę Monte Carlo w pomiarze ryzyka finansowego za pomocą VaR należy uwzględnić występowanie związków korelacyjnych między zmiennymi losowymi. W tym przypadku można zastosować różne algorytmy postępowania. Są to m.in. dekompozycje: Choleskiego, wartości własnych i wartości osobliwych. Artykuł stanowi przegląd możliwości wykorzystania metodologii Monte Carlo w szacowaniu wartości narażonej na ryzyko.

Wstęp

Jednym z najistotniejszych czynników wpływających na rozwój nauki o zarządzaniu ryzykiem, a jednocześnie będącym odpowiedzialnym za istotną dynamizację tego zjawiska w ostatnich latach, jest silny wzrost zmienności parametrów rynkowych. Również zachodzące procesy deregulacji i globalizacji rynków finansowych zmusiły podmioty gospodarcze do zweryfikowania dotychczasowych systemów zarządzania ryzykiem finansowym. Procesom tym towarzyszy powszechna liberalizacja przepisów prawnych na poziomie międzynarodowym oraz sekurytyzacja portfeli handlowych. Zjawiska te powodują istotny wzrost ryzyka funkcjonowania instytucji. Zarządzanie ryzykiem związanym z kształtowaniem się cen rynkowych stało się pierwszoplanowym zadaniem instytucji pragnących przetrwać na globalnym rynku usług finansowych.

Zarządzanie ryzykiem jest procesem złożonym, skupiającym w sobie konieczność identyfikacji, pomiaru oraz zabezpieczenia się przed skutkami ponoszonego ryzyka. Jednakże najistotniejszym elementem systemu zarządzania jest precyzyjny pomiar tego zjawiska. Podmioty mają do wyboru wiele metod pomiaru ryzyka: analizę duracji, modele symulacyjne, metodologię wartości narażonej na

ryzyko (VaR) i inne. Przełomem w metodologii pomiaru ryzyka było powstanie metody wartości narażonej na ryzyko, czyli VaR.

Metoda VaR miała swój początek w kwantyfikacji ryzyka rynkowego. Z czasem jednak zakres jej stosowalności zaczął się rozszerzać i stopniowo ogarniać coraz to nowe sfery zarządzania ryzykiem. Metoda ta zaczęła być postrzegana jako narzędzie umożliwiające kompleksowe zarządzanie ryzykiem w instytucjach. Metoda VaR zaakceptowana została również przez instytucje nadzorcze. W styczniu 1996 roku Bazylejski Komitet Nadzoru Bankowego uchwalił poprawkę do swoich standardów wymogów kapitałowych ze względu na ponoszone ryzyko, określającą sposób wykorzystania metodologii VaR w działalności banków. Na podejściu wartości zagrożonej bazuje również cały zestaw komplementarnych miar ryzyka np. metoda przeciętnej warunkowej straty, czy też analiza scenariuszy awaryjnych (ang. stress testing).

Metodologia VaR stanowi interesujące zagadnienie statystyczne i obliczeniowe.

Jedną z podstawowych kwestii dla ilościowych miar ryzyka jest występowanie w wielu wypadkach różnic pomiędzy empirycznymi rozkładami zmian stóp procentowych, a rozkładem normalnym. Niestety większość metod pomiaru bazuje na założeniu, że względne zmiany czynników ryzyka mają rozkład normalny.[1] Założenie to dalekie jest od rzeczywistości. Konsekwencją tego założenia jest niedoszacowanie wartości narażonej na ryzyko. Straty przekraczające jej wartość będą zdarzały się częściej, niż wynika to z teorii.

W artykule zostaną przedstawione możliwości zastosowania symulacji Monte Carlo w metodologii VaR

1.Metoda symulacji Monte Carlo (MC)

Metoda Monte Carlo (MC) została zaproponowana w czasie drugiej wojny światowej w pracach związanych z symulacją dyfuzji neuronów w materiale rozszczepialnym. W 1948 roku Fermi, Metropolis i Ulam uzyskali dzięki tej metodzie oszacowania wartości własnych równania Schrödingera. Pierwsze zastosowania obejmowały obliczenia związane z wykorzystaniem pierwotnej techniki obliczeniowej. Ze wzrostem możliwości technicznych efektywność metody Monte Carlo stawała się coraz bardziej oczywista i przydatna. Metoda Monte Carlo umożliwia generowanie zmiennych losowych skomplikowanych rozkładów gęstości prawdopodobieństwa. W matematyce wykorzystuje się ją między innymi do numerycznego obliczania wartości skomplikowanych całek oznaczonych. Ma ona również szerokie zastosowanie w świecie finansów, w statystyce i ekonomii, na przykład w procesie wyceny opcji.

---

[1] Best P.: Wartość narażona na ryzyko. Obliczanie i wdrażanie modelu VaR. Oficyna Ekonomiczna, Kraków 2000

Opisując szerokie możliwości zastosowania metod MC należy zaprezentować sposoby generowania liczb losowych. Literatura przedmiotu szeroko opisuje sposoby generowania liczb losowych na podstawie rozkładu jednostajnego. Wykorzystując rozkład jednostajny możemy aproksymować nieznane rozkłady zmiennej losowej. Na przykład zmienne losowe o rozkładzie wykładniczym, rozkładzie Gaussa-Laplace'a i inne. [2]W 1947 roku John von Neuman przedstawił metodę zwaną metodą odwracania dystrybuanty.

Niech zmienna losowa X jest zmienną o rozkładzie jednostajnym na przedziale [0,1], a F jest dystrybuantą dowolnego rozkładu i jest rosnąca . Zdefiniowano funkcje quasi-odwrotną :

$$F^{-1}(t) = \inf\{ x : t \leq F(x) \} \qquad (1)$$

Wtedy zmienna losowa $Y = F^{-1}(X)$ ma rozkład dany przez dystrybuantę F.

Metodologię Monte Carlo można wykorzystać również do symulacji wielu procesów stochastycznych, służących do modelowania zmian wartości wielu wskaźników finansowych.

Ze względu na typ procesu stochastycznego symulującego konfiguracje parametrów rozważa się różne grupy metod.

Literatura przedmiotu wymienia pięć podstawowych algorytmów służących do generowania próbki losowej z zadanej gęstości oraz aproksymacji całek wielokrotnych. Są to: metoda Laplace'a, metoda IS (importance sampling), metoda RS (rejection sampling), metoda SA (simulated annealing) oraz metoda SIR (sampling importance resampling).[3] Jeśli mechanizm symulacji jest łańcuchem Markowa, to mówimy o metodach Monte Carlo typu łańcuch Markowa (MCMC). Najważniejsze z tej grupy metod to: losowanie Gibbsa i algorytm Metropolisa-Hastingsa.[4] Metody Monte Carlo oparte na łańcuchach Markowa są dynamicznie rozwijającą się dziedziną metod numerycznych, znajdującą szerokie implementacje w wielu modelach ekonometrycznych, w tym w analizie bayesowskiej[5].

W artykule ograniczono się do omówienia metod MC z funkcja ważności oraz losowania Gibbsa.

Metoda Monte Carlo z funkcją ważności MCIS

Niech X – wektor losowy o gęstości w(x) , $x \in S$.

---

[2] Metody generowania liczb losowych na podstawie rozkładu jednostajnego przedstawia Gątarek D., Maksymiuk R., Krysiak M., Witkowski Ł. (2001) : Nowoczesne metody zarządzania ryzykiem finansowym. Warszawa, WIG-Press.

[3] Wymienione algorytmy zostały opisane w Ross S.M.(1997): Simulation. Second Edition.Academic Press, Berkeley

[4] Opis metod MCMC znajduje się w : Krzykowski G.(2000): Analiza wyników badań marketingowych. Metody symulacyjne. Wydawnictwo Uniwersytetu Gdańskiego

[5] Gamerman D.: Marcov Chain Monte Carlo . Stochastic simulation for Bayesian inference. Chapman & Hall, 1997

$$I(m) = \int_S \frac{m(x)\pi_0(x)}{w(x)} w(x)dx \tag{2}$$

$$\hat{I}_l(m) = \frac{1}{l}\sum_{i=1}^{l} \frac{m(X^{(i)})\pi_0(X^{(i)})}{w(X^{(i)})} \tag{3}$$

będzie estymatorem całki I(m), gdzie $X^{(1)},..., X^{(l)}$ są niezależnymi wektorami losowymi o gęstości w(x). $\hat{I}_l(m)$ dąży według prawdopodobieństwa do I(m), gdy $l \to \infty$.

Metoda losowania Gibbsa

Metoda losowania Gibbsa umożliwia generowanie próbki losowej z gęstości $\pi(x)$ o nośniku $S \subset R^n$ na podstawie gęstości warunkowych $\pi(x_i \mid x_j, j \neq i)$, $i=1,...,n$, gdzie $x = (x_1,...,x_n) \in S$. Łańcuch Markowa wygenerowany za pomocą tej metody nazywamy łańcuchem Gibbsa.
Aby wygenerować próbkę losową o gęstości π stosujemy algorytm:
Wybieramy stan początkowy $x^{(0)} = (x_1^{(0)},...,x_n^{(0)}) \in S$ w łańcuchu Markowa $\{X^{(k)}\}$.
W kroku algorytmu k = 1,2,.., generujemy stan $x^{(k)} = (x_1^{(k)},...,x_n^{(k)}) \in S$ w następujący sposób:

$$\begin{aligned} &x_1^{(k)} \text{ jest realizacją z } \pi(x_1 \mid x_2^{(k-1)},...,x_n^{(k-1)}) \\ &x_2^{(k)} \text{ jest realizacją z } \pi(x_2 \mid x_1^{(k)}, x_3^{(k-1)},..,x_n^{(k-1)}) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (4) \\ &x_n^{(k)} \text{ jest realizacją z } \pi(x_n \mid x_1^{(k)},...,x_{n-1}^{(k)}). \end{aligned}$$

Powtarza się, aż do uzyskania zbieżności algorytmu , a następnie aż do otrzymania zadowalającej aproksymacji obliczanych całek. Oszacowanie interesujących nas charakterystyk (wartości oczekiwanych, kwantyli) uzyskuje się poprzez ich odpowiedniki empiryczne, obliczone na podstawie realizacji uzyskanych w kolejnych cyklach Gibbsa po odrzuceniu cykli spalonych.

2. Wartość narażona na ryzyko - VaR

Value at Risk jest to strata wartości rynkowej (np. instrumentu, portfela, instytucji) taka, że prawdopodobieństwo osiągnięcia jej lub przekroczenia w zadanym przedziale czasowym jest równe zadanemu poziomowi tolerancji.

Przy mierzeniu wartości zagrożonej rozważa się jedynie potencjalne straty, a nie zyski, czyli interesujące są jedynie procentowe zmiany cen finansowych w dół. Stąd definicja tej miary oparta jest na jednostronnym przedziale ufności.
VaR zależy od dwóch parametrów, które powinien określić decydent:

- poziomu tolerancji (poziomu ufności)
- horyzontu czasowego (okresu przetrzymania)

Im niższy poziom tolerancji, tym większa wartość VaR.

Okres przetrzymania oznacza przedział czasu, w którym może mieć miejsce wyliczona potencjalna strata na portfelu.
Wartość narażoną na ryzyko określa następująca relacja:

$$P\{W \leq W_0 - VaR\} = \alpha \tag{5}$$

gdzie: W_0 – obecna wartość np. instrumentu, portfela czy instytucji
W - wartość np. instrumentu, portfela czy instytucji na koniec okresu
α – poziom tolerancji.

Można zauważyć, że VaR jest funkcją kwantyla rozkładu wartości odpowiadającego zadanemu poziomowi tolerancji, oznaczanego jako W_α :

$$P\{W \leq W_\alpha\} = \alpha \tag{6}$$

Porównując wzory (5) i (6) otrzymano związek:

$$W_\alpha = W_0 - VaR \tag{7}$$

Oznaczając kwantyl rozkładu stopy zwrotu odpowiadający zadanemu poziomowi tolerancji przez R_α, równość (7) została zmodyfikowana dla:

- zwykłych stóp zwrotu

$$VaR = - R_\alpha W_0 \tag{8}$$

gdzie:

$$R_\alpha = \frac{W_\alpha - W_0}{W_0} \tag{9}$$

- dla logarytmicznych stóp zwrotu

$$VaR = W_0 (1 - e^{R_\alpha}) \tag{10}$$

gdzie:

$$R_\alpha = \ln W_\alpha - \ln W_0 \tag{11}$$

Wybór metody szacowania wartości zagrożonej jest podstawową decyzją jaką podejmuje instytucja używająca VaR.
Najczęściej stosowane metody liczenia VaR to:

- podejście wariancji – kowariancji
- symulacja historyczna
- symulacja Monte Carlo
- podejście oparte na teorii wartości ekstremalnych.[6]

---

[6] Metody te wraz z empirycznymi przykładami zostały zaprezentowane w :Skrodzka W., Włodarczyk A.: „Wykorzystanie Value at Risk do pomiaru ryzyka kursu walutowego na polskim rynku kapitałowym.", II Ogólnopolska Konferencja Naukowa: „Efektywność zastosowań systemów informatycznych", Wydawnictwo Naukowo-Techniczne, Warszawa - Szczyrk 2002

3.Zastosowanie symulacji Monte Carlo do obliczania VaR

Z uwagi na swa elastyczność metoda Monte Carlo jest najdokładniejszą techniką wyznaczania VaR. Pozwala na uchylenie założeń stawianych przy wykorzystaniu metody parametrycznej, tj. nie wymaga stawiania założeń co do kształtu rozkładów czynników ryzyka, pozwala na nieliniowe oraz liniowe relacje między czynnikami ryzyka a kryterium pomiarowym. Podstawą koncepcji pomiaru opartej o metodę Monte Carlo jest generowanie losowego procesu dla analizowanej zmiennej, pokrywającego cały jej rozkład. Pierwszym krokiem tej metody jest wybór modelu stochastycznego opisującego zachowanie się cen rynkowych. W kolejnym kroku dokonuje się symulacji wszystkich czynników ryzyka w każdym z badanych momentów czasu. Ostatnim etapem jest rewaluacja portfela instrumentów finansowych z wykorzystaniem otrzymanych w wyniku symulacji parametrów. W wyniku wielokrotnego powtórzenia tego zabiegu otrzymuje się rozkład wartości portfela, z którego odczytuje się VaR.
Szczegółowe etapy wyliczeń przebiegają następująco :

1. Określenie zmienności i korelacji dla wszystkich czynników ryzyka
2. Utworzenie szeregu cen z prawidłowymi współczynnikami zmienności i korelacji dla czynników ryzyka
3. Wyznaczenie wartości własnych i wektorów własnych macierzy korelacji
4. Utworzenie skorelowanego szeregu cen
5. Wygenerowanie zmian portfela, uporządkowanie ich, wyznaczenie percentyla oraz odczytanie VaR portfela

Etap pierwszy, podobnie jak w metodzie wariancji–kowariancji, polega na wyliczeniu współczynników zmienności i korelacji dla każdego czynnika ryzyka w portfelu.

Symulacja Monte Carlo dla pojedynczego instrumentu

Tworzenie zestawu zmian cen dla pojedynczych składników polega na generowaniu liczb losowych z rozkładu jednorodnego i konwertowaniu ich w zestaw zmian cen o rozkładzie normalnym. Pierwszy krok w tworzeniu zestawu zmian cen o rozkładzie normalnym polega na generowaniu zestawu liczb losowych. Należy wyselekcjonować generator liczb losowych, który uwzględni współczynniki zmienności i korelacji charakterystyczne dla danego składnika aktywów. Poszukiwane liczby zawierają się w przedziale (0,1). Następnie przekształca się zestaw wygenerowanych liczb losowych w zestaw liczb o rozkładzie normalnym. Wybrany przedział pozwala na założenie, że wygenerowane liczby losowe są punktami

dystrybuanty rozkładu normalnego. Pozwala to na konwertowanie ich w rozkład normalny poprzez zastosowanie funkcji odwrotnej dystrybuanty.(równość (1)) Mnożąc wygenerowane zmiany cen przez odchylenie standardowe uzyskuje się szereg zmian cen o tej samej zmienności co modelowany instrument.

Symulacja Monte Carlo dla portfela złożonego z wielu instrumentów

W portfelach wieloskładnikowych zmiany cen muszą być skorelowane. Skorelowanie generuje się w oparciu o wektory i wartości własne, które opisują jak zmiany cen grup czynników ryzyka przesuwają się w stosunku do siebie.
Powyższy proces wymaga obliczenia wektorów własnych i wartości własnych a następnie generowania skorelowanych zmian cen dla wszystkich składników aktywów. Szeregi skorelowanych losowych zmian cen są tworzone dla każdego instrumentu za pomocą równania:

$$x_k = \sum_i \sqrt{\lambda}_i \cdot x^*_k \cdot v_{ki} \cdot \sigma_k \qquad (12)$$

x_k - skorelowana losowa zmiana ceny dla instrumentu k o rozkładzie normalnym i współczynniku zmienności dla tego instrumentu
λ_i - wartość własna dla i- tego instrumentu
x^*_k - zmiana losowa ceny z szeregu o rozkładzie normalnym
v_{ki} - k-ty element wektora własnego dla i- tego instrumentu
σ_k - współczynnik zmienności k- tego instrumentu.

Wygenerowane skorelowane zmiany cen stosujemy do portfela aktywów tak jak w symulacji historycznej, czyli sortujemy, a percentyl odpowiadający pożądanemu poziomowi ufności określa wartość VaR dla portfela.

4. Korelacja pomiędzy aktywami

Stosując metodę Monte Carlo w pomiarze ryzyka finansowego za pomocą VaR należy uwzględnić występowanie związków korelacyjnych między zmiennymi losowymi. W tym przypadku można zastosować różne algorytmy postępowania. Są to m.in. dekompozycje: Choleskiego, wartości własnych i wartości osobliwych [7].

Metoda Choleskiego

Metoda Choleskiego polega na rozłożeniu macierzy wariancji-kowariancji $K=[s_{ij}]$ na dwie macierze: trójkątną górną A i A' takie, że :

$$A' A = K \qquad (13)$$

Wówczas $y = A'\varepsilon$, (14)

[7] Bałamut T. (2002): Metody estymacji Value at Risk. Materiały i studia.Zeszyt 147 wydawnictwo NBP

gdzie ε- wektor niezależnych zmiennych losowych o rozkładzie N[0;1], ma wielowymiarowy rozkład normalny z macierzą wariancji–kowariancji K.

W przypadku dekompozycji Choleskiego istnieje algorytm wyznaczania macierzy A, dany przez wzór:

$$a_{ii} = \left[s_{ii} - \sum_{k=1}^{i-1} a_{ik}^2 \right]^{1/2}$$

$$a_{ij} = \frac{\left[s_{ij} - \sum_{k=1}^{i-1} a_{ik} a_{jk} \right]^{1/2}}{a_{ii}}, \quad j = i+1, i+2, \ldots, N \qquad (15)$$

Metoda ta nie jest skomplikowana obliczeniowo, ale wymaga aby macierz korelacji była półdodatnio określona, co nie zawsze jest spełnione.

Metoda wartości osobliwych

W metodzie wartości osobliwych przedstawiamy macierz K w postaci:

$$K = UDV, \qquad (16)$$

gdzie U,V- macierze ortogonalne ; D- macierz diagonalna z wartościami osobliwymi na przekątnej.

Dla dowolnej symetrycznej macierzy kwadratowej U = V . Wówczas

$$y = Q'\varepsilon, \qquad (17)$$

gdzie $Q = D^{1/2} V'$, ε - wektor niezależnych zmiennych losowych o rozkładzie N[0;1], ma wielowymiarowy rozkład normalny z macierzą wariancji–kowariancji K.

Metoda wartości własnych

W metodzie wartości własnych przedstawiamy macierz K w postaci :

$$K = C\Delta C' = Q'Q, \qquad (18)$$

gdzie C- macierz ortogonalna, której kolumnami są wektory własne

$C'C = I$; Δ – macierz diagonalna z wartościami własnymi na przekątnej ; $Q = \Delta^{1/2} C'$.

Wówczas

$$y = Q'\varepsilon, \qquad (19)$$

gdzie ε - wektor niezależnych zmiennych losowych o rozkładzie N[0;1], ma wielowymiarowy rozkład normalny z macierzą wariancji–kowariancji K.

Podsumowanie

Ocenę techniki pomiaru skali ryzyka za pomocą VaR z wykorzystaniem metod Monte Carlo na tle alternatywnych technik przedstawia tablica 1:

Tablica1.Porównanie technik wyznaczania VaR

Kryterium porównawcze	Metoda symulacji historycznej	Metoda wariancji-kowariancji	Metoda Monte Carlo
Rodzaje pozycji			
Wycena	pełna	niepełne	pełna
Nieliniowe aktywa	tak	nie	tak
Rodzaje rozkładów			
Historyczny	normalny	empiryczny	dowolny
Zmiany w czasie	nie	tak	tak
Powiązany z bieżącą ceną instrumentów finansowych	nie	tak	tak
Rynek			
Nienormalne rozkłady	tak	nie	tak
Możliwość uchwycenia ekstremalnych ruchów	częściowo	częściowo	tak
Wdrożenie			
Omija ryzyko modelowe	tak	częściowo	nie
Łatwość obliczeniowa	łatwy	łatwy	trudny
Najważniejsze wady	zmiany w czasie	pozycje nieliniowe	ryzyko modelowe
	zdarzenia ekstremalne	zdarzenia ekstremalne	

Źródło: Opracowanie własne.

Metoda Monte Carlo jest preferowana szczególnie wtedy, gdy w portfelu mamy instrumenty typu opcja. Pozwala ona bowiem brać pod uwagę różne źródła ryzyka, czasową zwrotów lub zmienną strukturę zmienności. Metoda kowariancji nie uwzględnia współczynnika Gamma oraz opcji egzotycznych, dla których ceny wyprowadzane są za pomocą metod numerycznych. Wadą metody Monte Carlo jest długi czas obliczeń oraz zależność wyników od modelu cen lub stóp zwrotu .

Zarówno symulacja historyczna jak i symulacja Monte Carlo są w stanie uwzględnić dowolny rodzaj opcji. Jest to cecha, która wyróżnia je na tle metody kowariancji. Jednakże obie metody różnią się co do założeń dotyczących zachowania się instrumentów finansowych.

Niewątpliwą zaletą symulacji jest dokładna wycena portfela w przypadku instrumentów nieliniowych. Należy jednak wspomnieć, że podobne rezultaty można uzyskać przybliżając wartość portfela z wykorzystaniem współczynników delta, gamma, theta.

Metoda Monte Carlo sprawdza się w sytuacji, gdy portfel zawiera pozycje bez analitycznych formuł wyceniających lub gdy chcemy dokładnie zbadać ryzyko pozycji..

Należy zaznaczyć, że symulacja Monte Carlo jest procedurą czasochłonną. Problemem nie jest wygenerowanie trajektorii ruchu cen, ale wycena portfela np. za pomocą wzoru Blacka-Scholsa w przypadku, gdy portfel składa się z wielu nieliniowych derywatów.

Podejściem alternatywnym do symulacji Monte Carlo jest deterministyczne generowanie zdarzeń do testowania. [8]Eliminuje ono możliwość niewygenerowania zdarzeń ekstremalnych oraz problem zbieżności związany z losowo wygenerowanymi zdarzeniami.

Literatura

1. Best P. (2000): Wartość narażona na ryzyko. Obliczanie i wdrażanie modelu VAR. Kraków, Dom Wydawniczy ABC.
2. Bałamut T. (2002): Metody estymacji Value at Risk. Materiały i studia.Zeszyt 147 wydawnictwo NBP.
3. Gątarek D., Maksymiuk R., Krysiak M., Witkowski Ł. (2001) : Nowoczesne metody zarządzania ryzykiem finansowym. Warszawa, WIG-Press.
4. Jajuga K. (2000): Miary ryzyka rynkowego – część trzecia. Rynek Terminowy 8.
5. Jajuga K.(2000) (red.): Metody ekonometryczne i statystyczne w analizie rynku kapitałowego. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.
6. Jajuga K.(2000): Elements of Extreme Value Theory and Some Applications. Wrocław: Klasyfikacja i analiza danych; Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.
7. Krzykowski G.(2000): Analiza wyników badań marketingowych. Metody symulacyjne. Wydawnictwo Uniwersytetu Gdańskiego
8. RossS.M.(1997): Simulation. Second Edition.Academic Press,Berkeley
9. Skrodzka W., Włodarczyk A.: „Wykorzystanie Value at Risk do pomiaru ryzyka kursu walutowego na polskim rynku kapitałowym.", II Ogólnopolska Konferencja Naukowa: „Efektywność zastosowań systemów informatycznych", Wydawnictwo Naukowo-Techniczne, Warszawa - Szczyrk 2002

[8] Symulacje deterministyczne opisuje Stein M. (1987): Large sample properties of simulations using Latin hypercube sampling. Technometrics, vol.29 (2)

Wioletta Skrodzka
Instytut Informatyki i Ekonometrii Zakład Matematyki Ekonomicznej
Politechnika Częstochowska
Wydział Zarządzania

BADANIE WŁASNOŚCI ŁAŃCUCHÓW MARKOWA NA PRZYKŁADZIE POLSKIEGO RYNKU FINANSOWEGO

Wioletta Maria SKRODZKA, Aneta WŁODARCZYK

Streszczenie: W dotychczasowych badaniach empirycznych autorzy zbudowali łańcuchy Markowa drugiego rzędu dla notowań stóp procentowych WIBOR 1M i 3M, indeksu WIG20 oraz kursów walutowych USD/PLN i EUR/PLN obrazujące strukturę zmienności tychże instrumentów w poszczególnych dniach tygodnia handlowego. Celem tego artykułu jest zbadanie własności jakimi odznaczają się skonstruowane przez autorów łańcuchy Markowa drugiego rzędu, co pozwoli w konsekwencji określić klasę procesu opisującego zmienność rozważanych instrumentów polskiego rynku finansowego. Algebraiczne podejście do badania macierzy stochastycznych umożliwia ujawnienie niektórych probabilistycznych własności łańcucha Markowa, takich jak: ergodyczność i stacjonarność. Artykuł ten stanowi uzupełnienie wcześniejszych badań autorów. Ma on na celu sprawdzenie założeń stosowalności teorii ergodycznych łańcuchów Markowa dla instrumentów polskiego rynku finansowego.

Wstęp

Począwszy od połowy lat sześćdziesiątych wiele uwagi poświęcono badaniu koncepcji rynku efektywnego. Wśród badaczy, którzy stworzyli podstawy do sformalizowania tej hipotezy należy wymienić E. Fama oraz L. Bacheliera, M.G. Kendala, M.F. Osborne'a i P. Cootnera. Dzięki badaniom przeprowadzonym na rynkach finansowych charakteryzujących się długą historią odkryto różne anomalie kalendarzowe świadczące o niedoskonałości rynku i przyczyniające się do nieefektywnej alokacji kapitału. Jednym ze znanych przykładów anomalii jest tzw. efekt weekendowy, czyli zależność stóp zwrotu od dnia tygodnia oraz przerw pomiędzy dniami roboczymi w weekendy oraz święta. Autorzy tego opracowania przeprowadzili do tej pory badania empiryczne mające na celu sprawdzenie czy anomalie kalendarzowe występują również na polskim rynku finansowym [7,8]. Do wykrycia i opisania prawidłowości rządzących procesem zmienności stóp procentowych WIBOR 1M i 3M, kursów wymiany złotego względem USD i EUR oraz indeksu WIG 20 wykorzystali oni metodę opartą na teorii łańcuchów Markowa, która dotychczas w niewielkim stopniu była wykorzystywana do badania zachowań polskiego rynku finansowego. Empiryczne badania struktury zmienności zależą od założeń dotyczących specyfiki procesu, który generuje poziomy cen. Przedstawiona w ich badaniach metoda omija problem związany z szacowaniem parametrów określonego rozkładu. Tworząc dyskretne stany symbolizujące poszczególne

notowania instrumentów finansowych, użyto metodologii łańcuchów Markowa drugiego rzędu, która daje nowe spojrzenie na to zagadnienie.

Autorzy artykułu opracowali aplikację powyższej teorii do badania procesu zmienności notowań stóp procentowych, indeksu giełdowego oraz kursów walutowych. Bazowy zbiór danych pogrupowano do trzech równolicznych stanów zmienności: H dla wysokiego, N dla normalnego i L dla niskiego stanu zmienności. W opracowaniu wykorzystano model prawdopodobieństwa rozkładu stanów zrównoważonych do badania dziennej struktury przejścia między stanami, otrzymany w oparciu o metodologię łańcuchów Markowa drugiego rzędu. W tym celu oszacowano z jakim prawdopodobieństwem następuje przejście pomiędzy parami sąsiednich stanów, np. HN i NL, gdzie HN wskazuje stan normalnej zmienności pojawiający się po stanie wysokiej zmienności, który przechodzi w stan NL. Do estymacji tychże prawdopodobieństw można zastosować klasyczną lub uogólnioną metodę najmniejszych kwadratów, bądź też metodę największej wiarygodności zgodnie z teorią przedstawioną w [4, 5]. W wyniku powyższego działania otrzymano P – macierz prawdopodobieństw przejścia drugiego rzędu:

Tablica 1. Schemat macierzy prawdopodobieństw przejścia drugiego rzędu.

STANY PRZESZŁE	STANY OBECNE								
	HH	HN	HL	NH	NN	NL	LH	LN	LL
HH	$P_{HH,HH}$	$P_{HN,HH}$	$P_{HL,HH}$	0	0	0	0	0	0
HN	0	0	0	$P_{NH,HN}$	$P_{NN,HN}$	$P_{NL,HN}$	0	0	0
HL	0	0	0	0	0	0	$P_{LH,HL}$	$P_{LN,HL}$	$P_{LL,HL}$
NH	$P_{HH,NH}$	$P_{HN,NH}$	$P_{HL,NH}$	0	0	0	0	0	0
NN	0	0	0	$P_{NH,NN}$	$P_{NN,NN}$	$P_{NL,NN}$	0	0	0
NL	0	0	0	0	0	0	$P_{LH,NL}$	$P_{LN,NL}$	$P_{LL,NL}$
LH	$P_{HH,LH}$	$P_{HN,LH}$	$P_{HL,LH}$	0	0	0	0	0	0
LN	0	0	0	$P_{NH,LN}$	$P_{NN,LN}$	$P_{NL,LN}$	0	0	0
LL	0	0	0	0	0	0	$P_{LH,LL}$	$P_{LN,LL}$	$P_{LL,LL}$

Źródło: Shiyun W., Guan L.K., Chang C.: A new methodology for studying intraday dynamics of Nikkei index futures using Markov chains. Journal of International Financial Markets, Institutions and Money 9/1999

Element w macierzy P, na przykład: $P_{HN,LH}$, wskazuje na prawdopodobieństwo przejścia od stanu H do N, pod warunkiem, że w poprzednim kroku nastąpiło przejście od L do H. Boczek tablicy wnosi informacje o

wcześniejszym stanie w łańcuchu Markowa drugiego rzędu, natomiast główka - o późniejszym stanie. Następnie zdefiniowano wektor π prawdopodobieństw osiągnięcia stanów zrównoważonych w łańcuchu Markowa drugiego rzędu: $\pi = (\pi_{HH}\ \pi_{HN}\ \pi_{HL}\ \pi_{NH}\ \pi_{NN}\ \pi_{NL}\ \pi_{LH}\ \pi_{LN}\ \pi_{LL})^T$.

Hipoteza o czasie kalendarzowym zakłada występowanie odmiennych rozkładów zmienności dla dni następujących po weekendzie. Zgodnie z tym wzorcem, rynek jest bardziej aktywny i zmienności są wyższe w poniedziałki niż w pozostałych dniach tygodnia. Dlatego też, stany wysokiej zmienności częściej powinny występować po stanach wysokiej zmienności w poniedziałki niż po stanach wysokiej zmienności w pozostałych dniach handlowych. Aby sprawdzić czy $\pi_{HH} > \pi_j$, gdzie $j \neq HH$, przetestowano hipotezę zerową mówiącą o tym, że w długim okresie prawdopodobieństwa dla wszystkich przejść między dwoma stanami są identyczne w obrębie danego dnia. Do zweryfikowania tak postawionej hipotezy posłużył test zgodności chi-kwadrat. Odrzucenie hipotezy zerowej oznacza, że w długim okresie występują odmienne rozkłady zmienności dla dni następujących po weekendzie w stosunku do dni roboczych, zatem na rynku finansowym można zaobserwować występowanie efektu weekendowego.

W dotychczasowych badaniach empirycznych autorzy zbudowali łańcuchy Markowa drugiego rzędu dla notowań stóp procentowych WIBOR 1M i 3M, indeksu WIG20 oraz kursów walutowych USD/PLN i EUR/PLN obrazujące strukturę zmienności tychże instrumentów w poszczególnych dniach tygodnia handlowego. Przyjęto założenia *ex ante* o jednorodności rozważanych łańcuchów. Interesującą propozycję weryfikacji hipotezy o jednorodności łańcucha Markowa przedstawiono w [2]. Celem tego artykułu jest zbadanie własności jakimi odznaczają się skonstruowane przez autorów łańcuchy Markowa drugiego rzędu, co pozwoli w konsekwencji określić klasę procesu opisującego zmienność rozważanych instrumentów polskiego rynku finansowego.

1. Procesy Markowa

Procesy Markowa są klasą procesów stochastycznych, w których przyszłość i przeszłość – przy ustalonej teraźniejszości – są niezależne, lub przyszłość zależy od przeszłości tylko przez teraźniejszość. W [1] podano następującą definicję łańcucha Markowa:

Niech proces stochastyczny X(t) będzie procesem z czasem dyskretnym, czyli ciągiem zmiennych losowych X_0, X_1, ... Zakładamy, że te zmienne losowe są typu skokowego. Przyjęto, że wartość zmiennej losowej X_n (n = 0, 1,...) określa stan procesu w chwili n.

Wartości zmiennych losowych X_n należą do skończonego zbioru $\{s_1, s_2, ..., s_r\}$, gdzie $s_i \neq s_j$, $s_i \in R$, i=1,2,...r.

Definicja 1

Skończonym łańcuchem Markowa (SŁM) o przestrzeni stanów $S=\{1,2,..,r\}$ nazywa się proces stochastyczny $\{X_n, n \in N\}$ taki, że dla każdego i_0, $i_1,..., i_{n-2}, i_{n-1}, i_n \in S$ oraz $n \in N$ zachodzi równość:

$$P(X_n = i_n \mid X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i_{n-1}) = P(X_n = i_n \mid X_{n-1} = i_{n-1}) \quad (1)$$

Warunek (1) nosi nazwę własności Markowa.

Prawdopodobieństwo warunkowe

$$P(X_n = i_n \mid X_{n-1} = i_{n-1}) = p_{in\ in-1}(n) \quad (2)$$

jest nazywane prawdopodobieństwem przejścia łańcucha ze stanu i_{n-1} w chwili n-1 do stanu i_n w chwili n.

1.1 Skończone jednorodne łańcuchy Markowa (SJŁM)

W [6] podano następujące definicje związane z klasą jednorodnych łańcuchów Markowa:

Definicja 2

Skończonym jednorodnym łańcuchem Markowa nazywa się skończony łańcuch Markowa o przestrzeni stanów $S=\{1,2,..,r\}$ taki, że dla każdego n, $\alpha \in N$ prawdziwa jest równość:

$$P(X_n = j \mid X_{n-1} = i) = P(X_{n+\alpha} = j \mid X_{n+\alpha-1} = i) = p_{ij} \quad (3)$$

Przy czym p_{ij} oznacza prawdopodobieństwo przejścia łańcucha ze stanu i do j w jednym kroku .

Zatem dla jednorodnego łańcucha Markowa prawdopodobieństwa przejścia pomiędzy stanami w jednym kroku są stałe w czasie i niezależne od czasu trwania procesu.

Własności łańcuchów Markowa:

$$\sum_{j \in S} p_{ij} = 1, \text{ dla } i \in S \quad (4)$$

$$p_{ij} \geq 0, \text{ dla } i,j \in S \quad (5)$$

Definicja 3

Bezwarunkowym rozkładem prawdopodobieństw stanu skończonego jednorodnego łańcucha Markowa (SJŁM) w momencie $n \in N$ nazywa się wierszowy wektor

$$d_n = [\, d_{n1}\ d_{n2} \ldots d_{nr}], \quad \text{gdzie } d_{nj} = P(X_n = j), \quad j \in S, \quad n \in N \quad (6)$$

przy czym $d_{nj} \geq 0$, dla $j \in S$, $n \in N$,

$$\sum_{j \in S} d_{nj} = 1, \text{ dla } n \in N.$$

Bezwarunkowe prawdopodobieństwo stanu d_{nj}, dla $j \in S$, $n \in N$ można obliczyć ze wzoru na prawdopodobieństwo całkowite:

$$d_{nj} = P(X_n = j) = \sum_{i \in S} P(X_{n-1} = i)\, P(X_n = j \mid X_{n-1} = i) = \sum_{i \in S} d_{n-1,i}\, p_{ij} \qquad (7)$$

Stosując macierzowe równanie rekurencyjne można otrzymać następującą zależność:

$$d_n = d_0 P^n \qquad (8)$$

Znajomość macierzy prawdopodobieństw przejścia P oraz rozkładu początkowego d_0 pozwala na wyznaczenie bezwarunkowych rozkładów prawdopodobieństw d_n łańcucha w dowolnej chwili $n \in N$, czyli pełnej charakterystyki procesu.

Definicja 4

Rozkładem stacjonarnym SJŁM o macierzy prawdopodobieństw przejścia P nazywa się wierszowy wektor $d = [d_j]$, $j \in S$, taki, że

$$dP = d \qquad (9)$$

przy czym $d_j \geq 0$, dla $j \in S$, $\sum_{j \in S} d_j = 1$

Na mocy twierdzenia Dooba[1] dla każdej macierzy stochastycznej P istnieje przynajmniej jeden wektor d spełniający równość (9) .

1.2. Klasyfikacja skończonych jednorodnych łańcuchów Markowa

1.2.1. Własności macierzy stochastycznej P

Zależnie od wartości własnych macierzy P, ciąg potęg P^n ma różne własności przy $n \to \infty$, które są podstawą do wnioskowania o zachowaniu się łańcucha w długim okresie czasu. Odpowiednio do wartości własnych i ich krotności można podzielić stochastyczne macierze P na grupy:

Tablica2. Klasyfikacja macierzy stochastycznych

	$\alpha_1 = 1$	$\alpha_1 > 1$
$\forall_i \in \{2,\ldots r\}$ $\lvert\lambda_i\rvert < 1$	Nierozkładalne i niecykliczne czyli **regularne**	Rozkładalne i niecykliczne
$\exists_i \in \{2,\ldots r\}$ $\lvert\lambda_i\rvert = 1$	Nierozkładalne i cykliczne	Rozkładalne i cykliczne

Źródło: Podgórska M., Śliwka P., Topolewski M., Wrzosek M.: Łańcuchy Markowa w teorii i w zastosowaniach. Oficyna Wydawnicza SGH, Warszawa 2000

[1] Wentzell A. D.: Wykłady z teorii procesów stochastycznych. PWN, Warszawa 1980

W [5, 6] podano następujące charakterystyki dla macierzy P:

- Macierz stochastyczną $P \in M(n,n)$ nazywa się macierzą nierozkładalną, jeśli wartość własna $\lambda_1 = 1$ jest pojedynczym pierwiastkiem równania charakterystycznego tej macierzy.
- Macierz stochastyczną $P \in M(n,n)$ nazywa się macierzą niecykliczną, gdy wartość własna
  $\lambda_1 = 1$ jest jedynym pierwiastkiem równania charakterystycznego o module 1.
- Macierz nierozkładalna i niecykliczna nosi nazwę macierzy regularnej.
- Macierz stochastyczną $P \in M(n,n)$ nazywa się macierzą przywiedlną (redukowalną), jeśli istnieje permutacja[2], która sprowadza P do postaci:

$$\widetilde{P} = \begin{bmatrix} B & 0 \\ C & D \end{bmatrix} \tag{10}$$

w której B,D są macierzami kwadratowymi.

- Macierz $E \in M(n,n)$ nazywa się ergodyczną, jeżeli jest postaci:

$$E = \begin{bmatrix} e_1 & e_2 & \dots & e_n \\ e_1 & e_2 & \dots & e_n \\ \dots & \dots & \dots & \dots \\ e_1 & e_2 & \dots & e_n \end{bmatrix} \tag{11}$$

1.2.2. Zbieżność szeregów potęgowych macierzy P

Twierdzenie Dooba i Frécheta dają podstawy do wnioskowania o zachowaniu łańcucha w długim czasie.

Twierdzenie 1 Dooba[3]

Dla każdej macierzy stochastycznej $P \in M(n,n)$ istnieje macierz stochastyczna $A \in M(n,n)$ taka, że:

$$\lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} P^i = A \tag{12}$$

przy czym

---

[2] Permutacją macierzy kwadratowej nazywa się permutację dowolnej liczby wierszy i kolumn o tych samych numerach.

[3] Podgórska M., Śliwka P., Topolewski M., Wrzosek M.: Łańcuchy Markowa w teorii i w zastosowaniach. Oficyna Wydawnicza SGH, Warszawa 2000

$$PA=AP=A=A^2 \tag{13}$$

Wzór Dooba dla poszczególnych macierzy P przyjmuje następującą postać:

$$\lim_{k\to\infty}\frac{1}{k}\sum_{t=1}^{k}P^t=\begin{cases}\lim\limits_{k\to\infty}P^k=E & \text{dla macierzy P regularnej}\\ E & \text{dla macierzy P nierozkladalnej, cyklicznej}\\ \lim\limits_{k\to\infty}P^k=A & \text{dla macierzy P rozkladalnej, niecyklicznej}\\ A & \text{dla macierzy P rozkladalnej, cyklicznej}\end{cases}$$

gdzie macierz A nie jest macierzą ergodyczną.

Wobec powyższych własności macierzy regularnej prawdziwe jest następujące twierdzenie [5, 6]:

Twierdzenie 2

Jeżeli macierz prawdopodobieństwa przejścia P skończonego jednorodnego łańcucha Markowa jest regularna, to łańcuch jest ergodyczny, a to oznacza, że istnieje granica

$$\lim_{n\to\infty}p_{ij}^{n}=e_j \quad \text{dla } i,j\in S \tag{14}$$

przy czym

$$\sum_{j\in S}e_j=1 \text{ oraz } \quad 0\le e_j\le 1 \tag{15}$$

Bezwarunkowy rozkład łańcucha ergodycznego zmierza wraz z czasem do rozkładu stacjonarnego e, który jest niezależny od początkowego rozkładu łańcucha d_0.

Twierdzenie 3

Dla łańcucha ergodycznego istnieje jeden rozkład stacjonarny. Jest nim rozkład ergodyczny e otrzymany jako rozwiązanie układu równań:

$$\begin{cases}eP=e\\ e1=1\end{cases} \tag{16}$$

Jeśli macierz P jest nieprzywiedlna to e>0 i możliwe jest znalezienie się łańcucha w każdym ze stanów, a jeśli jest przywiedlna to e≥0 i trafienie do niektórych stanów po osiągnięciu stacjonarności jest niemożliwe.

Tempo zbieżności kolejnych potęg macierzy stochastycznej do postaci ergodycznej określa współczynnik ergodyczności $\tau_1(P)$ lub $\tau_2(P)$:

Definicja 5[4]

[4]Podgórska M., Śliwka P., Topolewski M., Wrzosek M.: Łańcuchy Markowa w teorii i w zastosowaniach. Oficyna Wydawnicza SGH, Warszawa 2000

Współczynnikiem ergodyczności $\tau_1(P)$ macierzy stochastycznej P nazywa się jej drugą co do modułu wartość własną:

$$\tau_1(P) = \max\{|\lambda_i|\} \quad \text{dla } i \in \{2, 3, \ldots, r\} \tag{17}$$

gdzie $\lambda_1=1$.

Im mniejsza wartość $\tau_1(P)$ tym szybciej macierz P^n osiąga postać ergodyczną.

Definicja 6[5]

Współczynnikiem ergodyczności $\tau_2(P)$ macierzy stochastycznej P nazywa się wyrażenie:

$$\tau_2(P) = \min_{1 \le i, j \le r} \sum_{k=1}^{r} \min(p_{ik}, p_{jk}) \tag{18}$$

Współczynniki ergodyczności osiągają wartości z przedziału [0, 1].

Twierdzenie 4

Jeżeli macierz prawdopodobieństwa przejścia P skończonego jednorodnego łańcucha Markowa jest nierozkładalna i niecykliczna, to łańcuch jest ergodyczny w sensie Césaro, a to oznacza, że istnieje granica

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} = e_j \text{ dla } i, j \in S \tag{19}$$

przy czym spełniona jest własność (15).

W celu wyznaczenia wiersza e granicznej macierzy E należy rozwiązać układ równań identyczny jak (16).

Łańcuch ergodyczny w zwykłym sensie jest także ergodyczny w sensie Césaro.

Twierdzenie 5

Jeżeli macierz prawdopodobieństwa przejścia P skończonego jednorodnego łańcucha Markowa jest rozkładalna i niecykliczna, to łańcuch jest nieergodyczny. Dla takiego łańcucha zachodzi równość:

$$\lim_{n \to \infty} d_n = \lim_{n \to \infty} d_0 P^n = d_0 \lim_{n \to \infty} P^n = d_0 A \tag{20}$$

Wektor d_0A jest rozkładem stacjonarnym łańcucha o rozkładalnej i niecyklicznej macierzy prawdopodobieństwa przejścia P. W tym przypadku istnieje nieskończenie wiele rozkładów stacjonarnych, a każdy z nich zależy od początkowego rozkładu d_0.

Twierdzenie 6

Jeżeli macierz prawdopodobieństwa przejścia P skończonego jednorodnego łańcucha Markowa jest rozkładalna i cykliczna, to łańcuch nie jest ergodyczny ani w zwykłym sensie, ani w sensie Césaro.

---

[5] Iosifescu M.: Skończone procesy Markowa i ich zastosowania. PWN, Warszawa 1988

Ciąg bezwarunkowych rozkładów tego łańcucha jest zbieżny przy $n \to \infty$ tylko w sensie średniej

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} d_k = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} d_0 P^k = d_0 \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P^k = d_0 A \qquad (21)$$

a graniczny wektor d_0A jest zależny od początkowego rozkładu łańcucha d_0 i jest rozkładem stacjonarnym tego łańcucha.

1.2.3 Własności skończonego jednorodnego łańcucha Markowa

W zależności od typu macierzy prawdopodobieństw przejścia P można określić własności SJŁM:

Tablica 3. Własności SJŁM w zależności od typu macierzy prawdopodobieństw przejścia

	Macierz prawdopodobieństw przejścia P	Własności łańcucha
1	Nierozkładalna i niecykliczna (regularna) oraz przywiedlna	- łańcuch ergodyczny - rozkład graniczny e niezależny od rozkładu początkowego d_0 - jeden rozkład stacjonarny $e \geq 0$
2	Nierozkładalna i niecykliczna (regularna) oraz nieprzywiedlna	- łańcuch ergodyczny - rozkład graniczny e niezależny od rozkładu początkowego d_0 - jeden rozkład stacjonarny $e > 0$
3	Nierozkładalna i cykliczna oraz przywiedlna	- łańcuch ergodyczny w sensie Césaro - rozkład graniczny (w sensie średniej) e niezależny od rozkładu początkowego d_0 - jeden rozkład stacjonarny $e \geq 0$
4	Nierozkładalna i cykliczna oraz nieprzywiedlna	- łańcuch ergodyczny w sensie Césaro - rozkład graniczny (w sensie średniej) e niezależny od rozkładu początkowego d_0 - jeden rozkład stacjonarny $e > 0$
5	Rozkładalna i niecykliczna	- łańcuch nieergodyczny - rozkład graniczny e zależny od rozkładu początkowego d_0 - wiele rozkładów stacjonarnych
6	Rozkładalna i cykliczna	- łańcuch nieergodyczny - rozkład graniczny (w sensie średniej) zależny od rozkładu początkowego d_0 - wiele rozkładów stacjonarnych

Źródło: Podgórska M., Śliwka P., Topolewski M., Wrzosek M.: Łańcuchy Markowa w teorii i w zastosowaniach. Oficyna Wydawnicza SGH, Warszawa 2000

2. Badania empiryczne - analiza własności łańcuchów Markowa drugiego rzędu dla instrumentów podstawowych polskiego rynku finansowego

Artykuł ten stanowi uzupełnienie wcześniejszych badań autorów przedstawionych w [7, 8]. Ma on na celu sprawdzenie założeń stosowalności teorii ergodycznych łańcuchów Markowa dla instrumentów polskiego rynku finansowego.

Wykorzystując przedstawioną powyżej teorię łańcuchów Markowa zbadano zachowanie się stopy procentowej WIBOR 1M, 3M, kursów wymiany złotego względem USD i EUR, indeksu WIG 20.

Rozważane szeregi czasowe dotyczą poziomu stopy procentowej WIBOR 1M, 3M w okresie od 2 stycznia 1995 do 13 marca 2003, kursu USD/PLN i EUR/PLN w okresie od 4 stycznia 1999 do 28 lutego 2003 oraz indeksu WIG 20 w okresie od 4 stycznia 1999 do 7 czerwca 2002. Na ich podstawie zgodnie z teorią przedstawioną we wprowadzeniu dla poszczególnych dni tygodnia skonstruowano łańcuchy Markowa drugiego rzędu. W [7, 8] zbudowano macierze P prawdopodobieństw przejścia pomiędzy sąsiednimi stanami dla zmienności notowań tychże instrumentów w poszczególnych dniach tygodnia. Przykładowa macierz dla poniedziałku dla kursu USD/PLN została przedstawiona w tablicy 4:

Tablica 4. Poniedziałkowa macierz prawdopodobieństw przejścia dla USD/PLN

	HH	HN	HL	NH	NN	NL	LH	LN	LL
HH	0,515	0,212	0,273	0	0	0	0	0	0
HN	0	0	0	0,292	0,375	0,333	0	0	0
HL	0	0	0	0	0	0	0,125	0,5	0,375
NH	0,53	0,235	0,235	0	0	0	0	0	0
NN	0	0	0	0,3	0,5	0,2	0	0	0
NL	0	0	0	0	0	0	0,321	0,429	0,25
LH	0,36	0,24	0,4	0	0	0	0	0	0
LN	0	0	0	0,318	0,409	0,273	0	0	0
LL	0	0	0	0	0	0	0,258	0,387	0,355

Źródło: Obliczenia własne

Na podstawie wyznaczonych macierzy prawdopodobieństw przejścia dla łańcucha Markowa drugiego rzędu zgodnie z twierdzeniem Frécheta obliczono wektory prawdopodobieństw długookresowych π:

Tablica 5. Wektor prawdopodobieństw długookresowych dla USD/PLN

	Po	Wt	Sr	Cz	Pi
HH	0,17086	0,13418	0,13331	0,15802	0,16566
HN	0,07837	0,10162	0,11788	0,09243	0,11178
HL	0,09935	0,07370	0,07214	0,08619	0,07664
NH	0,11133	0,11129	0,10157	0,08028	0,07600
NN	0,16161	0,10856	0,14179	0,09271	0,09563
NL	0,09272	0,13443	0,10868	0,12606	0,13101
LH	0,06640	0,06404	0,08845	0,09834	0,11242
LN	0,12568	0,14409	0,09238	0,11391	0,09523
LL	0,09368	0,12809	0,14380	0,15207	0,13564

Źródło: Obliczenia własne

Zgodnie z klasyfikacją skończonych jednorodnych łańcuchów Markowa zawartą w tablicy 3 zbadano własności macierzy prawdopodobieństw przejścia P. Pozwoli to na pełną charakterystykę rozważanego procesu. Celem zbadania ergodyczności rozważanych łańcuchów Markowa drugiego rzędu na mocy twierdzenia 2 zbadano pod kątem regularności macierze P. W myśl teorii przedstawionej w tablicy 2 obliczono wartości własne, ich moduły i krotności dla poszczególnych macierzy stochastycznych. Wyniki obliczeń dla przykładowo wybranych macierzy prawdopodobieństw przejścia poniedziałkowych i piątkowych zostały zamieszczone w tablicach 6 - 8:

Tablica 6. Charakterystyki algebraiczne macierzy prawdopodobieństw przejścia dla stóp procentowych

WIBOR 1M				WIBOR 3M			
Poniedziałek		Piątek		Poniedziałek		Piątek	
Wartości własne macierzy	Moduły wartości własnych	Wartości własne macierzy	Moduły wartości własnych	Wartości własne macierzy	Moduły wartości własnych	Wartości własne macierzy	Moduły wartości własnych
1	1	1	1	1	1	1	1
0,629	0,629	0,638	0,638	0,612	0,612	0,524	0,524
-0,319	0,319	0,430	0,430	0,354	0,354	-0,166+0,086i	0,187
-0,066+0,241i	0,250	-0,315	0,315	-0,089+0,187i	0,207	-0,166-0,086i	0,187
-0,066-0,241i	0,250	-0,103+0,103i	0,145	-0,089-0,187i	0,207	0,217	0,217
0,258	0,258	-0,103-0,103i	0,145	-0,239	0,239	0,077+0,076i	0,108
0,180	0,180	0,027+0,098i	0,101	-0,154	0,154	0,077-0,076i	0,108
-0,004+0,033i	0,034	0,027-0,098i	0,101	0,134	0,134	-0,022	0,022
-0,004-0,033i	0,034	0,093	0,093	0,083	0,083	-0,035	0,035

Źródło: Obliczenia własne

Analiza wyników zawartych w tablicy 6 pokazuje, że rozważane macierze stochastyczne są regularne. Wskazuje na to wartość własna $\lambda_1 = 1$, która w przypadku każdej z tych macierzy jest pojedynczym pierwiastkiem równania charakterystycznego oraz jedynym pierwiastkiem równania charakterystycznego o module 1.

Tablica 7. Charakterystyki algebraiczne macierzy prawdopodobieństw przejścia dla kursów walutowych

USD/PLN				EUR/PLN			
Poniedziałek		Piątek		Poniedziałek		Piątek	
Wartości własne macierzy	Moduły wartości własnych	Wartości własne macierzy	Moduły wartości własnych	Wartości własne macierzy	Moduły wartości własnych	Wartości własne macierzy	Moduły wartości własnych
1	1	1	1	1	1	1	1
0,278+0,109i	0,299	-0,216+0,034i	0,219	-0,395	0,395	0,279	0,279
0,278-0,109i	0,299	-0,216-0,034i	0,219	0,462	0,462	0,135	0,135
-0,259	0,259	-0,063+0,149i	0,162	-0,123+0,247i	0,276	-0,091+0,075i	0,118
-0,016+0,212i	0,213	-0,063-0,149i	0,162	-0,123-0,247i	0,276	-0,091-0,075i	0,118
-0,016-0,212i	0,213	0,236+0,1i	0,257	0,197+0,028i	0,199	-0,025+0,135i	0,137
0,011+0,041i	0,042	0,236-0,1i	0,257	0,197-0,028i	0,199	-0,025-0,135i	0,137
0,011-0,041i	0,042	0,274	0,274	0,068	0,068	0,019+0,066i	0,069
0,082	0,082	0,065	0,065	-0,020	0,020	0,019-0,066i	0,069

Źródło: Obliczenia własne

Rozważane macierze opisujące zmiany kursów walutowych są nierozkładalne i niecykliczne, zatem regularne. Wnioskujemy o tym na podstawie ich wartości własnych oraz ich krotności. Wszystkie wartości własne tych macierzy są pojedyncze.

Analiza wyników zawartych w tablicy 8 pokazuje, że rozważana macierz stochastyczna jest regularna. Wskazuje na to wartość własna $\lambda_1 = 1$, która w przypadku każdej z tych macierzy jest pojedynczym pierwiastkiem równania charakterystycznego oraz jedynym pierwiastkiem równania charakterystycznego o module 1.

Tablica 8. Charakterystyki algebraiczne macierzy prawdopodobieństw przejścia dla indeksu giełdowego

WIG 20			
Poniedziałek		Piątek	
Wartości własne macierzy	Moduły wartości własnych	Wartości własne macierzy	Moduły wartości własnych
1	1	1	1
0,367	0,367	-0,201+0,107i	0,228
-0,1+0,25i	0,269	-0,201-0,107i	0,228
-0,1-0,25i	0,269	-0,1+0,079i	0,127
-0,190	0,190	-0,1-0,079i	0,127
0,07+0,078i	0,104	0,318	0,318
0,07-0,078i	0,104	0,107+0,078i	0,133
0,000	0,000	0,107-0,078i	0,133
0,000	0,000	0,139	0,139

Źródło: Obliczenia własne

Macierze prawdopodobieństw przejścia P rozważanych instrumentów polskiego rynku finansowego dla pozostałych dni tygodnia charakteryzują się podobnymi własnościami algebraicznymi.

Skoro macierz P skonstruowana dla poszczególnych instrumentów finansowych jest regularna to na mocy twierdzenia Dooba, jej kolejne potęgi zmierzają do macierzy ergodycznej. Wiersz macierzy ergodycznej wyznaczono na podstawie relacji (16). Przykładowy wektor stacjonarny dla kursu USD/PLN został zaprezentowany w tablicy 5. Z własności ergodyczności opisywanych łańcuchów Markowa wynika, że prawdopodobieństwa wystąpienia określonego stanu zmienności w długim czasie są takie same.

Tempo zbieżności kolejnych potęg danej macierzy stochastycznej do jej postaci ergodycznej zależy od wymiaru macierzy oraz liczby i rozmieszczenia jej niezerowych elementów. Jedną z miar tempa zbieżności jest współczynnik ergodyczności zdefiniowany poprzez związek (17). Na podstawie tablic 6 – 8 wyznaczono współczynniki ergodyczności:

Tablica 9. Współczynniki ergodyczności $\tau_1(P)$ macierzy stochastycznych P

	WIBOR 1M		WIBOR 3M		USD/PLN		EUR/PLN		WIG 20	
	Pon.	Pt.	Pon.	Pt.	Pon.	Pt.	Pon.	Pt.	Pon.	Pt.
$\tau_1(P)$	0,629	0,638	0,612	0,524	0,299	0,274	0,462	0,279	0,367	0,228

Źródło: Obliczenia własne

Analizując wyniki przedstawione w tablicy 9 można zauważyć, że postać ergodyczną najszybciej osiąga piątkowa macierz stochastyczna zmienności kursu USD/PLN.

Dla przykładowego łańcucha Markowa drugiego rzędu opisującego zmienność kursu USD/PLN, którego macierz prawdopodobieństw przejścia przedstawiono w tablicy 4 określono również moment osiągnięcia postaci ergodycznej przy zadanej dokładności 10^{-89}:

$$\left|P^{16} - E\right| = 1{,}68869\text{E-}90 \leq 10^{-89}$$

zatem macierz P^n osiąga postać ergodyczną dla n = 16.

Aby uzyskać pełną charakterystykę procesu zbadano również własność redukowalności macierzy P. Rozważane macierze okazały się nieprzywiedlne, nie istnieją bowiem permutacje tych macierzy sprowadzające je do postaci (10).

Na mocy tablicy 3 własności regularności oraz nieprzywiedlności macierzy P implikują następujące charakterystyki rozważanych łańcuchów Markowa drugiego rzędu:

- ergodyczność
- rozkład graniczny π jest niezależny od rozkładu początkowego d_0 i jest jedynym dodatnim rozkładem stacjonarnym.

Potwierdzają to również wyniki badań empirycznych zawarte w pracach [7, 8].

Podsumowanie

Algebraiczne podejście do badania macierzy stochastycznych umożliwia ujawnienie niektórych probabilistycznych własności łańcucha Markowa, takich jak: ergodyczność i stacjonarność. Własności te są szeroko wykorzystywane przy budowie stochastycznych modeli ekonometrycznych wielu zjawisk występujących w świecie finansów. Zagadnienia związane z ergodycznością oraz stacjonarnością zostały uwzględnione przez autorów we wcześniejszych pracach [7, 8]. Sprawdzenie założeń stosowalności teorii ergodycznych łańcuchów Markowa dla instrumentów polskiego rynku finansowego potwierdziło zasadność wykorzystania modeli rozkładu stanów zrównoważonych do badania anomalii kalendarzowych.

Literatura

1. Iosifescu M.: Skończone procesy Markowa i ich zastosowania. PWN, Warszawa 1988
2. Kot S.M.: Matematyczne modele procesów informacyjnych w systemie nauki. ZN AE w Krakowie, seria specjalna: Monografie nr 64, Kraków 1984
3. Kowalenko I.N.,Kuzniecow N.J., Szurienkow W.M.: Procesy stochastyczne. PWN, Warszawa 1989
4. Lee T.C., Judge G.G., Zellner A.: Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data. North Holland, Amsterdam 1997

5. Nykowski I. (red.): Prognozowanie struktury za pomocą łańcuchów Markowa. Prace i Materiały Instytutu Cybernetyki i Zarządzania SGPiS, Warszawa 1980
6. Podgórska M., Śliwka P., Topolewski M., Wrzosek M.: Łańcuchy Markowa w teorii i w zastosowaniach. Oficyna Wydawnicza SGH, Warszawa 2000
7. Skrodzka W., Włodarczyk A.: Analiza zachowań stóp procentowych WIBOR i kursów wymiany złotego z zastosowaniem łańcuchów Markowa drugiego rzędu. Referat złożony na VI Konferencję Naukową: Inwestycje finansowe i ubezpieczenia - tendencje światowe a rynek polski. Szklarska Poręba 2003
8. Skrodzka W., Włodarczyk A.: Wykorzystanie łańcuchów Markowa do badania zmienności indeksu WIG20, II Ogólnopolska Konferencja Naukowa: Rynek Kapitałowy. Skuteczne Inwestowanie. Wydawnictwo Uniwersytetu Szczecińskiego, Szczecin 2002
9. Wentzell A. D.: Wykłady z teorii procesów stochastycznych. PWN, Warszawa 1980

Wioletta Maria Skrodzka
Instytut Informatyki i Ekonometrii, Zakład Matematyki Ekonomicznej
Aneta Włodarczyk
Katedra Ekonometrii i Statystyki
Politechnika Częstochowska
Wydział Zarządzania

INFORMATYCZNE PODSTAWY PROCESÓW TESTOWANIA GRUP SPOŁECZNYCH

WIOLETTA SOŁTYSIAK

Streszczenie: W artykule są przedstawione przykładowe grupy społeczne które są poddane analizie statystycznej. Pokazane są narzędzia programowe za pomocą których możemy w sposób czytelny i rzetelny zaprezentować wyniki naszych rozważań pokazanych w postaci odpowiednich działań matematycznych czy statycznych oraz prezentacja wyników za pomocą wykresów i tabel w programie *Microsoft Excel i Statistica*.

Wstęp

Testować możemy wszystkie grupy społeczne począwszy od dzieci tych najmłodszych od pierwszego roku życia a skończywszy na dojrzałej jednostce społeczeństwa czy grupie. Ważnym elementem procesów testowania jest odpowiedni dobór i liczebność badanej próbki czy reprezentacji a także zmienna czyli cecha która może przybierać różne wartości. Grupy społeczne dzielimy najczęściej ze względu wiek, płeć, poziom wykształcenia, miejsce zamieszkania ale również ze względu na status zawodowy, można badać menedżerów i nauczycieli jak spędzają wolny czas, jakie mają zainteresowania, priorytety życiowe, bezrobotnych ile czasu poświęcają na poszukiwanie pracy (Derbis, 2000), badać i wnioskować możemy rzeczy, zjawiska społeczne, cechy osobowościowe, tematyka wyboru przedmiotu badań jest niewyczerpana.

W zależności od tego jakie uzyskamy odpowiedzi na kolejne pytania dotyczące analizowanego zagadnienia możemy zastosować różne metody analizowania wyników i formy ich prezentacji. Każda aplikacja arkusza kalkulacyjnego jak *Microsoft Excel, Gnumeric* czy program *Statistica* oferuje szereg narzędzi do analizy danych jak analizy wariancji, korelacje, regresje czy testy t.

Wybór aplikacji

Podczas planowania i testowania grup społecznych a następnie wykonywania analizy statystycznej za pomocą komputera należy wziąć pod uwagę szereg zaleceń. Większość popularnych programów komputerowych pracuje w środowisku graficznym jak *Excel, Statistica* a ich obsługa polega głównie na wybieraniu za pomocą myszki odpowiednich poleceń z menu czy przycisków pokazanych na ekranie, jest to bardzo pobieżny opis tychże aplikacji ale z programem *Microsoft Excel* zetknął się prawie każdy kto korzysta z *Windowsów*.

Program *Statistica* przeznaczony jest dla użytkowników zaawansowanych zajmujących się analizami statystycznymi, choć do końca nie należy się przy tym stwierdzeniu upierać ponieważ, sposób w jaki jest zbudowany program - posiada budowę modułową niezwykle ułatwiającą użytkownikowi wybór odpowiedniego narzędzia a następnie pracę z nim, jest również bardzo czytelny HELP wyjaśniający i prowadzący niemal krok po kroku nowego użytkownika po zawiłościach i pomagający interpretować otrzymane wyniki, są podane również przykłady zastosowań z wnioskowaniem statystycznym.

Można zastosować również inny sposób polegający na sterowaniu programem np. *SPSS/PC* pracujący pod *DOSem* za pomocą ciągu poleceń w języku zrozumiałym dla programu, zapisanych w pliku w pamięci stałej. Taki plik, zwany skryptem, może być łatwo poprawiany czy dostosowywany do różnych potrzeb.

W tym celu należy przygotować skrypt wykonujący analizę jeszcze przed wprowadzeniem danych, mając jedynie projekt struktury zbioru danych. Przy odrobinie doświadczenia możemy ten skrypt zbudować nawet nie mając dostępu do programu, który będzie go wykonywał. Jeżeli nie czujemy się dość pewnie, możemy kolejne polecenia skryptu wprowadzać sukcesywnie, obserwując wyniki ich działania, a dopiero w pełni „przećwiczone" polecenia umieszczać w ostatecznej wersji skryptu.

Dodatkowym argumentem jest oddzielenie koncepcyjnej pracy nad algorytmem analizy od zmagania się z interfejsem użytkownika, który niekoniecznie musi odpowiadać naszym upodobaniom. Do tworzenia skryptu używamy jedynie dowolnego, wybranego przez siebie edytora tekstu oraz dokumentacji opisującej składnię i działanie poszczególnych instrukcji programu.

Kryteria wyboru programu

W procesie wyboru programu dla opracowania programu statystycznego należy zwrócić uwagę na następujące kryteria. Zasadniczym kryterium jest niezawodność oprogramowania. Błąd w programie skutkujący jego zawieszeniem jest uciążliwy i może prowadzić do utraty danych. Błędy skutkujące nieprawidłowymi wynikami obliczeń, znacznie trudniejsze do wykrycia, zdarzają się, na szczęście, stosunkowo rzadko.

Drugie kryterium to zasób procedur obliczeniowych, które program oferuje. Oczywiste jest, że musi on wykonać wszystkie zaplanowane przez nas obliczenia. Nie należy jednak sztywno zakładać, że wszystkie zaplanowane prace trzeba wykonać za pomocą jednego programu. Często wygodniej jest posłużyć się kilkoma wyspecjalizowanymi programami do wykonania kolejnych etapów analizy i np. wprowadzić dane za pomocą jednego programu np. Microsoft Excel, za pomocą drugiego wykonać obliczenia Statistica, a przy użyciu trzeciego sporządzić służące do prezentacji wyników wykresy i tabele. Musimy wówczas zadbać o to, aby było możliwe sprawne przenoszenie danych pomiędzy używanymi programami.

Trzecie kryterium to możliwość przyjmowania przez program poleceń z zewnętrznego pliku (skryptu) zawierającego ciąg instrukcji nakazujących wykonanie operacji na danych. Jak wspomniano, wykonywanie analizy metodą interakcyjną, bez rejestrowania wykonywanych operacji może mieć zastosowanie co najwyżej dla zapoznania się z możliwościami programu i ewentualnie prostych, pojedynczych obliczeń.

Czwarte kryterium to dostępność programu. Renomowane programy do analiz statystycznych są zwykle zbyt drogie dla indywidualnego użytkownika. Dobrze, jeżeli mamy możliwość skorzystania z takiego programu na uczelni czy w pracy.

Warto także zapoznać się z możliwościami arkuszy kalkulacyjnych, w szczególności dostępnych za darmo, takich jak *StarCalc* czy *StarOffice*, (obecnie rozwijanego jako *Open Office), WingZ Pro, GNumeric* i wiele innych. Programy te umożliwiają wykonanie prostych analiz z użyciem bogatego zbioru statystyk.

Możliwości zastosowań wybranych programów

Podstawowe statystyki występujące niemal w każdej aplikacji oferującej obliczania statystyczne pozwalają obliczać wartości średnie, odchylenia standardowe, wariancje, współczynniki asymetrii oraz kurtozę zmiennych. Wartość średnia jest miarą przeciętnej lub tendencji centralnej rozkładu ciągłego. Odchylenie standardowe i wariancja są miarami rozproszenia (zmienności) danych. Współczynnik asymetrii jest miarą asymetrii rozkładu to znaczy stopnia skoncentrowania rozkładu po lewej lub prawej stronie wartości średniej. Kurtoza jest miarą spłaszczenia rozkładu to znaczy mówi w jakim stopniu rozkład jest wysmukły w okolicach wartości średniej. Podstawowe Statystyki pozwalają na obliczanie tablic rozkładu liczebności oraz do wizualizacji rozkładu w postaci histogramu. Po obliczeniu statystyk opisowych można tworzyć wykresy rozkładów zmiennych.

Za pomocą wykresów statystycznych możemy w przystępny sposób graficznie przedstawić nasze wyniki , mogą to być: histogramy, wykresy kołowe, oraz liczne wykresy specjalistyczne (np. twarze Chernoffa w programie *Statistica*) które reprezentują najstaranniej dopracowany typ wykresu obrazkowego. Przypadki są przedstawiane w sposób graficzny przy pomocy schematu twarzy, w taki sposób, że wartości względne zmiennych wybranych do wykresu są reprezentowane przez warianty poszczególnych cech twarzy).

W opracowaniu i analizie wyników badań pamiętać należy o konieczności wyważenia potrzeby zakresu stosowania metod statystycznych. Nie należy bezkrytycznie dokonywać różnorodnych obliczeń statystycznych. Powinniśmy mieć pewność co do zasadności i racjonalności postępowania oraz umiejętności interpretowania rezultatów zastosowanej metody które powinny pomóc badaczowi w analizie i opisie badanych zjawisk czy cech osobowościowych.

Na przykład porównując studentów płci żeńskiej i męskiej na jednym roku pod względem uzyskanych ocen, wówczas Płeć mogłaby być określona jako zmienna niezależna, średnia ocen jako zmienna zależna.

Bardzo często zmienną niezależną nazywamy czynnikiem. Eksperymenty badające jednocześnie działanie dwóch zmiennych niezależnych nazywa się eksperymentami dwuczynnikowymi albo o klasyfikacji dwuzmiennowej a gdy pojawią się trzy jest to wtedy klasyfikacja trójzmiennowa czy trójczynnikowy eksperyment.

Można badać zmienne i wzajemne relacje między nimi, tak jak istnieją one w naturze jak wzrost, iloraz inteligencji, oceny z egzaminów, ciśnienie krwi, poziom białych czy czerwonych krwinek u obu płci czy występujące choroby u osób w wieku podeszłym. Takie badania bywają czasem nazywane korelacyjnymi. Niektóre z ważniejszych zmiennych istniejących w naturze są zmiennymi nominalnymi, na przykład kolor włosów. Za pomocą zmiennej nominalnej możemy twierdzić, że jeden element lub cecha jest pod względem interesującej nas właściwości taki sam lub inny niż drugi element.

W tym momencie należy wspomnieć o pozostałych zmiennych jak porządkowych, przedziałowych i stosunkowych. Jest to ważna i budząca zainteresowanie statystyków klasyfikacja zmiennych opierająca się na różnicach między typami informacji jakich dostarczają operacje klasyfikowania lub pomiaru. Procedury analizowania zmiennych przedziałowych i stosunkowych tworzą najważniejszą klasę metod statystycznych.

Zmienna przedziałowa pozwala na formułowanie twierdzeń o równości i różności oraz „< niż " i „ > niż " , jak również twierdzeń o równości przedziałów. Zmienna przedziałowa nie ma punktu zerowego ustalamy go arbitralnie. Przykładem tej zmiennej z arbitralnie ustalonym punktem zerowym jest czas kalendarzowy. Natomiast zmienna stosunkowa, różni się tym od poprzednio omówionej, że pozwala na formułowanie takich samych twierdzeń jak zmienna przedziałowa plus twierdzeń o równości stosunków. Tzn., że jedna wartość zmiennej jest dwa lub więcej dłuższa, liczniejsza czy cięższa niż druga. W tym przypadku istnieje zero absolutne tzn. liczby odzwierciedlają odległość od pierwotnego początku. Najczęściej zmienne tego typu spotykamy w naukach fizycznych.

Natomiast zmienna porządkowa jest właściwością określoną przez operację która pozwala na uszeregowanie elementów grupy. Sprawdzają się tu twierdzenia o równości bądź różności elementów oraz „< niż" i „> niż". Przykładem może być tutaj uporządkowanie grupy np. pod względem podatności na infekcje czy wrażliwości jako cechy osobowościowej.

Przygotowując plan eksperymentu, należy:

- wybrać wartość zmiennej bądź zmiennych niezależnych, które mają być porównywane,
- właściwie dobrać osoby badane do eksperymentu ,
- zastosować reguły bądź procedury, za których pomocą osobom badanym zostają przyporządkowane konkretne wartości zmiennej niezależnej,

- określa, jakie obserwacje lub pomiary mają zostać dokonane u poszczególnych osób badanych
- należy podjąć decyzje dotyczące przeprowadzenia odpowiednich pomiarów wyników, które umożliwią dokonanie należytych porównań zastosowanych metod.

W życiu codziennym bardzo często spotykamy się ze zjawiskiem pobierania próbki może to być próbka kosmetyku, którą otrzymujemy w sklepie przy okazji robienia zakupów czy rolnik wyciągający z worka garść ziarna i sprawdzający jego jakość. Umiejętność pobierania próby jest bardzo istotną rzeczą w całym przedsięwzięciu badawczym. Niekiedy z uwagi na duże rozmiary wielu populacji może być rzeczą niemożliwą do zrealizowania. W takiej sytuacji badacz pobiera próbę czyli podgrupę czy podzespół, wybrany z populacji z pomocą odpowiedniej metody. Po pobraniu próby badacz opisuje jej właściwości po czym wyciąga wnioski. Wnioski o właściwościach populacji mogą zawierać błędy. Wielkość błędu można oczywiście oszacować za pomocą odpowiednich procedur. Tam, gdzie nie da się w żaden sposób oszacować błędu, uogólnienia wniosków wyciągniętych z danych z próby na populacje jest bezwartościowe.

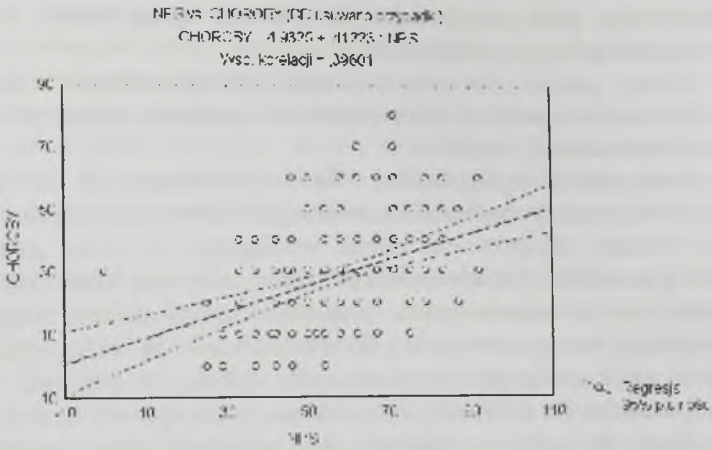
Procedury statystyczne stosowane do opisu próby albo dane dotyczące całej populacji określa się mianem statystyki opisowej. Mierzony iloraz inteligencji na dowolnej uczelni wszystkich jej studentów oraz gdy zmierzylibyśmy tylko u 100 studentów tej uczelni i obliczyli średni iloraz inteligencji tej próby to statystki te nazywać będziemy opisowymi, gdyż opisują właściwość tej próby.

Gdy dokonujemy pomiaru jakiejkolwiek wielkości to jesteśmy ograniczeni naszą ludzką niedoskonałością. Nie jesteśmy w stanie określić dokładnej danej wielkości, musimy zadowolić się jej miarą przybliżoną. Błąd próby jest to różnica między wartością w populacji czyli parametrem, a konkretną wartością z próby. Liczbową charakterystykę populacji, taką jest średnia lub odchylenie standardowe, nazywa się parametrem.

W celu opisania zmienności dowolnego zbioru pomiarów lub wartości posługujemy się zazwyczaj wariancją lub odchyleniem standardowym. Statystyki te opisują wielkość błędu próby.

Wyniki doświadczeń naukowych często ilustrowane są za pomocą wykresów punktowych. Jeżeli z wykresów tych wynika, że porównywane wielkości są wzajemnie powiązane.

Na wykresach 1 i 2 wykonanych za pomocą programu *Statistica* a dane surowe wprowadzone do programu *Microsoft Excel* widzimy sytuacje gdzie wielkości są bliskie rozkładowi, im punkty na wykresie są bliższe rozkładowi tym wzajemna zależność między odkładanymi względem siebie wielkościami jest mocniejsza. Wykres tych wielkości nie zawsze musi przechodzić przez początek układu współrzędnych. Jeżeli punkty na wykresie można przybliżyć przez jakąś regularną krzywą, to jest to argument, że odkładane wielkości na wykresie reprezentują w praktyce pewien rodzaj współzależności.



Wykres 1. Korelacja dodatnia

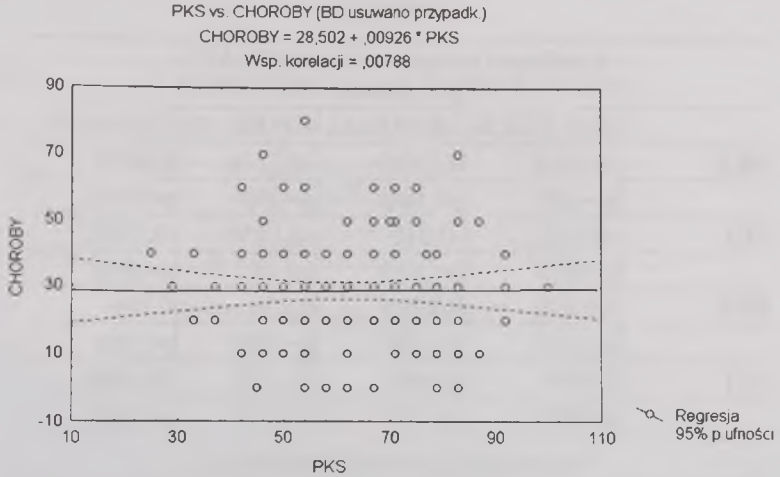
Nachylenie linii (ciągła prosta) wskazuje, w jakim stopniu Y w tym przypadku zmienna zależna Choroby kobiet w wieku podeszłym wzrasta wraz ze wzrostem X czyli zmiennej niezależnej Postawy w tym przypadku mamy Postawa Nie Pozytywna ku Sobie.

Mamy tu do czynienia z korelacją dodatnią, zmienne są zależne, ponieważ współczynnik korelacji liniowej Pearsona r wynosi 0,39601 i poziom istotności p=0,000 Wykres 1.

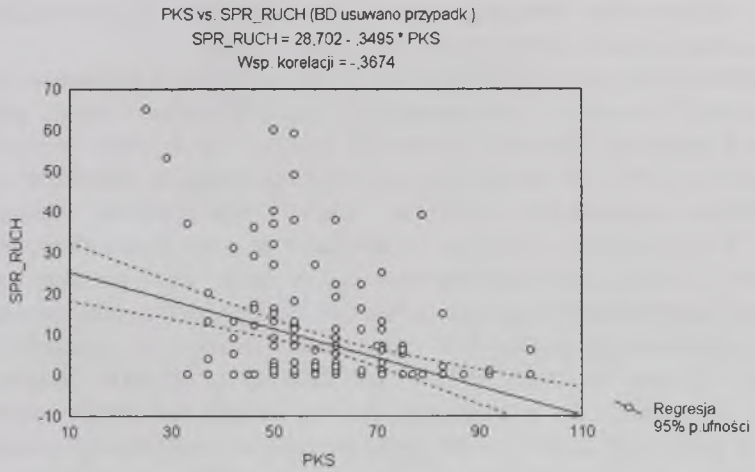
współczynnik korelacji liniowej Pearsona przyjmuje wartości z przedziału obustronnie domkniętego $r \in [-1;1]$. Wartości skrajne -1 oraz 1 uzyskuje w przypadku pełnej liniowej korelacji tzn., że wartości jednej zmiennej są funkcja liniową wartości drugiej zmiennej

-1 wtedy, gdy zachodzi korelacja ujemna natomiast 1 korelacja dodatnia. Im słabsza jest korelacja tym wartość bezwzględna wartości współczynnika jest bliższa zeru wykres 2.

gdzie r = 0,00788.



Wykres 2. Brak zależności korelacyjnych



Wykres 3. Korelacja ujemna

Wartość ujemna r = - 0,3674 wskazuje związek ujemny, czyli sytuację w której X maleje a Y wzrasta, Wykres 3.
W tym przypadku zmienna zależna Sprawność Ruchowa Kobiet w wieku podeszłym liczba badanej próbki N=172 kobiety wzrasta a Postawa Pozytywna ku Sobie maleje.
Zależności te są istotne statystycznie i wzajemnie na siebie oddziałują.

Tablica 1

Korelacje (kobiety z miasta) N=172
p < ,05000

Tabela 1.

	SPR RUCH	DEPRESJA	UPOŚL FI	CHOROBY
PKS	-0,3674	-0,2648	-0,2146	0,0079
	p=,000	p=,000	p=,005	p=,918
PKI	-0,2122	-0,215	-0,1456	-0,1782
	p=,005	p=,005	p=,057	p=,019
NPS	0,1375	0,3715	0,276	0,396
	p=,072	p=,000	p=,000	p=,000
NPI	0,0084	0,0197	0,1139	-0,0568
	p=,913	p=,797	p=,137	p=,459

Przykładowe zależności korelacyjne

W tabeli wyniki statystycznie istotne zostały pogrubione, choć wartość współczynnika poniżej 0,20 jest słaba, a o takiej zależności mówimy, że jest prawie nieznacząca. W powyższym przykładzie poziom wskaźnika korelacji mieści się w granicach zależności małej lecz istotnej. Zmienne korelują ze sobą i przyjmują odpowiednio korelacje ujemne lub dodatnie. W powyższej tabeli wystąpił parametr poziom istotności *p*, jaką pełni on rolę?

Istotność statystyczna wyniku jest estymowaną miarą stopnia, w jakim jest on "prawdziwy" (w sensie "reprezentatywny" dla populacji). Wartość poziomu p reprezentuje malejący wskaźnik rzetelności wyniku. Im wyższy poziom *p*, tym mniej możemy wierzyć, że zaobserwowana zależność między zmiennymi w próbie jest rzetelnym wskaźnikiem zależności między odpowiednimi zmiennymi w populacji. W szczególności poziom *p* reprezentuje prawdopodobieństwo błędu, związanego z przyjęciem zaobserwowanego wyniku jako prawomocnego, to znaczy jako "reprezentatywnego dla populacji". Na przykład poziom *p* równy 0,05 (tzn.1/20) wskazuje, że istnieje 5% prawdopodobieństwo, że zależność między zmiennymi wykryta w naszej próbie jest dziełem przypadku. Innymi słowy, gdybyśmy, zakładając, że w populacji nie ma żadnej zależności między tymi zmiennymi, powtarzali eksperymenty jeden po drugim, moglibyśmy oczekiwać, że w przybliżeniu co 20 powtórzeń eksperymentu trafiłby się jeden, w którym zależność między naszymi zmiennymi byłaby równa lub silniejsza od tej uprzednio przez nas zaobserwowanej. W wielu obszarach badawczych poziom p równy 0,05 jest traktowany jako graniczny akceptowalny poziom błędu.

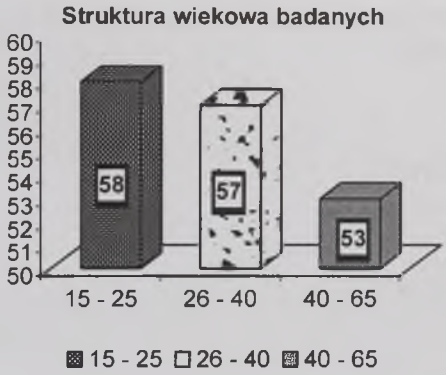
Badacz zwykle arbitralnie przyjmuje jakiś poziom istotności, powszechnie przyjmuje się je na poziomie 0,05 i 0,01.

Przykład analizy statystycznej pokazującej zależność między płcią a korzystaniem z promocji, użyję takich miar wielkości jak Średnia arytmetyczna, Moda, Mediana, Kwantyl I i Kwantyl III, Odchylenie standardowe. Badana próbka liczy 168 osób losowo wybranej społeczności obu płci w przedziale wiekowym 15 – 65 lat tabela i rysunek oraz obliczenia przy użyciu programu Microsoft Excel.

Tablica 2.

Wiek	Liczba badanych	Wskaźnik struktury
x_i	n_i	%
15 - 25	58	34,53
26 - 40	57	33,93
40 - 65	53	31,55
Razem	168	100 %

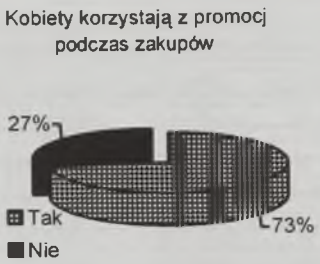
Struktura wiekowa wszystkich badanych



Tablica 3.

Korzystanie z promocji	Kobiety		Mężczyźni		Razem		Wskaźnik podobieństwa struktur
	Liczba badanych	Wskaźnik struktury	Liczba badanych	Wskaźnik struktury	Liczba badanych	Wskaźnik struktury	
x_i	n_i	%	n_i	%	n_i	%	%
Tak	60	73,17	50	58,13	110	65,47	58,13
Nie	22	26,82	36	41,86	58	34,52	26,82
Razem	82	100 %	86	100 %	168	100 %	84,95 %

Zależność między płcią a korzystaniem z promocji



Z wykresu możemy odczytać, iż ankietowane kobiety przy robieniu zakupów korzystają z promocji.
Policzymy teraz parametry statystyczne, które posłużą nam do syntetycznego opisu struktury zbiorowości kobiet.
Obliczenia
Średnia arytmetyczna

- $\overline{X} = \frac{60+22}{2} = 41$

Mediana (kwantyl drugi)

- $M_e = \frac{60+22}{2} = 41$

Odchylenia standardowe

- $S = \sqrt{\frac{4084}{2} - (41)^2} = 19$

Współczynnik zmienności

- $V = \frac{19}{41} * 100\% = 46{,}34$

Tablica 4.

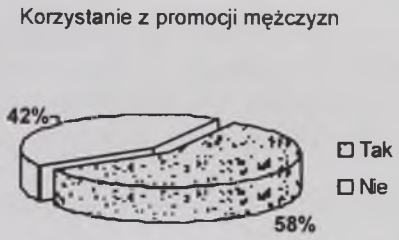
Zestawienie	Kobiety
Średnia arytmetyczna	41
Mediana	41
Odchylenia standardowe	19
Współczynnik zmienności	46,34

Tablica 5.

Odpowiedź	Mężczyźni Liczba osób
Tak	50
Nie	36
Razem	86

Korzystanie z promocji przez mężczyzn

Z wykresu możemy odczytać, iż ankietowani mężczyźni przy robieniu zakupów



również korzystają z promocji choć w mniejszym stopniu bo jest to 58% w stosunku do 75% kobiet .

Obliczenia

Średnia arytmetyczna

- $\overline{X} = \frac{50+36}{2} = 43$

Mediana (kwantyl drugi)

- $M_e = \frac{50+36}{2} = 43$

Współczynnik zmienności

- $V = \frac{7}{43} * 100\% = 16{,}27$

Tablica 6.

Zestawienie	Mężczyźni
Średnia arytmetyczna	43
Mediana	45
Odchylenia standardowe	7
Współczynnik zmienności	16,27

Odchylenia standardowe

➢ $$S = \sqrt{\frac{3796}{2} - (43)^2} = 7$$

Tablica 7.

Zestawienie	Kobiety	Mężczyźni
Średnia arytmetyczna	41	43
Mediana	41	43
Odchylenia standardowe	19	7
Współczynnik zmienności	46,36	16,27

Mediana dzieli zbiorowość na dwie równe części, połowa jednostek ma wartość cechy mniejsze lub równe medianie, a połowa – wartości równe bądź większe od mediany, stąd też mediana bywa nazywana wartością środkową.
Współczynnik zmienności jest ilorazem bezwzględnej miary zmienności cechy i średniej wartości tej cechy.
Odchylenie standardowe „S" określa przeciętne zróżnicowanie poszczególnych wartości cechy od średniej arytmetycznej.
Z powyższej analizy wynika, że:

- ✓ Średnia arytmetyczna wieku dla obu płci wyniosła 34,83 lat.
- ✓ Najwięcej ankietowanych osób mieści się w przedziale wiekowym od 15 do 25 lat.
- ✓ Mediana wyniosła w przypadku kobiet 41 z czego wynika, że połowa badanych ma nie więcej niż 41 lata, a druga połowa więcej niż 41 lat, natomiast dla mężczyzn Kwanyl drugi ma wartość 43.
- ✓ Odchylenie standardowe wyniosło odpowiednio 19 dla kobiet i 7 dla mężczyzn.
- ✓ Współczynnik zmienności dla odchylenia standardowego wyniósł 46,36 dla kobiet oraz 16,27 dla mężczyzn. W tym przypadku współczynnik zmienności wykazuje zróżnicowanie statystyczne istotne.

Wnioski

Jak należy rozumieć różnego rodzaju wykresy, tabele, a co najważniejsze, jak obiektywne mogą być te informacje?

W czasach rozwoju informacyjnego, bardzo ważną umiejętnością jest zdolność do oceny jakości informacji, a także zdolność ich właściwej interpretacji. Z pewnością wiedza z zakresu statystyki może znacznie ułatwić nabycie takich umiejętności.

W oparciu o odpowiedzi jakich udzielamy na kolejne pytania dotyczące natury rozwiązywanego problemu można zastosować różne metody analizowania danych jak i formy prezentacji wyników. Najbardziej atrakcyjnym sposobem

przedstawiania danych statystycznych jest ich prezentacja w postaci diagramów, postępuje się tak dlatego, że rysunkowa prezentacja analiz jest przez ludzi przyjmowana życzliwiej niż całe szeregi faktów czy kolumn liczbowych, ponieważ jest czytelna i przejrzysta, patrząc na diagram czy jest on słupkowy, kołowy czy obrazkowy zawsze widać która wartość zwyżkuje a która nie oraz jak całość danych dzieli się na poszczególne części i w jakiej są one wzajemnej relacji. Podobnie ma się z wykresami np. wykres liniowy jest doskonałym sposobem na wykreślenie tendencji wzrostowych i spadkowych dla zbiorów danych. Jest on niestety nadużywany w środkach masowego przekazu w reklamach, czasopismach czy ogłoszeniach, mogą wtedy wprowadzać w błąd. Dlaczego? Bardzo łatwo jest zmienić wygląd wykresu przez dobór skali na osiach. I dlatego też, wykresy powinny posiadać cechę rzetelności. Podsumowując, przystępując do pracy z wybraną aplikacją należy zwrócić uwagę na następujące punkty: środowisko pracy i wymagania sprzętowe oraz systemu operacyjnego, uruchamianie programu, podawanie poleceń i danych do programu, postać wynikowa tabelaryczna czy wykres, zasób statystyk i innych operacji, które posiada program i potrafi je wykonać w sposób zadawalający użytkownika oraz prezentacja wyników czy to w formie tekstowej za pomocą edytora tekstowego z rodziny Windows czy dla systemu *Linux* programy do grafiki wektorowej jak *XFig, Tgif, Sketch* czy też forma prezentacji multimedialnej za Microsoft PowerPointa gdzie wykresy zrobione mogą być w *Excelu* lub *Statistice*. Użycie programu komputerowego umożliwia w sposób szybki i skuteczny, ale także wiarygodny i czytelny przedstawienie wyników badań. Zastosowanie jednego z wymienionych programów to nie tylko oszczędność czasu, ale możliwość wyboru odpowiedniej metody obliczeń statystycznych, matematycznych, finansowych czy bazodanowych jak to ma miejsce np. w *Excelu*. Program *Statistica* to szeroki wachlarz narzędzi do obliczeń tak dla statystyka jak i dla zwykłego użytkownika komputera, który potrzebuje bardziej zaawansowanych narzędzi czy wyrafinowanych sposobów prezentacji danych i zademonstrowanie efektów swojej pracy. Odpowiedni wybór oprogramowania to pewność, że aplikacja dostarczy nam takich narzędzi które odzwierciedlą w sposób rzetelny wyniki badań czy analiz, ważnym kryterium jest również możliwość przyjmowania przez program poleceń z zewnętrznego pliku oraz ostatnie kryterium doboru oprogramowania to dostępność oprogramowania.

Literatura

1. Ferguson G.A., Takone Y., 1999, Statistical Analysis in Psychology and Education, PWN Warszawa
2. Krajewska A., Stachura A., Zimny M.T.,2002, Zastosowanie metod statystycznych w badaniach naukowych, APNiK Menos S.C. Częstochowa
3. Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski W., 1995, Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, PWN Warszawa

4. Derbis R., 2000, Doświadczanie codzienności Poczucie jakości życia swoboda działania Odpowiedzialność Wartości osób bezrobotnych, WSP Częstochowa
5. Łój G., 2003, Badania własne, WSP Częstochowa

Mgr inż. Wioletta Sołtysiak
Wyższa Szkoła Pedagogiczna w Częstochowie
e-mail: wsoltysiak@wsp.czest.pl

STATYSTYCZNE METODY BADANIA CEN KRUSZCÓW

Janusz SZOPA, Agnieszka ULFIK, Stefan NOWAK

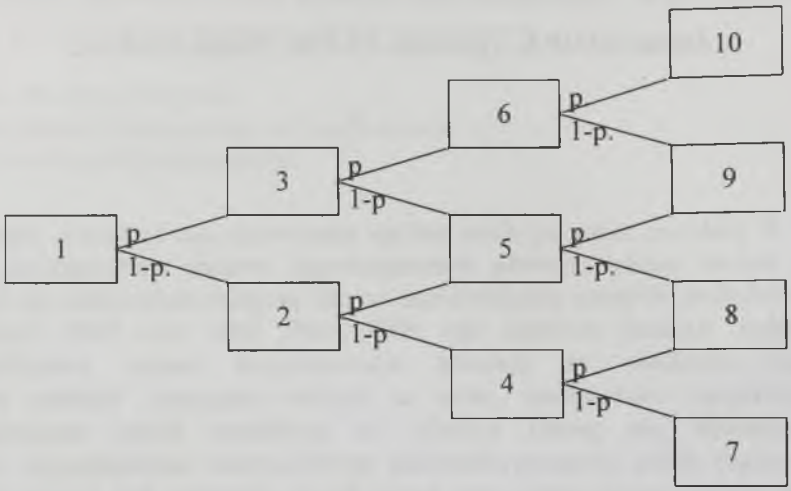
Wstęp

W praktyce znane są różne metody szacowania cen kruszców. Stosuje się między innymi model drzewka dwumianowego, metodę martyngałową, wzory Blacka-Scholesa. W pracy przedstawiono wyniki prognoz zachowania się cen złota oraz srebra. Analizie poddano lata 1999, 2000, 2001 oraz 2002. Otrzymane prognozy uzyskane za pomocą wymienionych metod zweryfikowano z rzeczywistymi wartościami, jakie te kruszce osiągnęły. Metoda drzewek dwumianowych jest prostą metodą, na podstawie której otrzymuje się wystarczająco dobre prognozy. Bardziej wyrafinowane matematycznie metody, takie jak metoda martyngałowa oraz wzory Blacka-Scholesa dają wyniki zbliżone do wyników uzyskanych metodą drzewek dwumianowych.

1. Metoda dwumianowa.

Zmiany wartości cen złota oraz srebra można prześledzić konstruując model drzewa dwumianowego, które przedstawia poziomy cenowe, jakie może osiągnąć dany rodzaj kruszcu. Przyjmuje się założenie, że w każdym okresie czasu Δt wartość, może wzrosnąć lub spaść, czyli przyjąć jedną z możliwych dwóch cen. Przy tym skok wartości w górę bądź w dół ma się odbywać o tę samą wartość. Badany przedział czasu $<0,t>$ jest wielokrotnością okresu Δt, tzn. $t = n \cdot \Delta t$ ($n = 1, 2,..., m$). Graficznie drzewko dwumianowe dla n=3 przedstawia rys.1.

Na rysunku przedstawiono dziesięć pozycji (wartości, węzłów) {1}- {10}, jakie w czasie wędrówki po drzewku może przyjąć badana wielkość. Założono, że ruch do góry lub w dół odbywa się odpowiednio z prawdopodobieństwem p i 1-p oraz że pozycją startową jest {1}. Pozycje końcowe {7}, {8}, {9} lub {10} można osiągnąć na osiem możliwych sposobów: {1,2,4,7}, {1,2,4,8}, {1,2,5,8}, {1,3,5,8}, {1,2,5,9}, {1,3,5,9}, {1,3,6,9}, {1,3,6,10}.



Rys. 1. Drzewko dwumianowe dla trzech okresów czasowych Δt.

Potrzebne przy wyznaczaniu zmian ceny instrumentu finansowego (skoku jego wartości po upływie okresu czasu Δt), wielkości (stopę zwrotu itp.) oblicza się według następujących wzorów:

1. Stopę zwrotu -x_i - według wzoru [1-4]:

$$x_i = \frac{S_i - S_{i-1}}{S_{i-1}} \tag{1}$$

gdzie: S_i - cena kruszcu w chwili i (np. dniu, minucie)
S_{i-1} - cena kruszcu w chwili poprzedniej.

W praktyce i zmienia wartość od drugiej do ostatniej (N) wartości.

2. Zmienność (chwiejność) ceny badanego kruszcu według wzoru:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{N} \tag{2}$$

Zmienność ta zawiera w sobie historię cen i jest wyznaczana od zadanej chwili czasu w przeszłości aż do chwili obecnej, do której prowadzi się obliczenia.

Jeżeli przedział czasu przeszłego ma długość T, to szacuje się zmienność dla przedziału czasu <0,t> jako wartość równą $\sigma\sqrt{\frac{t}{T}}$. Należy zwrócić uwagę, że czasy t i T muszą być wyrażone w tych samych jednostkach.

Jeżeli przez S oznaczymy cenę kruszcu w chwili t, to po upływie czasu Δt, $S \cdot u$ oznaczać będzie wzrost wartości, gdy nastąpi skok w górę, zaś $S \cdot d$ spadek wartości, gdy nastąpi skok w dół. Wielkości u, d oraz prawdopodobieństwo p obliczono według następujących wzorów:

$$u = e^{\sigma \cdot \sqrt{\Delta t}} \tag{3}$$

$$d = \frac{1}{u} \tag{4}$$

$$p = \frac{e^{r\Delta t} - d}{u - d} \tag{5}$$

gdzie: r - stopa procentowa wolna od ryzyka odnoszącą się do jednostek czasu w jakich rozważa się t i T, ponadto założono, że oprocentowanie jest ciągłe.

2. Zastosowanie metody dwumianowej do badania zmian wartości kruszców.

2.1. Wycena cen złota i srebra

Zaprezentowaną metodę drzewka dwumianowego wykorzystano do prognozy zmian wartości cen złota oraz srebra na koniec pierwszego kwartału 2003 roku na podstawie czterech przedziałów czasowych o różnej długości obejmujących wszystkie dane:

- z roku 2002,
- z lat 2001 i 2002,
- z lat 2000, 2001 i 2002,
- z lat 1999, 2000, 2001 i 2002.

W poniższej tablicy przedstawiono ilość notowań London Gold Fix oraz London Silver Fix w każdym wymienionym okresie. Notowania te są wyrażone w dolarach amerykańskich za uncję (1 uncja to 31,1 grama). Za r (czyli wolną od ryzyka stopę procentową w skali roku tej waluty) przyjęto 2%. Stopy zwrotu wyznaczono według zależności (1). We wzorze (2) przyjęto n = N, we wzorze (3), jako okres czasu, po którym następuje skok wartości akcji, przyjęto $\Delta t = 1/12$ roku.

Lata	Ilość danych	
	Złoto	Srebro
1999	251	252
2000	252	253
2001	251	253
2002	249	251

Tablica 1. Ilości danych dla złota i srebra

Wykorzystując wzory (3) – (5) otrzymano pozostałe dane do skonstruowania drzewek dwumianowych dla prognozowania cen srebra i złota.

W tablicach 2 i 3 przedstawiono wartości parametrów σ, u, d, p danych wzorami (2) – (5) dla różnych przedziałów czasowych. Wartości te wyznaczone zostały z wykorzystaniem wyznaczania stopy zwrotu według wzoru (1).

Tablica 2. Wartości parametrów danych wzorami (2) – (5) dla srebra

Przedział czasowy	σ	u	d	p
2002	0,187372	1,055579	0,947347	0,501893
2001 – 2002	0,186868	1,055426	0,947485	0,501971
2000 – 2002	0,280448	1,084326	0,922232	0,490062
1999 – 2002	0,374132	1,114051	0,897625	0,480733

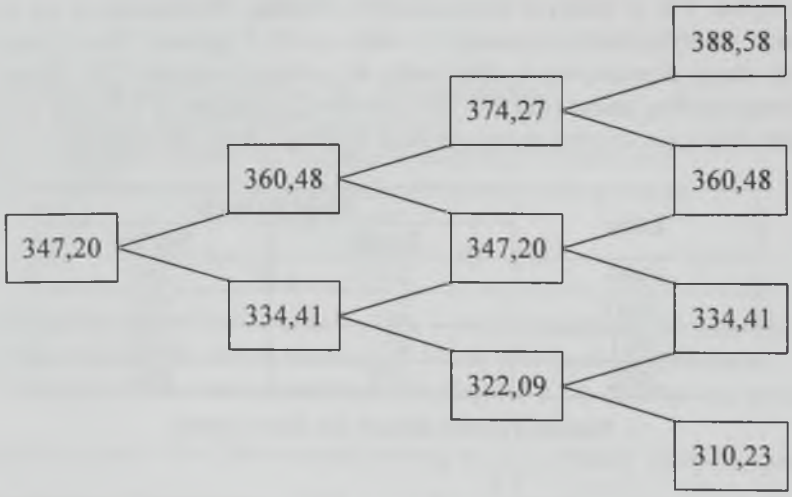
Źródło: obliczenia własne

Tablica 3. Wartości parametrów danych wzorami (2) – (5) dla złota

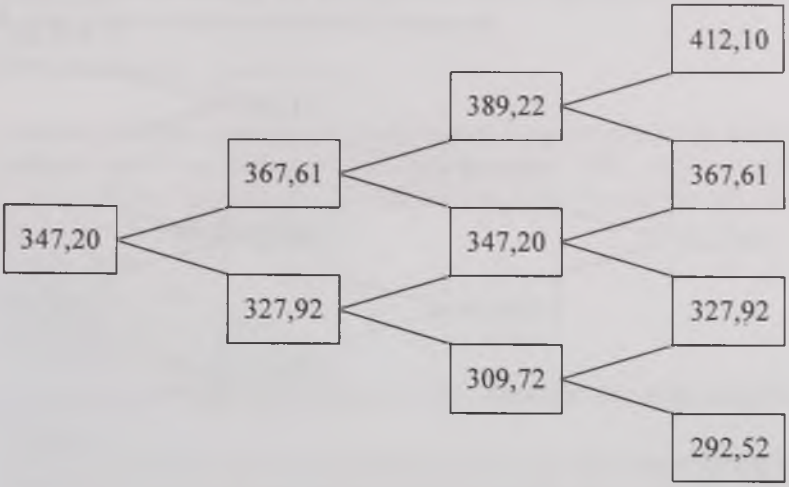
Przedział czasowy	σ	u	d	p
2002	0,130018	1,038246	0,963163	0,512834
2001 – 2002	0,197877	1,058785	0,944479	0,500316
2000 – 2002	0,245771	1,073525	0,93151	0,494016
1999 – 2002	0,292383	1,088068	0,91906	0,488781

Źródło: obliczenia własne

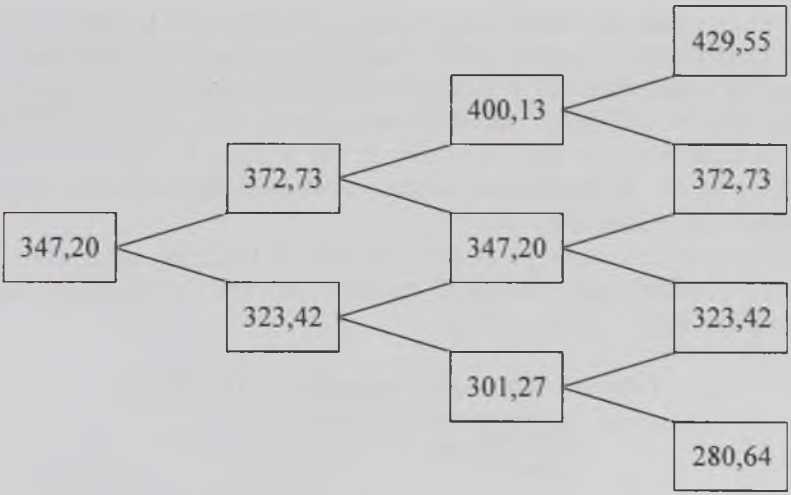
Poniżej przedstawiono modele drzew dwumianowych dla stopy zwrotu liczonej według wzoru (1) dla złota biorąc pod uwagę wymienione wcześniej cztery okresy czasowe.



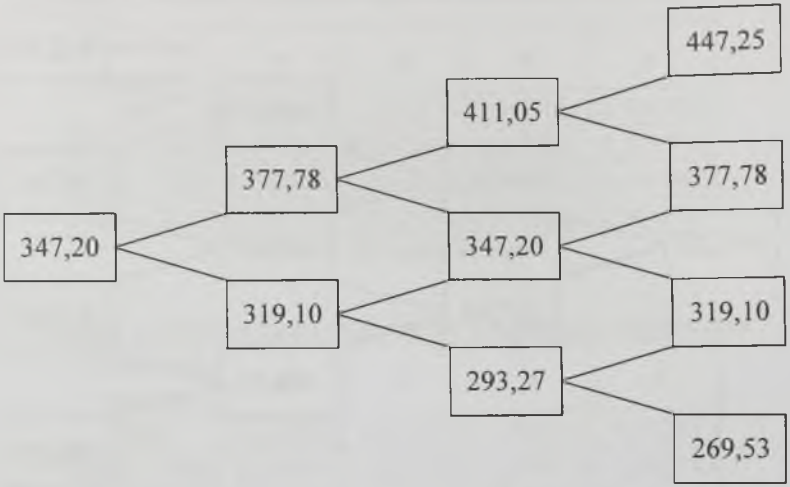
Rys. 2. Drzewko dwumianowe dla cen złota wyznaczone na podstawie danych z roku 2002.



Rys. 3. Drzewko dwumianowe dla cen złota wyznaczone na podstawie danych z lat 2001 – 2002.



Rys. 4. Drzewko dwumianowe dla cen złota wyznaczone na podstawie danych z lat 2000 – 2002.



Rys. 5. Drzewko dwumianowe dla cen złota wyznaczone na podstawie danych z lat 1999 – 2002.

Na powyższych schematach widać jak teoretycznie kształtowała się cena złota w pierwszym kwartale 2003 roku wyznaczona na podstawie okresów czasowych o różnej długości. W obliczeniach wykorzystano w pozycji {1} wartość ostatniego notowania fixingu popołudniowego ceny złota z końca 2002 roku jako wartość początkową oraz dane z tablicy 3. Na pozycjach {7} do {10} (według rys.1) znajdują się prognozowane wartości badanych instrumentów finansowych po okresie trzech miesięcy ($3 \cdot \Delta t$).

Wartości cen srebra i złota wyrażone są USD za uncję (31.1 grama). W tablicy 5 podano ceny badanych kruszców w dniu 31.03.2003r. (na koniec trzeciego kwartału).

Tablica 5. Notowania giełdowe z dnia 31.03.2003r.

Nazwa	Cena
Złoto	334,85
Srebro	4,46

Przeprowadzono także analizę drzewek dwumianowych dla fixingu cen srebra.

Biorąc pod uwagę wartości z tablicy 5, widać iż cena złota przy prognozie opierającej się tylko na roku 2002 mieści się dokładnie w pozycji nr 8, a przy prognozie opierającej się na dłuższej historii trafia w obszar znajdujący się między pozycjami nr 8 i 9. Przy wydłużaniu historii zauważyć możemy zwiększenie

przedziału, w jakim zawierają się pozycje od 7 do 10. Ponieważ przedział ten się zwiększa, występuje utrata precyzyjności prognozy.

2.2. Wycena opcji

Analizie poddano europejską opcję kupna i wykorzystano ją do prognozy zmian wartości opcji na koniec pierwszego kwartału 2003 roku na podstawie czterech przedziałów czasowych o różnej długości obejmujących wszystkie notowania fixingu złota oraz srebra:

- z roku 2002,
- z lat 2001 i 2002,
- z lat 2000, 2001 i 2002,
- z lat 1999, 2000, 2001 i 2002.

Za cenę wykonania przyjęto wartość pozycji {1} drzewka dwumianowego z końca grudnia 2002.

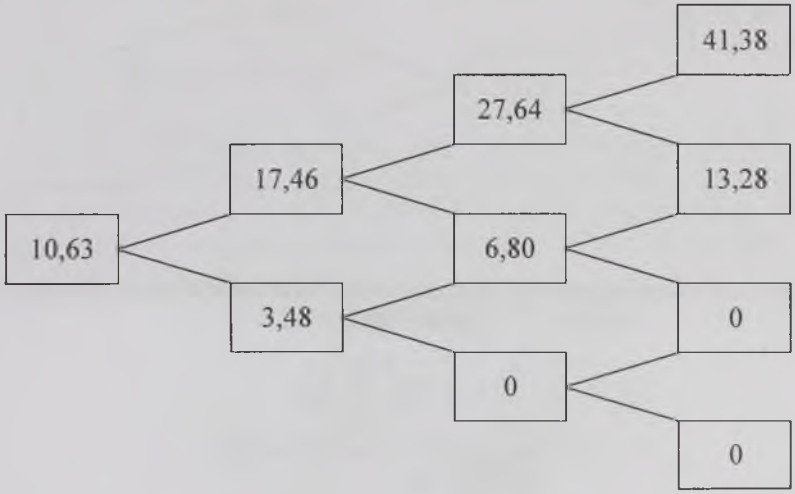
Cenę wykonania takiej opcji odejmuje się od wartości na pozycjach {7} do {10}. Jeśli różnica jest ujemna przyjmuje się wartość zero. Cofając się poprzez pozycje {4} – {2}, {2} – {3} oblicza się jej cenę opcji dla pozycji {1} stosując za każdym razem wzór [3, 4]:

$$C=(p\cdot C_u+(1-p)\cdot C_d)\cdot e^{-r\cdot\Delta t} \qquad (6)$$

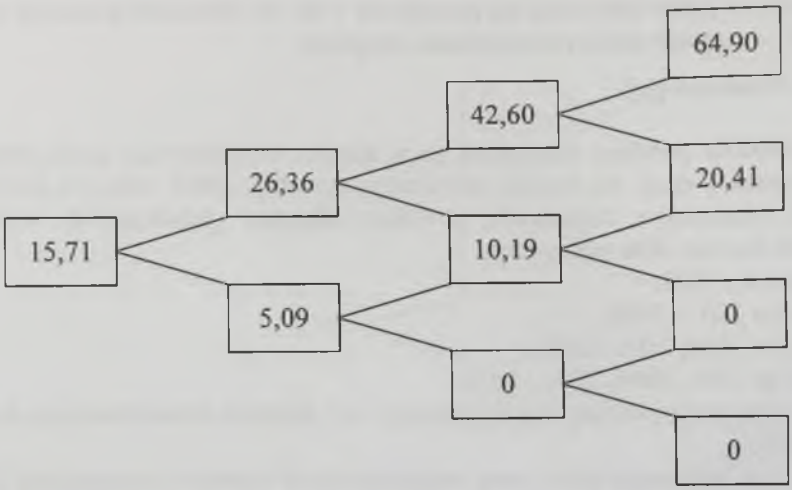
gdzie: C_u – wartość opcji z górnej pozycji,

C_d – wartość opcji z dolnej. pozycji.

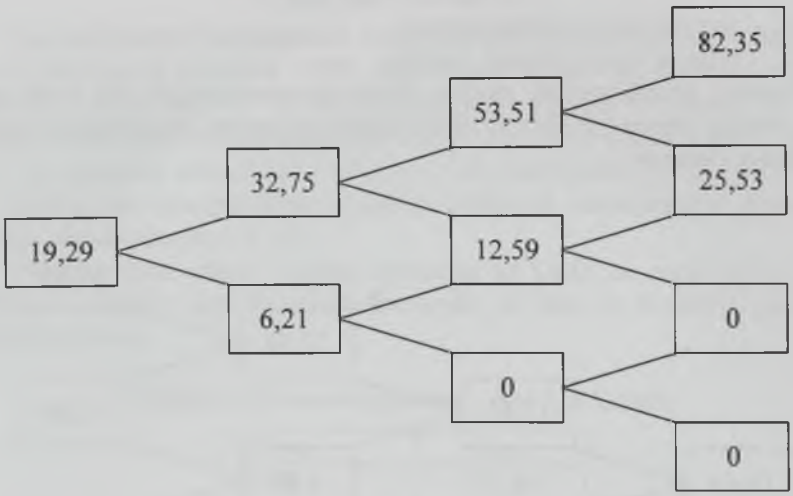
Poniżej przedstawiono modele drzew dwumianowych dla stopy zwrotu liczonej według wzoru (1) dla cen złota biorąc pod uwagę wymienione wcześniej cztery okresy czasowe.



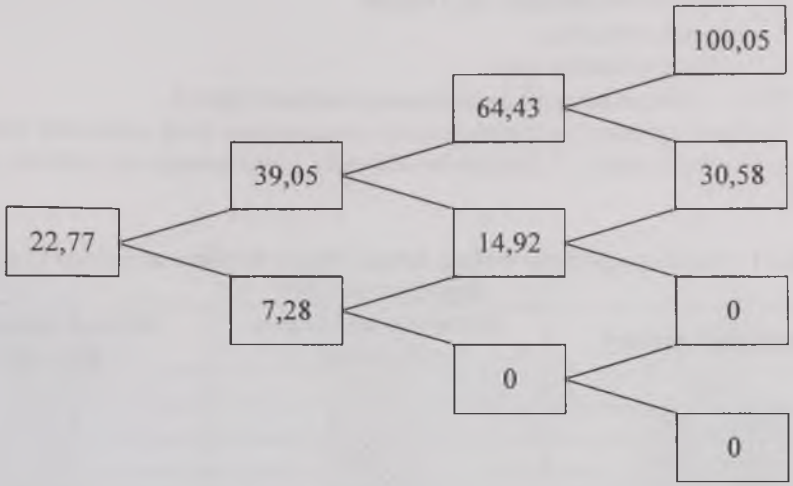
Rys. 6. Drzewko dwumianowe opcji dla ceny złota wyznaczone na podstawie danych z roku 2002.



Rys. 7. Drzewko dwumianowe opcji dla ceny złota wyznaczone na podstawie danych z lat 2001 – 2002.



Rys. 8. Drzewko dwumianowe opcji dla ceny złota wyznaczone na podstawie danych z lat 2000 – 2002.



Rys. 9. Drzewko dwumianowe opcji dla ceny złota wyznaczone na podstawie danych z lat 1999-2002.

Na powyższych schematach widać jak teoretycznie kształtowała się cena opcji złota w pierwszym kwartale 2003 roku wyznaczona na podstawie okresów czasowych o różnej długości. W obliczeniach wykorzystano w pozycji {1} wartość ostatniego notowania giełdowego z końca 2002 roku jako wartość początkową oraz dane z tabeli 3.

Zaobserwowano (wykresy dotyczące srebra nie jest zamieszczona w tej publikacji) iż oszacowanie wartości opcji zwiększało się przy uwzględnieniu dłuższej historii.

2.3. Metoda Blacka- Scholesa

Wartości opcji kupna oszacowano również stosując metodę Blacka–Scholesa [3, 4]. W modelu tym zakłada się, że zmiany cen instrumentu podstawowego są ciągłe, w przeciwieństwie do metody dwumianowej. Pozostałe teoretyczne założenia przyjęto za autorami [3] strona 171 i [4] strona 183.

Według tej metody

$$C = S \cdot \Phi(d_1) - X \cdot e^{-r \cdot n \cdot \Delta t} \cdot \Phi(d_2) \tag{7}$$

$$d_1 = \frac{\ln\left(\frac{S}{X}\right) + \left(r + \frac{\sigma^2}{2}\right) \cdot n \cdot \Delta t}{\sigma \cdot \sqrt{n \cdot \Delta t}} \tag{8}$$

$$d_2 = d_1 - \sigma\sqrt{n \cdot \Delta t} \tag{9}$$

gdzie: C – wartość europejskiej opcji kupna
S – bieżąca cena akcji
X – cena wykonania opcji
$\Phi(d)$ - oznacza wartość dystrybuanty rozkładu N(0,1).

Przyjęto zgodnie z wcześniejszymi oznaczeniami $S=X$, natomiast pozostałe dane z cytowanych tablic i diagramów. Wyniki zastosowania tej metody podaje tablica 4.

Tablica 4. Wycena opcji kupna według metody Blacka-Scholesa wyznaczona na podstawie danych z roku 2001.

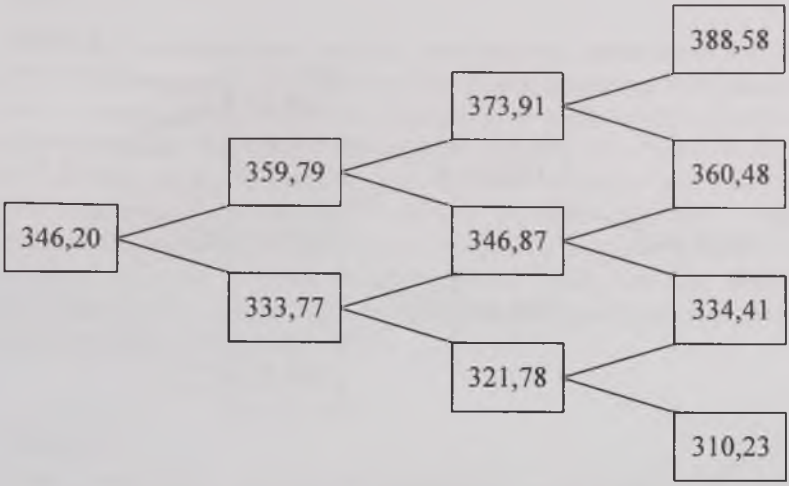
Przedział czasowy	wycena opcji kupna dla złota	wycena opcji kupna dla srebra
2002	9,8729	0,1857
2001 – 2002	14,5478	0,1853
2000 – 2002	17,8480	0,2719
1999 – 2002	21,0586	0,3585

Źródło: obliczenia własne

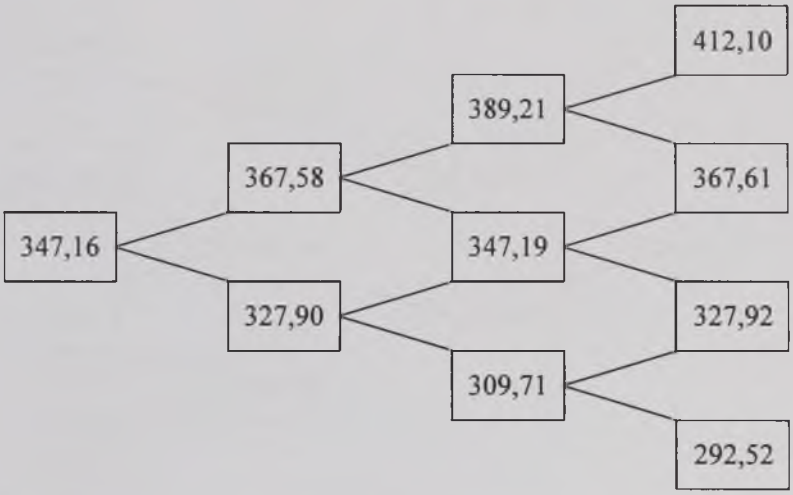
Porównując te wartości z wyceną wg prostej metody drzewa dwumianowego oraz przy zastosowaniu metody Blacka – Scholesa widać, że różnice są nieznaczne, co potwierdza sensowność stosowania metody drzewka dwumianowego do oceny zachowania się cen opcji kruszców,

2.4. Metoda martyngałowa

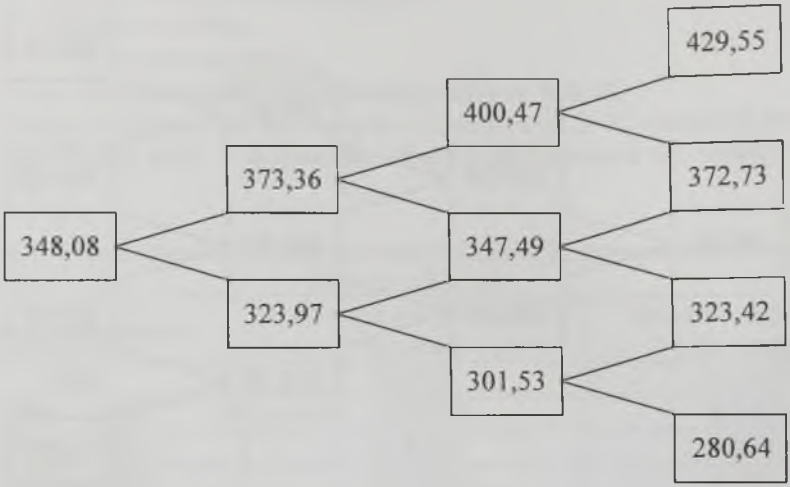
W [4] podano twierdzenie o reprezentacji martyngałowej i jego zastosowanie do teorii finansów. Założenia potrzebne do stosowania tej metody są takie same jak przy stosowaniu metody dwumianowej. Pozostałe założenia teoretyczne znajdują się w [4] strona 149. W szczególności można, korzystając z wartości instrumentu finansowego znajdujących się na końcu drzewek dwumianowych, dyskontować je do chwili początkowej. Biorąc pod uwagę wcześniejsze założenia dyskonto odbywać się będzie według oprocentowania ciągłego. Zatem szukany proces - $e^{-rt} \cdot S_t$, (martyngał) będzie wyznaczony względem r wynoszącego 2%. Biorąc jako pozycje 7 – 10 w drzewkach dwumianowych wartości wyznaczone w punkcie 1 (metodą dwumianową wg wzoru (1)) otrzymano następujące diagramy:



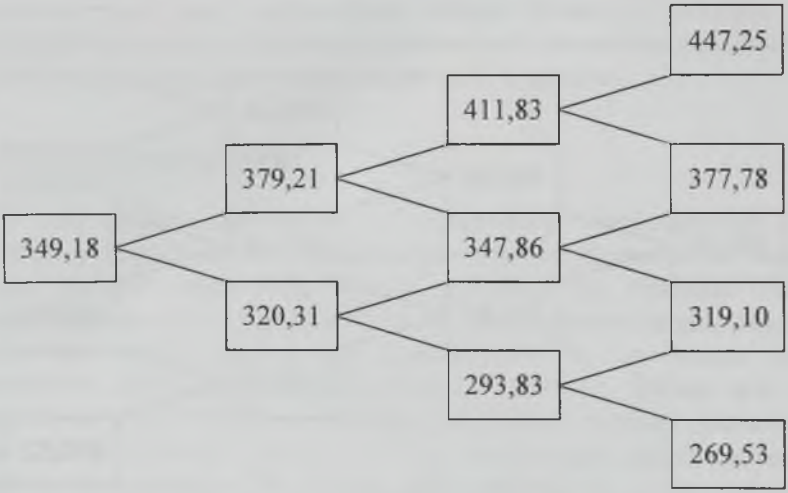
Rys. 10. Martyngał dla złota wyznaczony na podstawie danych z roku 2002.



Rys. 11. Martyngał dla złota wyznaczone na podstawie danych z lat 2001 – 2002.



Rys. 12. Martyngał dla złota wyznaczone na podstawie danych z lat 2000 – 2002.



Rys. 13. Martyngał dla złota wyznaczone na podstawie danych z lat 1999 – 2002.

Biorąc pod uwagę także obliczone wartości dla srebra widać, iż wartości prognozowane niewiele różnią się od wartości, jakie rzeczywiście były osiągnięte. Świadczy to o zbliżonym do „gry sprawiedliwej" zachowaniu fixingu cen srebra i złota.

Podsumowanie

W pracy przedstawiono metodę oszacowania zachowania się cen złota i srebra przy wykorzystaniu metody drzewka dwumianowego. Otrzymane wyniki porównano z metodą Blacka-Scholesa. Różnice przy zastosowaniu obu tych metod wynoszą kilka procent. Pokazuje to przydatność stosowania w praktyce dla badania cen złota i srebra prostej metody drzewka dwumianowego a jednocześnie przez porównanie końcowych rzeczywistych wartości cen tych kruszców – możliwości oszacowań europejskiej opcji kupna na te wielkości. Dla rozważanych wielkości skonstruowano również procesy martyngałowe. Prognozowane wartości cen kruszców były bliskie rzeczywistym wartościom występującym na Londyn Gold Fixing i Londyn Silver Fixingrów.

Literatura

1. Dziwago E. (1999). Procesy martyngałowe w wycenie europejskiej opcji kupna, Dynamiczne modele ekonometryczne. VI Ogólnopolskie Seminarium Naukowe, UMK Toruń
2. Szopa H., Szopa J. (2000). Impact of random conditions on projects in capital changes. The 3rd European Project Management Conference, Project Management Institute Jerusalem, Israel
3. Tarczyński W., Zwolankowski M. (1999). Inżynieria finansowa. Agencja Wydawnicza Placet, Warszawa
4. Weron A., Weron R. (1998). Inżynieria finansowa. Wydawnictwo Naukowo – Techniczne, Warszawa
5. http://www.kitco.com

Janusz Szopa
Instytut Ekonometrii i Informatyki
Wydział Zarządzania
Politechnika Częstochowska
Ul. Armii Krajowej 19b
42-200 Częstochowa
januszszopa@adm.pcz.czest.pl

AgnieszkaUlfik
Instytut Ekonometrii i Informatyki
Wydział Zarządzania
Politechnika Częstochowska
Ul. Armii Krajowej 19b
42-200 Częstochowa
ulfik@zim.pcz.czest.pl

Stefan Nowak
Instytut Ekonometrii i Informatyki
Wydział Zarządzania
Politechnika Częstochowska
Ul. Armii Krajowej 19b
42-200 Częstochowa
snowak@zim.pcz.czest.pl

TEORETYCZNE ASPEKTY WYKORZYSTANIA PRZEŁĄCZNIKOWYCH MODELI MARKOWA DO POMIARU RYZYKA WALUTOWEGO

Aneta WŁODARCZYK

Streszczenie: Modele przełącznikowe należą do klasy narzędzi odpowiednich do opisu dynamiki procesów, których charakterystyki podlegają skokowym zmianom w czasie, w związku z czym przy ich modelowaniu zachodzi potrzeba uwzględnienia odpowiednich skokowych zmian parametrów. Zmiana parametru modelu przełącznikowego następuje wraz ze zmianą reżimu determinującego siłę lub kierunek zależności między elementami rozważanego systemu. W licznych zastosowaniach modeli przełącznikowych procesem sterującym zmianami reżimu jest jednorodny łańcuch Markowa. Omawiany model można oszacować przy wykorzystaniu wersji algorytmu Expectations Maximization zaproponowanej przez Hamiltona. Analizując zachowanie kursów walutowych można zauważyć występowanie tzw. długich swingów, czyli okresów utrzymujących się tendencji aprecjacyjnych lub deprecjacyjnych kursu. W ten sposób zostaną wyodrębnione dwa reżimy: reżim odpowiadający deprecjacji kursu oraz reżim odpowiadający aprecjacji kursu. W artykule przedstawiono teoretyczne zagadnienia związane z budową i estymacją parametrów przełącznikowych modeli Markowa, które można wykorzystać do prognozowania zmian kursu walutowego, co stanowi podstawę pomiaru ryzyka walutowego.

Wstęp

Zmiany zachodzące na polskim rynku walutowym, zwłaszcza po roku 1995, związane z wprowadzeniem w życie nowego prawa dewizowego oraz płynnego kursu złotego sprzyjały większym i częstszym wahaniom kursów walut. Z tego też powodu w ostatnich latach w Polsce wzrasta rola zarządzania ryzykiem walutowym. Dyrektorzy finansowi coraz częściej uświadamiają sobie konieczność zabezpieczenia działalności finansowej przedsiębiorstwa przed niekorzystnymi zmianami kursu walutowego. W związku z tym, od dłuższego czasu można obserwować znaczne poszerzenie zakresu prowadzonych badań naukowych nad wykorzystaniem bogatego aparatu procesów stochastycznych do przewidywania zachowania się kursu wymiany w czasie. W opracowaniu tym autorka przedstawi propozycję zastosowania przełącznikowych modeli Markowa do opisania prawidłowości rządzących procesem zmienności kursu walutowego. W artykule zostaną przedstawione zagadnienia związane z budową i estymacją parametrów przełącznikowych modeli Markowa, które można wykorzystać do prognozowania zmian kursu walutowego, co stanowi podstawę pomiaru ryzyka walutowego.

Zmienność kursów wymiany w procesie pomiaru ryzyka walutowego

Jednym z najistotniejszych czynników wpływających na rozwój nauki o zarządzaniu ryzykiem, a jednocześnie będącym odpowiedzialnym za istotną dynamizację tego zjawiska w ostatnich latach, jest silny wzrost zmienności parametrów rynkowych. Nie można bowiem w sposób pewny ustalić jak ukształtują się w przyszłości poszczególne czynniki stanowiące podstawę teraźniejszych decyzji gospodarczych odnośnie zamierzeń inwestycyjnych w przedsiębiorstwie. Ryzyko rynkowe wynika z faktu, że dochód z inwestycji w instrument finansowy zależy od pewnych parametrów, które ustala rynek.[1] Z reguły tymi parametrami są różnego rodzaju ceny finansowe, a w szczególności: stopa procentowa, kurs walutowy, cena akcji, cena towaru itp. W zależności od tego, jaka cena finansowa wpływa na dochód z inwestycji, można wyróżnić różne rodzaje ryzyka rynkowego. Ryzyko kursów walut występuje wtedy, gdy instrument finansowy, w który się inwestuje, jest denominowany w innej walucie niż waluta kraju inwestora.[2] Wtedy zmiany kursu walutowego powodują, że stopy zwrotu wyrażone w dwóch różnych walutach nie są takie same. Zarządzanie ryzykiem finansowym jest procesem złożonym, skupiającym w sobie konieczność identyfikacji, pomiaru oraz zabezpieczenia się przed skutkami ponoszonego ryzyka. Jednakże najistotniejszym elementem systemu zarządzania ryzykiem jest precyzyjny pomiar tego zjawiska. Do oszacowania ryzyka walutowego według prostej czy bardziej zaawansowanej metody niezbędne są następujące informacje: prognoza ekspozycji walutowej netto, dane dotyczące przeprowadzonych transakcji zabezpieczających, bieżący kurs natychmiastowy i kurs forward oraz kurs prognozowany.[3]

W kontekście zarządzania ryzykiem walutowym zmienność stanowi miarę stopnia niepewności co do przyszłego kierunku zmian kursu walutowego[4]. Jeśli wzrasta zmienność, rośnie prawdopodobieństwo, że dany kurs wymiany znacznie zmieni swoją wartość w przyszłości. Prawidłowe oszacowanie parametru zmienności umożliwia zatem osiągnięcie większych dochodów, zmniejszenie ryzyka inwestycji czyli dostarcza ważnych informacji wykorzystywanych w procesie podejmowania decyzji inwestycyjnych. Według A. Dziewięckiego efektywność podejmowania decyzji na rynku walutowym jest uzależniona zarówno od znajomości czynników kształtujących kursy walut, jak i ciągłego monitorowania i analizowania sytuacji na rynku krajowym i międzynarodowym[5].

Kształtowanie się poziomu kursu walutowego wyjaśniają między innymi takie teorie ekonomiczne jak: teoria parytetu siły nabywczej (Purchasing Power Parity – PPP), prawo jednej ceny, teoria międzynarodowego efektu Fishera

---

[1] Ch. W. Smithson, C.W. Smith, Jr.D.S. Wilford: Zarządzanie ryzykiem finansowym. Dom Wydawniczy ABC, Kraków 2000

[2] Jajuga K. (red.): Metody ekonometryczne i statystyczne w analizie rynku kapitałowego. Wydawnictwo AE im. Oskara Langego, Wrocław 2000

[3] D. Bennett: Ryzyko walutowe. Dom Wydawniczy ABC, Kraków 2000

[4] Ch. W. Smithson, C.W. Smith, Jr.D.S. Wilford: Zarządzanie ryzykiem finansowym. Dom Wydawniczy ABC, Kraków 2000

[5] A. Dziewięcki: Inwestowanie na międzybankowym rynku walutowym. Rynek Terminowy 13, 3/2001

(International Fisher Effect - IFE), teoria parytetu stóp procentowych (Interest Rate Parity – IRP). Różny jest stopień zadowolenia z tych teorii[6]. Stosując kryterium czasu teorie kursu walutowego można podzielić na takie, które wyjaśniają zmiany w długim okresie (PPP), te które odnoszą się do okresu średniego (związane ze strefą celów) oraz zajmujące się kursem z punktu widzenia zmian krótkookresowych (teoria nadwrażliwości – overshoot)[7]. Zgodnie z powyższym podziałem poziom kursu wymiany w długim okresie jest determinowany przez wiele czynników gospodarczych, między innymi: tempo wzrostu PKB, tempo inflacji, interwencje banku centralnego, sytuację w bilansie płatniczym czy poziom rozwoju i strukturę gospodarki[8]. W krótkim czasie na kurs wymiany mają wpływ także czynniki psychologiczne i polityczne, które nagle mogą zwiększyć popyt na walutę, a tym samym wzmocnić jej kurs. Bardzo często uważa się, że zmiany kursu są efektem racjonalnych oczekiwań inwestorów. Oznacza to, że pojawienie się na rynku walutowym informacji o czynniku, który powinien wywołać wzrost kursu w przyszłości implikuje krótkookresową aprecjację kursu wymiany. Teoria racjonalnych oczekiwań oraz jej aplikacje w finansach została opisana przez M. Osińską[9]. Na kurs walutowy wywierają wpływ także czynniki nieprzewidywalne, takie jak na przykład zaburzenia na międzynarodowych rynkach walutowych, które mogą wywrzeć przejściowy, ale istotny wpływ na kurs wymiany. Należy także pamiętać, że inwestorzy krótkoterminowi w niewielkim stopniu kierują się w swych decyzjach oceną sytuacji w gospodarce i bilansie płatniczym danego kraju. Podejmują oni działania inwestycyjne na podstawie informacji i plotek pojawiających się na rynku walutowym, wykorzystując co najwyżej narzędzia analizy technicznej.

Omówione powyżej czynniki kształtujące poziom kursów walutowych, zarówno w długim, jak i krótkim okresie są uwzględniane przy konstrukcji różnych modeli ekonometrycznych, wykorzystywanych do prognozowania zmian kursów wymiany.

Przełącznikowy model Markowa

Modele przełącznikowe należą do klasy narzędzi odpowiednich do opisu dynamiki procesów, których charakterystyki podlegają skokowym zmianom w czasie, w związku z czym przy ich modelowaniu zachodzi potrzeba uwzględnienia odpowiednich skokowych zmian parametrów. Innymi słowy, w szeregu czasowym obserwacji takiego procesu zauważalne są okresy, w których wartości

---

[6] M. Dobija: Źródła wartości jednostki pieniądza. II Konferencja Naukowa: Rynek Kapitałowy. Skuteczne Inwestowanie. Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin 2002

[7] E. Chrabonszczewska, K. Kalicki: Teoria i polityka kursu walutowego. Oficyna Wydawnicza SGH, Warszawa 1996

[8] D. Bennett: Ryzyko walutowe. Dom Wydawniczy ABC, Kraków 2000

[9] M. Osińska: Ekonometryczne modelowanie oczekiwań gospodarczych. Wydawnictwo UMK, Toruń 2000

procesu generowane są przez różne reżimy (stany). Zmiana parametru modelu przełącznikowego następuje wraz ze zmianą reżimu determinującego siłę lub kierunek zależności między elementami rozważanego systemu. W modelach przełącznikowych przyjmuje się, że zarówno mechanizm sterujący zmianami w obrębie poszczególnych reżimów, jak i mechanizm zmiany reżimu jest losowy. Konsekwencją takich założeń jest niemożność jednoznacznego rozstrzygnięcia o obowiązującym reżimie jedynie na podstawie znajomości stanu procesu w ustalonej chwili. Zastosowania modeli przełącznikowych opierają się na ogólnym założeniu, że badane szeregi czasowe można modelować przy użyciu procesów stochastycznych zdefiniowanych jako ciągi zmiennych losowych o znanym typie rozkładu warunkowego w każdym reżimie. W licznych zastosowaniach modeli przełącznikowych procesem sterującym zmianami reżimu jest jednorodny łańcuch Markowa. Ta kategoria modeli przełącznikowych jest w literaturze określana mianem przełącznikowych modeli Markowa (*Markov switching model)*.[10]

Modele przełącznikowe pojawiły się w literaturze ekonometrycznej około 30 lat temu. Pierwsze wzmianki na temat modelu przełącznikowego można odnaleźć w pracy Goldfelda i Quandta (1973)[11], w której rozważano model regresji liniowej o współczynnikach zmieniających się wraz ze zmianą reżimu sterującego procesem. Zastosowania PMM są przedstawione w pracach dotyczących m. in. badania struktury czasowej stóp procentowych[12], analizy cykli gospodarczych[13] , zmian kursów walutowych [4, 5].

Przełącznikowy model Markowa zaproponowany przez Hamiltona (1989) jest ściśle związany z modelem TAR (*threshold autoregression, progowa autoregresja)*. Zasadnicza różnica między nimi polega na tym, że w PMM zmiany reżimu nie są determinowane poprzez poziom procesu (tzw. wartość progową), lecz poprzez nieobserwowalną zmienną stanu, która jest przeważnie modelowana jako łańcuch Markowa. Na przykład w [1] podano następujący opis tego modelu:

$$x_t = \begin{cases} \alpha_1 + \beta_1 x_{t-1} + \varepsilon_{1t} & \text{dla } s_t = 1 \\ \alpha_2 + \beta_2 x_{t-1} + \varepsilon_{2t} & \text{dla } s_t = 0 \end{cases} \qquad (1)$$

gdzie s_t jest nieobserwowalnym dwustanowym łańcuchem Markowa, z pewną macierzą prawdopodobieństw przejścia **P**. Dla obydwu reżimów, x_t jest procesem autoregresji rzędu pierwszego AR(1), jednak parametry tego procesu (zawierające wariancję składnika losowego) różnią się w poszczególnych reżimach, oraz zmiana w reżimie jest losowa oraz podlega autokorelacji.

---

[10] Hamilton J.D.: Time Series Analysis. Princeton University Press, Princeton, New Jersey 1994

[11] Goldfeld S.M., Quandt R.E.: A Markov Model for Switching Regressions. Journal of Econometrics 1/1973.

[12] Erlandsson U.G.: Regime Switches in Swedish Interest Rates. www.nek.lu.se/nekuer

[13] Goodwin T.H.: Business-Cycle Analysis with a Markov-Switching Model. Journal of Business & Economic Statistics 11/1993

Zmiany w reżimie są spowodowane przez inne czynniki niż czynniki występujące w szeregach, które aktualnie są modelowane (s_t określa reżim, a nie x_t), rzadko wiadomo w jakim reżimie jest proces (s_t jest nieobserwowalne), lecz po fakcie można często zidentyfikować w którym reżimie był proces, na pewnym stopniu ufności (s_t można oszacować, za pomocą procesu filtrującego Hamiltona [1989]). Ponadto, zmiany reżimu są identyfikowalne poprzez interakcje pomiędzy danymi a łańcuchem Markowa, a nie poprzez sprawdzenie *a priori* danych.
Trudności w pomiarze czynnika kształtującego w istotny sposób badane zjawisko uzasadniają traktowanie tej zmiennej jako nieobserwowalnej, oraz modelowanie jej jako łańcuch Markowa. Przełącznikowy model Markowa zaproponowany przez Engla, Hamiltona w [4] do badania zmienności kursu walutowego ma następującą postać:

$$\mathbf{y}_t = \mu_0(1 - s_t) + \mu_1 s_t + \varepsilon_t \quad , \qquad \varepsilon_t \sim N(\mathbf{0}, \mathbf{\Omega}_{s_t}) \tag{2}$$

Badane zjawisko przedstawiono jako wektorowy proces stochastyczny (y_1, y_2, ..., y_T). Zmienna nieobserwowalna s_t określa stan, w jakim znajduje się zmienna objaśniana w chwili t. Zmienna s_t jest łańcuchem Markowa o dwóch stanach i przyjmuje skalarne wartości {0, 1}. Pierwszy stan (reżim) określa sytuację, w której interesujące nas zdarzenie wystąpiło , drugi reżim oznacza pojawienie się zdarzenia przeciwnego. Gdy zmienna $s_t = 0$, zmienna objaśniana podlega rozkładowi normalnemu o parametrach μ_0, $\mathbf{\Omega}_0$; gdy $s_t = 1$ to zmienna objaśniana podlega rozkładowi normalnemu o parametrach μ_1, $\mathbf{\Omega}_1$. Innymi słowy proces stochastyczny generujący badane zjawisko jest mieszaniną dwóch rozkładów normalnych o różnych średnich.

Założono dla potrzeb tego modelu, że prawdopodobieństwa przejścia nie zmieniają się w czasie i wynoszą:

$$\begin{aligned} &P(s_t = 0/\, s_{t-1} = 0) = p_{00} \\ &P(s_t = 1/\, s_{t-1} = 0) = p_{01} = 1 - p_{00} \\ &P(s_t = 0/\, s_{t-1} = 1) = p_{10} = 1 - p_{11} \\ &P(s_t = 1/\, s_{t-1} = 1) = p_{11} \end{aligned} \tag{3}$$

Przyjęto zatem *a priori*, że rozważany proces s_t jest jednorodnym łańcuchem Markowa.

Przełącznikowy model Markowa ma kilka interesujących cech ze względu na cel przeprowadzanego badania. Ponieważ łańcuch Markowa określa zmiany stanu, proces może przełączać się gwałtownie, co naśladuje gwałtowne zmiany w poziomie zmiennej objaśniającej, wynikające na przykład ze zmian oczekiwań inwestorów na rynku walutowym.

Przełącznikowy model Markowa opisany równaniem (2) modeluje zmienną objaśnianą jako mieszaninę dwóch rozkładów normalnych, a to nie implikuje zależności, ze rozkład tej zmiennej jest normalny (Titterington i inni (1985)). Jest to zgodne z przeprowadzonymi do tej pory badaniami, które wykazały, że rozkład kursów walutowych oraz stóp zwrotu odbiega od rozkładu

normalnego.[14] Dodatkową zaletą wybranego modelu jest to, że należy on do klasy modeli nieliniowych.

Estymacja przełącznikowego modelu Markowa

Omawiany model można oszacować przy wykorzystaniu wersji algorytmu EM (Expectations Maximization) zaproponowanej przez Hamiltona w [6]. Jako wynik estymacji otrzymuje się wektor parametrów Θ, zawierający elementy następujących macierzy:

μ_j – (k×1) wektor średnich wartości zmiennej objaśnianej w stanie j, gdzie $j \in \{0,1\}$,

k – liczba składowych wektora $\mathbf{y}_t$

Ω - (k×k) macierz wariancji-kowariancji

p_{ij} – prawdopodobieństwo przejścia ze stanu i do stanu j

ρ– prawdopodobieństwo tego, że $\mathbf{y}_t$ był w stanie $s_t = 0$ w czasie t=1.

Dodatkowo, jako wynik estymacji otrzymuje się prawdopodobieństwo tego, że proces znajdował się w stanie s_t w chwili t:

$$P(s_t / \mathbf{y}_1, ..., \mathbf{y}_t; \Theta) \qquad (4)$$

Jeśli do oszacowania prawdopodobieństwa zostałaby wykorzystana informacja wykraczająca poza chwilę t (np. cała próba t=1, 2, .., T), to można otrzymać prawdopodobieństwo wygładzone:

$$P(s_t / \mathbf{y}_1, ..., \mathbf{y}_T; \Theta) \qquad (5)$$

Dysponując obserwacjami z okresów 1,2,3,..,T można zbudować funkcję wiarygodności dla rozważanego modelu przełącznikowego postaci:

$$L(y_1, y_2, ..., y_T, \Theta) = \prod_{t=1}^{T} (\sum_{j=1}^{r} P(S_t = j) \cdot P(Y_t = y_t / S_t = j)) \qquad (6)$$

Jednym ze sposobów maksymalizacji skonstruowanej dla potrzeb tego modelu funkcji wiarygodności jest zastosowanie odpowiedniej wersji algorytmu Expectations Maximization. Zastosowanie algorytmu EM do klasy modeli przełącznikowych wymaga wyprowadzenia zależności pomiędzy oszacowanymi na podstawie dostępnych informacji prawdopodobieństwami, że obserwacja $\mathbf{y}_t$ została wygenerowana prze reżim j

$$P(S_t = j / Y_t = y_t) = \frac{P(S_t = j) \cdot P(Y_t = y_t / S_t = j)}{\sum_{j=1}^{r} P(S_t = j) \cdot P(Y_t = y_t / S_t = j)} \qquad (7)$$

a warunkami narzuconymi na parametry poprzez układ równań wynikający z potrzeby maksymalizacji funkcji wiarygodności (6):

$$\frac{\partial L}{\partial p_{kl}}(\mathbf{y}_1, ..., \mathbf{y}_T, \Theta) = 0 \qquad \text{dla } k = 1, 2, ..., N \ , \ j = 1, 2, ..., r \ , \qquad (8)$$

---

[14] J. Szopa, A. Włodarczyk: Badanie statystycznych własności rozkładu stóp zwrotu kursu walutowego. II Ogólnopolska Konferencja Naukowa: Efektywność Zastosowań Systemów Informatycznych. WNT, Warszawa-Szczyrk 2002

$$\sum_{k=1}^{N} p_{kj} = 1 \qquad \text{dla } j = 1, 2, .., r.$$

Zależność ta pozwala na znalezienie w kolejnych iteracjach algorytmu coraz lepszych (w sensie kryterium związanego z funkcją wiarygodności) oszacowań wektora parametrów modelu. Po ustaleniu wartości początkowych dla wszystkich parametrów modelu, w każdej iteracji algorytmu EM wykonywane są dwa kroki. Pierwszy krok (*expectations)* polega na wyznaczeniu prawdopodobieństw $P(S_t = j / \mathbf{Y}_t = \mathbf{y}_t)$ zgodnie z wzorem (7). Krok drugi (*maximization)* prowadzi do wyznaczenia za pomocą układu równań (8) wektora parametrów maksymalizującego funkcję wiarygodności. Hamilton w [6] dowodzi, że estymator metody największej wiarygodności wektora Θ stanowi rozwiązanie następującego układu równań (przy założeniu: $\Omega_0 = \Omega_1 = \Omega$):

$$\mu_j = \frac{\sum_{t=1}^{T} \mathbf{y}_t p(s_t = j / \mathbf{y}_1,...,\mathbf{y}_T;\theta)}{\sum_{t=1}^{T} p(s_t = j / \mathbf{y}_1,...,\mathbf{y}_T;\theta)} \qquad j = 0,1 \tag{9}$$

$$\mathbf{\Omega} = \frac{\sum_{j=0}^{K}\sum_{t=1}^{T} (\mathbf{y}_t - \mu_j)(\mathbf{y}_t - \mu_j)^T p(s_t = j / \mathbf{y}_1,...,\mathbf{y}_T;\theta)}{T} \qquad K=1 \tag{10}$$

$$p_{ij} = \frac{\sum_{t=2}^{T} p(s_t = j, s_{t-1} = i / \mathbf{y}_1,...,\mathbf{y}_T;\theta)}{\sum_{t=2}^{T} p(s_{t-1} = i, / \mathbf{y}_1,...,\mathbf{y}_T;\theta)} \tag{11}$$

$$\rho = p(s_1 = 0 / \mathbf{y}_1,...,\mathbf{y}_T;\theta) \tag{12}$$

Procedurę rozpoczyna podanie dowolnych wartości początkowych wektora Θ, które służą do oszacowania wygładzonego prawdopodobieństwa (5). Z kolei oszacowane prawdopodobieństwo wykorzystuje się do estymacji parametrów wektora Θ opisanych równaniami (9) – (12). Estymacja ustaje w momencie spełnienia zadanego kryterium zbieżności. Aby zapobiec temu, że wynik estymacji zostanie zniekształcony poprzez niewielką liczbę bardzo wysokich obserwacji, wartości przekraczające próg trzech średnich absolutnych odchyleń od średniej z próby zostały zastąpione przez wysokość tego progu. Hamilton w [6] udowodnił,

że ciąg otrzymanych w ten sposób oszacowań jest zbieżny do lokalnego maksimum funkcji wiarygodności.

Aplikacja przełącznikowego modelu Markowa do prognozowania zmienności kursu wymiany

Analizując zachowanie kursów walutowych można zauważyć występowanie tzw. długich swingów (*long swings)*, czyli okresów utrzymujących się tendencji aprecjacyjnych lub deprecjacyjnych kursu. Zjawisko to wyjaśniane jest między innymi napływem informacji na rynek walutowy oraz ich subiektywną interpretacją przez poszczególnych uczestników rynku, którzy również kształtują kursy walutowe. Napływ informacji odbywa się w sposób nieregularny, w szczególnym przypadku seryjnie (co oznacza, ze mogą być one ze sobą skorelowane), a ich znaczenie i siła oddziaływania na zmianę odpowiednich kursów walutowych są zróżnicowane. W efekcie na rynku występują podokresy zmniejszonej i zwiększonej zmienności kursów. Zmienność kursów walutowych zależy również od oczekiwanych, nowych informacji (np. komunikatów o sytuacji makroekonomicznej i politycznej w danym kraju oraz na świecie). Terminy ich ogłaszania znane są zwykle wcześniej, a niepewność ich dotycząca przyczynia się do zwiększenia zmienności kursów wymiany.

Definiując przełącznikowy model Markowa należy w pierwszej kolejności wyznaczyć liczbę stanów, które generują wartości zmiennej objaśnianej, a następnie określić dynamikę ich zmian. Zgodnie z założeniami modelu Hamiltona opisanymi poprzez relacje (2,3), szereg empiryczny logarytmicznych stóp zwrotu dla notowań danego kursu wymiany należy zdekomponować na fazę wzrostu i spadku kursu. W ten sposób zostaną wyodrębnione dwa reżimy: reżim odpowiadający deprecjacji kursu oraz reżim odpowiadający aprecjacji kursu. Oznacza to przyporządkowanie zmiany wartości szeregu czasowego odpowiadającego danemu kursowi wymiany jednemu z dwóch reżimów. Przyjęto założenie, że stan obowiązujący w danej chwili kształtuje się zgodnie z łańcuchem Markowa, przy czym realizacje tego łańcucha są nieobserwowalne. Oznacza to istnienie dwóch rozkładów, z których pochodzą obserwacje kursu walutowego. Przykładowo: pierwszych k_1 obserwacji może być generowane przez reżim 0, następnych k_2 – przez reżim 1, kolejnych k_3 – przez reżim 0 i tak dalej. Wartości k_i dla i = 1, 2, .., K, gdzie K oznacza liczbę punktów przełączających, są nieznane. Nie dysponuje się bowiem wiedzą *a priori* na temat przyporządkowania poszczególnych obserwacji kursu walutowego odpowiadającym im reżimom.

Dla każdego reżimu należy wyznaczyć jego podstawowe charakterystyki: wartość oczekiwaną zmiany kursu (μ_0 i μ_1) i zróżnicowanie zmian kursu (σ_0 i σ_1), które przybliżą dynamikę zmian kursu w obrębie danego stanu. Należy oszacować również prawdopodobieństwa przejścia dla poszczególnych stanów (p_{00} i p_{11}). Estymację parametrów modelu przełącznikowego można przeprowadzić za pomocą algorytmu EM, który został opisany w poprzednim paragrafie. Wartości

prawdopodobieństw przejścia są punktem wyjścia do wyznaczenia średniej długości trwania okresów aprecjacji i deprecjacji kursu wymiany:

$$a = \frac{1}{1 - p_{00}} \tag{13}$$

$$d = \frac{1}{1 - p_{11}} \tag{14}$$

gdzie: a – średnia długość trwania okresu aprecjacji
d – średnia długość trwania okresu deprecjacji.

Podsumowanie

Przy modelowaniu krótkookresowych wahań kursu walutowego zwykle występuje potrzeba uwzględnienia uwarunkowań typu losowego, czyli rezygnacja z fundamentalnych teorii kursu walutowego na rzecz modelowania probabilistycznego. Wśród modeli stochastycznych na szczególną uwagę zasługują modele Markowa. Do głównych zalet modeli Markowa należy zaliczyć względną prostotę konstrukcji, łatwość wnioskowania, znane metody estymacji, a zwłaszcza zgodność własności tych modeli z obserwowanymi własnościami wielu rzeczywistych zjawisk. Zastosowanie modeli przełącznikowych opiera się na ogólnym założeniu, że badane szeregi czasowe można przedstawić jako ciągi zmiennych losowych o znanym typie rozkładu warunkowego w każdym reżimie. Znane z literatury propozycje modelowania kursu walutowego przy użyciu modeli przełącznikowych nie dostarczały wystarczająco dobrych prognoz przyszłych poziomów kursów wymiany, miedzy innymi ze względu na niską częstotliwość danych użytych do budowy modelu (dane kwartalne czy miesięczne [4, 5]). Autorka zamierza kontynuować rozpoczęte w tym opracowaniu rozważania nad wykorzystaniem modeli Markowa do badania zmienności kursu wymiany złotego. Kolejny etap jej pracy będzie związany ze skonstruowaniem na podstawie danych o wysokich częstotliwościach pomiaru przełącznikowego modelu Markowa, który następnie zostanie zastosowany do opisu trwania okresu deprecjacji i aprecjacji kursu złotego.

Literatura

1. Bennett D.: Ryzyko walutowe. Dom Wydawniczy ABC, Kraków 2000
2. Campbell J.Y., Lo A. W., MacKinlay A. C.: The Econometrics of Financial Markets. Princeton University Press, Princeton, New Jersey 1997
3. Chrabonszczewska E., Kalicki K.: Teoria i polityka kursu walutowego. Oficyna Wydawnicza SGH, Warszawa 1996
4. Engel C., Hamilton J.D.: Long Swings in the Dollar: Are They in Data and Do Markets Know It? American Economic Review 80/1990

5. Engle Ch.: Can the Markov switching model forecast exchange rates? NBER Working Paper no. 4210/ 1992
6. Hamilton J.D.: Analysis of Time Series subject to Changes in Regime. Journal of Econometrics 45/1990
7. Hamilton J.D.: Time Series Analysis. Princeton University Press, Princeton, New Jersey 1994
8. Lee T.C., Judge G.G., Zellner A.: Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data. North Holland, Amsterdam 1997

Mgr Aneta Włodarczyk
Politechnika Częstochowska
Wydział Zarządzania
Katedra Ekonometrii i Statystyki

ROZDZIAŁ 6

BEZPIECZEŃSTWO I EFEKTYWNOŚĆ

TEORIA I METODYKA PROWADZENIA TESTÓW NA POSZCZEGÓLNYCH ETAPACH PROJEKTU

Bogdan BEREZA-JAROCIŃSKI

Streszczenie: Choć pokusy tzw. integracji skokowej mogą być wielkie (po co testować kilkakrotnie na różnych poziomach, skoro może się udać zastosować tylko jeden, duży test systemowy?), to jednak wszystko wskazuje na to, że testowanie powinno odbywać się na wielu poziomach, których cele - i stosowane metody - są różne, pozwalają na znajdowanie błędów różnego typu oraz jak najwcześniejsze ich znajdowanie, co obniża koszty.
Podziałem testowania na etapy oraz opisem technik i celów specyficznych dla każdego z nich zajmuje się poniższe opracowanie.

1. Formalne metodyki testów - metody doboru przypadków testowych

Wprawdzie testowanie jest tą dziedziną inżynierii oprogramowania, gdzie intuicja, doświadczenie i praktyka mają niekiedy większe znaczenie niż teoria oraz techniki formalne, niemniej testowanie bez znajomości (i zastosowania) podstawowych choćby technik formalnych skazane jest na nieefektywność.

1.1. Podstawa: pokrycie wymagań

Nagminnie popełnianym błędem w wielu projektach informatycznych jest przystępowanie do testowania bez znajomości wymagań. Skazuje to testerów na wiele niekorzystnych konsekwencji.

- Nieznany jest zakres funkcjonalny testowanego oprogramowania, wobec czego nie wiadomo, co (jakie funkcje) należy testować, nie daje się zaplanować testów ani dokonać priorytetyzacji przypadków testowych. W tej sytuacji także decyzja o zakończeniu lub kontynuowaniu testowania jest raczej czystą loterią niż wnioskiem wyciągniętym z wyników testów.
- Nieznane są prawidłowe wyniki zadań testowych - decyzja w tej sprawie pozostawiona jest domyślności testera i podejmowana ad hoc. Prowadzi to zwykle do chaotycznych dyskusji na temat błędów: błąd czy cecha? Błąd istotny czy kosmetyczny? Tego rodzaju rozważania są czasochłonne, angażują wiele osób, mogą wywoływać konflikty i zwykle nie przynoszą rozstrzygnięć - gdy wymagania są nieznane.
- Pokrycie wymagań = ilość przetestowanych wymagań / ilość wszystkich wymagań. Ta miara, choć nieprecyzyjna, pozwala jednak oszacować jakość i zakres wykonanych testów. Gdy wymagania są nieznane, miara nie daje się obliczyć.

1.2. Testowanie statyczne i dynamiczne

Metodyki testowania dzieli się często na dwie podstawowe grupy:

- *techniki statyczne*: bez wykonywania gotowego programu, np. przeglądy, inspekcje dokumentacji, analiza statyczna (kodu źródłowego, modelu)
- *techniki dynamiczne*: program (lub jego fragment) wykonuje podane mu przypadki/zadania testowe na komputerze.

Niektóre sytuacje z trudem daje się dopasować do tego podziału. Np. wykonywany może być model lub prototyp (zapisane w innym języku niż ostateczny kod źródłowy), zaś kod binarny z kolei może być wykonywany w środowisku innym niż docelowe, przy użyciu symulatorów lub emulatorów.

1.3. Metody „czarnej skrzynki" i „szklanej skrzynki "

Kluczem do powodzenia w testowaniu jest właściwy wybór przypadków testowych. Użytkownicy - wiele setek lub tysięcy osób - będą aplikację wykorzystywać przez wiele lat, podczas gdy w projekcie informatycznym możemy na testowanie w najlepszym razie przeznaczyć czas liczony w tygodniach i zespół testowy mający kilkanaście osób. Dlatego konieczne jest dokonanie takiego wyboru przypadków testowych, by móc odnaleźć błędy najważniejsze (gdzie konsekwencje awarii byłyby największe) lub najbardziej widoczne (gdy awarie - nawet niegroźne - występują w funkcjach często przez użytkowników wykorzystywanych).

Istnieje wiele technik wyboru przypadków testowych w sposób systematyczny i kontrolowany. Dzieli się je niekiedy na dwie grupy:

- Techniki "czarnoskrzynkowe" (*black-box*): zadania testowe wybiera się uwzględniając głównie dane wejścia/wyjścia, zaś aplikację/system traktując jak "czarna skrzynkę" o nieznanej strukturze i zawartości.
- Techniki "szklanoskrzynkowe": zdania testowe wybiera się biorąc pod uwagę konstrukcję systemu/programu (np. pokrycie kodu źródłowego przy wykonywaniu zadań testowych).

1.4. Analiza dynamiczna

Nie wszystkie możliwe błędy programów da się jednoznacznie i ściśle opisać w specyfikacjach wymagań i funkcjonalnych. Szczególna klasa błędów to niekorzystne/błędne skutki uboczne, których rodzajów może być wiele i które mogą się ujawniać w trudnych do przewidzenia miejscach.

Poszukiwanie niepożądanych/błędnych skutków ubocznych w czasie wykonywania programu nazywamy *analizą dynamiczną*. Różnymi metodami (zwykle przy użyciu pracujących "w tle" specjalnych narzędzi) poszukuje się "przecieków" pamięci, omyłkowego zniszczenie danych lub kodu programu, niespójności danych, zakleszczeń i innych błędów czasu rzeczywistego.

1.5. Analiza pokrycia kodu

Ważną pomocniczą techniką testową, pozwalającą na oszacowanie jakości (kompletności) wykorzystanego zestawu testów, jest *analiza pokrycia kodu źródłowego*.

Pomiary pokrycia wykonuje się z reguły przy użyciu narzędzi. Narzędzia te działają dwufazowo: najpierw dokonują instrumentacji kodu źródłowego, potem mierzą pokrycie podczas wykonywania zadań testowych.

Wyróżnia się różne typy pokrycia kodu: pokrycie instrukcji, pokrycie decyzji, pokrycie warunków logicznych itd.

Interpretacja wyników pomiaru pokrycia kodu nie jest zadaniem łatwym. Trzeba pamiętać, że 100% pokrycia kodu nie musi oznaczać 100% pokrycia wymagań! Żadne testy strukturalne nie są w stanie ujawnić np. braku niektórych funkcji czy ich niepełnej implementacji.

Pomiar pokrycia kodu wspomaga identyfikację potencjalnych zagrożeń: istnienie tzw. "martwego kodu", funkcji nieprzetestowanych czy takich (np. kodu obsługującego przerwania), których dynamiczne przetestowanie byłoby trudne i kosztowne, a wobec tego korzystniejsze może być zastosowanie np. inspekcji ich kodu.

1.6. Metody statystyczne

Potencjalnie, metody statystyczne są bardzo obiecujące. W praktyce ich trudność polega na tym, że niełatwe i zwykle bardzo kosztowne jest stworzenie zgodnego z rzeczywistością, pełnego modelu rozkładu prawdopodobieństwa różnych czynności użytkownika czy różnych trybów pracy programu.

Jeśli jednak taki model (choćby przybliżony) jest dostępny, umożliwia generowanie i dobór kolejności wykonywania zadań testowych na podstawie modelu zachowań użytkownika. Na tej podstawie daje się z wysokim poziomem istotności przewidywać MTBF, przewidywać ilość pozostałych jeszcze błędów przy pomocy analizy trendu.

1.7. Testowanie różnych atrybutów jakości

Zwykle przedmiotem specyfikacji wymagań jest *funkcjonalność* projektowanego programu. Nie mniej ważne, choć często traktowane zbyt powierzchownie lub wręcz zapominane, są tzw. *atrybuty niefunkcjonalne* lub *właściwości* programów, takie jak wydajność (czas odpowiedzi, obciążenia), użyteczność, pielęgnowalność, adaptowalność, bezpieczeństwo, łatwość instalacji.

Nawet jeśli atrybuty niefunkcjonalne zostały pominięte lub niedostatecznie zdefiniowane w specyfikacji wymagań, powracają one z reguły jako przedmiot testów systemowych i akceptacyjnych, co warto wziąć pod uwagę.

2. Fazy testowania na poszczególnych etapach projektu

2.1. Przeglądy oraz inspekcje

Przeglądy oraz inspekcje - zwykle jedyne metody testowe dające się zastosować wobec dokumentacji zanim powstanie kod oprogramowania, nie zawsze zaliczane są do "testowania".

Mimo to są to techniki bardzo przydatne i skuteczne: jedyne, dostępne we wczesnych fazach projektu, jedyne dające się zastosować wobec nieformalnej dokumentacji.

Choć przeglądy nie są algorytmicznymi, formalnymi technikami, istnieje wiele przykładów na to, że - właściwie zastosowane - bywają zadziwiająco skuteczne. Jednocześnie są często odbierane jako nudne i budzące niechęć.

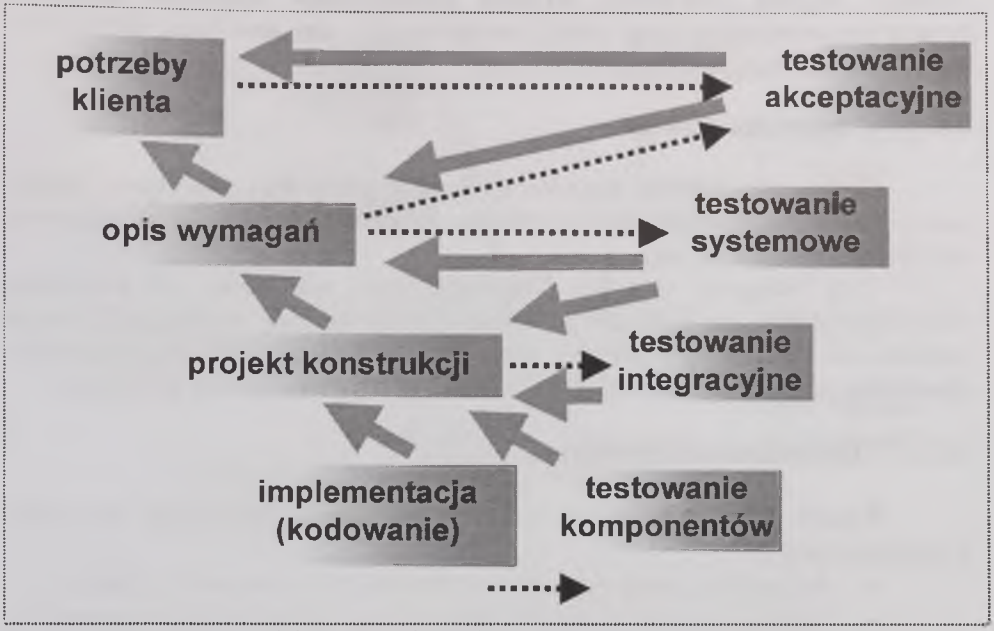
Najskuteczniejsze są przeglądy formalne, tak zwane *inspekcje*, które jednak dają się zastosować tylko do niektórych dokumentów.

2.2. Testowanie wymagań

Wyróżnia się cztery znaczenia tego pojęcia:

- Kontrolę zgodności opisu (specyfikacji) wymagań z potrzebami i oczekiwaniami użytkownika (walidację wymagań)
- Kontrolę spójności, kompletności, jednoznaczności itd. opisu wymagań
- Generowanie zadań testowych na podstawie specyfikacji wymagań (co da się wykonać automatycznie, jeśli specyfikacja wymagań napisana jest w jakimś języku formalnym)
- Test gotowego systemu wobec wymagań (test akceptacyjny, ostatnia faza walidacji).

2.3. Model "V"



Rys. 1. Uproszczony model "V" uwzględniający planowanie i wykonanie testów

2.4. Testowanie komponentów

Testowanie komponentów - pierwsza faza testowania dynamicznego - ma szczególne bogactwo rozmaitych nazw: testowanie jednostkowe, testowanie modułów, podstawowe, konstrukcyjne itd.

Testy komponentów są często wykonywane przez konstruktorów (programistów), często także nieformalne i nieudokumentowane. Zarazem jednak ich jakość i skuteczność ma kluczowe znaczenie dla sprawnej integracji. Testy komponentów pozwalają też wcześnie identyfikować błędy, co - jak już wiemy - z reguły znacznie obniża koszty ich usuwania.

2.5. Testowanie integracyjne

Uwaga: integracja to *konstrukcja*, a nie test (ale często niesłusznie są mieszane).

Nazwa *testowanie integracyjne* bywa różnie rozumiana. Niektórzy zaliczają do niej przede wszystkim poprzedzające integrację *testowanie interfejsów*, zaś inni - następujące zaraz po integracji testowanie powstających większych całości.

Kiedy integracja sprawia wiele kłopotów, często staje się to przyczyną przekonania, że zawodzi testowanie. Należy pamiętać, że szukanie błędów w trakcie integracji to czynność konstrukcyjna, a nie test!

Test integracyjny zależny jest od wybranej strategii integracji. Np. strategia odgórna (top-down) wymaga konstruowania *namiastek* (zaślepek, programów zastępczych; ang. *stubs*), zastępujących - aby umożliwić testowanie - nieistniejące jeszcze funkcje niższego rzędu.

2.6. Testowanie systemu

Testowanie całości systemu to przede wszystkim testowanie funkcji całościowych (*end-to-end*) oraz atrybutów niefunkcjonalnych (bezpieczeństwa, wydajności, pojemności itp).

Przy *integracji skokowej* ("big-bang") test systemowy jest pierwszym formalnym testem, co z reguły powoduje wiele trudności: wielka ilość błędów utrudnia ich lokalizację; konieczne okazują się spóźnione, gruntowne przeróbki; spiętrzenie problemów następuje przy bliskim już terminie dostawy do klienta!

2.7. Testowanie konfiguracyjne

Istnieje kilka definicji czynności wykonywanych w ramach testowania konfiguracyjnego:

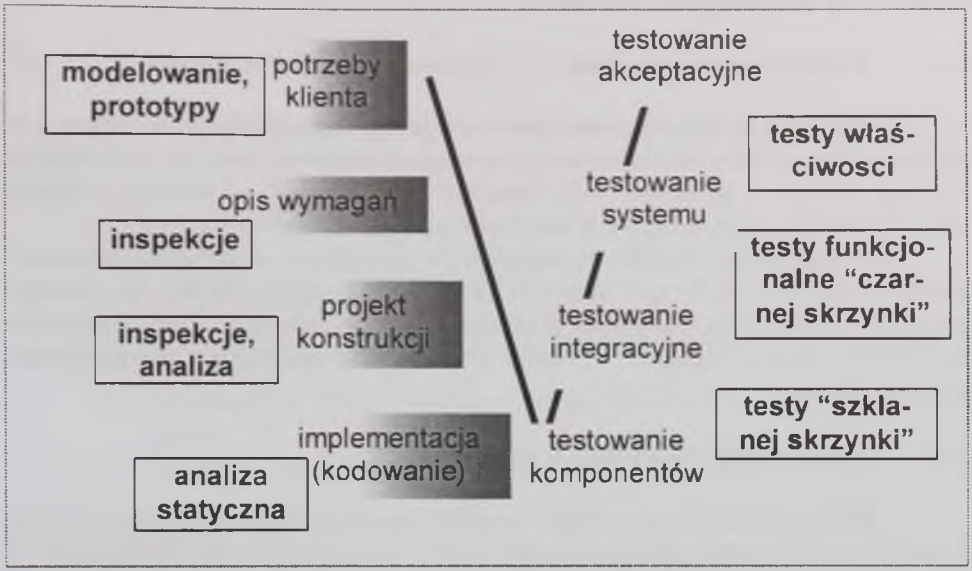
- Różne konfiguracje sprzętowe systemu (np. z różnymi zasobami)
- Ten sam system w różnym środowisku, np. aplikacja internetowa na różnych klientach, wersjach systemu operacyjnego, wersjach i ustawieniach przeglądarek
- Test aplikacji we współdziałaniu z różnymi innymi aplikacjami ("integracja wysokiego rzędu")

2.8. Test w trakcie użytkowania i konserwacji systemu

Technicznie test w czasie pielęgnacji systemu niczym nie różni się od testu we wcześniejszych fazach. Istnieją natomiast istotne różnice organizacyjne i logistyczne:

- Test na działającym systemie
- Krótki czas od awarii do naprawy
- Konieczność zachowania nienaruszonych danych
- Brak środowiska testowego
- Wdrożenie nowej wersji wymaga czasem planów rezerwowych (np. obie wersje czasowo równolegle, możliwość odwrotu)

2.9. Techniki testowania na różnych etapach projektu



Rys. 2. Techniki testowania na różnych etapach projektu informatycznego

3. Testowanie na poziomie wymagań

3.1. Test, weryfikacja, walidacja, certyfikacja

Różne normy mają różne definicje, niemniej ich sens jest zwykle zbliżony i dający się opisać następująco:

- test: porównanie wyniku faktycznego z oczekiwanym (niekiedy utożsamiany z testowaniem dynamicznym)
- weryfikacja: porównanie wyniku etapu "n+1" z wymaganiami etapu "n"
- walidacja: porównanie wyniku (zwłaszcza końcowego) z wymaganiami lub ich opisem
- certyfikacja: porównanie ze standardem

3.2. Walidacja systemu

Walidacja jest „udowodnieniem", że oprogramowanie spełnia funkcje wymagane przez użytkownika.

Walidacja jest procesem wielostopniowym. Osiąga się ją częściowo poprzez walidację specyfikacji wymagań, poprzez weryfikację kolejnych kroków

w trakcie konstrukcji jak i wreszcie poprzez walidację gotowego systemu (test systemu, test akceptacyjny, wdrożenie).

3.3. Zatwierdzenie wymagań

Walidacja wymagań: skontrolowanie, że opis (specyfikacja) wymagań jest rzeczywiście zgodna z oczekiwaniami i potrzebami użytkownika, nie jest zdaniem prostym. Jedną z podstawowych trudności jest zwykle rozbieżność między językiem użytkownika a językiem konstruktora.

Istnieje wiele metod, pozwalających uporać się z tymi trudnościami: modelowanie, np. w formie przypadków użycia; prototypowanie, np. samego interfejsu użytkownika; metodyki przyrostowe i zaangażowanie zleceniodawców na każdym etapie; śledzenie wymagań (*traceability*) ułatwiające uwzględnienie późnych zmian.

3.4. Konstrukcja testowalnych wymagań

Błędy znalezione na początku projektu, np. w fazie specyfikacji wymagań, z reguły są wielokrotnie tańsze do naprawienia i usunięcia niż błędy znalezione w testach systemowych, nie mówiąc już o tych, które na jaw wyjdą dopiero po wdrożeniu czy instalacji w środowisku produkcyjnym. Dlatego tak istotne jest, by same wymagania poddać w miarę rygorystycznemu testowaniu.

Ważnym atrybutem wymagań jest to, na ile są "testowalne" tj. czy są wystarczającą podstawą do tworzenia przypadków testowych. Warunki, jakie muszą spełniać testowalne wymagania, to m.in. aby były pojedyncze (nie złożone), spójne wewnętrznie, spójne terminologicznie, jednoznacznie sformułowane i mierzalne (przy zaangażowaniu realnych środków).

3.5. Brak lub zmienne wymagania

W przemysłowej rzeczywistości bardzo często musimy sobie radzić w sytuacji (częściowego) braku wymagań i ich dużej zmienności, wobec których techniki formalne, algorytmiczne są bezradne. Temu celowi mogą służyć następujące metody:

- Testowanie “eksploracyjne”
- Metodyki przyrostowe i spiralne
- Aktywne poszukiwanie wymagań przez zespół testowy
- Określenie zadań testowych jako *de facto* wymagań
- Uproszczony opis wymagań, np. w XP (*Extreme Programming*)
- Udział użytkowników w testowaniu

4. Związek miedzy testowaniem a pozostałymi etapami projektu

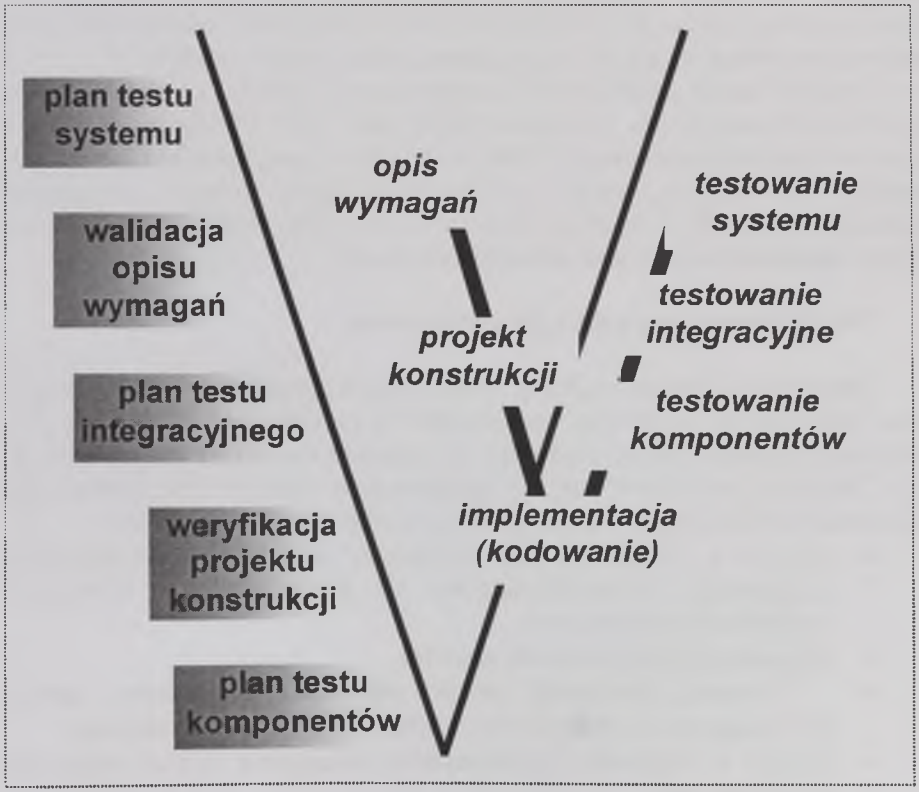
4.1. Niebezpieczeństwa traktowania testu jak „etapu"

Traktowanie testowania jako "etapu" odziedziczyliśmy z modelu kaskadowego, gdzie testowanie pojawia się na końcu procesu wytwarzania programu, kiedy wszystko inne jest już gotowe.

Takie podejście niesie ze sobą wiele niebezpieczeństw:

- Testowanie kumuluje się pod koniec projektu, kiedy brakuje czasu
- Zła jakość dostaw przedłuża integrację
- Koszt naprawy błędów jest dramatycznie wyższy w późniejszych fazach
- Zaniechanie testu w fazie wymagań powoduje kumulowanie się błędów
- Spóźnione jest także planowanie testowania

4.2. Modele „V " i „W "



Rys. 3. Model "W"

4.3. Definicje testowania

Sposób zdefiniowania, czym jest testowanie, może na pierwszy rzut oka wydać się problemem czysto akademickim, nie mającym wiele wspólnego z realiami produkcji oprogramowania. Tak jednak nie jest - przyjęta definicja testów ma często bardzo konkretny wpływ na to, w jakim celu i w jaki sposób testy się wykonuje.

Oto kilka przykładów definicji testowania, które odzwierciedlają różne "filozofie" podejścia do testów i ich roli w procesie i projekcie:

- Test = śledzenie i naprawa błędów
- Test = udowodnienie, że działa
- Test = próba udowodnienia, że nie działa (w celu znalezienia błędów)
- Test = pomiar i ocena ryzyka istnienia dalszych błędów
- Test = sposób podejścia

4.4. Niebezpieczne obszary „ziemi niczyjej ”

Testowanie jest dziedziną, której nie daje się skutecznie zrealizować za zamkniętymi drzwiami. Znalezione błędy (awarie) nie przyczynią się do polepszenia jakości aplikacji, jeśli informacja na ich temat nie zostanie skutecznie przekazana programistom, tak by ich przyczyny mogły zostać usunięte.

Jest kilka takich sąsiadujących z testowaniem obszarów "ziemi niczyjej", o której nie wolno zapomnieć. *Integracja (*często jako “test i integracja”) odciąga testerów od właściwego testowania. *Śledzenie błędów* wymaga procesu, organizacji i narzędzi, i nie może być “wojną” ani domeną wyłącznie testerów. *Zarządzanie konfiguracją i zmianami*: w chaosie może się czasem udać zbudować produkt, ale test jest w warunkach chaosu jest niemal bezużyteczny.

4.5. Wyniki testowania a decyzje wdrożeniowe

Testowanie wzbudza niekiedy wiele niepotrzebnych emocji. Zwłaszcza wyraźne staje się to w okresie bezpośrednio poprzedzającym wdrożenie czy wypuszczenie produktu, kiedy pojawiają się rozmaite naciski na zespół testowy, aby nie "utrudniać" projektowi sukcesu znajdowaniem błędów. Aby uniknąć tego rodzaju zbędnych konfliktów, warto uświadomić sobie następujące zasady:

- decyzja o wdrożeniu jest *biznesowa*, nie techniczna! (powinna uwzględniać ryzyko techniczne, ale także ryzyko i wymagania rynkowe czy kontraktowe)
- odpowiada za nią kierownik *projektu*...
- ... natomiast kierownik *testów* odpowiada za dostawę danych pozwalających na maksymalnie trafną ocenę ryzyka technicznego
- decyzja o wdrożeniu (wypuszczeniu) uwzględnia ryzyko techniczne oraz biznesowe.

LITERATURA

1. Beizer, B. *Software System Testing and Quality Assurance*, 1996
2. Black, R. *Managing the Testing Process*, 2002
3. Fewster, M., Graham, D. *Software Test Automation*, 1999
4. Gilb, T., Graham, D. *Software Inspection*, 1993
5. Hetzel, B. *The Complete Guide To Software Testing*, 1993
6. Kaner, C., Falk. J., Nguyen, H.Q. *Testing Computer Software*, 1993
7. Koomen, T., Pol, M. *Test Process Improvement*, 1999
8. Myers, G. *The Art of Software Testing*, 1979
9. Patton, R. *Testowanie oprogramowania*, MIKOM, 2002
10. Wiszniewski, B. *Weryfikacja, walidacja i testowanie* [w:] Górski, J. (red.) *Inżynieria oprogramowania*, MIKOM, 2000
11. Wiszniewski B. et. al. *Od projektu do programu* [w:] Szejko, St. (red.) *Metody wytwarzania oprogramowania*, MIKOM, 2003
12. BS 7925-1, *Vocabulary of Terms in Software Testing*
13. BS 7925-2, *Software Component Testing*
14. IEEE Std 829-1998, *Standard for Software Test Documentation*
15. IEEE Std 1008-1987, *Standard for Software Unit Testing*
16. IEEE Std 1012-1998, *Standard for Software Verification and Validation*

Bogdan Bereza-Jarociński

bbj Test www.bbj.com.pl

ul. Nutki 2 m. 8

PL 02-785 Warszawa

Tel.: +46-709-714 293

e-mail: bbj@bbj.com.pl

ZARZĄDZANIE KONFIGURACJĄ OPROGRAMOWANIA A ZARZĄDZANIE WERSJAMI ELEMENTÓW KONFIGURACJI

Krzysztof GÓRBIEL

Streszczenie: Celem pracy jest usystematyzowanie wiedzy dotyczącej zarządzania konfiguracją. Po zdefiniowaniu najważniejszych pojęć takich jak konfiguracja, element konfiguracji, wersje, rewizje, warianty opisane zostaną elementy składowe zarządzania konfiguracją. W efekcie umożliwi nam to wzajemną lokalizację takich pojęć jak zarządzanie wersjami czy zarządzanie zmianami względem całości procesu zarządzania konfiguracją.

Wstęp

Zarządzanie konfiguracją jako odrębna dziedzina inżynierii oprogramowania powstała na przełomie lat siedemdziesiątych i osiemdziesiątych. Widoczny stawał się wtedy tzw. kryzys oprogramowania. Wiele przedsięwzięć informatycznych kończyło się niepowodzeniem lub dużym opóźnieniem, czego nie można było przypisać jedynie błędom programistów. Okazało się, że proces wytwarzania oprogramowania czy architektura systemu, na które do tej pory nie zwracano większej uwagi są niezwykle istotne. Odpowiedzią programistów na te problemy stał się burzliwy rozwój metod inżynierii oprogramowania, wśród których ważne miejsce zajęło zarządzanie konfiguracją. Często jest ono utożsamiane jedynie z zarządzaniem wersjami elementów konfiguracji (wersjonowaniem) czy też z zarządzaniem zmianami, co jest jednak dużym uproszczeniem. Celem referatu, będzie więc usystematyzowanie pojęć związanych z zarządzaniem konfiguracją oraz syntetyczne opisanie jego elementów składowych.

1. Definicja zarządzania konfiguracją

Pod pojęciem zarządzania konfiguracją rozumieć będziemy wszystkie czynności, których celem jest modyfikacja oprogramowania lub jej przygotowanie i usprawnienie. Takie ujęcie problemu jest tożsame z szeroko stosowanym w literaturze angielskojęzycznej pojęciem Software Configuration Management.

Najczęściej spotykaną i powszechnie akceptowaną definicją zarządzania konfiguracją jest definicja podana w standardzie IEEE nr 729 z 1983r [4]:
Zarządzanie konfiguracją jest procesem identyfikującym i definiującym elementy systemu, kontrolującym ich zmiany przez cały okres życia produktu, raportującym status elementów i propozycji zmian oraz weryfikującym kompletność i poprawność tych elementów.

Definicja ta traktuje, więc zarządzanie konfiguracją jako proces trwający przez cały cykl życia produktu, a nie jak jest to spotykane w niektórych

opracowaniach jako proces wspierający fazę rozwoju oprogramowania do momentu oddania go użytkownikowi (dla odróżnienia od fazy konserwacji). W definicji tej na uwagę zasługuje też wyodrębnienie czterech składników zarządzania konfiguracją:

- Identyfikacja konfiguracji i jej elementów składowych: opis struktury i elementów systemu w jednoznaczny sposób poprzez numerację wersji elementów.
- Kontrola zmian: zarządzanie procesem wprowadzania zmian poprzez tworzenie wersji bazowej systemu.
- Raportowanie statusu elementów systemu: opisywanie produktu i procesu za pomocą statystyk obliczonych na podstawie zebranych danych.
- Weryfikacja kompletności i poprawności konfiguracji: sprawdzenie czy produkt jest złożony z odpowiednich elementów.

Zawartość powyższej lista jest powszechnie akceptowana, jednak niektórzy autorzy dodają do niej dodatkowe elementy. Przykładem może być jej uzupełnienie o takie elementy jak [1]:

- Budowa systemu z elementów składowych, a więc zarządzanie procesem budowy oprogramowania poprzez optymalny dobór narzędzi i procedur.
- Wypuszczanie oprogramowania na rynek
- Zarządzanie procesem wytwarzania oprogramowania poprzez zapewnienie poprawnego wykonywania procedur, polityki firmy czy metodyki prowadzenia projektu.
- Praca zespołowa – kontrola wielu programistów pracujących nad jednym produktem

Ze względu na kluczowe znaczenie przedstawionych składników zarządzania konfiguracją opisane one zostaną bardziej szczegółowo w rozdziale czwartym po wyjaśnieniu niezbędnych pojęć.

2. Podstawowe pojęcia

2.1. Element konfiguracji a konfiguracja oprogramowania

Konfigurację systemu zdefiniować możemy jako funkcjonalną i fizyczną charakterystykę sprzętu, oprogramowania i rozwiązań organizacyjnych zapisaną w technicznej dokumentacji lub rzeczywiście osiągniętą w wytworzonym produkcie. Oznacza ona też kolekcję elementów sprzętowych, programistycznych i organizacyjnych utworzoną zgodnie z ustalonymi procedurami i w określonym celu [7]. Konfiguracja oprogramowanie jest więc częścią tak rozumianej konfiguracji systemu i składa się z elementów konfiguracji oprogramowania (w dalszej części pracy przez konfigurację rozumieć będziemy konfigurację oprogramowania, a przez element konfiguracji, element konfiguracji oprogramowania).

Elementem konfiguracji nazywamy natomiast wyodrębnioną jednostkę oprogramowania, sprzętu lub ich kombinację traktowaną jako pojedynczy element w procesie zarządzania konfiguracją [2]. Elementy konfiguracji tworzą hierarchię – elementy wyższego poziomu składają się z innych elementów. W ten sposób zarządzanie konfiguracją może operować nie tylko na modułach, ale też np. podsystemach. Elementem najwyższego poziomu jest cała konfiguracja.

W każdym systemie zarządzania konfiguracją do elementów konfiguracji zaliczamy moduły z kodem źródłowym, jednakże często uwzględniane są również inne elementy związane z przebiegiem prac programistycznych. Przykładami elementów konfiguracji są [5]:

- Moduły z kodem źródłowym
- Pliki systemowy z danymi
- Skrypty używane do budowy systemu z elementów składowych
- Specyfikacja systemu
- Specyfikacja interfejsów
- Plany testów
- Procedury testowe
- Dane testowe
- Wyniki testów
- Dokumentacja dla użytkownika
- Plan zapewnienia jakości
- Plan zarządzania zmianą
- Narzędzia budowy systemu (kompilatory, analizatory kodu)
- Systemy operacyjne
- Wewnętrzne procedury i standardy

Oczywiście powyższa lista przedstawia jedynie możliwe elementy konfiguracji, nie jest ona jednak wyczerpująca, nie zawsze też wszystkie wymienione elementy należy objąć zarządzaniem konfiguracją. Wybór elementów następuje w procesie identyfikacji konfiguracji z uwzględnieniem specyfiki projektu, co zostanie opisane w dalszej części pracy.

Elementy konfiguracji można podzielić na trzy kategorie:

- Elementy źródłowe – są to wszelkie dokumenty i fragmenty programu stworzone ręcznie przez programistę np. przy użyciu interaktywnego edytora. Elementy te nie mogą być, więc wygenerowane automatycznie.
- Elementy pochodne – elementy generowane automatycznie najczęściej z elementów źródłowych. Elementy pochodne nie muszą być przechowywane gdyż można je odtworzyć z elementu źródłowego znając wersję programu generującego. Ze względu na usprawnienie pracy często zdarza się jednak, że najczęściej wykorzystywane elementy są przechowywane w specjalnie do tego przeznaczonych obszarach dyskowych.

- Narzędzia generujące elementy pochodne z elementów źródłowych, przykładami których są kompilatory, programy formatujące wydruki, itp.

Główną troską zarządzania konfiguracją są tylko elementy źródłowe oraz programy generujące. Wyjątkiem od tej reguły są elementy pochodne po dokonaniu na nich ręcznej modyfikacji. W takiej sytuacji element ten staje się elementem źródłowym oraz jest starannie przechowywany.

Aby uzyskać przejrzystość organizacyjną każdy element konfiguracji powinien być kontrolowany przez wyznaczone do tego celu osoby decydujące o jego zmianach. Wyróżnić możemy trzy poziomy kontroli elementów [2]:

- Kontrola autora. Dotyczy ona programistów dokonujących zmiany w tworzonym oprogramowaniu do momentu przejścia napisanych fragmentów przez testy modułów lub dokumenty przed akceptacją kierownictwa.
- Kontrola kierownika projektu odpowiedzialnego za łączenie niskopoziomowych elementów konfiguracji w większe jednostki oprogramowania. Kierownik projektu jest odpowiedzialny za poszczególne moduły lub produkty po ich przetestowaniu i zaakceptowaniu.
- Kontrola rady nadzorującej przebieg rozwoju oprogramowania. Głównym jej zadaniem jest akceptacja wersji bazowych systemu oraz ich zmian. Rada ta powinna składać się z wysokiej klasy specjalistów w zakresie budowy systemów. W małych projektach składa się ona często jedynie z kierownika projektu i przedstawiciela klienta, w dużych natomiast jest ona powiększona o kierowników poszczególnych zespołów czy przedstawiciela działu zapewnienia jakość (QA).

2.2. Rewizje, wersje a produkt końcowy

Aby skutecznie identyfikować zmiany w systemie niezbędne jest rozróżnienie pomiędzy ich różnymi rodzajami, co znajduje odzwierciedlenie w nazewnictwie zmodyfikowanych elementów.

Pojęciem wyjściowym jest wersja, oznaczająca nowopowstały element konfiguracji na bazie elementu pierwotnego. W zależności od jej dalszej ścieżki rozwoju możemy wyróżnić następujące rodzaje wersji:

- Rewizja – zastępuje wersję pierwotną. Dalszy rozwój elementu następuje na bazie tej rewizji.
- Wariant elementu – umożliwia równoległy rozwój więcej niż jednej wersji elementu. Warianty często oferują taką samą funkcjonalność, ale zaprojektowane są dla innych warunków działania (np. inna platforma sprzętowa, inna wersja językowa). Ze względu na trudność śledzenia rozwoju wielu wariantów ich liczba powinna być możliwie mała, co można osiągnąć poprzez tworzenie uniwersalnego kodu. Oczywiście prowadzi to do większej złożoności systemu, ale dzięki temu osiągamy zmniejszenie złożoności zarządzania konfiguracją. Dlatego też decyzja o wprowadzeniu nowego wariantu powinna wyważyć potencjalne korzyści i straty z tym związane.

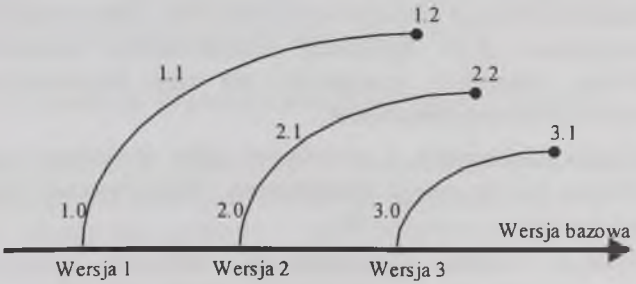
Dla opisanie rozwoju systemu jako całości używamy pojęcia – dystrybucja (release), oznaczającą formalnie zaakceptowaną konfigurację oddaną do dyspozycji użytkownika. Dystrybucja składa się, więc z wielu wersji. Czasem dla określenia całego systemu oddanego użytkownikowi stosuje się też pojęcie wydanie.

2.3. Wersja bazowa

Wersją bazową (baseline, baza kodu) nazywamy formalnie przetestowany i zaakceptowany element stanowiący podstawę dalszego rozwoju, który może być zmieniony jedynie w wyniku formalnego procesu (powstanie nowej wersji) [5]. Pomimo, że definicja ta może być stosowana do poszczególnych elementów, najczęściej wersja bazowa jest utożsamiana z całym systemem, mówimy więc tutaj o wersji bazowej produktu. Rozwój systemu stanowi, więc proces przechodzenia pomiędzy kolejnymi wersjami bazowymi zwanymi często punktami kontrolnym projektu (project milestones) odpowiadającymi kolejnym fazom rozwoju (analiza wymagań, projekt, implementacja, integracja, testowanie, wdrożenie). Wczesne wersje bazowe zawierają, więc jedynie dokumenty projektowe wzbogacane kolejno o kod źródłowy poszczególnych modułów i wyniki testów.
Wyróżnić możemy trzy rodzaje wersji bazowych [5]:

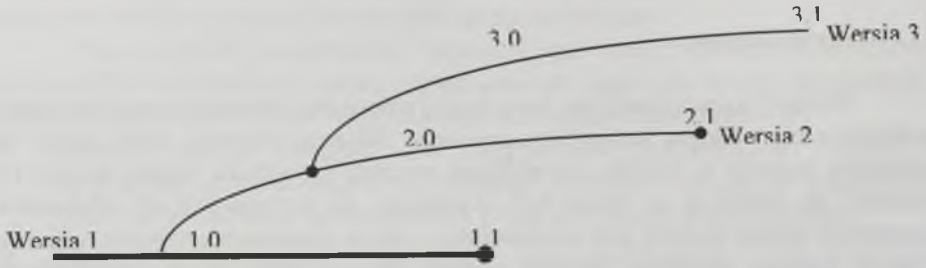
- Funkcjonalna składająca się z opisu zaakceptowanych przez klienta wymagań. Opisuje ona, więc jedynie co system ma robić.
- Alokowana (allocated) identyfikująca elementy konfiguracji oraz ich specyfikacje. Jej celem jest koordynacja prac nad systemem.
- Produktowa będąca implementacją poprzednich dwóch. Zawiera ona obok dokumentów projektowych również kod źródłowy.

W przypadku rozwoju systemu w oparciu o jedną wersję bazową wszelkie prace nad poszczególnymi częściami systemu dokonywane są bezpośrednio w wersji bazowej lub też jako nowe wersje systemu (np. wersja w produkcji, wersja rozwojowa) a następnie scalane z wersją bazową. W ten sposób wersja bazowa ewoluuje przez wszystkie fazy rozwoju łącznie z fazą konserwacji, co jest przedstawione na rys. 1 [6].



Rys. 1. Model rozwoju systemu z jedną wersją bazową

Alternatywnym sposobem rozwoju systemu jest model opierający się kolejnych promocjach od wersji rozwojowej do wersji produkcyjnej czasem z etapami pośrednimi. W przypadku takim nie mamy już do czynienia z jedną wersją bazową, ale wieloma wersjami stworzonymi w konkretnym celu. Model ten przedstawia rys. 2 [6].



Rys. 2. Model rozwoju systemu oparty o promocje kolejnych wersji

W porównaniu z poprzednim modelem, model oparty na promocjach wykazuje dwie wady. Po pierwsze zmiana aktualnej wersji bazowej jest często trudna do przeprowadzenia ze względu na problemy w komunikacji, a po drugie przeniesienie rezultatów prac programistów z jednej wersji do drugiej jest czasochłonne i narażane na błędy. Z tego też względu model z jedną wersją bazową jest najczęściej modelem preferowanym.

2.4. Repozytorium elementów konfiguracji

Centralnym elementem systemu zarządzania konfiguracją jest repozytorium, w którym przechowywane są wszystkie wersje elementów konfiguracji. W zależności od przeznaczenia i dostępności, biblioteki elementów możemy podzielić na [2]:

- Biblioteki dynamiczne – używane przez programistów do przechowywania nowoutworzonych lub zmodyfikowanych elementów. Kontrolowane są wyłącznie przez programistów.
- Biblioteki kontrolowane – służą do zarządzania wersjami bazowymi i do kontroli wprowadzanych w nich zmian, które muszą być zaakceptowane w formalnym procesie rozwoju wersji bazowej. Repozytorium to jest tworzone i zarządzane przez specjalnie wyznaczonego administratora (software librarian), natomiast programiści nie mają bezpośredniego dostępu do przechowywanych elementów.
- Biblioteki archiwalne – stosowane jako archiwum poprzednich wersji bazowych jak np. wersji produkcyjnej. Kontrolowane przez administratora repozytorium.

Głównym celem powyższych bibliotek jest uniemożliwienie nieautoryzowanego dostępu oraz jednoczesnej aktualizacji tego samego elementu

przez dwóch programistów. Zapewnione jest to poprzez przyznanie programistom odpowiednich praw oraz różne mechanizmy kontroli dostępu.

3. Elementy składowe zarządzania konfiguracją

3.1. Identyfikacja konfiguracji

Właściwa identyfikacja elementów konfiguracji jest warunkiem efektywnego przebiegu pozostałych czynności związanych z zarządzaniem konfiguracją. Jeżeli elementy nie są łatwo rozpoznawalne niemożliwa staje się efektywna kontrola wprowadzonych zmian, raportowanie postępów prac, czy też weryfikacja kompletności i bezbłędności konfiguracji. Identyfikacja konfiguracji składa się z trzech kroków:

3.1.1. Wybór elementów konfiguracji

Wybór elementów konfiguracji następuje podczas analizy wymagań lub na wczesnym etapie projektowania systemu. Ogólną zasadą jest taki wybór elementów, aby mogły być one niezależnie projektowane, implementowane i testowane. Innymi kryteriami wyboru elementów są [4]:

- Krytyczność fragmentu oprogramowania dla wydajności systemu, bezpieczeństwa lub ryzyka związanego z projektem.
- Duża złożoność elementu wynikająca m.in. z wdrożenia w nim najnowszych technologii.
- Element służy jako interfejs z innymi fragmentami systemu lub innymi systemami.
- Duże prawdopodobieństwo częstych modyfikacji elementu po wdrożeniu systemu.
- Fragment oprogramowania jest zainstalowany na innej platformie sprzętowej niż pozostałe elementy systemu
- Wielokrotne użycie tego samego elementu

Zidentyfikowane elementy powinny tworzyć logiczną całość, a jednocześnie powinny umożliwiać stosowanym narzędziom łatwość dostępu i sposobu operowania. Z tego też względu najczęściej elementem najniższego poziomu staje się plik, a w systemach bazodanowych rekord.

3.1.2. Oznaczenie elementów poprzez nadanie unikatowej nazwy

Stosowana konwencja nazewnictwa i numerowania powinna umożliwiać łatwą identyfikację typu elementu (kod źródłowy, dokument), miejsca w hierarchii systemu (element najniższego poziomu – plik/moduł/produkt) oraz jego wersji. Najczęściej stosuje się w tym celu oznaczenia złożone z ciągu znaków oznaczających kolejno typ elementu, nazwę produktu/modułu/pliku a następnie liczbę oznaczającą numer wersji. Przykładem poprawnego oznaczenia elementu

może być BS-ACC-ROU-FT4-7.5.5 co oznacza dystrybucję 7 w wersji 5 rewizji 5 programu (ROU – routine) o nazwie FT4 produktu Accounting systemu bankowego BS. Oczywiście tak rozbudowane nazwy stosuje się rzadko w codziennych czynnościach. Najczęściej poszczególne elementy pogrupowane są w mniejsze grupy (produkty/moduły), w ramach których jako oznaczenia używa się jedynie nazwy i typu elementu oraz jego wersji (często bez numeru dystrybucji). Stosowany schemat nazewnictwa spełniać musi też zasadę rozszerzalności, co oznacza, że dodanie dowolnej liczby nowych elementów nie wymaga zmiany oznaczenia istniejących.

3.1.3. Opis elementów

Opis elementów umożliwia późniejszą weryfikację poprawności konfiguracji i stanowi podstawę późniejszej kontroli zmian. Kod źródłowy będący pod kontrolą zarządzania konfiguracją opisuje się w dołączonych na początku programu komentarzach lub w systemowym opisie pliku, co ułatwia późniejszą kontrolę wprowadzonych zmian.

Właściwa identyfikacja konfiguracji wymaga uwzględnienia płaszczyzny produktowej i płaszczyzny wersji dowolnych elementów. Aspekt produktowy konfiguracji opisuje system bez uwzględnienia możliwości istnienia wielu wersji jednego elementu. Płaszczyzna wersji mówi natomiast o numerze wersji. Ażeby zapewnić możliwość jednoznacznej identyfikacji każdego atomowego elementu konfiguracji oznacza się go dwoma identyfikatorami – jednoznacznym identyfikatorem obiektu oraz unikatowym dla danego elementu systemu identyfikatorem wersji. W ten sposób powstaje przestrzeń wersji będąca przedmiotem wszystkich działań związanych z zarządzaniem konfiguracją.

Ze względu na wagę właściwego przedstawienia wszystkich wersji konfiguracji zarządzanie konfiguracją często utożsamia się tylko z zarządzaniem wersjami. Jako statyczny opis systemu stanowi ono podstawę dalszych działań, celem których jest modyfikacja konfiguracji, a nie jedynie jej opisanie.

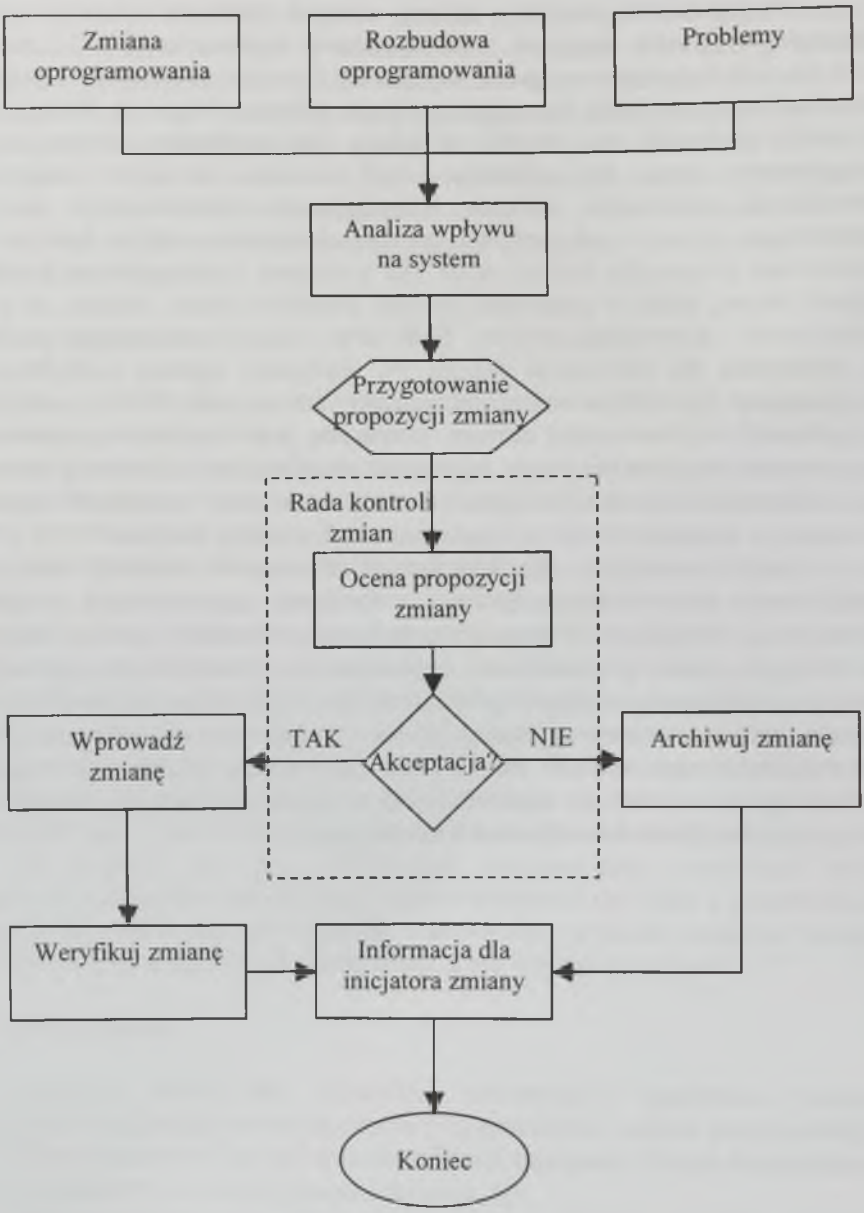
3.2. Kontrola zmian

Kontrola zmian jest najbardziej zewnętrznym przejawem procesu zarządzania konfiguracją polegającym na przygotowaniu, ocenie, przydzielaniu i implementacji propozycji zmian oraz aktualizacji związanej z nimi dokumentacji. Czynności związane z kontrolą zmian obejmują [5]:

- Definicja procesu wprowadzania zmian
- Ustalenie polityki i procedur zapewniających realizacje zdefiniowanego procesu
- Utrzymanie wersji bazowych
- Przetwarzanie propozycji zmian
- Opracowanie formularzy używanych do proponowania zmian
- Kontrola wprowadzania nowych dystrybucji systemu

Wprowadzenie dowolnej zmiany stanowi mini-cykl rozwoju systemu. Obejmuje, więc takie etapy jak analiza, projekt, implementacja czy testowanie. Cykl ten jest rozpoczynany przez najczęściej formalną propozycję modyfikacji funkcjonalności produktu lub naprawy błędu (Change Request). W pierwszym przypadku pochodzi ona zwykle od klienta lub analityków, natomiast błędy przedstawiane mogą być również przez testerów. Kolejnym etapem jest identyfikacja elementów systemu wymagających implementacji oraz opis koniecznych zmian i ich wpływu na funkcjonowanie całego systemu. Tak opracowana propozycja zmiany musi być następnie zaakceptowane przez radę kontroli zmian, która w przypadku małych projektów może składać się tylko z jednej osoby – kierownika projektu. Rada może zdecydować o przyjęciu zmiany, jej odrzuceniu lub odroczeniu decyzji. Po akceptacji, żądanie modyfikacji jest przypisywane do wykonawcy – programisty lub zespołu. W celu weryfikacji poprawności wprowadzonej zmiany konieczne jest dokładne przetestowanie poprawionego modułu oraz testy regresyjne całego systemu. Ostatnią czynnością jest ustawienie znacznika stabilności, a więc oznaczenie wszystkich elementów konfiguracji składających się na najnowszą stabilną wersję systemu.

Łatwo zauważyć, że kluczowym elementem kontroli zmian jest zdefiniowanie odpowiedniego procesu a następnie jego realizacja za pomocą procedur czy formularzy. W procesie kontroli zmian wyróżnić możemy etapy takie jak inicjacja, ocena, przydzielenie, implementacja i weryfikacja wprowadzonej zmiany, co schematycznie przedstawiono na rys. 3 [5]. Celem tak zdefiniowanego procesu jest zapewnienie wysokiej jakości oprogramowania poprzez sprawne wprowadzanie sugerowanych zmian oraz zapewnienie właściwego oznaczania zmienionych elementów, co stanowić będzie podstawę do śledzenia postępów prac nad projektem i kontroli kompletności konfiguracji.



Rys. 3. Ogólny proces kontroli zmian

3.3. Raportowanie statusu elementów

Celem generowanych raportów jest powiązanie wprowadzanych zmian ze zmodyfikowanymi elementami konfiguracji oraz rejestrowanie wszystkich zmian w wersji bazowej. Raportowanie statusu obejmuje takie czynności jak [5]:

- Określenie typu raportów i sposobu rejestrowania zmian. Typy i szczegółowość stosowanych raportów zależy przede wszystkim od wielkości projektu i potrzeb kierownictwa oraz ewentualnych wymogów formalnych, co powinno zostać sprecyzowane w planie zarządzania zmianą stworzonym na wczesnym etapie projektu.
- Śledzenie statusu elementów konfiguracji. Przykładowymi stanami elementów mogą być: nowy, przetestowany, zintegrowany (z resztą systemu). Dzięki temu programiści są w stanie określić czy możliwa jest już budowa/testowanie modułów współpracujących z danym elementem, natomiast kierownicy projektów są w stanie kontrolować postępy prac nad projektem
- Śledzenie statusu propozycji zmian umożliwiają kontrolę zgłoszonych poprawek do oprogramowania. Każda propozycja zmiany jest opisywana za pomocą takich atrybutów jak autor, opis zmiany, data wprowadzenia, status, który może przybierać takie wartości jak „oczekująca", „zaakceptowana", „anulowana", „przydzielona", „gotowa do testów", „przetestowana", „wdrożona". Bieżące aktualizowanie tych danych umożliwia sprawne raportowanie przebiegu procesu rozwoju oprogramowania, co jest podstawą sprawnego zarządzania projektem.
- Generowanie raportów przedstawiających postępy prac za pomocą ilości zmian oraz ich statusu w ujęciu zagregowanym. Porównanie liczby zaakceptowanych i wdrożonych zmian mówi, bowiem, o stabilności systemu oraz może być podstawą decyzji związanych z terminem wypuszczenia nowej dystrybucji systemu.
- Dokumentowanie czynności związanych z zarządzaniem zmianą. Najważniejsze jest tu śledzenie zawartości poszczególnych zmian, dzięki czemu możliwe jest stworzenie powiązania między elementami systemu a wprowadzonymi zmianami. Stanowi to później podstawę do budowy nowej wersji bazowej oraz umożliwia weryfikację kompletności konfiguracji.

3.4. Weryfikacja kompletności i poprawności konfiguracji

Weryfikacja ta polega na sprawdzeniu zgodności wytworzonego oprogramowania ze specyfikacją systemu, założeniami umowy czy przyjętymi standardami. Celem weryfikacji jest kontrola czy wszystkie elementy zostały poprawnie wdrożone, przetestowane i opisane, a wszystkie propozycje zmian uwzględnione. Weryfikacja ta następuję w dwóch fazach (niekoniecznie sekwencyjnych):

- Funkcjonalna weryfikacja konfiguracji sprawdzająca zgodność produktu z wymaganiami funkcjonalnymi. Dokonywana jest ona poprzez ocenę

stosowanych procedur, przegląd wyników testów, raportów oraz innych dokumentów powstałych w czasie realizacji projektu.

- Fizyczna weryfikacja konfiguracji, której celem jest ustalenie czy wszystkie elementy oprogramowania, danych i dokumentacji znajduję się rzeczywiście w wytworzonym produkcie i że są one w odpowiedniej wersji.

Weryfikacja kompletności i poprawności konfiguracji obejmuje czynności takie jak [5] definicja procedur weryfikacyjnych i harmonogramu prac, wyznaczenie osoby odpowiedzialnej za przeprowadzenie weryfikacji, weryfikacja określonych wersji bazowych oraz generowanie raportów z wynikami weryfikacji.

4. Zakończenie

Opisane w pracy elementy zarządzania konfiguracją oprogramowania pozwalają na uporządkowanie terminologii związanej z procesem nadzoru nad wprowadzanymi zmianami. Zarządzanie wersjami równoważne z identyfikacją konfiguracji, czy zarządzanie zmianami (kontrola zmian) nie są, więc tożsame z procesem zarządzania konfiguracją. Stanowią one jedynie wraz z raportowaniem statusu elementów i weryfikacją poprawności zbudowanego systemu jego elementy składowe. Na uwagę zasługuje również fakt, że opisywane składniki nie powinny być traktowane niezależnie od siebie. Wręcz przeciwnie powiązane są one silnymi zależnościami. Identyfikacja konfiguracji jest, bowiem warunkiem efektywnego wykonania pozostałych funkcji, kontrola zmian decyduje o sprawnym wprowadzaniu niezbędnych modyfikacji, a raportowanie postępów prac umożliwia wydajne zarządzanie projektem. Dzięki powyższym czynnościom oraz dokładnej weryfikacji dostarczony system jest zgodny z oczekiwaniami klienta natomiast cały proces wytwarzania oprogramowania może być łatwiej kontrolowany.

Literatura

1. Dart. S; "Spectrum of Functionality in Configuration Management Systems"; SEI Technical Report CMU/SEI-90-TR-11; December 1990.
2. „Guide to Software Configuration Management – ESA PSS-05-09"; ESA Board for Software Standardisation and Control; 1995.
3. IEEE-Std-729-1983 za B. Appleton; „SCM Definition"; http://www.enteract.com/~bradapp/acme/scm-defs.html
4. *"NASA Software Configuration Management Guidebook"*; National Aeronautics & Space Administration; 1995.
5. *"Software Configuration Management Technologies and Applications"*; Software Technology Support Center, Hill Air Force Base Utah; 1999.
6. Wingerd L., Seiwald C; "High-Level Best Practices in Software Configuration Management"; Springer Lecture Notes in Computer Science; No. 1439.
7. *Guide to Software Engineering Body of Knowledge*, IEEE 2001.

Mgr Krzysztof Górbiel
Szkoła Główna Handlowa w Warszawie, Katedra Informatyki Gospodarczej
Al. Niepodległości 162
e-mail: kgorbi@sgh.waw.pl

WYBRANE ASPEKTY POLITYKI BEZPIECZEŃSTWA INFORMACJI W E-PRZEDSIĘBIORSTWIE

Dorota JELONEK

Streszczenie: Polityka bezpieczeństwa definiuje podstawowe zasady ochrony informacji w przedsiębiorstwie. Celem artykułu jest przedstawienie istoty e-przedsiębiorstwa, wskazanie roli informacji w zarządzaniu, oraz identyfikacja zagrożeń bezpieczeństwa informacji. Stworzenie podstaw dla wdrożenia właściwych metod i procedur ochrony informacji jest podstawą kształtowania polityki bezpieczeństwa każdego e-przedsiębiorstwa.

... "Panowanie nad bezpieczeństwem technologii informatycznych stanie się kluczem do potęgi w XXI wieku"...

1. Istota e-przedsiębiorstwa

Wszystkie przedsięwzięcia biznesowe, które wykorzystują technologię Internetu określane są jako e-biznes (ang. e-business) lub biznes elektroniczny. Jest on transformacją kluczowych procesów biznesowych przy czym, rozważane są tutaj wszystkie procesy biznesowe nie tylko te odnoszące się do procesów sprzedaży i zakupu produktów lub usług.

W gospodarce elektronicznej można wyróżnić dwóch partnerów kooperacji: pojedynczego klienta (ang. consumer) lub całe przedsiębiorstwo (ang. business). Zatem ze względu na to, kim są partnerzy uczestniczących w transakcjach można wyróżnić cztery podstawowe modele e-biznesu:

- przedsiębiorstwo – klient (B2C),
- przedsiębiorstwo – przedsiębiorstwo (B2B),
- klient – klient (C2C),
- klient – przedsiębiorstwo (C2B).

Model typu przedsiębiorstwo – klient dotyczy przede wszystkim handlu elektronicznego, czyli sprzedaży bezpośredniej ostatecznemu odbiorcy. Może to być także świadczenie usług na odległość. Najczęściej usługi te mają charakter informacyjny, komunikacyjny lub związane są z telepracą. Wzrost zainteresowania tą formą elektronicznego biznesu obserwujemy zwłaszcza wśród małych i średnich przedsiębiorstw.

Relacje przedsiębiorstwo-przedsiębiorstwo dominują w e-biznesie i obejmują około 80% wszystkich zawieranych transakcji. Skuteczną podstawą realizacji współpracy przedsiębiorstw są systemy ekstranetowe. Dostęp do sieci mają tylko partnerzy biznesowi danego przedsiębiorstwa. Pozwala ona realizować wszystkie procesy biznesowe począwszy od prac nad projektem kontraktu, dyskusji nad

szczegółami, podpisaniem umowy kontraktu, składaniem zamówień na towar od dostawców, odbiór faktur i realizację należności finansowych. Zastosowanie rozwiązań ekstranetowych pozwala na:

- wymianę informacji z partnerami biznesowymi w formie elektronicznej,
- pomaga ograniczyć koszty komunikacji,
- pozwala udostępniać dealerom bazy danych i umożliwia składanie zamówień z wykorzystaniem elektronicznych formularzy,
- ułatwia zarządzanie e-przedsiębiorstwem o rozproszonym charakterze

Model klient – klient (ang. consumer to consumer), dotyczy transakcji zawieranych przez osoby prywatne. Najczęściej przejawami tego modelu są aukcje, drobne ogłoszenia oraz wymiana przedmiotów hobbistycznych.

W modelu klient-przedsiębiorstwo (ang. consumer to business) kierunek kooperacji jest skierowany od pojedynczego klienta do przedsiębiorstwa.

We wszystkich przedstawionych modelach jednym z podmiotów jest przedsiębiorstwo, które w tym sektorze zostało określone jako *e-przedsiębiorstwo* (szerzej [3]). E-przedsiębiorstwo musi doskonalić zasady współpracy z klientami oraz partnerami biznesowymi. Warunkiem skutecznej współpracy jest elastyczność jej zasad i umiejętność adaptowania ich do zmieniających się potrzeb. Podstawowym warunkiem prowadzenia przez e-przedsiębiorstwo działalności biznesowej jest szerokie stosowanie rozwiązań sieciowych i to nie tylko Internetu, ale także rozwiązań typu Ekstranet czy Intranet. Rozważając istotę e-przedsiębiorstwa należy podkreślić, iż przedsiębiorstwa te echuje dążenie do realizacji wszystkich procesów biznesowych drogą elektroniczną, zaś w działalności wykorzystują jeden z przedstawionych modeli biznesowych. Ponadto w szerokim zakresie korzystają z mutimedialnych możliwości sieci w realizacji kontaktów międzyludzkich.

W biznesie elektronicznym do realizacji kontaktów międzyludzkich wykorzystywane są te media elektroniczne, które zapewniają kontakt wzrokowy i głosowy. Pozwala to na wspólną pracę kilku osób np. nad projektem.

Kolejną cechą charakterystyczną e-przedsiębiorstwa jest wykorzystywanie dokumentów elektronicznych, czyli docelowe zastąpieniem wszystkich dokumentów papierowych, zarówno formalnych jak i nieformalnych dokumentami elektronicznymi i wykorzystaniu poczty elektronicznej do obrotu nimi. Dokumenty nieformalne w zależności od potrzeb mogą być przechowywane w postaci plików, lub publikowane na stronach WWW. Na dokumentach formalnych wymagany jest podpis bądź podpisy upoważnionych osób. W tym przypadku ważnym w świetle prawa jest podpis elektroniczny, który jest szczególną metodą szyfrowania. Celem szyfrowania nie jest utajnianie treści dokumentu, ale uwiarygodnienie nadawcy tego dokumentu.

Stosowanie elektronicznych dokumentów ułatwia ich rozpowszechnianie, gdyż znika bariera odległości, skraca się czas ich przesyłania, zaś koszty są minimalne. Kolejne zalety dokumentów elektronicznych to łatwość ich przetwarzania, modyfikowania, gromadzenia, wyszukiwania i archiwizowania.

Na rynku oprogramowania dostępnych jest wiele systemów zarządzania obiegiem dokumentów i zarządzania informacjami w całym przedsiębiorstwie.

Wykorzystanie ich pozwala zwiększyć szybkość działania i podejmowania trafnych decyzji biznesowych oraz pozyskać zaufanie klientów przedsiębiorstwa.

2. Znaczenie informacji dla e-przedsiębiorstwa

Informacja zawsze towarzyszyła procesom zarządzania e-przedsiębiorstwem jednak nigdy nie była tak ceniona jak to ma miejsce obecnie. Jest postrzegana jako podstawa i tworzywo tych procesów i cenny zasób przedsiębiorstwa. Rola informacji w zarządzaniu e-przedsiębiorstwem polega na redukcji niepewności, oszacowaniu wielkości ryzyka podejmowanych decyzji oraz ocenie szans i zagrożeń związanych z funkcjonowaniem przedsiębiorstwa na rynku. W sferze nauk o zarządzaniu informacja oznacza wiedzę potrzebną do osiągnięcia celów organizacji.

Dla potrzeb zarządzania e-przedsiębiorstwem niezbędne są wszechstronne informacje, obejmujące zarówno historyczną perspektywę na wyniki przedsiębiorstwa (informacje retrospektywne), oraz możliwość rzetelnej i wiarygodnej oceny stanu bieżącego jak również prognozowania zachowań procesów biznesowych w przyszłości (informacje prospektywne). O ich użyteczności decyduje fakt czy spełniają one podstawowe wymogi standardów jakościowych informacji. Jako najważniejsze cechy informacji przyjęto[4]: neutralność, istotność, wiarygodność, porównywalność i typ własności. Przez neutralność informacji rozumiemy jej niezależność i to w możliwie najszerszym obszarze. Standard istotności wymaga, aby informacja była znacząca lub użytecznie związana z działaniem, które powinno przynieść oczekiwane rezultaty. Na istotność informacji wpływa jej aktualność i zrozumiałość dla decydenta. Aktualność to dostosowanie informacji do czasu jej użytkowania. Informacja powinna być wiarygodna i wolna od przekłamań. Standard typu własności informacji, czyli określenie praw dostępu do niej jest szczególnie istotny dla każdego e-przedsiębiorstwa. Standard ten jest skorelowany z typem ochrony informacji, czyli ustaleniem sposobów, które utrudniają dostęp oraz ewentualną modyfikacją informacji przez osoby nieuprawnione. Zarówno typ własności jak i typ ochrony informacji nabierają coraz istotniejszego znaczenia w związku z dynamicznym rozwojem technologii informacyjnej, która zapewniając szeroki dostęp do informacji powoduje też narastanie konfliktów wokół ustalenia praw dostępu do niej.

3. Identyfikacja zagrożeń

Identyfikacja zagrożeń powinna obejmować zagrożenia wewnętrzne, zewnętrzne oraz uwzględniać zagrożenia spowodowane wypadkami losowymi. Wykryte zagrożenia można uporządkować według kilku kategorii (por. [5], [6]).

1. Zagrożenia fizyczne:

- kradzież lub fałszowanie informacji,
- kradzieżach sprzętu,

- kradzież nośników i oprogramowania,
- celowe niszczenie,
- wandalizm oraz działania typu cybergraffiti,
- oszustwa komputerowe

2. Zagrożenia technologiczne:
- awarie sprzętowe,
- awarie systemowe i błędy programów,
- blokadach działania systemów przez zawirusowanie lub wywołanie przeciążeń,

3. Zagrożenie związane z transmisją danych:
- celowy podsłuch,
- nieautoryzowany dostęp do systemu.

4. Zagrożenia ze strony użytkowników:
- pomyłki użytkowników
- akty zemsty byłych lub aktualnie zatrudnionych pracowników,
- wykorzystanie sprzętu lub oprogramowania niezgodnie z przeznaczeniem.

5. Zdarzenia losowe:
- żywioły - ogień, woda, pioruny.

Na etapie identyfikacji zagrożeń należy także ocenić na ile jest prawdopodobne wystąpienie danego zagrożenia oraz ustalić, jaki może być jego wpływ na działalność e-przedsiębiorstwa. Analiza ryzyka to systematyczny podział na kategorie zagrożeń danych i środków im przeciwdziałających oraz opracowanie planu działania, który większość zasobów (technicznych i pozatechnicznych) skieruje przeciw najbardziej prawdopodobnemu ryzyku. Celowe jest przede wszystkim uwzględnianie priorytetów zagrożeń. Podstawowym celem prowadzonej analizy ryzyka jest opracowanie planu ochrony, który powinien zapewnić stopień bezpieczeństwa proporcjonalny do wagi chronionej informacji.

Zagrożenia wewnętrzne związane są z oprogramowaniem, sprzętem komputerowym i ludźmi są charakterystyczne dla większości przedsiębiorstw. Dotyczą: zniszczenia, utraty, modyfikacji, kopiowani i zdradzenia treści [6].

E-przedsiębiorstwo jest szczególnie narażone na zagrożenia zewnętrzne, związane ze specyfiką ich działania i wykorzystaniem Internetu. Sieć ta postrzegana jest jako jedno z największych zagrożeń bezpieczeństwa w nowoczesnych przedsiębiorstwach. Zagrożenia zewnętrzne dotyczą [2]:

- autoryzacji (np. używanie przez jedną ze stron nieprawdziwej tożsamości całkowitego podawanie nieprawdziwych danych, trudności całkowitego identyfikacji klienta),
- dostępności (np. ataki z zewnątrz, których celem jest zablokowanie serwera i uniemożliwienie korzystania z niego przez innych użytkowników),
- płatności (np. regulowanie opłat z wykorzystaniem kradzionych kart kredytowych)
- ochrony prywatności (e-przedsiębiorstwo musi zagwarantować swoim klientom, że ich dane osobowe, a nawet adresy e-mail nie zostaną ujawnione osobom trzecim).

Osiągnięcie całkowitego bezpieczeństwa informacji zwłaszcza tych gromadzonych w systemach komputerowych jest w praktyce nieosiągalne. Związane jest to z faktami, iż systemy komputerowe często są systemami otwartymi i zaimplementowanie w nich stuprocentowego bezpieczeństwa mogłoby spowodować, że straciłyby swój charakter. Ponadto koszt wprowadzenia absolutnego bezpieczeństwa mógłby być tak wysoki, iż przerósłby wartość samego systemu.

Ocena bezpieczeństwa informacji w systemie komputerowym powinna ustalić i zidentyfikować luki i niedociągnięcia a następnie przyczynić się do podniesienia poziomu ochrony informacji przechowywanych i przetwarzanych w systemie informatycznym przedsiębiorstwa.

4. Elementy bezpieczeństwa w e-przedsiebiorstwie

Bezpieczeństwo informacji ma charakter złożony. Rozważania w tym zakresie mogą dotyczyć:

- istoty informacji i jej specyficznej postaci, często niejasnej i niezrozumiałej dla wielu osób. Pracownicy nie znając wagi informacji często nie są świadomi, że nastąpiła kradzież,
- systemów, w których informacja jest wytwarzana, przetwarzana, gromadzona i rozpowszechniana,
- środowiska, w którym te systemy działają. Należy ocenić w aspektach bezpieczeństwa pomieszczenia, okablowania czy nawet zasilania systemu,
- personelu, który korzysta z tych systemów i który często bywa niefrasobliwy, i nieobliczalny w swoich działaniach, łamiąc zasady bezpiecznego użytkowania systemów informatycznych,
- zgodności z uregulowaniami prawnymi.

Zapewnienie bezpieczeństwa informacji nie jest zatem trywialne a jednocześnie jest jednym z najważniejszych wymogów w działalności każdego e-przedsiębiorstwa. Stosowane zabezpieczenia można podzielić na dwie grupy [2]:

- narzędzia podstawowe,
- narzędzia zaawansowane

Stosowanie narzędzi podstawowych jest niezbędnym, ale niestety niewystarczającym działaniem. *Ochrona antywirusowa* gwarantuje uodpornienie systemu informatycznego na wirusowe infekcje, które mogą być zagrożeniem dla działalności e-przedsiębiorstwa. Stosowanie *urządzeń UPS* zabezpiecza system przed awariami zasilania. Bardzo ważne są też procedury związane z dostępem do serwerów tzw. *procedury bezpieczeństwa*. Obejmują hasła dostępu, regularne tworzenie kopii bezpieczeństwa, ograniczenie dostępu do serwera tylko do wybranych numerów IP. *Pojedyncza rejestracja* zapewnia prosty i bezpieczny dostęp do danych, natomiast strategię wielopoziomową zapewnia *autoryzacja i uwierzytelnienie*.

Do narzędzi zaawansowanych zaliczymy *zapory ogniowe* (ang. firewall), które zapewniają ochronę serwera i sieci przedsiębiorstwa przed atakami z zewnątrz. Oprogramowanie instalowane jest zazwyczaj na wydzielonvm serwerze

pełniącym rolę pośrednika między sieciami. Do jego podstawowych zadań należy monitorowanie wszystkich wchodzących i wychodzących informacji i odmawianie żądań dostępu uznanych za niebezpieczne. Większość systemów firewall dla zwiększenia skuteczności łączy w sobie dwie lub więcej różnych technik obronnych, z których najprostsze obejmują filtrowanie pakietów, stosowanie procedur identyfikacji użytkownika, uzupełnienie programów obsługujących pewne protokoły (np. FTP lub Telnet) o mechanizmy zabezpieczające lub całkowite wyłączenie wybranych usług

Efektywnym sposobem zabezpieczeń jest *szyfrowanie połączeń*. Pomocna jest Wirtualna sieć prywatna - VPN (ang. Virtual Private Network), która wykonuje szyfrowanie danych przed wysłaniem ich za pośrednictwem sieci publicznej. W celu zapewnienia poufności przesyłanych danych stosuje się *szyfrowanie*, czyli utajnienie wiadomości w taki sposób, aby niemożliwe stało się jej odczytanie przez osoby postronne. Podstawowe metody wykorzystują szyfrowanie z kluczem symetrycznym lub z kluczem asymetrycznym (więcej w [6]). Zastosowanie metod kryptograficznych pozwala na [2]:

- zachowanie tajności,
- uwierzytelnienie (autentyfikację) nadawcy,
- zachowanie integralności przesyłki.

Weryfikacja źródła informacji pozwala na jednoznaczne uwierzytelnienie nadawcy, natomiast zachowanie integralności daje nam pewność, że nikt nie ingerował w treść wiadomości. Metody kryptografii niesymetrycznej wykorzystano w *podpisie elektroniczny*.

Gwarancje bezpieczeństwa transakcji w środowisku biznesowym uzyskami dzięki wykorzystaniu *certyfikatów*. Jest on potwierdzeniem tożsamości danej firmy. Klienci łącząc się z danym serwisem są pewni, że połączyli się z właściwą firmą i nikt się pod nią nie podszywa. Certyfikat pozwala także firmie na identyfikację klienta, użytkownika Internetu. Uzyskać certyfikat można w upoważnionej do tego celu firmie np. Centrum Certyfikacji Certum; http://www.certum.pl

W zależności od tego, do jakich celów wykorzystuje się certyfikaty możemy je podzielić na [Cyfrowy...

- certyfikaty urzędów certyfikacji - potwierdzają tożsamość Centra Certyfikacji,
- certyfikaty serwerów - uwiarygodniające serwery – użytkownik ma pewność, że połączył się z właściwym serwerem,
- certyfikaty osobiste – wykorzystywane np. do przesyłania poczty elektronicznej. certyfikat gwarantuje, że wiadomość nie została zmieniona przez osoby trzecie i potwierdza tożsamość nadawcy.
- certyfikaty producentów oprogramowania służące cyfrowemu podpisywaniu oprogramowania np. rozprowadzanego przez Internet.

Certyfikaty pozwalają firmom zawierającym transakcje na drodze elektronicznej zyskać pewność, że jest to transakcja bezpieczna i zgodna z prawem natomiast partner jest tym za kogo się podaje.

Bezpieczeństwo danych to także ich ochrona przed utratą z przyczyn losowych. Okresowo wykonywane kopie zapasowe danych (backup) oraz archiwizo-

wanie wybranych zasobów sieciowych pozwala odtwarzać dane uszkodzone lub usunięte z powodu awarii systemu lub błędu użytkownika.

Podsumowanie

Problematyka bezpieczeństwa ma charakter złożony i interdyscyplinarny. Jej aspekty: techniczny, fizyczny, organizacyjny, osobowy, kulturowy i prawny są ze sobą wzajemnie powiązane. Dla e-przedsiębiorstwa środowisko zagrożeń cechuje duża zmienność i trudne jest jego rozpoznanie i identyfikacja. Zapewnienie bezpieczeństwa danych w systemach informatycznych powinno być sumą rozwiązań organizacyjnych i technicznych. Należy zatem podjąć działania ewidencyjne i uporządkować zasoby (informacje, oprogramowanie, środki techniczne, infrastrukturę, personel, itp.) oraz ustalić zasoby informacji wrażliwych, które należy chronić. W ocenie wartości pewnych zasobów pomocne jest wykorzystanie metody "co się stanie jeśli...".Następnie należy przeprowadzić identyfikację i analizę zagrożeń. Uzyskane dane pozwolą ustalić cele polityki bezpieczeństwa e-przedsiębiorstwa. Zachowując właściwe relacje między wartością dóbr chronionych i kosztem zabezpieczeń można ustalić metody, techniki oraz środki zabezpieczające poszczególne zasoby informacyjne. Kolejny etap wdrażania systemu zabezpieczeń powinien dotyczyć odpowiednich zakupów i podjęcia niezbędnych czynności techniczno-organizacyjnych. Działania organizacyjne obejmują przeszkolenie personelu i kształtowanie odpowiedzialnej i wrażliwej postawy wobec pojawiających się zadrożeń bezpieczeństwa danych e-przedsiębiorstwa. Osiągnięty stan bezpieczeństwa wymaga ciągłego nadzoru i prac doskonalących będących odpowiedzią na nowe zagrożenia. Zapewnienie bezpieczeństwa pozwoli e-przedsiębiortwu bez przeszkód realizować ustalone cele i zadania. .

Literatura

1. Cyfrowy certyfikat, BizReporter z 5 kwietnia 2001, nr 4, http://biz.reporter.pl/2001/04/b601.htm
2. Gregor B., Stawiszyński M.: E-commerce, Oficyna wydawnicza Branta, Bydgoszcz- Łódź, 2002.
3. Jelonek D. Przewaga konkurencyjna e-przedsiębiorstwa. w: Ekonomika i Organizacja Przedsiębiorstwa, nr 3, marzec 2003.
4. Jelonek D.: Systemy komputerowego wspomagania monitorowania otoczenia przedsiębiorstwa, Wydawnictwo Wydziału Zarządzania Politechniki Częstochowskiej, Częstochowa 2002.
5. Klisz L., BanasikowskaJ.: Bezpieczeństwo transferu danych w wirtualnej organizacji, w: Systemy Wspomagania Organizacji SWO'2000, Praca zbiorowa pod red. Gołuchowskiego J. i Sroki H., Wydawnictwo Akademii Ekonomicznej w Katowicach, 2000.

6. Wstęp do systemów informacyjnych zarządzania w przedsiębiorstwie, praca zbiorowa pod redakcją A. Nowickiego, Wydawnictwo Politechniki Częstochowskiej, Częstochowa 2002.

dr Dorota Jelonek
Wydział Zarządzania
Politechnika Częstochowska
Al. Armii Krajowej 19b
42-200 Częstochowa
tel. (0-34) 32-50-391, 32-50-388
e-mail: jelonek@zim.pcz.czest.pl

THE ELECTRONIC REPORTS ARCHIVING IN THE REMOTE STUDY PROCESS

BARBARA KRÓLIKOWSKA

Summary: The example of the electronic archiving in the remote study can be: COLD (Computer Output to Laser Disc) – it processes data streams, which are sent to printer, it indexes and compresses them as a digital file on a magnetic disc; COOL (Computer Output On-line) – it builds the knowledge base by the connection of electronic reports with source documents. It secures an access to documents on a single user or group of users level. It enables an enclosing of notes to reports with an access control.

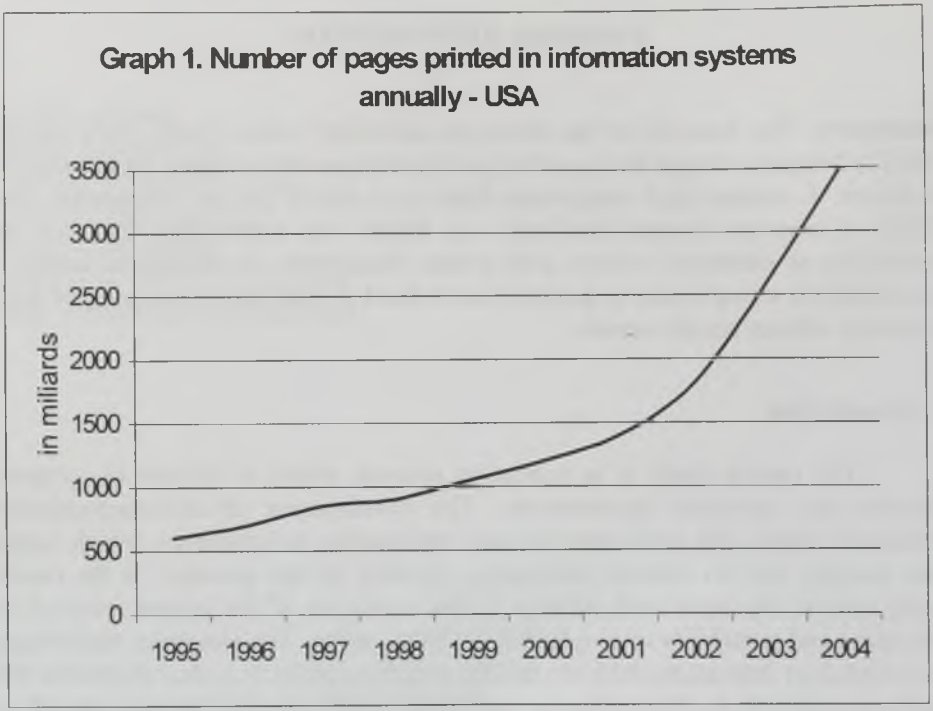
1. Introduction

The remote study is an education process, which is realised in computer systems and networks environments. The development of telecommunication techniques makes the application of new information technologies, which supply data security and its correct processing, possible in this process. In the remote study process the additional element is the assurance of the proper level of the education and possibility of acquired knowledge rating. The electronic archiving is the technology that assists both the teacher and the student. It makes electronic data exchange possible by the replacing traditional handbooks and lectures run off on the duplicator with the electronic library. Reports in the form of credit lists, semester records, checking tests can be stored in an electronic form and can be checked and printed when it is required. The electronic archiving has an effect on the reduction of education costs. The profitability of the usage of this technology increases with the development of computer networks, spread of Internet and the extension of the capacity of external memory carriers[1].

2. COLD Technology

The traditional way of a data presentation and data storage in a paper form is ineffective. Gartner Consulting Group from Stanford has published that about 600 milliards paper pages was produced in American companies in 1995. It cost approximately 300 milliards USD. Number of printed pages increases constantly (graph 1). Logical consequence of this is growing costs of information systems use. Costs connected with printing are one of the main categories of costs in an organisation. It is estimated that 60% of reports are conscientiously stored and used incidentally, 30% are put on microfilms in COM technology (Computer Output on Microfilm Technology). Microfilms just as paper media can be easily damaged.

The way of a restoring is also a problem in this environment. Remaining 10% of printouts simply get lost [5].



The usage of paper reports raises costs of the process. According to Coopers & Lybrand 19 copies of a paper report are made in a medium size firm. One from twenty disappears. Search for the lost document costs on the average 120 USD and restoring it - twice as much.

In the teaching process paper documents: checking tests, semester records, materials use during lectures and classes are essential documents. The cyclical generation of these documents takes teachers a lot of time. The administrative department that serves the teaching process has difficult task related to efficient control and service of realized operations. Considering a value of this information in teaching, COLD technology secures an efficient communication and control of data.

This technology made data stream processing which are sent to printer possible in range:

- indexing,
- data compressing,
- storing on the magnetic medium in a digital file form.

Archiving is realized on WORM or CCW (Continuous Composite WORM) optical discs. These discs guarantee persistence and security of recording. In system, that serves the teaching process, the access to data should be secured to make changes

in marks or in checking tests impossible. The COLD technology has an enlarged range of protection, which assigns an access to select pages or single lines of printout. Data can be protected even against a system administrator. The suitable security system enables the identification of user who has an access to data and who can make changes. Stored documents are secure in this situation.

The COLD system reflects data in time. This is statistical representation of this what happened in a particular moment. Exam records and semester projects made by students can be the example of this.

Looking through traditional records is very strenuous. The COLD technology optimises access to archived information. Reports are indexed. Thanks to it looking for them is quick and easy. Thanks to indexes it is possible to search common data, for example: marks of different subjects for one student or examination session results. The COLD technology makes electronic data exchange possible, for example: STUDENT – DEANERY, TEACHER – STUDENT.

Data in an electronic form enable generation of result lists, semester reports without using the archive. The student-friendly teaching process means advantages that are hard to calculate but are very important in a school process of working.

The COLD technology can be implement in the client/server environment or mainframe environment. The number of COLD applications grows up most quickly in client/server environment. According to M. Grisby from Imerge Consulting (graph 2) client/server has a large calibration, indexing possibility and open system architecture.

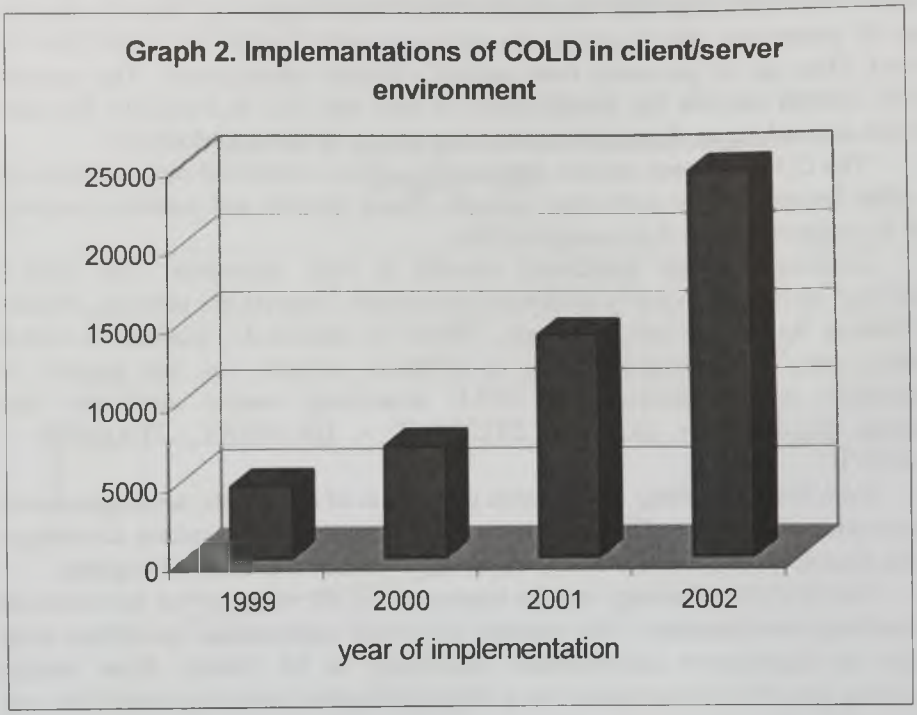
The open client/server architecture makes possible:

- generation of electronic reports from different applications,
- service of organisations that are geographical distracted,
- service of many users in organisation.

The COLD technology serves also Internet environment as source of data: e-mail files, www pages and electronic transaction.

The remote study process, which uses the COLD technology, can be supported by On-Line Analytical Processing application (OLAP). The result of an analysis initialises a suitable action, for example: an urging e-mail to the student with information about an overdue payment or exam.

The COLD solution enables a use of the existing net in the organisation. It enables also connection of users in any environment: WINDOWS NT, UNIX independently from hardware platform (figure 1).



The COLD technology brings following advantages (on the ground of [5]):

- savings by replacing paper and microfilms by magnetic media,
- possibility of finding information in reports,
- on-line access for many distributed users,
- print on demand any report or its part,
- integration with Intranet applications that enables data analysis and building report warehouses,
- electronic reports ca be input data for other information systems,
- terminals don't need special software for access to COLD reports; this access is secured by Internet viewer,
- possibility of processing in night hours,
- savings of space on servers.

The main condition of the realisation of the remote study is an information system in the network environment that assists students and teachers. The school administration demands first of all obtainment of required information in order to warrant efficient service for the deanery. The new generation of the COLD system is the COOL technology, which secures knowledge management in an organisation

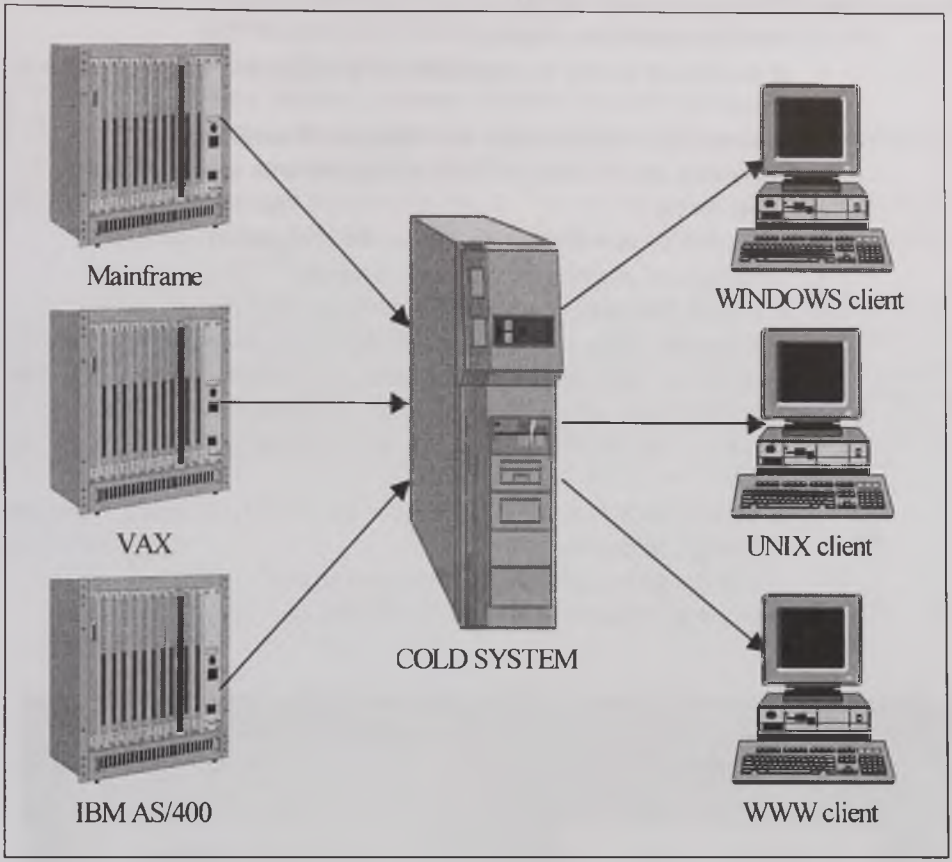


Figure 1 The COLD system environment (on the ground of [5])

3. COOL Technology

Knowledge known as knowledge of information connections , is indispensable in a service of the remote teaching. A number of information increases very fast. It is rated that this number will double to year 2020 every 73 days [4]. The knowledge Management (KM) is the element that determines success[3]. The COOL system can be classified as KM technology.

The COOL system uses the client/server technology independently from hardware platform to build a knowledge archive. During producing electronic reports it makes relations between them and source documents. User interface should work in Graphical User Interface (GUI) mode.

Range of the COOL system activity [4]:

- it makes electronic reports,
- it secures an access to documents on a single user or group of users level,
- it keeps a journal of events and changes of restrictions,
- it enables an enclosing of notes to reports with restricted access for other users,
- it renders source documents connected with report accessible,
- it indexes in accordance with user's needs,
- it finds information coming form different reports,
- it integrates data warehouses and WFM, for example: on base of analysis of documents of payment the system discovered, that student X, who always paid on time, is overdue with payment for 3 weeks – the result of analysis will initiate suitable activity of WFM.
- it serves such formats of printouts as: AFP (Advanced Function Printing), Metacode,
- it enables taking advantage of Internet stores.

The COOL system consists of 4 modules (figure 2).

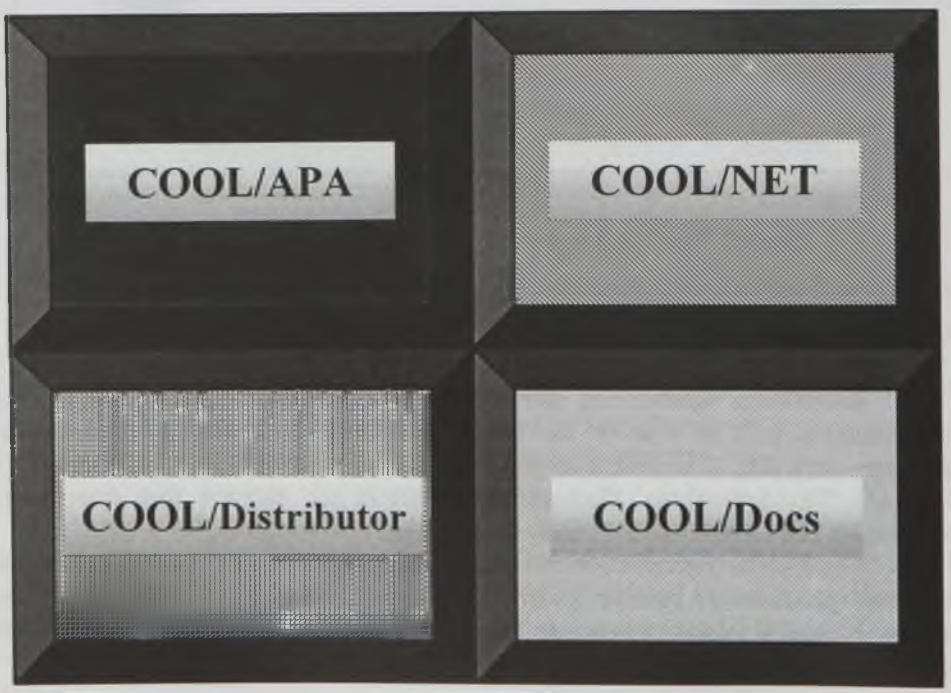


Figure 2.The COOL system structure (on the ground of [4]).

COOL/APA – realizes functions connected with service of All Points Addressable (APA) files, which enable use of forms during looking through a report,

COOL/Net – throws Internet, Intranet, Extranet supplies open,

COOL/Distributor –serves electronic memorization of reports and throws data coming from different sources open to the end user,

COOL/Docs – manages documents and throws documents from any environment open: reports, OLAP analyses, EDI standard, Internet environment.

Chicago Merchant Exchange (CME) has used the COOL technology since 1994. Everyday from 900 megabytes to several gigabytes of data are archived there. Information requirements are realised practically at once. Another example of the efficiency of proposing electronic archiving technologies is EVEREN Capital Corporation (a broker company from Chicago) that has been using COLD and COOL since 1996. There are 550 users in 9 branches who use them.

4.Conclusions

The documents' structure and scheme of their transmission is less complicated in the distance learning process than in the financial environment.

The possibility of application of automatic processing – OLAP (On Line Analytical Processing) in COLD and COOL technologies is worthy of attention. It secures a complete analysis of student's work results. It enables mention the student of the research area for his/her thesis, on the base of achieved results of study.

It will simplify preparing students' files before the defence of a thesis. OLAP can be helpful for the teacher in preparing classes or lectures and materials for them.

The important problem in the distance learning process is the control of a student knowledge. The electronic archiving can enable keeping a record of tests and analysing them considering a compatibility of different subjects.

The board of curriculum, supported by the electronic archiving, will prepare optimal program of study. The incoherent program contents of proposed subjects will be avoided in this way.

The data security is a very important element in the remote learning process. Cryptographic algorithms, data replication and access groups guarantee the security.

The electronic archiving technology secures also a proper co-operation with Internet in the remote learning process. This enables expanding of the teaching process on any geographical regions. This will also secure disabled students a connection with school. In the future it will enable a packet transmission in the GPRS technology.

The electronic archiving is necessary in an organisation, which has the workflow management system (WMF). At university or other schools the WFM

system solves the problem of data archiving– exams' records, thesis, students' files. This mass data storage is effective and secure. Stored data can be quickly searched.

The electronic archiving brings a lot of advantages, but sometimes it is complicated and expensive. Often it brings also some negative experience [2], because there isn't used a proper archiving technology in an organisation. The electronic archiving is indispensable in the distance learning process The COOL is the most suitable technology for this environment.

Bibliography

1. Królikowska B., Technologie informatyczne stosowane w systemach wspomagania decyzji, Zeszyty Naukowe, STUDIA INFORMATICA nr 12/2000, US SZCZECIN
2. Maślana R., Blaski i cienie elektronicznej archiwizacji, INFORMATYKA nr 12/1998
3. Mertens P., Wieczorek H.W., Data X-Strategien, Springer-Verlag Berlin-Heidelberg, 2000
4. COOL – Computer Output On Line, Computron POLAND, 2000
5. COLD – Computer Output to Laser Disc, Computron POLAND, 2000

Dr Barbara Królikowska
Instytut Informatyki w Zarządzaniu Uniwersytet Szczeciński
Ul. Mickiewicza 64, Tel.(091)444-19-15
e-mail: bkrol@uoo.univ.szczecin.pl

ROZWIĄZANIA BUSINESS CONTINUITY I DISASTER RECOVERY - CIĄGŁOŚĆ PROWADZENIA BIZNESU I ZABEZPIECZANIE ZASOBÓW DANYCH

JERZY KUROWSKI

Streszczenie: W artykule omówiona jest problematyka *Business Continuity* i *Disaster Recovery*. Na wstępie wyjaśniono podstawową terminologię z tego zakresu. W dalszej części przedstawione zostały wiodące pakiety do planowania (*Business Continuity Planning*) oraz wybrane rozwiązania organizacyjne, sprzętowe i programowe do zabezpieczania i odtwarzania zasobów informacyjnych przedsiębiorstw.

Wstęp

Nie jest obecnie możliwe prowadzenie poważnej działalności gospodarczej bez wykorzystywania nowoczesnych metod gromadzenia, przetwarzania i analizy informacji. W każdym przedsiębiorstwie widoczne jest powiązanie procesów biznesowych z technologiami oraz systemami informacyjnymi. Coraz więcej aplikacji funkcjonuje w trybie 24 x 365 (24 godziny na dobę przez wszystkie dni roku). Można tu przykładowo wymienić systemy: handlu elektronicznego (*e-commerce*), rezerwacji biletów, centrów obsługi klientów (*Call Centers*), firm transportowych, energetyki, centrów autoryzacji kart płatniczych itd. W większości przypadków dopuszczalne przerwy w ciągłości pracy wynoszą kilka godzin.

Postęp w zakresie technologii zapewnił możliwość zakupu sprzętu komputerowego o bardzo wysokiej niezawodności. Jednak posiadanie nawet najlepszych serwerów i aplikacji oraz zespołu świetnych specjalistów nie zapewni firmie gwarancji, że obsługujący podstawowe procesy biznesowe system informacyjny będzie funkcjonował w sposób ciągły. Ze względu na występowanie klęsk żywiołowych, aktów terroryzmu oraz awarii systemów komputerowych przedsiębiorstwa coraz bardziej doceniają potrzebę przygotowania się do takich zdarzeń.

Zapewnienie ciągłości pracy w wielu przypadkach jest warunkiem dalszego istnienia firmy i utrzymania się jej na rynku.

Dla zapewnienia firmie wymaganego poziomu ciągłości prowadzenia biznesu niezbędne jest posiadanie odpowiednich technologii informacyjnych i rozwiązań klasy ***Business Continuity***, których podstawowym zadaniem jest pomoc w zapewnieniu możliwie wysokiego poziomu bezpieczeństwa eksploatacji krytycznych systemów informacyjnych.

Taka funkcjonalność uzyskiwana jest poprzez m. in.:

- efektywne zarządzanie infrastrukturą IT,
- ochronę systemów pamięci przed utratą zgromadzonych danych,
- zabezpieczenie zasobów sieci przedsiębiorstwa przed atakami hackerów

i wirusami komputerowymi.

Musimy wiedzieć, co nam grozi i jak się przygotować na wypadek poważnej awarii lub katastrofy. Zwykle celem jest maksymalne zabezpieczenie się przed awarią, a w przypadku katastrofy możliwie szybkie odtworzenie zasobów niezbędnych do wznowienia pracy podstawowych systemów informacyjnych.

Spróbujmy zdefiniować podstawowe pojęcia, które będą występowały w dalszej części tekstu.

1. Podstawowa terminologia

Business Continuity opisuje procesy i procedury wprowadzone przez organizację dla zapewnienia kontynuacji istotnych funkcji biznesowych podczas i po katastrofie.

Z kolei dziedzina planowania ciągłości pracy (***Business Continuity Planning***) poszukuje:

- sposobów uniknięcia przerw w realizacji krytycznych usług,
- metod sprawnego przywrócenia pełnego funkcjonowania w możliwie krótkim czasie po wystąpieniu awarii lub katastrofy.

Analiza zagrożeń (***Risk Analysis***) identyfikuje najważniejsze funkcje i zasoby niezbędne do funkcjonowania przedsiębiorstwa, a następnie określa prawdopodobieństwo ich awarii lub utraty. Po określeniu ryzyka definiowane są cele oraz strategie dotyczące uniknięcia określonych zagrożeń i minimalizacji wpływu zdarzeń losowych. W ramach analizy należy najpierw przygotować listę krytycznych funkcji biznesowych oraz zasobów wraz z przypisanymi im priorytetami. Następnie określa się prawdopodobieństwo wpływu specyficznych zagrożeń na istotne funkcje i zasoby biznesowe.

Nazwą ***Disaster Recovery*** określa się całość działań służących do przywrócenia ciągłości funkcjonowania przedsiębiorstwa czy centrum obliczeniowego po wystąpieniu katastrofy (poważnej awarii sprzętu, klęski żywiołowej, ataku terrorystycznego itp.). Plan *Disaster Recovery* to ukierunkowany na zasoby, systemy i technologie IT plan przygotowany dla przywrócenia operatywności określonych systemów, aplikacji czy komputerów (często w alternatywnej lokalizacji - w tzw. centrum zapasowym) po wystąpieniu nagłego zdarzenia (awarii lub katastrofy). Plan odnosi się do zdarzeń katastroficznych, których wystąpienie uniemożliwia przez dłuższy czas dostęp do normalnych lokalizacji i systemów informacyjnych. Obejmuje on analizę procesów biznesowych i potrzeb z zakresu zapewnienia ciągłości ich prowadzenia. Funkcja *Disaster Recovery* realizuje więc część zadań należących do dziedziny *Business Continuity*.

2. Zagrożenia dla ciągłości pracy systemów informacyjnych

Tragiczne wydarzenia niedalekiej przeszłości spowodowały większe zainteresowanie organizacji i przedsiębiorstw zapewnieniem sobie ochrony na

wypadek wystąpienia zagrożeń i w związku z tym wdrożeniem skutecznych technologii i produktów z zakresu *Business Continuity*.

Prawie wszystkie duże światowe organizacje i firmy mogą być narażone na wszystkie (lub większość) możliwych zagrożeń.

1/ awarie związanie z zasobami, infrastrukturą lub systemami informacyjnymi:
- awarie sprzętu (np. serwery, pamięci, sieć),
- usterki oprogramowania systemowego,
- błędy aplikacji,
- błędy obsługi,
- wirusy komputerowe,

2/ katastrofy naturalne (np. powodzie, trzęsienia ziemi),

3/ katastrofy spowodowane przez inne czynniki (np. pożary, ataki terrorystyczne).

Według niedawno opublikowanego raportu firmy Sungard opracowanego na podstawie ankiet dotyczących okresu ostatnich 5 lat deklarowane przyczyny „katastrof" były następujące:
- 56 % - awarie sprzętu i oprogramowania oraz systemów zasilania i telekomunikacji,
- 24 % - katastrofy naturalne (pożar, powódź, trzęsienie ziemi itd.),
- 20 % - katastrofy spowodowane przez działanie człowieka (np. ataki terrorystyczne – np. 11 września 2001...).

Tradycyjnie uważano problem utrzymania ciągłości pracy systemów informacyjnych wyłącznie za zadanie działów IT. Aktualnie ze względu na złożoność procesów biznesowych, powszechność technologii przetwarzania rozproszonego oraz coraz popularniejsze wprowadzanie rozwiązań biznesu elektronicznego całe kierownictwo firmy oraz wszyscy pracownicy muszą uczestniczyć w planowaniu, budowie i utrzymywaniu systemów dla *Business Continuity*. Kierownictwo przedsiębiorstwa czy organizacji (na czele z Prezesem lub Dyrektorem Generalnym) jest odpowiedzialne za konsekwencje przerw w prowadzeniu działalności oraz za ewentualną utratę ważnych zasobów. Coraz częściej wprowadzane są rygorystyczne przepisy dotyczące bezpieczeństwa informacji (np. Ustawa o ochronie danych osobowych). Czasem dotyczą one określonych użytkowników informatyki (np. sektora administracji czy instytucji finansowych).

Największe potrzeby w zakresie zapewnienia ciągłości operacyjnej wykazują: agendy rządowe, instytucje finansowe (w tym banki) i ubezpieczeniowe, firmy telekomunikacyjne, producenci i dostawcy energii oraz podmioty z sektora służby zdrowia. Do tego grona w ostatnich latach dołączyły firmy kategorii ISP i ASP (dostawcy Internetu oraz usług aplikacyjnych) oraz branża szeroko rozumianego handlu i biznesu elektronicznego. Jednak potrzeba przygotowania planu na wypadek katastrofy lub poważnej awarii dotyczy właściwie wszystkich firm przynajmniej średniej wielkości.

Koszty braku ciągłości funkcjonowania przedsiębiorstwa są różne dla różnych sektorów. Ocenia się (według badań z rynku amerykańskiego), że największe straty za godzinę przestoju występują w następujących branżach:
- usługi brokerskie,

- firmy energetyczne,
- obsługa kart kredytowych,
- firmy telekomunikacyjne,
- przemysł,
- usługi finansowe,
- ubezpieczenia,
- handel.

Dla zapewnienia ciągłości procesów biznesowych niezbędna jest ciągłość funkcjonowania infrastruktury informatycznej przedsiębiorstwa (komputerów, sieci, systemów, aplikacji). Należy także pamiętać, że jedną z podstawowych wartości każdej firmy są zgromadzone w jej systemach informacyjnych zasoby baz danych. Jednym z krytycznych elementów ochrony przedsiębiorstw przed katastrofami dotykającymi przedsiębiorstw lub ataków na ich systemy informacyjne i sieci jest możliwość odzyskania danych i aplikacji. Rozwiązania z zakresu zarządzania systemami pamięci zapewniają szeroki zakres funkcji i mechanizmów, które umożliwiają organizacjom odzyskanie i odtworzenie danych oraz aplikacji (stanu kluczowych systemów biznesowych i administracyjnych). Systemy zarządzania zasobami pamięci wykorzystywane są także przy tworzeniu i eksploatacji zapasowych (zwykle zdalnych) centrów przetwarzania danych zapewniających dostęp do danych lub kompletnego środowiska aplikacyjnego w przypadku katastrofy lub poważnej awarii podstawowych zasobów struktury informacyjnej przedsiębiorstwa. Oprogramowanie do zarządzania systemami pamięci obejmuje m. in. produkty służące do: replikacji oraz wirtualizacji danych, wspomagania systemów macierzy dyskowych i bibliotek taśmowych, obsługi procedur zabezpieczania i odtwarzania (***Backup / Restore***) oraz archiwizacji danych, zarządzania sieciami pamięci (SAN, NAS).

Badania firmy Gartner pokazały, że średnio tylko 4% wartości budżetu centrów przetwarzania danych przeznaczanych jest na cel związany bezpośrednio z tematem *Disaster Recovery*. Zwykle ochrona ośrodka obliczeniowego jest włączana do kompleksowego planu *Business Continuity* przedsiębiorstwa czy organizacji, w ramach których funkcjonuje centrum danych. Wiele firm posiada plany działania na wypadek wystąpienia zdarzeń, których się spodziewały (braki w zaopatrzeniu, czasowa niedostępność systemów komputerowych). Plany takie obejmują zwykle tylko statyczny wizerunek przedsiębiorstwa. W niewielu przypadkach firmy dysponują skutecznym planem odnoszącym się do wszystkich istotnych zagrożeń oraz miejsc ich wystąpienia. Często istniejące w przedsiębiorstwie plany są nieaktualne i nie mogą być skutecznie wykorzystane w przypadku wystąpienia rzeczywistego zagrożenia.

3. Wpływ e-Biznesu i sieci WWW na technologie Business Continuity oraz Disater Recovery

Firma IDC (zajmująca się doradztwem oraz prowadzeniem analiz) szacuje, że w roku 2004 ilość użytkowników Internetu na świecie przekroczy 877 mln.

Mają oni w przyszłym roku w ciągu każdej minuty wydawać 3.1 mln USD na zamawiane poprzez globalną Sieć towary i usługi. Te liczby świadczą o skali handlu elektronicznego.

Wymagania wobec ciągłości obsługi w rozwiązaniach bazujących na technologiach sieci Internet są jeszcze wyższe i bardziej złożone niż w przypadku systemów tradycyjnych. Spadek wydajności serwera WWW (czy portalu Internetowego) lub jego awaria powodują znaczne zmniejszenie obrotów oraz masowe odejścia klientów do konkurencji. Klasyczne plany *Business Continuity* nie obejmowały zagrożeń związanych z pracą w sieci WWW - koncentrowały się głównie na problemach scentralizowanych centrów danych (nie były odpowiednie dla środowisk sieciowych i przetwarzania rozproszonego).

Nowymi i bardzo istotnymi zagrożeniami związanymi z pracą w sieci Internet są próby wtargnięcia intruzów do zasobów sieci oraz rozsyłane wirusy komputerowe. Takie działania stanowią coraz poważniejsze zagrożenie dla ciągłości pracy systemów. Centrum CERT/CC gromadzi i publikuje statystyki dotyczące takich incydentów. W latach 1998 – 2002 łącznie zarejestrowana 182 463 przypadki. Ich ilość bardzo szybko rosła i wynosiła kolejno: 3734, 9859, 21756, 52658, 82094.

W okresie ostatnich kilku lat dał się zauważyć silny wpływ wprowadzania rozwiązań biznesu elektronicznego na programy i technologie wykorzystywane w procedurach zapewnienia ciągłości operacyjnej przedsiębiorstw oraz centrów przetwarzania danych. Eksploatacja systemów E-biznesu i powszechne korzystanie z sieci w sposób istotny zwiększyły poziom ryzyka wystąpienia katastrof czy też istotnych awarii. Firmy, które oferują aplikacje pracujące 24 godziny na dobę w sieci Internet muszą zapewnić sobie maksymalny poziom *Business Continuity*. W tych przypadkach dopuszczalny czas przestoju liczony jest nie w godzinach, ale w minutach. Te wszystkie czynniki zwiększają także wymagania wobec oprogramowania służącego do wspomagania procesów zabezpieczeń przed utratą danych oraz osób odpowiedzialnych za planowanie i realizację funkcji zapewnienia ciągłości prowadzenia biznesu. Niezbędna jest dodatkowa wiedza z dziedziny technologii sieci komputerowych oraz aplikacji klasy *Web-enabled*. Analizy przeprowadzone w wielu firmach wykazały, że ze względu na wprowadzenie w nich rozwiązań biznesu elektronicznego muszą one w szybkim tempie dokonać istotnych modyfikacji w posiadanych planach postępowania na wypadek katastrof.

Ze względu na fakt przepływu w systemach E-biznesu w krótkim czasie wielkich ilości danych (np. realizacji tysięcy transakcji na sekundę) w przypadku awarii lub katastrofy następują znacznie większe straty niż w systemach wykorzystujących wcześniejsze technologie. Według raportu ogłoszonego w USA przez firmę Strategic Research Corp. koszt jednej godziny przestoju w firmie należącej do sektora usług brokerskich może wynieść ok. 6.5 mln USD, a w przypadku awarii systemu autoryzacji kart płatniczych ok. 2.6 mln USD. Ta sama firma oceniła straty wynikające z godzinnej awarii systemu obsługującego bankomaty na ok. 14 500 USD dla każdego „niesprawnego" wówczas urządzenia ATM.

Koszt zapewnienia dostępności (niezawodności) dla przedsiębiorstwa dramatycznie rośnie jeżeli wymagamy poziom niezawodności zbliżony do 100 %. W takim przypadku musi być gwarantowana dostępność krytycznych systemów i sieci niezależnie od możliwych zdarzeń. Problemem jest zdefiniowanie stopnia równowagi pomiędzy względnym kosztem przestoju oraz kosztem zapewnienia dostępności krytycznych procesów biznesowych.

Dla uzyskania wysokiego poziomu ciągłości prowadzonych operacji biznesowych firmy i instytucje zwykle podpisują umowy z wiodącymi dostawcami technologii i usług informatycznych.

4. Międzynarodowe standardy dotyczące ciągłości operacyjnej

A/ norma ISO 9002 - standard jakości dla firm zajmujących się produkcją oraz usługami; Dotyczy ona m. in. ochrony w systemach informacyjnych oraz szerzej rozumianego bezpieczeństwa produkcji - w tym także planowania ciągłości systemów IT przedsiębiorstwa.

B/ ISO/IEC 17799 : 2000 (opublikowany w grudniu 2000 r.) "*2000 Information Technology - Code of practice for information security management*". Jedna z sekcji dokumentu (11) dotyczy bezpośrednio zarządzaniem ciągłością operacyjną przedsiębiorstw. Dodatkowe zalecenia zawarte są w sekcjach: 3 (Polityka Bezpieczeństwa) oraz 7 (Ochrona Fizyczna i Środowiskowa).

C/ ISO/IEC Technical Report 13335 – „*Guidelines for the Management of IT Security" 13335-2 - „Managing and Planning IT Security*" - raporty techniczne obejmują wymagania stawiane wobec bezpieczeństwa (w tym ciągłości operacyjnej biznesu).

5. Grupy rozwiązań dla Business Continuity i Disaster Recovery

1/ Usługi:
- konsultacje / doradztwo,
- tworzenie planów postępowania na wypadek awarii / katastrofy,
- centra zapasowe (*Recovery Sites*) i udostępnianie zasobów.

1/ Rozwiązania systemowe:
- klastry serwerów,
- replikacja danych
- sieci pamięci masowych,

2/ Oprogramowanie:
- do planowania (BCP - *Business Continuity Planning*),
- do zabezpieczania oraz archiwizacji danych.

6. Oprogramowanie do planowania - Business Continuity Planning

Zadaniem oprogramowania BCP jest wspomaganie przedsiębiorstw lub organizacji w zakresie planowania, opracowywania, testowania oraz utrzymania procedur niezbędnych do zapewnienia ciągłości ich operacyjnej na wymaganym poziomie.

Większość aplikacji tej kategorii tworzy kwestionariusze ułatwiające analizę stanu bezpieczeństwa oraz planowanie niezbędnych działań. Wprowadzane do nich dane są gromadzone w odpowiednich bazach zintegrowanych z infrastrukturą IT firmy.

Narzędzia klasy BCP umożliwiają organizację procesu planowania, a następnie aktualizacji procedur w miarę zmieniających się wymagań lub możliwości.

Kategorie oprogramowania BCP:

- do analizy zagrożeń - BIA (*Business Impact Analysis*),
- narzędzia do budowy planów BC,
- bazy danych BCP,
- narzędzia do pracy grupowej BCP.

Typowe wymagania wobec środowiska pracy oprogramowania BCP to system operacyjny Windows (zwykle serwer oraz 1 - kilka stacji roboczych) wraz z narzędziami pomocniczymi (Word, Excel, Access) oraz bazą SQL Server lub architektura UNIX klient/serwer wraz z relacyjną bazą danych (np. DB2 czy Oracle).

Typy planów tworzonych przez oprogramowanie BCP i ich cele:

1/ ***Disaster Recovery Plan*** – przeniesienie krytycznych funkcji biznesowych i aplikacji do zapasowej lokalizacji,

2/ ***Business Resumption Plan*** (*workspace recovery*) – wznowienie podstawowych funkcji w dotkniętej katastrofą lub alternatywnej lokalizacji,

3/ ***Business Recovery Plan*** - wznowienie funkcjonowania krytycznych procesów biznesowych w zapasowej lokalizacji,

4/ ***Contingency Plan*** - zarządzanie zewnętrznymi zdarzeniami mającymi wpływ na funkcjonowanie przedsiębiorstwa.

W zależności od stanu przygotowania firmy (planowanie po raz pierwszy lub aktualizacja planu) oprogramowanie BCP można wykorzystywać w różny sposób. Zawsze należy pamiętać, że nawet najlepsze oprogramowanie nie zastąpi człowieka specjalisty (lub wręcz zespołu ludzi) i stanowi tylko wsparcie.

6.1 Liderzy oprogramowania BCP w opinii firmy Gartner

Najlepsi producenci branży posiadają kilkudziesięcioletnie doświadczenie w zakresie planowania procedur zabezpieczeń przedsiębiorstw przed różnymi zagrożeniami i katastrofami.

SunGard

Firma jest liderem rynku BCP. Aktualnie w ponad 50 krajach posiada ona ponad 20 tys. klientów. Oferta kierowana jest przede wszystkim do sektora finansowego oraz przedsiębiorstw w szerokim zakresie wykorzystujących technologie biznesu elektronicznego. Warto wspomnieć, że z ich oprogramowania korzysta 47 z 50 największych światowych instytucji finansowych.

PreCovery 3.2 jest narzędziem do planowania klasy BCP pracującym w architekturze Klient / Serwer. Zapewnia automatyzację wszystkich etapów budowy planu oraz ocenę skutków ewentualnej katastrofy.

Podstawowe cechy funkcjonalne:

- analiza wymagań przedsiębiorstw i generowanie raportów dotyczących różnych scenariuszy,
- obsługa budowy i modyfikacji planów BCP (technika „*drag-and-drop*"),
- specyfikacja automatycznie uruchamianych procedur importu i eksportu danych,
- wyszukiwanie w zasobach określonych danych na podstawie złożonych kryteriów (technologia filtrów).

ePlanner 3.2 to narzędzie BCP realizujące wszystkie funkcje oprogramowania PreCovery, ale wyposażone w interfejs Web'owy (wykorzystanie przeglądarki Internetowej) z 128-bitowym szyfrowaniem SSL. Wykorzystuje motor bazy danych pakietu PreCovery (pełna przenaszalność bazy).

ePlanner ASP 3.2 - wersja oprogramowania udostępniana jako usługa ASP (*Application Services Provider*).

Revolution Rel. 4.0 jest narzędziem BCP bazującym na najnowszych technologiach sieci WWW. Zapewnia ono obsługę budowy, administrowania oraz konserwacji wszystkich elementów planu BCP. Zainteresowane firmy mogą zakupić to oprogramowanie lub korzystać z niego w trybie ASP.

Wybrane możliwości:

- dostępność wzorcowych planów BCP
- kontrola funkcji realizowanych przez wszystkich użytkowników,
- korzystanie ze standardu SSL,
- tworzenie baz w wersjach treningowych i produkcyjnych,
- dołączanie dokumentów z opisem do wybranych elementów planu BCP,
- prowadzenie dziennika zmian planów,
- wiele wersji językowych,
- dostosowywanie „szaty graficznej" planu do wymagań przedsiębiorstwa (kolorystyka, logo itp.).

ComPAS Rel 6.0 jest narzędziem BCP dostępnym w wersji 1-stanowiskowej lub sieciowej.

Wybrane istotne komponenty / funkcje:

- Plan Wizard – obsługa procesu tworzenia nowych planów BCP i modyfikacji istniejących,
- Report Manager – wyprowadzanie (drukowanie) raportów ze zbudowanego planu,

- Asset Management – zarządzanie środkami trwałymi (analiza możliwości zastępowania sprzętu w przypadkach awarii lub katastrofy).

Produkty Revolution i ComPAS zostały przejęte z firmy Comdisco.

Strohl

Baza klientów obejmuje ponad 1000 organizacji w ponad 60 krajach świata. Wśród nich są m. in. 4 z 5 największych producentów samochodów, 3 z 5 największych gigantów ubezpieczeniowych, 6 z 10 największych banków czy 9 z 10 liderów rynku usług telekomunikacyjnych.

Podstawowy pakiet o nazwie **LDRPS** (*Living Disaster Recovery Planning System*) - oferowany aktualnie w wersji 9.1 - obejmuje analizę zagrożeń, planowanie oraz obsługę bazy BCP. Produkt współpracuje z bazami danych SQL Server, Sybase oraz Oracle. Istnieje możliwość wykorzystania wbudowanych w pakiet funkcji pracy grupowej. Dostępne są wersje przeznaczone dla pojedynczych użytkowników oraz tzw. korporacyjne z interfejsem Web-owym. W miarę rozwoju przedsiębiorstwa plan jest automatycznie modyfikowany. Dane są przenoszone, aktualizowane i uzupełniane. Bardzo wygodnym wbudowanym w pakiet narzędziem jest Planning Assistant zapewniający „nawigację" poprzez wszystkie etapu budowy planu BCP.

Wybrane cechy funkcjonalne LDPRS 9.1:

- Dostępność wzorcowych planów opracowanych przez ekspertów z zakresu BCP. Wzorce takie pomagają zdefiniować zadania, procesy i zasoby uwzględniane w tworzonym planie. Można je dostosowywać do konkretnych specyficznych wymagań danego przedsiębiorstwa. Korzystanie z nich znacznie przyspiesza proces budowy planu BCP.
- Możliwość korzystania z wielu wersji językowych (opisy ekranów, funkcje pomocy).
- Automatyczna realizacja importu (aktualizacji) wybranych danych zawartych w bazie planu (według zdefiniowanego harmonogramu).

Zapewnienie aktualności planu BCP.

- Automatyzacja procesu drukowania planu w różnych formatach.
- Przypominanie użytkownikom o wymaganych modyfikacjach planu.
- Obsługa funkcji udostępniania i zatwierdzania planów.
- Zarządzanie dostępem użytkowników do określonych planów, ekranów, pól czy zapisów (dostosowywanie zakresu do indywidualnych wymagań).

Wśród innych produktów firmy Strohl warto m. in. wymienić: **BIA Professional** oraz **Incident Manager**.

RSM McGladrey

Podstawową działalnością firmy RSM International jest doradztwo finansowe. Swoją ofertę z zakresu oprogramowania klasy BCP kieruje głównie do sektora przedsiębiorstw średniej wielkości (posiada ponad 600 biur w 75 krajach świata). Podstawowe produkty to **BCPS** (*Business Continuity Planning System*) oraz **BRPS** (*Business Recovery Planning System*). Obejmują one funkcje analizy zagrożeń, budowy planu ciągłości operacyjnej, obsługi bazy danych BCP oraz pracy grupowej w zakresie planowania.

CSCI (Computer Security Consultants Inc.)

Najpopularniejszy pakiet firmy to **RecoveryPAC**, który jest wykorzystywany przez ponad 1 400 średnich i dużych przedsiębiorstw oraz urzędów w 40 krajach (głównie Ameryka Płn. i Europa). Wśród klientów najwięcej jest agencji rządowych oraz firm z sektora finansowego. Oprogramowanie obejmuje moduły BCP Builder oraz BCP Database. Istnieje możliwość zintegrowania oprogramowania z dodatkowym narzędziem klasy BIA (**RiskPAC**). Pakiet RecoveryPAC może pracować w środowiskach: Windows 95/98/NT/ME/2000.

Business Protection Systems (BPSI)

Firma zajmuje się wyłącznie oprogramowaniem grupy BCP. Zestaw narzędzi oraz usługi oferowane są firmom wszelkiej wielkości (także prowadzącym działalności w sieci Internet). BPSI ma podpisaną strategiczną umowę partnerską z firmą IBM. Podstawowy produkt o nazwie **Business Protector** dostępny jest w trzech wersjach: Express, Professional oraz Gateway.

LBL Technology Partners

Firma ta wcześniej świadczyła finansowe usługi doradcze, a następnie rozszerzyła zakres działalności o wprowadzanie nowych technologii. Jej oferta z zakresu oprogramowania BCP kierowana jest przede wszystkim do agencji rządowych, firm prawniczych oraz służby zdrowia. Podstawowym produktem jest **LBL Contingency Planner**.

Przedsiębiorstwa zainteresowane wprowadzeniem rozwiązań klasy Business Continuity mogą zlecić opracowanie odpowiedniego dla nich planu wyspecjalizowanej firmie zewnętrznej lub na bazie jednego z przedstawionych wyżej pakietów narzędzi BCP przygotować własne rozwiązanie (zawsze później można poprosić o audyt). Drugi ze sposobów podejścia powinien dotyczyć jedynie tych przedsiębiorstw, które są odpowiednio przygotowane pod względem kadrowym.

7. Zapasowe centra przetwarzania danych

Firmy oferujące usługi w zakresie DR i BC lub dostawcy sprzętu proponują m. in.: szybkie wymiany uszkodzonych urządzeń (w tym serwerów i systemów pamięci) na sprawne o identycznej lub większej wydajności, zdalne udostępnianie zasobów (serwery, macierze dyskowe itd.), usługi centrów zapasowych wyposażonych w sprzęt umożliwiający w bardzo krótkim czasie podjęcie eksploatacji.

Komercyjne centra zapasowe (***recovery sites***) umożliwiają kontynuacje usług przetwarzania danych czy obsługi sieci w przypadku katastrofy (lub poważnej awarii). Prowadzące je firmy oferują podpisanie kontraktu gwarantującego wsparcie w ściśle określonym czasie zgodnie z wcześniej zaplanowanymi i przetestowanymi procedurami.

Kategorie centrów zapasowych:
A/ tzw. gorące (***hot site***) - w pełni wyposażony w niezbędne dla systemów klienta platformy sprzętowe i programowe ośrodek gotowy do podjęcia pracy bezpośrednio po wystąpieniu katastrofy (lub maksymalnie w ciągu kilku godzin); Takie zabezpieczenie jest niezbędne dla klientów pracujących w czasie rzeczywistym, firm ISP czy ASP itp.

B/ tzw. zimne (***cold site***) - pomieszczenia z pełną infrastrukturą (m. in. telekomunikacyjną) przygotowane do zainstalowania niezbędnego dla konkretnego klienta systemu komputerowego (sprzęt jest zwykle instalowany na podstawie osobnej umowy z producentem / dostawcą); Oczywiście podjęcie eksploatacji w tym przypadku wymaga czasu (np. kilku dni). Opcja ta nie jest akceptowana przez użytkowników realizujących przetwarzanie transakcyjne w trybie *on-line* czy eksploatujących systemy e-Biznesu.

C/ ruchome (***mobile site***) - zestawy sprzętu komputerowego oraz telekomunikacyjnego zamontowane w naczepach samochodowych; Minusem rozwiązania jest ograniczony zakres instalowanego sprzętu (ze względu na małą przestrzeń), natomiast plusem jest możliwość ulokowania centrum w położeniu najbardziej dogodnym dla klienta.

7.1. Technologie przenoszenia danych i oprogramowania do centrów zapasowych

Można wyróżnić kilka podstawowych grup rozwiązań:
A/ kopie wszystkich istotnych danych (zwykle na taśmach / kasetach) są przechowywane w osobnym budynku oddalonym od lokalizacji ośrodka przetwarzania danych lub np. w depozycie bankowym; Niektóre firmy oferują możliwości przechowywania nośników w specjalnych klimatyzowanych, chronionych pomieszczeniach.

B/ dane przekazywane są w czasie rzeczywistym od klienta do ośrodka zapasowego (tzw. gorącego); Przepisywane dane mogą pochodzić z macierzy dyskowych pracujących w trybie bezpośredniego podłączenia do serwera (SAS) lub z sieci pamięci masowych (SAN lub NAS). Ze względu na ochronę danych wymagane jest aby zapis był realizowany na dyskach dedykowanych dla konkretnego klienta.

Firma Gartner wśród liderów usług z zakresu zapewnienia ochrony danych i ciągłości operacyjnej wymienia przede wszystkim Comdisco, IBM oraz SunGard. Firmy te utrzymują m. in. specjalne wyposażone w sprzęt komputerowy i oprogramowanie oraz bogatą infrastrukturę teleinformatyczną ośrodki, które mogą służyć ich klientom jako zapasowe centra danych – przygotowane do niezwłocznego przejęcia usług przetwarzania danych i eksploatacji krytycznych aplikacji biznesowych użytkowników. Usługi takie oferowane są przede wszystkim dla sektora finansowego (np. banki) oraz usług i handlu w sieci Internet.

Zapasowe centra danych (*Hot Sites*) oferowane są od początku lat 80-tych XX wieku. W tym czasie skorzystało z nich wiele firm, których podstawowe lokalizacje uległy np. zniszczeniu na skutek katastrof (zwykle takich zdarzeń losowych jak pożar, huragan czy powódź). Firma Gartner liderem w zakresie usług Disaster Recovery Services określa firmę IBM, która przeprowadziła do tej pory bezbłędnie ponad 400 operacji odtwarzania zasobów. Głównymi konkurentami IBM na terenie Ameryki Północnej (przede wszystkim w zakresie platformy mainframe) są firmy Comdisco oraz SunGard. Wśród firm europejskich Gartner wskazuje na Guardian plc oraz Safetynet, które głównie zajmują się platformami klasy midrange.

Usługi **IBM** Business Continuity and Recovery Services (w ramach **IBM Global Services**) prowadzone są m. in. za pośrednictwem udostępniania klientom ponad 100 wyposażonych centrów przetwarzania danych. Znajdują się one na terenie całego świata (w Ameryce Płn., rejonie Azji/Pacyfiku, na Środkowym Wschodzie, w Afryce, Ameryce Łacińskiej oraz na Karaibach). Jeden z ośrodków działa i oferuje usługi w Warszawie. Lokalizacje trzech podstawowych centrów (Major Recovery Centers) firmy IBM w USA to Gaithersburg (Maryland), Boulder (Kolorado) oraz Sterling Forest (Nowy Jork). Poza nimi są w USA 4 centra regionalne (Regional Recovery Centers) oraz 9 tzw. lokalnych centrów dostępu (Regional Local Access Suites). Ośrodki te spełniają nawet najbardziej złożone wymagania klientów (przykładowo po ataku huraganu Floyd w 1999 roku IBM w swoich centrach udostępnił 46 klientom 96 różnych konfiguracji serwerów). Należy podkreślić fakt, że firma IBM w wielu swoich centrach regionalnych wspiera także platformy innych producentów (HP, Sun, Unisys, Digital Equipment, Tandem czy Compaq) oraz zapewnia ciągłość eksploatacji systemów ERP pochodzących m. in. z firm Oracle, SAP, Baan czy PeopleSoft..

Firma **Comdisco** oferuje swoje usługi przy wykorzystaniu aktualnie 54 tzw. *global data centers* - zlokalizowanych w różnych rejonach świata ośrodków zapasowych (m. in. w USA, Kanadzie, Francji, Wielkiej Brytanii, Niemczech, Singapurze i Malezji) oraz łączącej je globalnej sieci o wysokiej wydajności bazującej na infrastrukturze SONET (Synchronous Optical Network) zainicjowanej w grudniu 1993 roku przez AT&T przy wykorzystaniu łączy 155 mbps oraz technologii znajdowania dróg alternatywnych FASTAR (Fast Automatic Restoration). Na terenie USA 28 ośrodków pełni rolę centrów ogólnokrajowych.

SunGard Recovery Services (część firmy SunGard Data Systems) oferuje swoim 5.5 tys. klientom ponad 100 tys. metrów kwadratowych powierzchni wyposażonych centrów przetwarzania danych na terenie Ameryki Płn. (ponad 30 lokalizacji, w tym 26 w USA). Każdy z ośrodków przygotowany jest do przejęcia usług na różne platformy sprzętowe i systemowe. Firma udostępnia także tzw. zimne centra (powierzchnie przygotowane do instalacji sprzętu) oraz serwery zamontowane w specjalnych nadwoziach samochodowych (centra mobilne).

8. Systemy replikacji danych

Najpopularniejsze wydajne rozwiązania replikacji danych wykorzystują technologie zdalnego kopiowania pomiędzy systemami pamięci (dysk – dysk) na poziomie woluminu dyskowego (macierz wynikowa pracuje w ośrodku zapasowym). Operacje realizowane są zwykle bez utraty wydajności serwera obsługującego np. bazę danych przedsiębiorstwa lub bankowy system transakcyjny. Takie rozwiązania mają wiele zalet, z których podstawową jest objęcie kopiowaniem całego środowiska aplikacyjnego utrzymywanego w ośrodku. Większość technologii wykorzystuje tryb kopiowania synchronicznego. Dla zapewnienia wysokiej wydajności transferu (i uniknięcia obniżenia wydajności aplikacji produkcyjnych) stosuje się zwykle pomiędzy ośrodkiem podstawowym i rezerwowym bardzo szybkie łącza światłowodowe (zwykle odległości wynoszą od kilku do ok. 100 km).

Firma Gartner jako wiodące rozwiązania synchronicznej replikacji danych typu *disk-to-disk* dla procedur Disaster Recovery na poziomie woluminu dyskowego (*disk volume level*) wymienia technologie /w kolejności alfabetycznej firm/: DRM (Compaq), SRDF (EMC), HRC (Hitachi), Continuous Access XP (HP) oraz PPRC (IBM).

Alternatywą mogą być rozwiązania bazujące na systemach host oraz kopiowaniu na poziomie bloków dyskowych. Przykładem są np. Geographic Remote Mirror for AIX (IBM) czy Volume Replicator (Veritas Software). Natomiast dla środowiska Windows popularne są rozwiązania operujące na poziomie zbiorów (*file-based*) pochodzące np. z firm Legato Systems oraz NSI Software. Jeszcze inną grupę stanowią replikacje baz danych na podstawie zawartości logu (*log-based*). Produkty takie oferują m. in.: DataMirror, Microsoft, Oracle.

Przedstawię krótko kilka wiodących technologii wykorzystywanych do replikacji danych centrów obliczeniowych.

8.1. Macierze ESS (Shark) oraz technologia PPRC firmy IBM

IBM Enterprise Storage Server (ESS) jest systemem pamięci masowej przeznaczonym do zastosowań wymagających wysokiej niezawodności, wydajności oraz skalowalności. Zasoby dyskowe mogą być dynamicznie przydzielane do wielu podłączonych serwerów (w tym S/390 – zSeries, AS/400 – iSeries, UNIX, Windows NT/2000).

Dostępne w macierzach **ESS** (znanych też pod nazwą SHARK) technologie to:

- Peer-to-Peer Remote Copy (PPRC) – synchroniczna kopia danych może być tworzona lokalnie lub w lokalizacji zdalnej (np. w centrum zapasowym),
- Extended Remote Copy (XRC) – asynchroniczna kopia danych systemu OS/390 (z/OS) w zdalnej lokalizacji poprzez łącza telekomunikacyjne,
- FlashCopy – natychmiastowa kopia danych.

System umożliwia także dynamiczne przydzielanie zasobów pamięci różnym systemom dołączonym do ESS oraz udostępnia funkcje współdzielenia przestrzeni dyskowej. Zastosowanie systemu SHARK umożliwia konsolidację danych przedsiębiorstwa pochodzących z różnych platform. Model IBM 2105-F10

zapewnia pojemność użyteczną w standardzie macierzy RAID5 wynoszącą od 420 GB do 3.3 TB, natomiast w modelu F20 do 22.4 TB. Podłączenie może być zrealizowane za pośrednictwem do 32 interfejsów standardu SCSI / ESCON lub do 16 FC / FICON.

Rys. 1. Macierz dyskowa IBM ESS (SHARK)

PPRC to funkcja systemu pamięci ESS umożliwiająca odtwarzanie danych po wystąpieniu katastrofy. Dane są kopiowane z jednej na drugą macierz ESS (umieszczoną w zdalnej lokalizacji) w czasie rzeczywistym w trybie RAID-1 (*mirroring*). Przy tej operacji nie są dodatkowo obciążane serwery, a kopia realizowana jest na poziomie woluminu. Operacja może być realizowana dla środowisk OS/390 (z/OS), z/VM, VSE/ESA, AIX, Sun, HP-UX, OS/400, Windows NT/2000, Numa-Q. Technologia ta zapewnia możliwość synchronicznego kopiowania danych na odległość do 103 km. Warto zauważyć, że PPRC może być także wykorzystywane w rozwiązaniu GDPS (*Geographically Dispersed Parallel Sysplex*). W przypadku współpracy z platformą mainframe technologia PPRC jest zarządzana przez polecenia systemu operacyjnego OS/390 (z/OS), natomiast w środowiskach otwartych (UNIX, Windows) za pośrednictwem oprogramowania firmy IBM o nazwie **StorWatch ESS Specialist**. Administrator może wykorzystywać do obsługi przeglądarkę Internetową.

W lutym 2003 firma IBM zaanonsowała kilka nowości:

- interfejs API kompatybilny z nowym otwartym standardem Bluefin dla ESS Shark (zapewnienie łatwego zarządzania zasobami w sieciach pamięci wykorzystujących produkty pochodzące od wielu dostawców),

- rozszerzony zakres obsługi funkcji Disaster Recovery w środowiskach Linux na platformach mainframe (S/390 i zSeries) – możliwość korzystania z technologii FlashCopy oraz PPRC,
- nowe napędy dyskowe (wzrost wydajności do 50% -72.8GB, 15K rpm).

Specyfikacja Bluefin / SMIS (*Storage Management Initiative Specification*) ma dostarczyć wspólnego interfejsu dla zarządzania zasobami pamięci w środowiskach rozwiązań heterogenicznych. Bazuje ona na technologiach rozproszonego modelu DMTFCIM (*Distributed Management Task Force Common Information Model*) oraz WBEM (*Web Based Enterprise Management*). Korzystając z nowej specyfikacji mogą być w jednym standardzie tworzone aplikacje korzystające z różnych urządzeń. Szereg firm partnerskich IBM grupy ISV wspiera inicjatywę Bluefin / SMIS i prezentowało już gotowe rozwiązania wykorzystujące ten standard (m. in. BMC, Computer Associates, InterSAN, McDATA, Tivoli Software, VERITAS). Aktualnie standard Bluefin / CIM określa się często nazwą SMI (*Storage Management Initiative*).

W ostatnim roku firma IBM wprowadziła na rynek dwa nowe rozwiązania wspierające wykorzystywanie środowiska TotalStorage ESS:

- Enterprise Storage Expert (**ESS Expert**) V2.1.1 – narzędzie przeznaczone do wspomagania zarządzania wydajnością systemów pamięci rodziny ESS – Enterprise Storage Server - (udostępnianie administratorowi statystyk pracy, monitorowanie zasobów, informowanie o stanie wspieranych systemów, pomoc w zcentralizowanym zarządzaniu wieloma zdalnymi systemami zlokalizowanymi na terenie przedsiębiorstwa, optymalizacja podsystemów dyskowych i taśmowych),
- Enterprise Tape Library Expert (**ETL Expert**) V1.2 – narzędzie przeznaczone do wspomagania zarządzania podsystemami pamięci taśmowych (takich jak: Magstar 3494 Tape Library, Magstar Virtual Tape Server, Magstar Peer-to-Peer Virtual Tape Server). Administrator ma dostęp do szerokiego zakresu statystyk dotyczących wydajności (w tym funkcja Health Monitor). Narzędzie może współpracować z oprogramowaniem ESS Expert. Dzięki wykorzystaniu architektury Web'owej istnieje możliwość monitorowania podsystemów dyskowych i taśmowych pracujących w dowolnych lokalizacjach na świecie (z zapewnieniem odpowiednich technologii ochrony i bezpieczeństwa).

8.2. Technologie oferowane przez Hitachi Data Systems

Macierze dyskowe **Lightning 9900 Series** firmy HDS charakteryzują się wysoką wydajnością i niezawodnością. Pojemność użyteczna pojedynczego systemu (przy organizacji RAID5) wykorzystywanego dla środowisk otwartych: Windows NT/2000, Red Hat Linux, UNIX (AIX, HP-UX, Sun Solaris) wynosi dla modelu 9910 do 2.4 TB, a dla 9960 do 27.5 TB (dla platformy S/390 wielkości te wynoszą odpowiednio 2.3 oraz 25.7 TB).

Aktualnie firma HDS oferuje serię macierzy **Hitachi Freedom Storage Lightning** 9900 V Series (modele 9970V i 9980V) zapewniającą znaczne zwiększenie wydajności pracy zasobów o pojemności ponad 140 TB przy

wykorzystaniu architektury Hi-Star, oprogramowania **Freedom Storage Software** oraz środowiska zarządzania HiCommand.

HDS oferuje kilka rozwiązań służących do replikacji danych. Technologia HRC (*Hitachi Remote Copy*) zapewnia zdalną replikację danych platformy S/390 (zSeries) w trybie synchronicznego transferu pomiędzy systemami pamięci Lightning 9900 lub macierzy 7700E do zdalnej lokalizacji (odległość limitowana jest możliwościami kanałów ESCON). Dla systemów otwartych przygotowana została technologia HORC (*Hitachi Open Remote Copy*).

Oprogramowanie **Hitachi ShadowImage** umożliwia jednoczesną duplikację obrazów woluminów logicznych (LVI) oraz jednostek logicznych (LUN) w sposób nieprzerwany bez zakłócania pracy serwera systemu (bez spadku jego wydajności). W ramach jednego systemu Lightning 9900 możemy tworzyć do 10 kopii woluminu. Rozwiązanie dostępne jest dla platformy mainframe (S/390) oraz systemów otwartych. Zarządzanie realizowane jest za pośrednictwem interfejsu graficznego lub komend wprowadzanych z konsoli.

Następcą rozwiązania HORC jest technologia Hitachi TrueCopy przeznaczone dla pamięci **Lightning** 9900 oraz **Freedom Storage** 7700E. Służy ono do ciągłej aktualizacji i utrzymywania kopii krytycznych danych biznesowych w oddalonej lokalizacji (np. ośrodku zapasowym) bez obciążania jednostki centralnej systemu. Transfer odbywa się poprzez kanały ESCON lub FC (*Fibre Channel*). Od niedawna dostępne jest też oprogramowanie **NanoCopy** (bazujące na technologii TrueCopy), które umożliwia replikację danych pomiędzy praktycznie dowolną ilością systemów podstawowych oraz zapasowych (w oddalonych lokalizacjach). Wykorzystywana jest technologia *snapshots* (okresowym pobieraniu stanu pamięci bez zakłócania pracy systemu). Funkcja TrueCopy może pracować w trybie synchronicznym oraz asynchronicznym. Technologia Hitachi TrueCopy wykorzystywane jest m. in. w ramach rozwiązania Oracle Database Disaster Recovery (zabezpieczania zasobów bazy Oracle).

Pamięci dyskowe Hitachi mogą być także wykorzystywane w rozwiązaniach klasy GDPS.

8.3. Technologia SRDF firny EMC

Firma **EMC** w swojej oferuje dwie rodziny macierzy dyskowych: **Symmetrix** oraz **CLARiiON**. Rozwiązanie Symmetrix serii z8000 zapewnia w pojedynczym systemie pojemność do 70 TB. Może ono współpracować z różnymi platformami, w tym mainframe, Windows NT/2000, UNIX oraz Linux. W ramach macierzy CLARiiON serii CX oferowane są modele grup 200, 400 i 600. Proces replikacji na systemach CLARiiON CX400 i CX600 może być realizowany automatycznie przy wykorzystaniu oprogramowania **EMC Replication Manager** i technologii EMC SnapView. Można też korzystać z oprogramowania **SAN Copy** do przenoszenia zasobów macierzy w ramach sieci.

Firma EMC do replikacji danych oraz aplikacji na platformach systemów Symmetrix standardowo oferuje technologię **SRDF** (*Symmetrix Remote Data Facility*). Umożliwia ona w czasie rzeczywistym kopiowanie danych do jednej lub

kilku lokalizacji niezależnie od oddalenia od centrum podstawowego. Replikacja lokalna może być wykonywana z wykorzystaniem kanałów FC. Dla rozwiązań zdalnych wykorzystywane są m. in. technologie ATM, DWDM (*Dense Wave Division Multiplexing*) oraz IP (także za pośrednictwem bezpiecznych sieci VPN).

Rozwiązanie SRDF firmy EMC umożliwia migrację danych pomiędzy centrami w sposób bardzo szybki i bez przerywania pracy serwerów produkcyjnych. W obu lokalizacjach możemy wykorzystywać technologie RAID1 (*mirroring*) lub RAID5. Administrowanie procesem realizowane jest za pośrednictwem oprogramowania **EMC ControlCenter Symmetrix Data Mobility Manager** lub **SRDF Host Component**. Rozwiązanie SRDF jest kompatybilne z technologią GDPS.

W lutym 2003 firma EMC zaanonsowała nową serię systemów pamięci o nazwie **Symmetrix DMX** bazującą na architekturze *Direct Matrix Architecture* zapewniającej postęp w zakresie wydajności, niezawodności oraz zmniejszenia kosztów przechowywania wielkich objętości danych. Zamiast tradycyjnych technologii szyny (*bus*) oraz przełączników architektura DMX proponuje stosowanie macierzy dedykowanych połączeń bezpośrednich „punkt-punkt" (*messaging matrix*). Każda z dróg dostępu do danych jest dedykowana (bez korzystania z techniki współdzielenia zasobów). Wykorzystywany jest też specjalny wielki bufor pamięci (*DMX cache*) z wyróżnionymi 32 regionami (obsługuje do 32 jednoczesnych transferów). W sumie prowadzi to do znacznego wzrostu wydajności systemów. Architektura DMX może obsługiwać aktualnie w 1 systemie do 2048 napędów dyskowych. Dostępne są obecnie trzy modele systemów pamięci Symmetrix DMX Series:

- Symmetrix DMX800 (montowane w stojakach) o pojemności od 1.2 do 17.5 TB (*cache* 4 – 32 GB) – dla „systemów otwartych",
- Symmetrix DMX1000 – o pojemności „szafy" 3.5 do 21 TB (*cache* 4 – 64 GB) – dla systemów „otwartych" i mainframe,
- Symmetrix DMX2000 – w 2 „szafach" mieszczą do 42 TB pojemności (*cache* do 128GB).

Systemy Symmetrix DMX są w pełni zgodne ze specyfikacją SMI (*Storage Management Initiative*) znanej wcześniej pod nazwą Bluefin / CIM. Mogą być one zarządzane przez oprogramowanie EMC ControlCenter (dostępnych jest też 60 produktów wspomagających zarządzanie macierzami Symmetrix z 37 firm).

9. Oprogramowanie wspomagające zarządzanie systemami i zasobami (w tym do zabezpieczania i odtwarzania danych)

Oprogramowanie do zarządzania systemami pamięci obejmuje m. in. produkty służące do replikacji oraz wirtualizacji danych, narzędzia do wspomagania systemów macierzy dyskowych oraz bibliotek taśmowych, narzędzia do obsługi procedur zabezpieczania i odtwarzania (*backup/restore*) oraz archiwizacji danych, oprogramowanie do zarządzania systemami pamięci masowych oraz sieciami pamięci (SAN, NAS). Według firmy Gartner obroty w

zakresie nowych licencji na oprogramowanie do zarządzania systemami pamięci (*Storage Management*) wyniosły w roku 2000 ok. 5.6 mld USD. Jako liderzy tego sektora wymieniani są następujący producenci: EMC, IBM (Tivoli), Computer Associates oraz Veritas Software. Wszystkie te firmy mają udział w rynku oceniany na po kilkanaście procent (łącznie ok. 65 %). Innymi liczącymi się dostawcami oprogramowania tej grupy są m. in. Legato Systems, HP, Sun, BMC czy Microsoft. Najpopularniejszymi na rynku produktami z dziedziny zarządzania zasobami pamięci według badań ankietowych są: Backup Exec, NetBackup (Veritas), ARCserve (CA), OmniBack, Data Protector (HP), Tivoli Storage Manager (IBM), TimeFinder (EMC), NetWorker (Legato), Resolve RSM (BMC), Solstice Disk Suite (Sun Microsystems), Retrospect (Dantz).

Można też wymienić grupę innych producentów, którzy weszli na rynek oprogramowania przeznaczonym dla obsługi rozwiązań sieci pamięci masowych (EMC, Network Appliance) wykorzystujących protokół NDMP (*Network Data Management Protocol*) – np. firmy: BakBone (NetVault), CommVault (Galaxy) czy Workstation Solution (Quick Restore).

VERITAS

Veritas Software osiągnął swoją wysoką pozycję w segmencie oprogramowania dla procedur backup m. in. dzięki zakupom kilku firm. W 1996 r. przejęta została firmy OpenVision posiadająca udany pakiet NetBackup przeznaczony dla środowisk UNIX (Veritas wprowadził następnie na rynek wersję NetBackup dla wsparcia platformy Windows NT). Kolejne przejęcie dotyczyło Seagate Software NSMG wraz z pakietem Backup Exec dla środowisk Windows NT oraz NetWare. Veritas przejął także kilka innych mniejszych firm (Telebackup Systems, Winward Technologies i NuView). Warto dodać, że Microsoft dostarcza swoim klientom łącznie z systemem Windows 2000 produkt Logical Disk Manager opracowany także przez firmę Veritas....

Według różnych firm analitycznych produkty rodzin VERITAS **NetBackup** oraz **Backup Exec** obsługują 30 - 40 % procedur Backup/Recovery dla platform Windows NT oraz NetWare. Od stycznia 2003 oferowanym rozwiązaniem dla obsługi funkcji składowania i odtwarzania danych dla środowisk Windows jest m. in. VERITAS **Backup Exec 9.0 for Windows Servers**. Produkt wyróżnia się wysoką wydajnością (m. in. dzięki wykorzystywaniu Windows Change Journal) oraz bardzo funkcjonalnym graficznym interfejsem użytkownika. Wykorzystywane mogą być także dodatkowe zaawansowane opcje: Advanced Open File (zabezpieczenie lokalnych lub zdalnych plików używanych w trakcie operacji składowania danych), Intelligent Image (szybkie składowanie danych z serwerów obsługujących wielkie ilości plików), SAN Shared Storage (przydzielanie wymaganych zasobów pamięci taśmowych). Podstawowe usprawnienia w stosunku do wcześniejszych wersji (8.6) produktu Backup Exec to:

- szybsza obsługa środowisk systemu poczty elektronicznej MS Exchange, możliwość indywidualnego składowania folderów i załączników poczty,
- wsparcie dla platform MS Windows Server 2003 i technologii .NET,
- bardzo szybka instalacja (ok. 10 minut),

- zastosowanie interfejsu „Anywhere Internet" umożliwiającego administratorom monitorowanie i zarządzanie procedurami składowania danych z dowolnego PC dołączonego do sieci Internet.

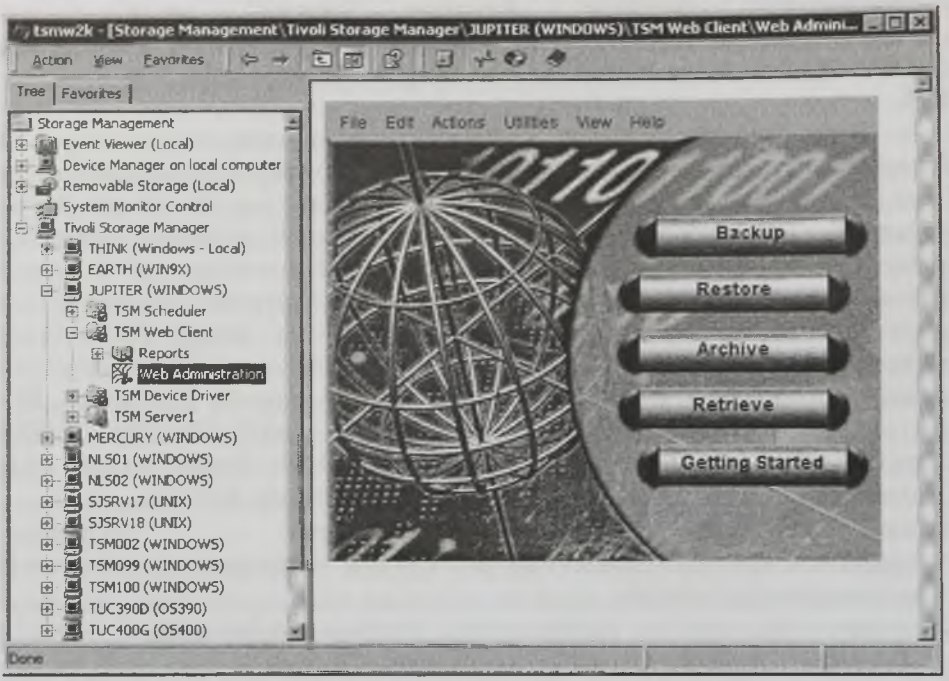
IBM

Firma IBM oferuje bardzo szeroki zakres sprzętu oraz oprogramowania do obsługi procedur zabezpieczania danych. Produkt oferowany wcześniej pod nazwą Tivoli Disaster Recovery Manager aktualnie włączony jest do pakietu **Tivoli Storage Manager Enterprise Edition**. Obejmuje on m. in. procedury związane z zabezpieczeniem danych przed katastrofą (*Disaster Preparation and Recovery)* oraz archiwizacją HSM (*Hierarchical Storage Management*) realizowaną w specjalnej technologii ułatwiającej dostęp do „zrzuconych" danych. Pakiet Tivoli Storage Manager w wersji 5.2 dostępny jest dla wszystkich popularnych platform serwerowych, w tym: Windows Server 2000 / 2003, Linux/x86 (2.4 kernel), HP-UX (11.0, 11.11), Sun Solaris (8 i 9), AIX (5.1, 5.2), z/OS (OS/390), OS/400. Klienci mogą być dodatkowo instalowani w środowiskach: Novell NetWare (5.1, 6), Macintosh, SGI IRIX UNIX (6.5), Tru64 UNIX, Windows NT/ Me.

Podstawowe cechy pakietu **IBM Tivoli Storage Manager**:

- zcentralizowane administrowanie procedurami składowania i odtwarzania oraz archiwizacji danych,
- wsparcie dla szerokiego zakresu sprzętu i platform systemowych,
- korzystanie z „inteligentnych" procedur obsługi danych,
- obsługa zasobów pamięci w architekturze hierarchicznej (Tivoli Storage Manager for Space Management) oraz sieci pamięci masowej SAN(Tivoli Storage Manager for Storage Area Networks).

Zarządzanie zasobami (z automatyzacją funkcji) może być realizowane w sposób bardzo elastyczny: jednocześnie przez kilku administratorów, kilku serwerów z jednego punktu sieci, poprzez sieć WWW.



Rys. 2.Przykładowy ekran klienta pakietu IBM Tivoli Storage Manager

W przypadku, gdy wymagane jest zabezpieczenie transferu danych w sieci pomiędzy klientem a serwerem pakietu Tivoli Storage Manger możliwe jest użycie opcji szyfrowania. Inteligentne funkcje obsługi danych wykorzystują m. in. koncepcję tzw. „*storage pools*" (dane są przenoszone dynamiczne na różne rodzaje nośników – pamięci dyskowe, napędy optyczne, kasety). Możliwe jest zabezpieczanie całości zasobów, konkretnych dysków lub przyrostowe składowanie danych (*progressive incremental backup*). Zapewniona jest możliwość restartowania procedur dzięki wykorzystywaniu specjalnych znaczników kontrolnych (*checkpoints*). Zawarte w zasobach pamięci dane należące do konkretnych klientów grupowane są w zespoły logiczne i mogą być zabezpieczane na wydzielonych nośnikach magnetycznych lub optycznych. Oprogramowanie Tivoli zapewnia także możliwość automatycznego usuwania z systemu plików nieaktualnych (po upływie zdefiniowanego okresu ważności).

CA

Rodzina rozwiązań firmy CA (**Computer Associates**) z zakresu zarządzania pamięcią o nazwie **BrightStor** obejmuje szereg komponentów obsługujących procesy i zadania zapewniające dostępność danych i aplikacji. Podstawę stanowi oprogramowanie do zabezpieczania i odtwarzania (*backup / recovery*) danych oferowane dla pełnego zakresu platform i środowisk systemów operacyjnych (od urządzeń przenośnych i komputerów PC, poprzez serwery instalowane na poziomie wydziałów i przedsiębiorstw, do systemów przetwarzania danych klasy mainframe).

W ramach tej grupy rozwiązań firma CA oferuje technologie oraz produkty służące do zarządzania danymi oraz do ich zabezpieczania. Operacje te mogą być realizowane na wszystkich popularnych platformach sprzętowych i systemowych oraz w odniesieniu do wszystkich dostępnych architektur systemów pamięci masowych.

Oferta **BrightStor** obejmuje następujące podstawowe grupy rozwiązań do zarządzania:

- Data Availability – stałą i niezawodną dostępnością danych,
- Media Management – systemami oraz urządzeniami pamięci,
- SAN i NAS Storage Management – rozwiązaniami sieciowych pamięci masowych,
- Storage Resource Management – zasobami pamięci.

Firma CA dostarcza też narzędzie o nazwie **Storage Management Portal**, które umożliwia zcentralizowane elastyczne zarządzanie złożonymi infrastrukturami i systemami pamięci. Administrator za pośrednictwem przeglądarki Internetowej może w wygodny sposób z dowolnej lokalizacji kontrolować pracę wszystkich wyżej wymienionych rozwiązań.

Z punktu widzenia zapewnienia maksymalnego poziomu ciągłości funkcjonowania infrastruktury IT i obsługiwanych procesów biznesowych największe znaczenie mają następujące produkty grupy BrightStor:

- **BrightStor Enterprise Backup**
  BrightStor ARCserve
  BrightStor Mobile Backup - zapewniają dostępność danych, aplikacji i serwerów poprzez obsługę (na różnych platformach) zaawansowanych procedur składowania i odtwarzania danych oraz technologii *Disaster Recovery*,
- **BrightStor CA-Disk Backup and Restore**
  BrightStor VM: Backup - procedury składowania danych i archiwizacji na poziomie zbiorów dla platformy *mainframe*; (dane mogą być przenoszone pomiędzy różnymi systemami pamięci (w tym np. dyskami magnetycznymi i optycznymi, taśmami).
- **BrightStor CA-FAVER VSAM Data Protection** - szybkie i efektywne procedury składowania, odtwarzania oraz reorganizacji dla zbiorów VSAM,
- **BrightStor High-Avalibility Manager** - obsługa procesów replikacji danych w trybie *on-line* pomiędzy różnymi platformami (zapewniających stałą dostępność danych oraz odporność systemów na wystąpienie awarii).

Rozwiązania **BrightStor Enterprise Backup** (przeznaczone dla platform UNIX i Windows) oraz **BrightStor ARCserve Backup** (pracujące w środowiskach Windows, Linux i Novell NetWare) wykorzystują agentów programowych, którzy mogą być instalowani w takich specjalnych środowiskach jak Microsoft Exchange czy baza danych Oracle (dla wsparcia specyficznych

funkcji tych pakietów). Umożliwiają oni realizację zabezpieczania danych nawet podczas pracy tych aplikacji (niezależnie od stanu obciążenia systemu).

Bardzo ważną rolę dla zapewnienia wysokiej efektywności pracy systemów IT przedsiębiorstwa mogą spełniać zaawansowane rozwiązania z grupy **BrightStor Storage Resource Manager**. Realizują one funkcje zarządzania procesami składowania oraz odtwarzania danych oraz udostępniają użytkownikom zaawansowane narzędzia realizujące zadania monitorowania, alarmowania, harmonogramowania, realizacji analiz (audytu) oraz raportowania. Dzięki zastosowaniu takiego oprogramowania można uzyskać m. in. automatyzację procesów obsługi systemów pamięci oraz zapewnić integrację zadań z zakresu zarządzania zasobami. Następuje optymalizacja ich wykorzystywania na obszarze całego przedsiębiorstwa (w tym systemów pamięci i plików, rozwiązań NAS, aplikacji i baz danych).

Wymienione wyżej rozwiązania poza obsługą procesów zabezpieczania i archiwizacji systemów plików i danych aplikacyjnych obsługują także procesy *Disaster Recovery*, w których kompleksowe informacje składowane są w formatach umożliwiających bezzwłoczne i kompletne odtworzenie stanu środowiska (na podstawie tzw. *snapshot*'u - czyli stanu pamięci w określonym momencie bez zakłócania pracy systemu).

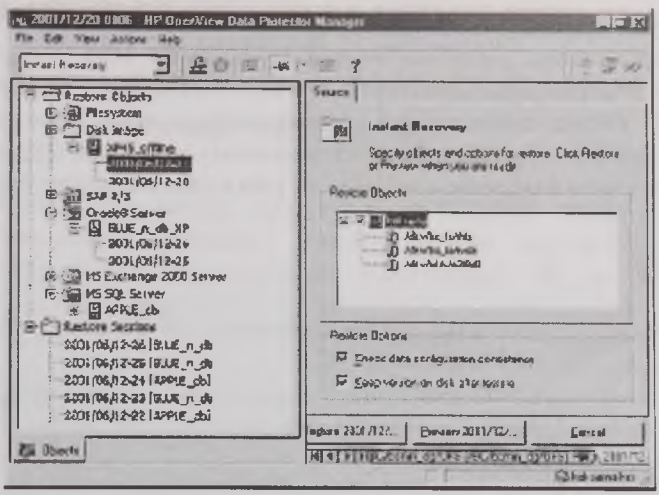
HP

Inny z liderów rynku IT – firma HP – w swojej ofercie posiada duży zakres sprzętu do zabezpieczania i archiwizacji danych (wraz z niezbędnym oprogramowaniem zarządzającym). Do rodziny oprogramowania **HP OpenView** zaliczany jest m. in. pakiet **OmniBack II** (wersje 4.x). Zapewnia on realizację funkcji automatyzacji ochrony danych oraz zarządzania mediami pamięci. OmniBack II udostępnia szereg technologii, które eliminują tzw. *backup windows* (przerwy w pracy serwera z pełną wydajnością ze względu na obsługę procedur zabezpieczania danych).

Pakiet oferowany jest dla wielu platform oraz może być integrowany z rozwiązaniami i systemami pamięci masowych –np. HP Surestore Virtual Array, HP Surestore XP Disk Array czy EMC Symmetrix.

Od połowy 2002 roku dostępne jest także kolejne rozwiązanie (dla platform Windows, HP-UX oraz Sun Solaris) o nazwie **HP OpenView Storage Data Protector 5.0**. Zapewnia ono znaczne rozszerzenie możliwości w stosunku do Omniback II m. in. poprzez:

- zwiększenie dostępności informacji (poziomu ciągłości prowadzenia biznesu),
- automatyzację i uproszczenie procedur zarządzania ochroną zasobów,
- zwiększenie wydajności infrastruktur IT,
- wyższą skalowalność, wsparcie różnorodnych platform, systemów i aplikacji oraz bieżących standardów technologii informacyjnych.



Rys. 3. Przykładowy ekran HP OpenView Data Protector Manager

Wykorzystując rozwiązanie **Storage Data Protector** oraz technikę Instant Recovery wspierającą technologie *mirroring* i *snap-shot* można w ciągu kilku – kilkunastu minut zabezpieczyć zasoby pamięci o wielkości terabajtów. W jednym produkcie mamy zawarte procedury zabezpieczania na nośnikach dyskowych oraz taśmowych. Agenci pakietu mogą być instalowani w wielu różnorodnych środowiskach aplikacyjnych (SAP, Baan i inne) oraz zapewniać efektywne zabezpieczanie zasobów baz danych (Oracle, MS SQL Server, Informix, Sybase) czy systemów pracy grupowej i poczty elektronicznej (np. MS Exchange, Lotus Notes). Agenci zabezpieczania danych mogą być instalowani na różnorodnych platformach systemowych, w tym: Windows 95/98/ME/NT/2000/XP, Novell NetWare, HP-UX, Sun Solaris, Linux (Red Hat, Debian , SuSE), Caldera OpenLinux, IBM AIX, SGI IRIX, SNI SINIX, SCO (OpenServer, Unixware), Compaq Tru64 i innych z zapewnieniem współpracy za pośrednictwem technologii NFS/*shared disk*, CIFS oraz NDMP NAS *filer*.

Oprogramowanie wspiera nowoczesne technologie sieci pamięci masowych SAN / NAS oraz szereg wykorzystywanych w aktualnych systemach protokołów (iSCSI – SCSI over TCP/IP, FCIP – FC over IP, iFCP – Internet FC Protocol).

Rozwiązanie Storage Data Protector jest kompatybilne z wcześniejszym Omniback II w zakresie wykorzystywania nośników (ten sam standard zapisu na taśmach), skryptów oraz procedur.

10. Zamiast podsumowania

Tragedia, która rozegrała się w Nowym Jorku 11 września 2001 roku stanowiła bardzo ciężką próbę dla wielu firm i organizacji zlokalizowanych w

bliźniaczych wieżach WTC oraz ich bezpośrednim pobliżu. Firmy posiadające dobrze opracowane plany na wypadek katastrofy oraz mające przygotowane zastępcze lokalizacje podstawowych zasobów niezbędnych do kontynuacji biznesu wraz z ich obsługą informatyczną i zabezpieczające swoje dane poniosły straty, ale uniknęły upadku. Niektóre z firm jednak musiały zakończyć działalność.... nie były przygotowane na katastrofę o takiej skali jaka miała miejsce na Manhattanie.

Literatura

1. Kurowski J. - Rozwiązania klasy Business Continuity, Praca w reżimie ciągłym. Telenet Forum , 07/2002 http://www.telenetforum.pl/index_2.php?show=archiwum&art_id=118
2. Materiały informacyjne dostawców technologii i produktów Business Continuity i Disaster Recovery.

Jerzy Kurowski
ZETO OLSZTYN Sp. z o.o.
10-005 OLSZTYN, ul. Pieniężnego 6/7
tel. 089 – 5272113, kurowski@zeto.olsztyn.pl

KLASTER PC. SPOSÓB NA EFEKTYWNE WYKORZYSTANIE ZASOBÓW

MAREK WIERZBICKI

Streszczenie: Artykuł wprowadza czytelnika w kwestię bezpiecznego i efektywnego wykorzystania serwerów. Zawiera kolejno wyjaśnienie jakimi cechami musi charakteryzować się serwer, co to jest klaster i jak zorganizowany jest klaster przełączeniowy Microsoftu. Po wprowadzeniu najważniejszych pojęć przedstawione są domyślna oraz inne, bardziej efektywne konfiguracje, które umożliwiają przyspieszenie pracy, zwiększenie funkcjonalności systemu informatycznego bądź obniżenie ceny klastra.

Wiele jest w Polsce firm, w których serwery pracują przez 24 godziny na dobę. Osoby korzystające z nich prawie zawsze wymagają, aby zasoby tych serwerów były dostępne bez przerwy, a ich awaryjność była minimalna. Jednak od przekonania użytkowników do rzeczywistej niezawodności wiedzie trudna droga usiana wydanymi pieniędzmi. Można podać różne przykłady, kiedy niezawodność pracy serwera jest krytycznym parametrem określającym możliwości przedsiębiorstwa. Praca na trzy zmiany nie zawsze oznacza konieczność posiadania idealnie niezawodnego serwera. Jeśli jednak proces produkcyjny jest sterowany komputerowo, a zatrzymanie go spowoduje straty nieproporcjonalnie duże w porównaniu do czasu trwania przerwy, klaster staje się nieodzowny. Można też sobie wyobrazić przykłady firm nie pracujących przez 24 godziny, w których nawet krótka przerwa jest nie do zaakceptowania. Nikt nie wróci do banku, w którym są ciągłe problemy z wypłatą pieniędzy, ze względu na „awarię systemu komputerowego".

1. Sposoby podnoszenia niezawodności

Prostym sposobem podnoszenia niezawodności funkcjonowania sprzętu jest stosowanie redundancji, czyli nadmiarowości elementów tworzących komputer. Dwa zasilacze, dwie karty sieciowe czy kontrolery dyskowe są powszechnie stosowane nawet w dość tanich serwerach. Jednak dwa procesory to już nie zawsze nadmiarowość procesorów (niewiele jest na świecie konstrukcji, w których można wymienić procesor bez zakłócenia pracy komputera). A dwie lub więcej płyt głównych w jednej obudowie to raczej kilka komputerów, a nie redundancja (konstrukcję taką nazywa się czasami mianem serwera *blade*).

W przypadku niektórych elementów, które nie są krytyczne dla pracy systemu stosuje się różne technologie, które umożliwiają wymianę uszkodzonych części w trakcie pracy. *Hot-plug* to technologia, która umożliwia jawne zatrzymanie pracy jakiegoś elementu i wymianę go w trakcie pracy komputera. Rozwiązanie takie może dotyczyć na przykład karty sieciowej. Oczywiście sens

stosowania tej technologii istnieje tylko wtedy, gdy karta taka jest zdublowana i zatrzymanie jednej z nich nie powoduje utraty komunikacji serwera z siecią. *Hot-swap* to technologia umożliwiająca wymianę elementu w czasie pracy (pod napięciem). Rozwiązania takie stosuje się do zasilaczy czy dysków twardych. Często technologia ta jest pochodną stosowania innych konstrukcji, mających na celu podniesienie niezawodności komputera. Tak więc wymiana twardego dysku może odbyć się w przypadku, gdy dyski pracują w trybie *mirror* (lustrzane kopie), bądź tworzą macierz typu RAID.

Mimo tego, że większości awarii można w ten sposób zapobiec, może wystąpić sytuacja, gdy element mimo wymiany na sprawny nie podejmie stabilnej pracy. Poza tym istnieją w komputerach elementy, które nie posiadają żadnej nadmiarowości stwarzając przez to nadmierne zagrożenie przerwy w pracy.

1.1. Serwer PC

Poszczególne rozwiązania stosowane w celu podniesienia bezawaryjności pracy komputera nie czynią z niego jeszcze serwera. W czasach zamierzchłej przeszłości informatycznej nikt nie myślał o używaniu komputerów typu PC w roli serwerów. Jednak od początku lat dziewięćdziesiątych tendencja ta powoli się odwracała. Obecnie platforma intelowska (należy włączyć w to również procesory AMD) obfituje w podzespoły dedykowane do produkcji serwerów. Elementy te wyróżniają się, w stosunku do używanych w zwykłych PC-tach, zarówno niezawodnością pracy jak i szczegółami konstrukcyjnymi, lepiej predystynującymi je do pracy serwerowej (itd. lepiej dopracowaną wielowątkowścią). Mimo tego, że elementy te są droższe od standardowych, zbudowane z nich serwery są znacznie tańsze, niż komputery bazujące na przykład na platformie RISC. W związku z tym każda poważna firma komputerowa ma w swojej ofercie serwery bazujące na platformie Intel.

Poza wykorzystaniem specjalizowanych elementów serwer, w stosunku do zwykłego PC-ta, wyróżnia się on kilkoma ważnymi cechami. Najważniejsze to stosowanie technologii, która umożliwia wewnętrzną kontrolę pracy poszczególnych podzespołów. Dobrym przykładem są macierzowe interfejsy obsługi dysków twardych, które nie tylko zapewniaja większą szybkość i bezpieczeństwo, ale same są w stanie wykryć ewentualna awarię jednego z nadmiarowych elementów macierzy i przekonfigurować się w czasie działania tak, że nie zmieni to warunków pracy serwera. Prawie zawsze serwery wyposażone sa też w zasilacze, umożliwiające jednoczesne zasilanie komputera z dwóch różnych źródeł zasilania. Regułą też jest, że zdublowane elementy pracują w technologii *hot-swap* bądź *hot-plug*, chodź nie zawsze jest to warunek spełniany w sposób jawny. Osobnym, niefizycznym aspektem serwera jest gwarancja udzielana przez producenta i czas reakcji na ewentualną awarię. Niczym wyrafinowanym nie jest tu 4 godzinny okres, który upływa od zgłoszenia awarii do pojawienia się serwisu firmowego w serwerowni.

2. Klaster bądź farma komputerów

Dalsze podnoszenie niezawodności i skracanie potencjalnego czasu przestoju może się odbywać wyłącznie poprzez zwielokratnianie całych komputerów i łączenie ich w hybrydy (klastry bądź farmy) widoczne z zewnątrz jako jedna maszyna. Generalnie istnieją trzy sposoby łączenia komputerów w celu podniesienia niezawodności całego układu. Klasyczny, znany z systemów UNIXowych i pochodnych, to farma komputerów, widziana jako jeden wielki zasób jednej instancji systemu operacyjnego. Dwa dwuprocesorowe komputery są widziane jako cztery procesory dostępne dla systemu. Awaria jednego z komputerów powoduje wyłącznie spadek mocy całego systemu, bez utraty wykonywanych zadań. Ostatnimi czasy bardzo modne stało się wykorzystanie komputerów bazujących na platformie Intela do budowy serwerów typu *blade* (oczywiście można używać w tym celu również procesorów typu RISC, lecz zaletą procesorów intelowskich jest ich niska cena). Są to małe komputery, skonstruowane w postaci kart wpinanych do wspólnej, bardzo szybkiej magistrali sieciowej. Formalnie każda z kart jest pełnoprawnym komputerem, przystosowanym do wykorzystania wewnętrznej sieci. Zestaw tych komputerów połączony jest programowo tak, aby zarówno z zewnątrz jak i wewnętrznie był traktowany jako jeden komputer. Cały serwer pracuje poprawnie bez względu na to, czy w obudowie wstawione jest 16, 128 czy 1024 karty. Jedyna różnicą jest szybkość wykonywania określonego zadania. System posiada mechanizmy, które umożliwiają wykrycie awarii i powierzenie pracy wykonywanej przez zepsuty fragment klastra innemu elementowi. W razie ewentualnej awarii system pamięta zadanie przydzielone zepsutej karcie i ponownie przydziela to zadanie innej.

Drugie, typowo sprzętowe rozwiązanie to fizyczne zdublowanie komputera (itd. HITACHI-MARATON). Wszystkie elementy muszą być identyczne nie tylko z punktu widzenia statycznego (takie same procesory, pamięci czy karty rozszerzeń) ale i z punktu widzenia dynamiki pracy (taki sam czas wykonania poszczególnych zadań. Wszystkim zarządza fizyczny kontroler klastra. To on dubluje polecenia dla obu maszyn w sposób niewidoczny dla użytkownika. System operacyjny jest instalowany bliźniaczo na obu maszynach. Wszystkie polecenia użytkownika i działania systemowe są wykonywane niezależnie przez dwa identyczne komputery. Jednak na koniec fizyczny kontroler klastra odrzuca efekt działania jednego z nich (użytkownik końcowy widzi tylko jeden komputer). Awaria aktywnej maszyny powoduje rozpoczęcie wykorzystania efektów działania drugiej z nich, bez żadnego opóźnienia. Jest to klaster czysto sprzętowy, zupełnie niezależny od systemu operacyjnego (system operacyjny nawet nie wie, że nastąpiło przełączenie sprzętu). Rozwiązanie to ma wadę w postaci wysokiej ceny specjalizowanego sprzętu i jego nieefektywnym wykorzystaniu (pracują 2 komputery, a w praktyce tylko jeden jest używany). Jednak zaletą jest idealnie bezstratny model pracy (awaria komputera nie wymaga żadnych ponownych działań – wynik „zastępczy" dostępny jest od razu). Ponadto awaria nie powoduje utraty mocy komputera.

Ostatni sposób, to dostępna w Windows usługa klastrowa, która umożliwia przełączanie zasobów pomiędzy kilkoma komputerami, widocznymi w sieci stale pod tą sama nazwą i adresem. Główną cechą tego typu rozwiązania jest zapewnienie maksymalnej dostępności zasobów serwera dla użytkowników końcowych. Jakkolwiek to samo zadanie nie może być wykonywane jednocześnie przez dwa różne komputery, istnieją proste sposoby na zwiększenie wydajności serwera, dzięki wykorzystaniu specyficznych cech tego typu klastra.

3. Klaster według microsoftu

Klasyczny klaster pracujący pod kontrolą Windows 2000 składa się z 2 do 4 komputerów połączonych ze sobą wewnętrzną siecią *heartbeat*, która umożliwia kontrolę procesów sterujących zachowaniem całego klastra. Dodatkowo, specjalnie w celu wymiany i przechowywania ważnych informacji, klaster posiada wydzielony dysk zwany *quorum*, który poza bieżącymi informacjami zawiera historię działania maszyny. Informacje te są wykorzystywane w przypadku awarii jednego z komputerów (drugi jest w stanie na tej podstawie odtworzyć stan wyłączonej maszyny tuż przed wyłączeniem). Poszczególne fizyczne maszyny to węzły klastra, który z zewnątrz jest widziany jako jedna maszyna, posiadająca własną nazwę i adres IP, różny od poszczególnych węzłów. Każdy z węzłów jest też widoczny jako samoistny komputer. Jednak nie to stanowi o sile i możliwościach tej hybrydy.

Zanim przystąpię do opisu rozwiązania, które stanowi największą siłę w klastrach Microsoftu, skupię się na dość trudnej kwestii zasobów dyskowych. Klaster pracujący pod kontrolą Windows powinien być wyposażony we współdzielone zasoby pamięci zewnętrznej (jest to warunek konieczny możliwości utworzenia klastra). Zasoby takie (macierze dyskowe, choć nie widzę powodów, żeby ograniczać to tylko do dysków) należy skonfigurować w taki sposób, aby mogły być dostępne z obu maszyn, bez żadnych sprzętowych przełączeń (tą kwestią zajmuje się producent klastra). Cechy takie niesie ze sobą interfejs SCSI lub FiberChannel. Dyski (w praktyce ze względów bezpieczeństwa najczęściej są to macierze dyskowe – problem ten rozwinę dalej) są udostępniane poszczególnym węzłom klastra na wyłączność, jednak z możliwością przewłaszczenia. To znaczy zarówno na żądanie jak i na skutek awarii dysk X: może przestać należeć do jednego węzła i stać się własnością innego. Jednoczesny dostęp dwóch różnych komputerów do tego samego zasobu w sposób niezależny jest niemożliwy. To znaczy dane zapisane przez jeden z węzłów są niewidoczne dla drugiego tak długo, dopóki jest on właścicielem dysku. Dopiero po przekazaniu dysku drugiemu węzłowi będzie on w stanie zobaczyć te dane. Oczywiście po oddaniu praw własności drugiemu węzłowi, pierwszy przestaje widzieć dane zgromadzone na tym dysku.

Rozwiązanie bazujące na przełączaniu zasobów dyskowych zamiast współdzielenia ich wymusza sposób pracy klastra. Nie jest to rozwiązanie, które dzięki rozproszeniu pracy udostępnia zwiększenie wydajności (co jak dalej pokażę

nie jest do końca prawdą), ale zapewnia maksymalną ciągłość pracy. W przypadku awarii, która wymusza wyłączenie komputera, czyli nie może być usunięta poprzez wykorzystanie nadmiarowości jego elementów, zadanie pierwszego węzła może przejąć drugi. Jeśli konfiguracja jest dokładnie taka sama (a zapewnia to usługa klastrowa) i zasoby danych są identyczne (a są, gdyż są to fizycznie te same dane przełączone do drugiego komputera), wtedy praca może być podjęta po niezwykle krótkiej (czasami niezauważalnej) przerwie, w tym samym miejscu, w którym została przerwana. Warto zwrócić uwagę, że zasoby dyskowe pozornie nie są redundantne, więc stanowią potencjalne zagrożenie dla stabilności systemu. Radą na to jest stosowanie sprzętowych macierzy RAID, które mają nadmiarową liczbę dysków i są zorganizowane w ten sposób, że awaria jednego z nich nie powoduje utraty żadnych informacji. Biorąc pod uwagę możliwość wymiany dysku w czasie pracy i podwójne kontrolery dyskowe wraz z podwójnymi magistralami od macierzy do komputera można uznać, że jest to rozwiązanie bezpieczne. Warto pamiętać, że macierze są konfigurowane najczęściej w trakcie budowy klastra, więc jego przeznaczenie powinno być przemyślane i przeanalizowane przed zakupem (do problemu tego powrócę dalej jeszcze raz).

3.1. Serwery wirtualne

Po ogólnym wyjaśnieniu jak działa sprzęt tworzący klaster przystąpię teraz do opisu środków wykorzystywanych w celu realizacji nieprzerwanej pracy klastra. Jak to już wcześniej zaznaczałem, klaster działający pod kontrolą Windows 2000 jest klastrem przełączeniowym, to znaczy w każdej chwili pracuje tylko część jego zasobów, a reszta nie jest wykorzystywana (ominięcie tego ograniczenia opisuję dalej). W przypadku awarii następuje wyłączenie zepsutej maszyny i podstawienie na jej miejsce drugiej. Aby po przełączeniu się maszyn były one z zewnątrz widziane dokładnie tak samo, należało zastosować rozwiązanie, które zapewniłoby utrzymywanie stałego adresu IP i nazwy komputera. Najprostszym rozwiązaniem było zastosowanie wirtualnych serwerów pracujących w systemie, w sposób niezależny od ustawień sprzętowych. Idea tego rozwiązania bazuje na wydzieleniu z systemu operacyjnego części jego zasobów i przydzieleniu tych zasobów na własność procesowi, który będzie traktowany z zewnątrz jak osobna instancja systemu operacyjnego. Oczywiście nie polega to na instalowaniu Windowsa w innym Windowsie. Jest to po prostu usługa klastrowa, która potrafi wydzielić kilkanaście typów zasobów i przyporządkować je do określonego serwera wirtualnego. Każdy taki wirtualny serwer ma swój adres i nazwę sieciową niezależne od maszyny na której pracuje i od całego klastra.

Podstawą istnienia każdego wirtualnego serwera są zasoby dyskowe. To znaczy serwery takie mogą istnieć nawet bez adresu sieciowego (co oczywiście nie ma większego sensu), ale nie mogą istnieć bez dysku. I nie może to być fragment dysku (partycja) tylko cały dysk (cały w sensie urządzenia SCSI). Tak więc projektując klaster, należy przewidzieć odpowiednią ilość zasobów dyskowych, tak by nie stanąć w pewnym momencie przed problemem niemożności utworzenia kolejnego serwera wirtualnego. W teorii liczba niezależnych dysków powinna być

równa przewidywanej liczbie serwerów plus jeden na *quorum*. W praktyce liczba potrzebnych dysków może się okazać większa. Jeśli zamierzamy wykorzystywać wirtualny serwer jako podstawę pracy serwera SQL, wtedy warto zwiększyć liczbę dostępnych dysków. Klasyczna sugestia mówi, że dla każdego serwera SQL potrzeba przynajmniej jednego dysku typu RAID5 na dane i jednego typu RAID10 na logi. Jeśli serwer jest bardzo obciążony (a z reguły właśnie takie są instalowane na klastrach) wtedy dane możemy umieścić na większej liczbie dysków (logi mogą pozostać na jednym). Tak więc w zależności od planowanego przeznaczenia konfiguracja macierzy dyskowej może być bardzo rozbudowana.

Poza dyskami wszystkie pozostałe zasoby mogą być dzielone między wirtualnymi serwerami w sposób dynamiczny. To znaczy kilka serwerów może korzystać z tej samej karty sieciowej, procesora czy pamięci operacyjnej. Oczywiście zasoby te są odpowiednio dzielone, ale podział ten może następować poprzez zmianę parametrów w systemie operacyjnym (nie trzeba w tym celu zmieniać konfiguracji sprzętowej). Jeśli na przykład na jednej fizycznej maszynie pracują dwa serwery wirtualne, które służą za środowisko pracy różnych programów (o zmiennym obciążeniu procesora), wtedy w zależności od potrzeb procesor jest wykorzystywany naprzemiennie przez różne serwery (tak jak to się dzieje w przypadku zwykłych programów). Podobnie jest z zasobami pamięci czy dostępem do sieci, choć izolacja między serwerami wirtualnymi jest większa, niż pomiędzy zwykłymi programami pracującymi w obrębie standardowego systemu operacyjnego.

3.2. Zachowanie awaryjne

Wiemy już jakie elementy systemowe są używane do zapewnienia nieprzerwanej pracy. Teraz pozostaje wyjaśnić sposób działania, który zapewni jak największą dostępność zasobów serwera. Wyobraźmy sobie klaster, na którym posadowiono jeden serwer wirtualny, a na serwerze tym zainstalowano serwer SQL. Dopóki fizyczna maszyna, na której właśnie pracuje ten serwer (a przez to i SQL) sprawuje się dobrze, nic się nie dzieje. W przypadku jakiejś nieprzewidzianej awarii i przerwy w pracy następuje przerzucenie wirtualnego serwera z jednej maszyny na drugą. Jeśli przełączenie jest przez system przewidziane, wtedy następuje zatrzymanie pracy serwera na jednej maszynie i podjęcie jej na drugim węźle. Gdy pierwszy węzeł psuje się bez ostrzeżenia, wtedy przełączenie trwa trochę dłużej. Po pierwsze drugi węzeł klastra musi zauważyć, że ten pierwszy przestał pracować. Dopiero kiedy to się stanie, usiłuje odtworzyć ostatni stan serwera wirtualnego i ruszyć z pracą dokładnie w tym samym miejscu na drugiej maszynie. Wydłużenie czasu przełączania wynika z konieczności wykrycia fizycznej awarii maszyny i odtwarzania ostatniego stanu pracy. Wszystkie dane, zarówno w czasie planowanego jak i nieplanowanego przełączenia, są pamiętane na wydzielonym w tym celu dysku *quorum*. Nawet jeśli przełączenie wyniknęło z awarii, czas przez który serwer nie jest widoczny (przestał pracować na jednym komputerze, a nie zaczął jeszcze pracy na drugim) nie jest większy niż kilka

sekund. Uznaje się, że dobre serwery klastrowe mają zawodność znacznie poniżej 5 minut w roku, czyli pracują przez ponad 99.999% czasu.

Wróćmy jednak do przełączenia serwera z jednej maszyny na drugą. Jeśli aplikacje korzystające z serwera pracującego na klastrze są odpowiednio napisane, użytkownik końcowy nie jest w stanie zauważyć, że coś się stało. Dla aplikacji klienckiej różnica powinna być jedynie taka, że wynik, który standardowo pokazywał się na ekranie po sekundzie, teraz wyjątkowo pokazał się po kilku sekundach (jeśli pytanie było zadane akurat w chwili przełączania). Użytkownik tej aplikacji nie ma żadnej świadomości o tym, że nastąpiło przełączenie serwera (na przykład SQLowego) z jednej maszyny na drugą. Taka sytuacja jest bardzo korzystna w przypadku instytucji czy firm wymagających pracy bez żadnych przerw (zarówno banków jak i firm handlowych, czy stosujących zaawansowane aplikacje sterujące procesami produkcyjnymi).

Należy jednak zwrócić uwagę na kilka niebezpieczeństw czających się w Microsoftowym sposobie podejścia do niezawodności pracy. Otóż bazowanie niezawodności całego klastra na infrastrukturze sieciowej wymaga bardzo dużej niezawodności tej infrastruktury. Mowa tu zarówno o sprzęcie (trzy karty sieciowe w każdym węźle klastra nie są niczym dziwnym) jak i o serwerze domenowym. Utrata nazwy domenowej serwera wirtualnego może zakończyć się przeniesieniem tego serwera na drugi węzeł klastra, mimo tego, że żadna awaria sprzętu nie nastąpiła. Problemy mogą się też pojawić jako skutek działania programów, które nie do końca zostały przystosowane do pracy w trybie wirtualnym. Generalnie każdą aplikację można uruchomić na serwerze wirtualnym (chyba, że specjalnie została napisana w taki sposób, aby to uniemożliwić). Jednak niektóre zachowania mogą wprowadzać dodatkowe problemy. Na przykład MSSQL 2000, jakkolwiek w wersji Enterprise dość dobrze dostosowany do pracy klastrowej, w czasie przełączania przerywa wykonywanie procedur składowanych i już nie podejmuje ich dokańczania. Podobnie dzieje się z aplikacjami serwerowymi, które tracą połączenie w skutek przełączenia. Przykładem niech będą aplikacje, które ustanawiają komunikację klient-serwer korzystając z innego portu TCP/IP, niż później odbywają standardową pracę. W takim przypadku po przełączeniu serwer czeka na nawiązanie nowych połączeń, a klienci oczekują, że serwer odpowie na wcześniej zainicjowane połączenia. Oczywiście problemy te nie są problemami dostępności serwera, lecz wynikają ze złej konstrukcji programów (nieprzystosowania ich do warunków pracy).

4. Efektywne wykorzystanie zasobów klastra

Opisując sposób działania klastra kilkukrotnie wspominałem, że rozwiązanie Microsoftu w założeniu nie służy do podziału obciążenia na większą liczbę komputerów. Jeśli pozostaniemy w obrębie przykładowego SQL serwera, nie występuje sytuacja, że część pracy wykonywana jest na jednym komputerze, a część na drugim. Jeśli uruchamiamy na przykład procedurę składowaną, wtedy wykona się ona na tylko jednym fizycznym komputerze. Mimo to można

zaprojektować klaster, wraz z jego funkcjonalnością w taki sposób, aby w pewien sposób rozłożyć pracę na 2 komputery (lub więcej, jeśli klaster składa się z większej liczby węzłów). Wymaga to czasami zmian w podejściu do sposobu realizacji zaplanowanych celów. Najczęściej zmiany takie wpływają pozytywnie na pracę firmy nie tylko ze względu na możliwość równoległego przetwarzania danych, ale i lepszą ich organizację. Dalsze rozważania na temat efektywnego wykorzystania zasobów klastrów będą dotyczyły rozwiązania z 2 maszynami, lecz bez trudu można je rozszerzyć do 4 węzłów.

4.1. Prosta eliminacja niewykorzystanych zasobów

Klasyczne rozwiązanie (naturalne przy pierwszym kontakcie z klastrem) przewiduje pracę serwera w trybie *active-pasive*. Oznacza to, że jeden z komputerów fizycznych obsługuje jeden serwer wirtualny. Drugi pracuje w stanie gotowości, bez żadnego obciążenia. Ma on za zadanie przejąć ewentualne obciążenie w przypadku awarii pierwszej maszyny. W takim układzie jedna z maszyn jest całkowicie niewykorzystana. Pozornie rozwiązanie takie jest bardzo dobre, gdyż praca na „biegu jałowym" zapewnia możliwość natychmiastowego przejęcia obciążenia. Oczywiście jest to tylko pozorne, gdyż całe obciążenie musi zostać przeniesione z jednej maszyny na drugą. Gdyby praca podzielona była na 2 komputery, wtedy ewentualna awaria dotyczyłaby tylko części zasobów. Ponadto w przypadku awarii przeniesienie obciążenia na drugą maszynę nie musi dotyczyć wszystkich procesów (część z nich pracuje bezawaryjnie właśnie na tej drugiej maszynie). Opisane rozwiązanie legło u podstaw pracy w architekturze *active-active*. W trybie tym na klastrze instaluje się 2 niezależne serwery wirtualne. Jedną z cech wirtualnego serwera jest domyślna fizyczna maszyna, na której pracuje ten serwer (oczywiście, jeśli jest to możliwe, czyli maszyna ta jest sprawna). Parametry te ustawiamy tak, aby w trakcie standardowej pracy serwery wirtualne pracowały na różnych węzłach, co zapewni rozłożenie obciążenia. W przypadku awarii jednej z maszyn następuje przełączenie i krótkotrwała praca pod zdwojonym obciążeniem (dwa serwery na jednym węźle).

Rozwiązanie takie wymaga oczywiście modyfikacji ideii korzystania z klastra. Jeśli korzystamy z niego, jako z serwera baz danych, architektura taka doskonale się sprawdza w przypadku używania różnych typów danych: zarówno on-line jak i zagregowanych. Wiele analiz (roczna sprzedaż, rentowność, wiarygodność kredytowa, itd., itd.) może być wykonywana na danych nieznacznie nieaktualnych (sprzed godziny, a często nawet z wczorajszego dnia). Można więc zorganizować pracę w następujący sposób: jeden serwer obsługuje aktualną pracę firmy, a drugi zapytania dotyczące danych historycznych. W nocy (a dokładniej wtedy, gdy obciążenie obu serwerów jest najmniejsze) następuje uaktualnianie danych drugiego serwera, najczęściej z ich wstępnym przetwarzaniem. Aktualizacja z agregacją wykonywana jest tak, aby ułatwić i przyspieszyć tworzenie przekrojowych analiz. Oczywiście rozwiązanie takie jest przydatne tylko wtedy, gdy jesteśmy w stanie oddzielić bieżące wykorzystanie danych od działań przekrojowych. W obrębie SQL servera można wykorzystać jego składowe

instalując na jednym z komputerów sam serwer bazodanowy, a na drugim z nich serwer analiz OLAP (wymaga to jednak zastosowania pewnych sztuczek, gdyż z niewiadomych powodów Microsoft nie przewidział w swoim pakiecie takiego oczywistego rozwiązania). Można też próbować podzielić serwer ze względu na strukturę organizacyjną bądź lokalizacyjną. Jeden z serwerów przeznaczyć na przykład do obsługi centrali, a drugi na oddziały regionalne. Jeśli i taki podział jest niemożliwy należy wykorzystać drugi serwer do pracy z innym programem, na przykład do zarządzania pocztą, czy innym obciążającym zasoby procesem.

4.2. Modyfikacje sprzętowe

Według moich doświadczeń trudno znaleźć firmę, w której nie da się zastosować podziału zasobów serwera w taki sposób, aby architektura *active-active* była niemożliwa do zastosowania. Zdarzają się jednak takie przypadki, gdy klaster jest instalowany w już istniejących warunkach, z trudno modyfikowalnym oprogramowaniem specjalizowanym bądź niszowym. Takie przypadki powinno się zauważyć na etapie projektowania klastra. Rozwiązaniem, które uchroni przyszłego użytkownika klastra przed nieefektywnym wydawaniem pieniędzy, jest niesymetryczna architektura *active-passive*. Jeśli zasobów nie da się podzielić na dwa jednocześnie pracujące serwery wirtualne, można się zdecydować na klaster niesymetryczny. Jeden węzeł, ten który będzie domyślnie komputerem aktywnym, powinien być maszyną znacznie silniejszą od komputera „zapasowego", który przez większość swojego życia będzie tylko czekał na krótkotrwałe obsłużenie ewentualnej awarii. Niesymetryczne rozwiązanie pozwoli nam zaoszczędzić przynajmniej 20 procent kosztów, co przy cenie klastra nie jest małą sumą. Alternatywnym efektem (jeśli firmie nie zależy tak bardzo na oszczędnościach) może być podniesienie wydajności głównej maszyny. Zamiast kupować 2 serwery czteroprocesorowe można kupić jeden ośmioprocesorowy i drugi czteroprocesorowy, ale o niższej wydajności (niższa szybkość, mniej pamięci).

Na niesymetrycznym klastrze *active-passive* instalujemy tylko jeden serwer wirtualny. W czasie standardowej pracy funkcjonuje on na silniejszej maszynie. Po awarii zostaje przeniesiony na słabszy węzeł i działa tam aż do usunięcia przyczyny wyłączenia komputera głównego. Przy takim rozwiązaniu, już na etapie zakupu serwera, należy zadbać o właściwą ochronę gwarancyjną. W przypadku serwerów symetrycznych przełączenie między poszczególnymi węzłami nie wiąże się z żadnym niebezpieczeństwem spowolnienia pracy. Klaster niesymetryczny po przełączeniu się na komputer zapasowy może spowolnić pracę firmy. Należy więc zadbać, aby spowolnienie takie trwało jak najkrócej.

4.3. Inne rozwiązania

Poza zaproponowanymi, najciekawszymi według mnie, rozwiązaniami można oczywiście stosować inne, bardziej lub mniej klasyczne. Najczęściej doświadczenie z klastrami rozpoczyna się od symetrycznej architektury *active-passive*, czyli symetryczny klaster z jednym serwerem wirtualnym. Rozwiązanie

takie może być przejściowe do momentu przejścia na architekturę *active-active*. Należy jednak pamiętać, aby przewidzieć odpowiednią konfigurację macierzy dyskowej, umożliwiające w przyszłości taką migrację. Inne, bardziej zaawansowane rozwiązanie to niesymetryczna architektura *active-active*, czyli dwa serwery wirtualne domyślnie pracujące na różnych maszynach (ten bardziej obciążony na silniejszym węźle). Bardzo rzadko stosuje się też rozwiązanie *passive-passive*, w zasadzie tylko w przypadku budowy klastra z klastrów. Jest to jednak wyższa szkoła jazdy i pominę opis tego rozwiązania. Oficjalna literatura sugeruje też możliwość pracy hybrydowej. Klaster jest tam wykorzystywany na 2 sposoby. Standardowo pracują na nim serwery wirtualne. Dodatkowo węzły klastra są używane jako zwykłe pojedyncze serwery (pełniąc rolę na przykład print serwerów czy serwerów DNS). Jakkolwiek oficjalna literatura podaje takie rozwiązanie jako możliwe osobiście uważam je za nie najlepsze. Usługę klastrową można zastosować też dla pojedynczego komputera, wyłącznie w celu uruchomienia serwerów wirtualnych. Umożliwia to na przykład przeniesienie zasobów kilku rzeczywistych serwerów na kilka wirtualne, a jeden rzeczywisty. Rozwiązanie takie zapewnia w pewien sposób wzrost efektywności, który wynika z faktu, że obciążenie różnych komputerów niemal zawsze jest nierównoczesne. Zgromadzenie zasobów czterech jednakowych komputerów w jednym, który jest tylko 3 razy silniejszy niemal zawsze powoduje zwiększenie wydajności każdego z nich. Jest to rozwiązanie dość ciekawe ze względu na efektywność wykorzystania sprzętu, ale nie mające z niezawodnością nic wspólnego.

5. Podsumowanie

W dzisiejszym artykule starałem się przekazać czytelnikom podstawową wiedzę na temat podnoszenia niezawodności pracy z wykorzystaniem klastrów tworzonych z użyciem Windows 2000. Z konieczności jest to wiedza bardzo ogólna. Same materiały szkoleniowe Microsoftu na ten temat to kilkaset stron dokumentacji. Wiele też można dowiedzieć się z artykułów technicznych niezależnych ekspertów. Jednak prawdziwą wiedzę na temat możliwości i przydatności klastra można zdobyć najwcześniej po kilku miesiącach rzeczywistej pracy. Kupując klaster warto więc zostawić sobie furtkę w postaci możliwości wymiany bądź rozbudowy klastra w ciągu pewnego okresu od daty zakupu. Biorąc pod uwagę cenę klastra warto też poświęcić trochę czasu i pieniędzy na analizę przeprowadzoną przez niezależnych ekspertów. Mimo trudnych początków warto jednak pamiętać, że efektem końcowym jest podniesienie niezawodności pracy firmy, co niemal zawsze przekłada się na finansowe zyski.

Literatura

1. *Implementing Microsoft Windows 2000 Clustering*, Microsoft, 2087ACP
2. Zawadzki M., *Wespół w zespół*, PCkurier 1 maja 2000
3. Krysowski O., *Większa moc mniejszym kosztem*, Teleinfo 18 listopad 2002

Marek Wierzbicki, MOTTE
ul. Aleksandrowska 106/22
91-224 ŁÓDŹ
motte@polbox.com
marek.wierzbicki@azymut.pl

ZABEZPIECZENIE INFORMACJI PRZESYŁANYCH Z ZASTOSOWANIEM KRYPTOGRAFII I INFRASTRUKTURY KLUCZA PUBLICZNEGO

Franciszek WOŁOWSKI

Streszczenie:
Elektroniczna wymiana informacji jest obecnie bardzo pożądana, lecz wymaga zapewnienia starannych i silnych zabezpieczeń. Wymagania te może spełnić kryptografia klucza publicznego i związana z nią infrastruktura PKI. Usługi kryptograficzne oferowane przez tą nową technologię zapewniają poufność, integralność, wiarygodność oraz niezaprzeczalność. Aby jednak skorzystać z dobrodziejstw tych nowych metod należy uzyskać certyfikat a zwłaszcza wiedzę o możliwościach jego użycia. Gdy uda się przebrnąć przez bariery techniczną i psychologiczną możemy liczyć na znaczne podniesienie poziomu bezpieczeństwa i wymierne korzyści materialne.

1. Wprowadzenie

Potrzeba wymiany informacji poprzez publiczne sieci rozległe w obecnej dobie jest bardzo pożądana i powszechna. Zwłaszcza firmy zdają sobie sprawę, że jeśli nie zaistnieją w internecie mogą w ogóle zniknąć z rynku gdyż transakcje realizowane klasycznymi metodami są o wiele bardziej kosztowne od tych, które są realizowane przez internet a ponadto są uciążliwe i zabierają wiele czasu. Również dla osób fizycznych korzystanie z internetu jest bardzo wygodne i atrakcyjne.
Chcemy używać internetu do:

-Wysyłania poczty elektronicznej
–Dokonywania zakupów
–Dystrybucji oprogramowania
–Przesyłania danych wrażliwych lub prywatnych/osobowych

Ale mamy pewne obawy - w jaki sposób:

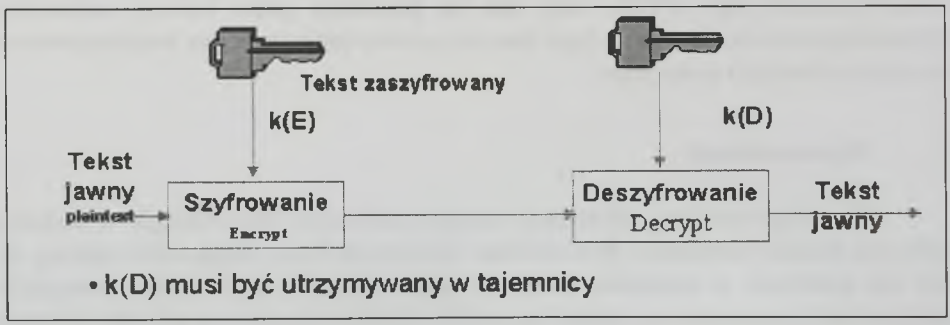
–Stwierdzić, że osoba jest tą za którą się podaje?
–Stwierdzić, że połączyłem się z właściwym sprzedawcą/handlowcem?
–Zapewnić prywatność wymiany informacji
–Stwierdzić, czy naruszono integralność informacji?
–Dowieść, że ktoś przesłał nam określoną wiadomość

Nowe metody kryptograficzne w oparciu, o które realizowany jest podpis elektroniczny mogą zasadniczo wpłynąć na wszystkie czynności związane z wymianą informacji gdyż w sposób bardzo istotny może obniżyć koszty jej przekazywania, skrócić czas niezbędny na ich przekazanie a co chyba najistotniejsze zabezpieczyć procesy związane z wymianą informacji zapewniając, że informacja nie zmieni ani treści ani formy i że osoby niepowołane nie będą mogły zapoznać się z jego treścią. Aby jednak można było w sposób bezpieczny

stosować podpis elektroniczny trzeba uzyskać certyfikat, pozyskać umiejętności jego wykorzystania (dobrze byłoby zapoznać się z podstawowymi zasadami, bez zbędnych szczegółów) oraz przystosować swoje środowisko. To wszystko jest przedmiotem niniejszej publikacji.

2. Podstawy kryptografii w (łatwo strawnej) pigułce:

Możliwość składania podpisu elektronicznego została udostępniona przez **kryptografię klucza publicznego**. Systemy kryptograficzne są przedmiotem studiów dwóch działów nauki: **kryptografii** i **kryptoanalizy**, które wspólnie nazywane są **kryptologią**. Tematyka kryptoanalizy, czyli nauki o metodach łamania szyfrów, która wnosi bardzo istotny wkład do kryptologii gdyż weryfikuje poprawność i siłę mechanizmów stworzonych przez kryptografię - nie zostanie poruszona w niniejszej publikacji.



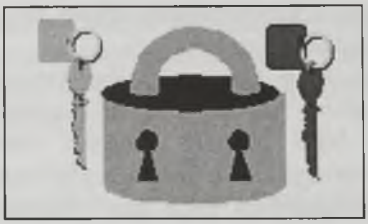
Procesy szyfrowania i deszyfrowania

Słowo kryptografia pochodzi z greckiego „Ukryta wiadomość (słowo)" Kryptografia to nauka, która udostępnia różnego rodzaju metody/systemy szyfrowania, które przez wiele wieków właściwie dawały możliwość uzyskania jedynie **poufności,** czyli zapobiegały możliwości zapoznania się z treścią wiadomości przez osobę, dla której wiadomość ta nie została przeznaczona.
Już Cezar pisał do Cicero i jego przyjaciół w Rzymie ponad 2000 lat temu, używając szyfru w którym każda litera jawnego tekstu była zastępowana przez trzecią literę (sposób szyfrowania =ALGORYTM) alfabetu łacińskiego (ciąg liter o znanym uporządkowaniu= KLUCZ), słowo CAESAR zastępowane jest przez FDHVDU. Problemem w tym przypadku był fakt, że każdy z przyjaciół Cezara znał zarówno algorytm jak i klucz a więc jeśli któryś z nich uzyskał dostęp do informacji dla niego nie adresowanej mógł ją odczytać. Aby temu zapobiec Cezar powinien wymienić się z każdym z przyjaciół innym kluczem a więc innym uporządkowaniem wszystkich liter. A więc bezpieczeństwo tego typu systemu kryptograficznego polega na tym aby każda para respondentów wymieniła się innym kluczem, który będą z sobą współdzielić i starannie go chronić przed innymi. Te systemy kryptograficzne nazywa się **systemami symetrycznymi** gdyż

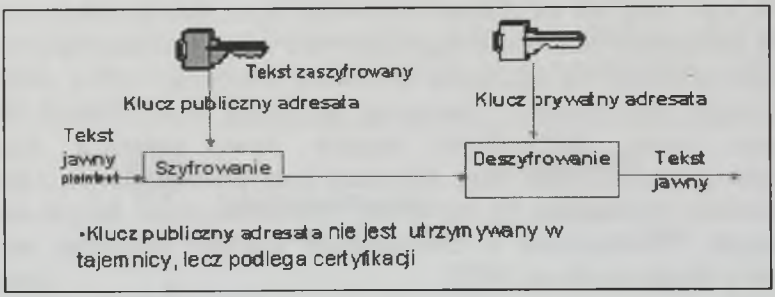
zarówno do szyfrowania jak i deszyfrowania używa się tego samego klucza, czyli takiego samego wyrazu/zdania (ciągu znaków).

Obecnie najpowszechniej używanym algorytmem jest system DES, który w 1977 r. został przyjęty przez Narodowe Biuro Standardów USA (ang. National Bureau of Standards – NBS obecnie National Institute of Standards and Technology - NIST. jako amerykańska norma szyfrowania danych, opublikowana w dokumencie FIPS 46. Algorytm DES jest przeznaczony do szyfrowania i odszyfrowywania 64-bitowych bloków danych. Blok wyjściowy (zaszyfrowany) ma taką samą długość jak blok wejściowy (poddawany szyfrowaniu). stał się obiektem wielu badań dotyczących jego bezpieczeństwa. W trakcie tych badań nie znaleziono żadnych cech świadczących o jego ewentualnych słabościach, z uwagi jednak na zbyt małą jak na obecne możliwości obliczeniowe przestrzeń kluczy, zaleca się zaniechanie stosowania tego algorytmu w jego podstawowej postaci. W jego miejsce proponuje się najczęściej użycie tzw. potrójnego DES-a, polegającego na trzykrotnym zaszyfrowaniu wiadomości za pomocą trzech różnych kluczy. W najbliższym czasie do praktyki wejdzie nowa generacja algorytmów symetrycznych o nazwie AES (ang. Advanced Encryption Standard), Algorytmom tym postawiono wymagania, by szyfrowały 128-bitowe bloki danych za pomocą 128-bitowych, 192-bitowych i 256-bitowych kluczy. Algorytmy te zostały wyłonione w drodze konkursu NIST.

Mimo swojej skuteczności oraz ogromnego zapotrzebowania na ochronę informacji, zasięg kryptosystemów symetrycznych pozostał przez tysiąclecia mocno ograniczony do dosyć hermetycznych kręgów, ponieważ głównym problemem związanym z ich użyciem jest konieczność wcześniejszego (poprzedzającego transmisję wiadomości) uzgodnienia klucza pomiędzy każdą parą potencjalnych korespondentów. Ze względu na to, że klucz ma być znany tylko danej parze nadawca-odbiorca, więc jego uzgodnienie musi odbywać się przy użyciu specjalnego, bezpiecznego kanału łączności (np. przy użyciu kuriera). Stwarza to ogromne problemy, zwłaszcza w dużych sieciach, gdyż liczba koniecznych do uzgodnienia kluczy rośnie proporcjonalnie do kwadratu liczby użytkowników sieci (dokładna zależność to:N(N-1)/2 gdzie N oznacza liczbę użytkowników. Przy 100 użytkownikach liczba koniecznych do wymiany kluczy wynosi 4500). Dodatkowo sprawę komplikuje fakt, że ze względów bezpieczeństwa klucze te powinny być możliwie często zmieniane. Problemem też jest zapewnienie właściwej ochrony tym sekretnym kluczom.



Analogia do kłódki



Proces szyfrowania i deszyfrowania w systemie asymetrycznym

Te zasadnicze niedogodności usunęła kryptografia klucza publicznego, która zupełnie zmieniła podejście do szyfrowania, ponieważ używa dwu kluczy stanowiących wzajemnie zależną parę a ich szczególna cecha to to, że jeśli wiadomość zostanie zaszyfrowana przy użyciu jednego z tych kluczy to rozszyfrowanie może nastąpić tylko przy pomocy tego drugiego klucza. To tak jakby kłódka z dwoma kluczami, jednym się tą kłódkę zamyka, ale otworzyć można tylko tym drugim. Jeden z tych kluczy nazwano kluczem publicznym, gdyż udostępnia się go osobom, z którymi chce się wymieniać informacje a drugi to klucz prywatny gdyż pod żadnym pozorem nie można go udostępniać i trzeba zapewnić mu ochronę na miarę wartości/ważności transakcji, do jakich chcemy go stosować. Jednakże i w tym przypadku jest zasadnicza różnica pomiędzy dwoma systemami kryptograficznymi, gdyż za ochronę klucza prywatnego bierze jednoznaczną odpowiedzialność tylko jedna strona. W przypadku, bowiem kryptografii symetrycznej naruszenie ochrony sekretnego klucza może nastąpić przez obie strony a wykrycie, kto był za to odpowiedzialny może nastręczać trudności.

Najbardziej znanym i rozpowszechnionym algorytmem kryptosystemu asymetrycznego jest RSA (od nazwisk twórców: Rivest, Shamir i Adelman). Kryptosystem RSA jest systemem opublikowanym w 1978 roku. Jako jeden z nielicznych kryptosystemów asymetrycznych oparł się próbom ataku i stał się de facto standardem, używanym powszechnie m.in. do tworzenia podpisów cyfrowych.

W algorytmie RSA ciąg bitów podawanych na wejście przekształcenia szyfrującego jest traktowany jako duża (rzędu kilkuset cyfr dziesiętnych) liczba całkowita. Rolę klucza prywatnego pełni trójka dużych liczb całkowitych SK=(d,p,q), rolę klucza publicznego para liczb PK=(e,N).
Liczby p i q są dużymi liczbami pierwszymi.
Liczba N jest iloczynem liczb p i q, $N=p\times q$.
Liczby e i d związane są zależnością: $e\times d=1 \bmod ((p-1)(q-1))$.

Algorytm jest przeznaczony do szyfrowania i odszyfrowywania n-bitowych bloków danych (gdzie n oznacza długość liczby N). Blok wyjściowy (zaszyfrowany) ma taką samą długość jak blok wejściowy (poddawany szyfrowaniu), tzn. n bitów. Przyjęta obecnie za bezpieczną i powszechnie używana wartość parametru *n* wynosi 1024 bity.

Przekształcenie szyfrujące wyznaczone przez klucz publiczny jest proste obliczeniowo i polega na podniesieniu liczby reprezentującej wiadomość M do potęgi e określonej przez klucz publiczny, modulo liczba N stanowiąca drugi element klucza publicznego.

Wykonanie przekształcenia odwrotnego, tzn. odtworzenie wiadomości M na podstawie znajomości szyfrogramu C i klucza publicznego PK=(e,N), a bez znajomości klucza prywatnego SK=(d,p,q), jest obliczeniowo niemożliwe przy aktualnym stanie techniki obliczeniowej. Znajomość klucza prywatnego pozwala na "zamknięcie zapadki" i wykonanie przekształcenia odwrotnego (odszyfrowującego) w czasie porównywalnym do czasu potrzebnego na wykonanie przekształcenia prostego (szyfrującego).

3. Usługi kryptograficzne

Zastosowanie kryptosystemu asymetrycznego umożliwia uzyskanie nie tylko poufności, ale stwarza również odpowiednie metody do praktycznego uzyskania **integralności,** czyli zapewnienia, że wysłana wiadomość dotrze do adresata w postaci niezmienionej, **wiarygodności,** czyli zapewnienia, że wiadomość zastała wysłana przez osobę, która się podpisała pod tą wiadomością oraz **niezaprzeczalności,** czyli że osoba, która wysłała/otrzymała wiadomość nie może zaprzeczyć, że dokonała tej czynności.

Wymienione powyżej cztery rodzaje zabezpieczeń nazywa się **usługami kryptograficznymi**.

Poufność uzyskuje się w ten sposób, że nadawca szyfruje wiadomość kluczem publicznym adresata a ten rozszyfrowuje ją swoim kluczem prywatnym.

Z uwagi jednak na małe szybkości szyfrowania i deszyfrowania w kryptosystemach asymetrycznych – są one do tysiąca razy wolniejsze od tych samych operacji kryptosystemu symetrycznego- najczęściej stosuje się szyfrowanie i deszyfrowanie hybrydowe co oznacza że operacje szyfrowania i deszyfrowania są wykonywane w kryptosystemach, symetrycznych stosując jednorazowy (sesyjny) klucz. Ten klucz jest chroniony w trakcie jego przekazywania poprzez szyfrowanie kluczem publicznym odbiorcy a odbiorca rozszyfrowuje go swoim kluczem prywatnym w ten sposób uzyskując klucz sesyjny w postaci jawnej i może rozszyfrować otrzymaną wiadomość.

Każda z usług chroniących przed zmodyfikowaniem lub sfałszowaniem przesyłanej wiadomości, czyli przed tzw. podsłuchem aktywnym (usługi integralności, uwierzytelnienia i niezaprzeczalności) polega na dołączeniu do transmitowanej w niezmienionej postaci wiadomości pewnego ciągu bitów. Ciąg ten, będący zawsze funkcją wiadomości, otrzymywany jest w różny sposób, ma różne długości i różne nazwy dla każdej z trzech podstawowych usług: integralności, uwierzytelnienia i niezaprzeczalności.

Dla usługi ***integralności*** ciąg ten to **skrót wiadomości** (*ang. message digest*). Używa się też nazw **znacznik integralności wiadomości** lub **znacznik wykrywalności modyfikacji** i ich angielskich akronimów **MIC** (*ang. Message Integrity Code*), **MDC** (*ang. Modification Detection Code*). Znacznik ten jest otrzymywany jako rezultat działania omówionej silnej funkcji skrótu h(M).
Realizacja usługi integralności polega:

- po stronie nadawcy - na dołączeniu do wiadomości obliczonego na jej podstawie skrótu;
- po stronie odbiorcy - na porównaniu skrótów wyliczonych po stronie nadawczej i odbiorczej - MDC i MDC'.

Ze względu na własności silnej funkcji skrótu wykrywana jest w ten sposób zmiana choćby jednego bitu przesyłanej wiadomości.

Realizowana w opisany powyżej sposób usługa integralności nie zapewnia żadnego zabezpieczenia przed możliwością podszycia się przez intruza pod innego użytkownika sieci. Każdy z użytkowników sieci (a więc w szczególności intruz) może wygenerować skrót do utworzonej przez siebie fałszywej wiadomości, w której przedstawi się on jako ktoś zupełnie inny. Dlatego, usługa integralności nie jest nigdy realizowana samodzielnie, lecz zawsze jako część składowa usług uwierzytelnienia i niezaprzeczalności.

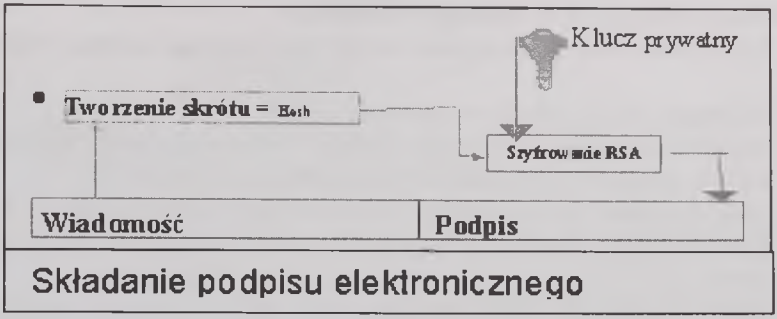
Usługa ***uwierzytelnienia*** polega na dołączeniu do wiadomości ciągu bitów zwanego **znacznikiem uwierzytelnienia wiadomości** (*ang. Message Authentication Code*). Bardzo często zamiast pełnej nazwy używa się akronimu **MAC**.

MAC jest obliczany jako funkcja wiadomości oraz tajnego klucza znanego tylko nadawcy i odbiorcy. Tylko nadawca i odbiorca są w stanie obliczyć prawidłową wartość znacznika, odpowiadającą danej wiadomości.

Weryfikacja autentyczności wiadomości polega na obliczeniu przez odbiorcę znacznika uwierzytelnienia jako funkcji wiadomości odebranej M' i porównaniu znaczników wyliczonych po stronie nadawczej i odbiorczej.

Niezgodność tych znaczników świadczy o tym, że albo wiadomość, albo dołączony do niej znacznik zostały zmienione w czasie transmisji.

Realizacja ***niezaprzeczalności*** polega po stronie nadawczej na dołączeniu do wiadomości obliczonego na jej podstawie podpisu cyfrowego i weryfikacji tego podpisu po stronie odbiorczej.



Upraszczając można powiedzieć, że usługi integralności, uwierzytelnienia i niezaprzeczalności uzyskuje się stosując podpis elektroniczny.

4. Podpis elektroniczny

Podpis elektroniczny to zaszyfrowana wiadomość - lub częściej jej skrót - kluczem prywatnym. Tym razem szyfrowanie dokonywane jest przy pomocy klucza prywatnego a więc każdy kto posiada klucz publiczny, stanowiący parę z tym kluczem prywatnym może dokonać rozszyfrowania wiadomości.. Czyli występuje tu analogia do podpisu odręcznego gdzie tylko osoby, które znają ten podpis mogą go zidentyfikować lecz same nie mogą go stworzyć. Podpis ten jest nie do podrobienia, identyfikuje jednoznacznie osobę, która złożyła ten podpis. W odróżnieniu od podpisu odręcznego podpis elektroniczny odnosi się do całego dokumentu (nie tylko strony na której został złożony) i dotyczy tylko tego dokumentu. Podpis tej samej osoby pod innym dokumentem jest całkowicie innym podpisem choć złożony przy użyciu tego samego klucza prywatnego zapewnia jednoznaczną identyfikację osoby składającej podpis. Jedyną właściwością jakiej nie posiada podpis elektroniczny to możliwość wnioskowania o charakterze i upodobaniach osoby składającej podpis.

Ustawa z dnia 18 września 2001 r. Definiuje podpis elektroniczny *jako „dane w postaci elektronicznej, które wraz z innymi danymi, do których zostały dołączone lub z którymi są logicznie powiązane, służą do identyfikacji osoby składającej podpis elektroniczny"*

Podpis cyfrowy związany z daną wiadomością, zapamiętany w osobnym pliku może służyć do późniejszego udowodnienia autorstwa wiadomości wobec osoby trzeciej - sędziego. Weryfikacja przeprowadzana jest w sposób analogiczny do sprawdzania integralności i autentyczności wiadomości przez odbiorcę. Pozytywna weryfikacja wyklucza możliwość sfałszowania wiadomości przez kogokolwiek, wliczając w to odbiorcę wiadomości.

Najczęściej obecnie stosowane metody generowania podpisu obejmują trzy fazy:

1. obliczenie silnego skrótu wiadomości - MDC;
2. rozszerzenie MDC do postaci EMDC o długości równej długości bloku wejściowego asymetrycznego przekształcenia szyfrującego;
3. zaszyfrowanie rozszerzonego skrótu EMDC (otrzymanego w p.2) przy użyciu przekształcenia szyfrującego wyznaczonego przez klucz prywatny nadawcy.

Skrót wiadomości h(M) otrzymywany jest w wyniku przekształcenia wiadomości przy użyciu funkcji skrótu h. Użycie funkcji skrótu pozwala podpisywać sam skrót wiadomości MDC=h(M), zamiast wiadomości, co znacznie poprawia wydajność algorytmów, zwłaszcza dla długich wiadomości

Skrót wiadomości, który podlega szyfrowaniu kluczem prywatnym w celu otrzymania podpisu elektronicznego, uzyskuje się stosując specjalną metodę matematyczną, od której wymaga się spełnienia następujących, niezwykle ostrych, wymagań:

- funkcja *h* jest określona dla wiadomości *M* o dowolnej długości;
- w wyniku przekształcenia *h* otrzymujemy skrót wiadomości *MDC=h(M)* o ustalonej długości *nh* bitów;
- skrót wiadomości *h(M)* jest łatwy do obliczenia;
- jest obliczeniowo niemożliwe odtworzenie wiadomości *M* na podstawie znajomości jej skrótu *h(M)*;
- funkcja *h* jest wolna od kolizji, tzn. znając funkcję *h* jest obliczeniowo niemożliwe wygenerowanie takiej pary wiadomości *(M,M')*, dla której *M_M'* i *h(M)=h(M')*.

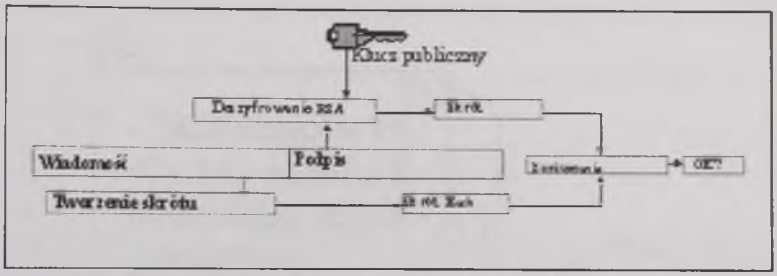
Funkcja skrótu nie zależy od żadnego tajnego parametru i stąd jej wartość może być wyliczona przez każdego użytkownika.
Skrót wiadomości można stworzyć, posługując się technikami opisanymi w ISO/IEC 10118-1, 10118-2 i 10118-3.

Adresat otrzymuje, więc przesyłkę zawierającą samą wiadomość i podpis, czyli zaszyfrowany skrót wiadomości.

5. Weryfikacja podpisu elektronicznego i Integralności

Weryfikacja podpisu elektronicznego i Integralności polega na porównaniu dwu skrótów, skrótu uzyskanego z rozszyfrowania podpisu kluczem publicznym składającego podpis i skrótu uzyskanego z operacji skracania dokonanej na otrzymanym tekście wiadomości. Jeśli te skróty są identyczne oznacza to, że

podpis złożyła osoba, do której należy klucz publiczny, którym rozszyfrowano skrót, oznacza to również, że otrzymana wiadomość nie została nawet w najmniejszym stopniu zmieniona.



Weryfikacja podpisanej wiadomości

Opisanej wcześniej sformalizowanej metodzie tworzenia podpisu odpowiada następujący sposób jego weryfikacji:

1. na podstawie otrzymanej wiadomości M' jest obliczany silny skrót wiadomości MDC';
2. silny skrót z wiadomości odebranej jest rozszerzany do EMDC' wg takiego samego algorytmu, jaki był używany po stronie nadawczej;
3. z podpisu dołączonego do wiadomości, jest odtwarzany (poprzez odszyfrowanie podpisu z użyciem klucza publicznego nadawcy) rozszerzony, silny skrót z wiadomości obliczony po stronie nadawczej (EMDC);
4. skróty wyliczane po stronie nadawczej EMDC i odbiorczej EMDC' są porównywane ze sobą.

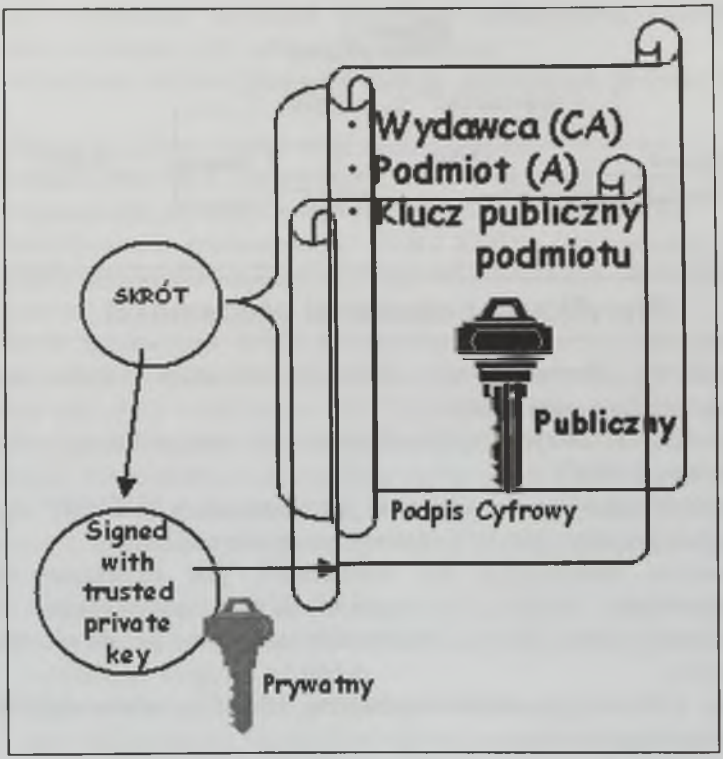
Jeżeli skróty są różne to może to świadczyć m.in. o tym, że:

a) wiadomość lub dołączony do niej podpis zostały zmienione w czasie transmisji lub
b) nadawca wiadomości nie jest tym za kogo się podaje.

6. Certyfikat cyfrowy/elektroniczny (Digital Certificate).

Jak widać użycie pary kluczy znakomicie zabezpiecza wymianę informacji. Pojawia się jednak problem zaufania do osoby, która przekazuje nam klucz publiczny, zwłaszcza, gdy realizuje to zdalnie i do tego jest to osoba nam nie znana. Czyli pojawia się problem: **Jak stwierdzić, że osoba jest tą za którą się podaje?** Aby móc jednoznacznie stwierdzić tożsamość osoby, która przedkłada swój klucz publiczny, który ma służyć do zabezpieczania wymiany informacji, musi zaistnieć trzecia strona, która będzie cieszyć się zaufaniem, najlepiej zaufaniem powszechnym a przynajmniej zaufaniem stron, które mają zamiar zabezpieczać proces wymiany informacji stosując klucze publiczne. Podstawowym zadaniem trzeciej strony w tym przedmiocie jest poświadczanie (w formie

certyfikatu) związku pomiędzy kluczem publicznym a identyfikatorem jego właściciela.. Głównym zagrożeniem jest bowiem wystawienie certyfikatu poświadczającego nieprawdziwy związek i umożliwiającego atak zwany maskaradą.



Trzecia strona weryfikuje, więc tożsamość osoby i składa swój podpis elektroniczny jako poświadczenie pozytywnej weryfikacji. Jeśli ta Zaufana Trzecia Strona (ang. Trusted Third Party - TTP) jest jeszcze zgodnie z ustawą o podpisie elektronicznym kwalifikowanym podmiotem świadczącym usługi certyfikacyjne to taki certyfikat jest certyfikatem kwalifikowanym i zgodnie z ustawą o podpisie elektronicznym jego właściciel może składać podpisy równoważne z odręcznymi stosując swój klucz prywatny stanowiący parę z kluczem publicznym zawartym w tym certyfikacie.

Według ustawy o podpisie elektronicznym Kwalifikowany certyfikat zawiera co najmniej następujące dane:

1) numer certyfikatu,
2) wskazanie, że certyfikat został wydany jako certyfikat kwalifikowany do stosowania zgodnie z określoną polityką certyfikacji,
3) określenie podmiotu świadczącego usługi certyfikacyjne wydającego certyfikat i państwa, w którym ma on siedzibę oraz numer pozycji w rejestrze kwalifikowanych podmiotów świadczących usługi certyfikacyjne,

4) imię i nazwisko lub pseudonim osoby składającej podpis elektroniczny; użycie pseudonimu musi być wyraźnie zaznaczone,
5) dane służące do weryfikacji podpisu elektronicznego,
6) oznaczenie początku i końca okresu ważności certyfikatu,
7) poświadczenie elektroniczne podmiotu świadczącego usługi certyfikacyjne, wydającego dany certyfikat,
8) ograniczenia zakresu ważności certyfikatu, jeżeli przewiduje to określona polityka certyfikacji,
9) ograniczenie najwyższej wartości granicznej transakcji, w której certyfikat może być wykorzystywany, jeżeli przewiduje to polityka certyfikacji lub umowa, o której mowa w art. 14 ust. 1.

Ustawa o podpisie elektronicznym
certyfikat - elektroniczne zaświadczenie, za pomocą którego dane służące do weryfikacji podpisu elektronicznego są przyporządkowane do osoby składającej podpis elektroniczny i które umożliwiają identyfikację tej osoby,
kwalifikowany certyfikat - certyfikat spełniający warunki określone w ustawie, wydany przez kwalifikowany podmiot świadczący usługi certyfikacyjne, spełniający wymogi określone w ustawie

Najpowszechniej stosowana struktura pól certyfikatu jest określona standardem ITU X509v3 poniżej przedstawiono istotne pola

Wersja (version) - określa wersję certyfikatu.

Nr seryjny (Serial number) – jest to niepowtarzalny(w ramach domeny) identyfikator certyfikatu, nadany przez urząd CA

Sygnatura (Signature) – pole to zawiera identyfikatory algorytmu i funkcji skrótu (haszującej), które posłużyły do podpisania certyfikatu przez urząd CA.

Wystawca (Issuer) – w polu tym przechowywana jest nazwa wyróżniająca (DN[1]) urzędu CA, który wygenerował certyfikat.

Okres ważności (Validity) – jest to przedział czasu, w którym urząd CA gwarantuje, że certyfikat będzie ważny i gotowy do użycia (Oczywiście zgodnie

[1] DN – *Distinguished Name*

z jego przeznaczeniem i umową zawartą pomiędzy CA a odbiorcą końcowym). Dane zawarte w tym polu są sekwencją dwóch dat:

notBefore (nie wcześniej niż) – data rozpoczynająca okres ważności

notAfter (nie później aniżeli) – data kończąca ten okres

Podmiot (Subject) – zawarta tu jest nazwa wyróżniającą (DN) podmiotu będącego właścicielem certyfikatu.

Informacja o kluczu publicznym podmiotu (Subject Public KeyInfo) - w polu tym obok klucza publicznego właściciela certyfikatu znajduje się także identyfikator algorytmu, który posłużył do wygenerowania tegoż klucza.

Unikalny identyfikator wystawcy (Issuer Unique Identifier) – w przypadku, gdy występuje pewna liczba urzędów CA o tej samej nazwie, unikalny identyfikator zawarty w tym polu ma umożliwić jednoznaczną identyfikację urzędu.

Unikalny identyfikator podmiotu (SubjectUniqueIdentifier) - w przypadku, gdy występuje pewna liczba posiadaczy certyfikatów o tej samej nazwie, unikalny identyfikator zawarty w tym polu ma umożliwić jednoznaczną identyfikację podmiotu.

Rozszerzenia (Extensions) – pole to zostało wprowadzone wraz z 3 wersją X.509. Dzięki niemu elastyczność oraz funkcjonalność certyfikatów wzrosła w sposób znaczący. Rozszerzenia można zdefiniować jako zbiór dodatkowych informacji, które dodane do certyfikatu mówią nam o jego przeznaczeniu, ograniczeniach i innych cechach. Znaczące jest to, że obok rozszerzeń standardowych (Standard Extensions), użytkownicy mogą definiować swoje własne rozszerzenia (Private Extensions), przez co funkcjonalność certyfikatu bardziej zbliżona jest do ich oczekiwań. Każde rozszerzenie może być oznaczone jako krytyczne(critical) lub niekrytyczne(non-critical). Jest to o tyle istotne, że rozszerzenia krytyczne muszą być bezwzględnie przetwarzane przez aplikacje/systemy. Jednak gdy zajdzie sytuacja, w której system nie będzie w stanie rozpoznać rozszerzenia, to certyfikat zostanie odrzucony. Rozszerzenia oznaczane jako niekrytyczne nie sprawiają takich problemów, gdyż w przypadku ich nierozpoznania są po prostu ignorowane.

Maksymalna liczba rozszerzeń jest definiowana w czasie projektowania struktury certyfikatu. Tylko jedna instancja danego rozszerzenia może wystąpić w certyfikacie.

Każde rozszerzenie jest sekwencją trzech informacji:

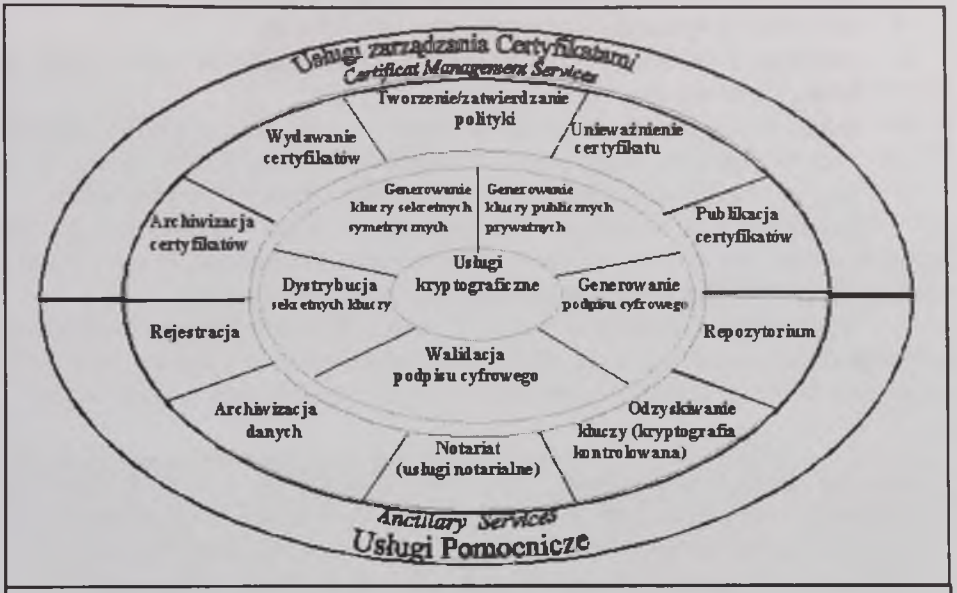
- identyfikatora – extnID
- oznaczenia ważności – critical/non-critical
- wartości – extnValue

7. Budowa systemu zaufania - Zaufana Trzecia Strona

Zaufana Trzecia Strona w ustawie o podpisie elektronicznym została nazwana „**podmiotem świadczącym usługi certyfikacyjne„. W** naszym środowisku podmioty te zwane są **centrami certyfikacji** a w różnego rodzaju

publikacjach urzędami certyfikacji z angielskiego Certification Authority. Ta ostatnia nazwa stworzyła akronim CA, który jest bardzo często używany.

•Według definicji zamieszczonej w normie PN ISO/IEC 13888-1 Zaufana Trzecia Strona - to serwis lub organizacja lub jej agent, który świadczy jedną lub więcej usług bezpieczeństwa i **posiada zaufanie innych podmiotów** w odniesieniu do tych usług.



Funkcjonalność PKI

Posiadanie zaufania to nie jakaś bliżej nieokreślona i niezdefiniowana własność, lecz systematycznie i żmudnie zbudowany system.

By być wiarygodną, Zaufana Trzecia Strona powinna dostarczyć dowody, że:

- Wykonywane usługi i cała działalność realizowana jest zgodnie z istniejącym układem legislacyjnym, normami i standardami zarówno krajowymi jak i międzynarodowymi
- stosuje odpowiednią politykę bezpieczeństwa,
- przeprowadzono szacowanie ryzyka,
- szacowanie ryzyka jest powtarzane regularnie w celu zapewnienia zgodności z wymaganiami dotyczącymi poufności, integralności, dostępności i niezawodności zasobów,
- ryzyko zostało ograniczone do akceptowalnego poziomu
- stosowane zasady, regulacje i procedury są zgodne z założonym poziomem wiarygodności
- stosowany jest system obsługi incydentów

- istnieje plan ciągłości działania, jego założenia są stosowane i jest on bieżąco aktualizowany w miarę zmieniających się warunków zewnętrznych jak i wewnętrznych
- właściwie zarządza się zmianami i konfiguracją
- zasady odpowiedzialności są jasno sformułowane i zrozumiałe,
- jakość stosowanych procesów, działań i praktyk została zweryfikowana,
- spełnione są wymagania organizacyjne i personalne,
- interfejsy i procedury do komunikacji z klientami są odpowiednie do funkcji oraz są prawidłowo stosowane,
- należycie wypełnia swoje zobowiązania wynikające z umowy zawartej z użytkownikiem. Wiarygodność Zaufanej Trzeciej Strony opiera się na tym, że poziom świadczonych usług jest sprawdzalny i weryfikowalny.

Spełnienie tych wymagań jest sprawdzane przez wyznaczony do tego organ, w naszej rzeczywistości jest to Ministerstwo Gospodarki Pracy i Polityki Społecznej, jak również przez niezależnego audytora zewnętrznego.

Ministerstwo Gospodarki Pracy i Polityki Społecznej po przeprowadzeniu audytu podmiotu udziela mu akredytacji w postaci wpisu do rejestru podmiotów świadczących usługi certyfikacyjne oraz wystawia zaświadczenie certyfikacyjne.

Ustawa o podpisie elektronicznym:
zaświadczenie certyfikacyjne - elektroniczne zaświadczenie, za pomocą którego dane służące do weryfikacji poświadczenia elektronicznego są przyporządkowane do podmiotu świadczącego usługi certyfikacyjne lub organu, o którym mowa w art. 30 ust. 1, i które umożliwiają identyfikację tego podmiotu lub organu,

Zaświadczenie to z punktu widzenia technicznego jest po prostu certyfikatem, wystawionym przez jednostkę nadrzędną w infrastrukturze klucza publicznego jednostce podrzędnej. W ten sposób tworzona jest techniczna część PKI a użytkownik ma możliwość weryfikacji poświadczeń składanych przez podmiot świadczący usługi certyfikacyjne. Wyjaśnienia wymaga jeszcze pojęcie „poświadczenie". Zgodnie z ustawą o podpisie elektronicznym podpisy są składane przez osoby fizyczne, którym wydano certyfikat i które mają klucz prywatny związany z kluczem publicznym na podstawie, którego ten certyfikat został wydany. Podmioty świadczące usługi certyfikacyjne też jak wskazano powyżej mają swój certyfikat i związany z nim klucz prywatny i mogą składać podpisy, dla odróżnienia jednak podpis taki w ustawie przyjął nazwę „poświadczenie elektroniczne"

Definicja w ustawie tego pojęcia została przedstawiona w następujący sposób: **poświadczenie elektroniczne** - dane w postaci elektronicznej, które wraz z innymi danymi, do których zostały dołączone lub logicznie z nimi powiązane, umożliwiają identyfikację podmiotu świadczącego usługi certyfikacyjne lub

organu wydającego zaświadczenia certyfikacyjne, oraz spełniają następujące wymagania:

a) są sporządzane za pomocą podlegających wyłącznej kontroli podmiotu świadczącego usługi certyfikacyjne lub organu wydającego zaświadczenie certyfikacyjne bezpiecznych urządzeń służących do składania podpisu elektronicznego i danych służących do składania poświadczenia elektronicznego,

b) jakakolwiek zmiana danych poświadczonych jest rozpoznawalna,

Poziom zaufania, jakim użytkownicy mogą darzyć TTP może być zwiększony, jeśli instytucja certyfikująca TTP jest akredytowana według określonego wzoru zasad właściwego z punktu widzenia wykonywanych zadań. Akredytacja zapewnia, że procedury różnych jednostek certyfikujących TTP są podobne i wyniki certyfikacji różnych instytucji certyfikujących są porównywalne. Akredytacja jest zdefiniowana w ISO/IEC Guide 2. Akredytacja instytucji certyfikującej TTP oznacza, że jest ona powszechnie uznawana jako kompetentna i godna zaufania w zakresie świadczenia usług certyfikacji TTP. Wynika z tego, że akredytacja instytucji certyfikujących TTP jest dodatkowym środkiem gwarantowania jakości usług świadczonych przez TTP, ponieważ organizacje akredytujące są niezależne i działają zgodnie z powszechnie akceptowanymi zasadami.

Instytucja akredytująca ocenia proceduralne i techniczne aspekty systemu zarządzania instytucji certyfikujących TTP zgodnie z ISO Guide 61 lub innymi podobnymi wzorami zasad takimi, jak seria norm europejskich 450xx.

Akredytacja instytucji certyfikującej TTP jest środkiem zapewniającym jakość pracy instytucji certyfikującej, nie mówi jednak nic o usługach świadczonych przez daną TTP. TTP definiuje usługę, którą świadczy, a instytucja certyfikująca TTP stwierdza, jak dobrze jest ona realizowana.

8. Dokumentacja systemu z którą powinien zapoznać się użytkownik

Zanim użytkownik podejmie decyzję o złożeniu wniosku o wydanie certyfikatu powinien zapoznać się z publikowanymi dokumentami zaufanej trzeciej strony a zwłaszcza umową między TTP a podmiotami korzystającymi z jej usług, czasem zwaną umową subskrybencką, polityką certyfikacji (certification policy CP) oraz regulaminem świadczenia usług certyfkacyjnych (certification practice statement CPS).

Umowy między TTP a podmiotami korzystającymi z jej usług

Formalne umowy między TTP a podmiotami korzystającymi z jej usług powinny jasno określać zarówno odpowiedzialność TTP, jakość dostarczanych usług, jak i odpowiedzialność podmiotów.

Umowa powinna wyjaśniać zasady zarządzania i organizacji TTP oraz stosowane procedury. TTP powinna opublikować regulamin świadczenia usług, w którym

opisano, czego podmioty mogą oczekiwać od usług TTP tak, aby jasno zdefiniować zasady działania i wymagania stawiane TTP, jakość usługi, problemy etyczne oraz opłaty subskrybcyjne.
Umowa powinna jasno określać, w jaki sposób TTP zapewnia zgodność swoich działań z obowiązującym prawem. Umowa powinna wskazywać jurysdykcję, która ma zastosowanie oraz jurysdykcję dla rozstrzygania sporów.

Przypadkowe i rozmyślne błędy popełnione przez TTP mogą powodować znaczące szkody w działalności biznesowej. Aby osiągnąć wystarczający poziom zaufania do stosowania usług TTP, umowa powinna określać ograniczenie odpowiedzialności TTP z tytułu roszczeń swoich użytkowników. Roszczenia z tytułu szkody powinny być pokryte z odpowiedniej polisy ubezpieczeniowej. Wymagane ubezpieczenie powinno być zdefiniowane w umowie miedzy TTP a jej użytkownikami.

Umowa powinna zawierać listę wszystkich spraw odnoszących się do podziału odpowiedzialności między TTP a jej użytkownikami tak, aby użytkownicy mogli mieć dostęp do odpowiedniej profesjonalnej porady w celu uzyskania właściwego wsparcia prawnego w przypadku problemu wynikającego ze świadczenia i korzystania z usług TTP.

Umowa powinna określać zamierzony zakres zastosowania usługi i jej parametry oraz umożliwiać przerwanie świadczenia usługi, jeśli któraś ze stron umowy korzysta z niej niewłaściwie lub nielegalnie.
Umowa powinna zawierać klauzule jasno wskazujące, że dla rozstrzygania sporów miedzy TTP a jej użytkownikami może być powołany niezależny i bezstronny arbiter.

Umowa powinna określać, w jaki sposób będzie chroniona prywatność danych osobowych i innych wrażliwych informacji, oraz okoliczności, w których dane te mogą być ujawnione.

Polityka certyfikacji i regulamin świadczenia usług certyfikacyjnych

Gdy urząd certyfikacji wydaje certyfikaty, przedstawia użytkownikowi tego certyfikatu deklarację, że określony klucz publiczny jest powiązany z określonym podmiotem Jednakże, zakres zaufania użytkownika certyfikatu do deklaracji CA powinien być przedmiotem oceny tego użytkownika. Różne certyfikaty są wydawane zgodnie z różnymi praktykami i procedurami oraz mogą być przeznaczone dla różnych celów i zastosowań.

W zaleceniu X.509 zdefiniowano politykę certyfikacji jako „**zbiór zasad, który wskazuje zastosowanie certyfikatu w określonej grupie użytkowników i/lub klasy aplikacji o wspólnych wymaganiach bezpieczeństwa**". Certyfikat zgodny z X.509, wersja 3 może zawierać wskazanie polityki certyfikacji, możliwej do zastosowania przez użytkownika certyfikatu dla stwierdzenia, czy w określonym przypadku może mieć zaufanie do certyfikatu, czy nie.
Polityka certyfikacji, która musi być rozpoznawalna zarówno przez certyfikat wystawcy, jak i użytkownika certyfikatu jest reprezentowana w certyfikacie za pomocą jednoznacznego, zarejestrowanego identyfikatora obiektu. Proces

rejestrowania jest zgodny z normami ISO/IEC i zaleceniami ITU-T. Strona, która rejestruje identyfikator obiektu, publikuje także tekstowy opis polityki certyfikacji, umożliwiając w ten sposób użytkownikowi zapoznanie się z jej treścią. Każdy pojedynczy certyfikat zwykle zawiera odniesienie do jednej polityki certyfikacji, lub w miarę możliwości jest wydawany zgodnie z pewną, niewielką liczbą różnych polityk.

Polityki certyfikacji stanowią ponadto podstawę dla certyfikacji wzajemnej CA, zgodnie z regulaminem świadczenia usług certyfikacyjnych (CPS). Każdy CA jest certyfikowany w odniesieniu do jednej lub kilku polityk certyfikacji, które urząd ten wdrożył. W przypadku, gdy jeden CA wystawia certyfikat innemu CA, wystawiający CA musi określić zbiór polityk certyfikacji, w zakresie, których darzy zaufaniem CA będący podmiotem certyfikatu (taka ocena może opierać się na certyfikacji, w odniesieniu do branych pod uwagę polityk). Zdefiniowany zbiór polityk certyfikacji jest wtedy wskazywany przez wystawiający CA w certyfikacie dla drugiego CA. Układ logiczny przetwarzający ścieżkę certyfikacji, zgodny z X.509, wykorzystuje wskaźniki polityk certyfikacji w celu poprawnego określenia modelu zaufania.

CPS jest szczegółową deklaracją CA odnoszącą się do stosowanych przez urząd praktyk, które powinny być rozumiane i konsultowane przez subskrybentów i użytkowników certyfikatów. Jakkolwiek, poziom szczegółowości regulaminów jest bardzo różny, to jednak zwykle są one bardziej szczegółowe niż definicje polityk certyfikacji. Regulaminy mogą być obszernymi, rozbudowanymi dokumentami, zawierającymi precyzyjny opis oferowanych usług, szczegółowe procedury zarządzania związanego z cyklem życia certyfikatów, zatem poziom szczegółowości, który wiąże regulamin z określonym wdrożeniem oferowanych usług.

Szczegółowość w regulaminie jest konieczna ze względu na potrzebę pełnej oceny wiarygodności w przypadku, gdy brak jest schematu akredytacji lub innych uznawanych metod oceny jakości. Jednakże, szczegółowy regulamin nie stanowi wyłącznej podstawy do osiągnięcia stanu współdziałania między CA prowadzonymi przez różne instytucje. Natomiast polityki certyfikacji służą najlepiej ufającym stronom dla zdecydowania, czy dany certyfikat odpowiada ich potrzebom i zastosowaniom. CA z jednym regulaminem świadczenia usług certyfikacyjnych może spełniać wymagania wielu polityk certyfikacji (używanych w różnych zastosowaniach i/lub przez różne grupy użytkowników). Także wiele różnych CA, których regulaminy nie są identyczne, mogą spełniać wymagania tej samej polityki certyfikacji.

9. Odpowiedzialność stron

TTP powinna zdefiniować zakres przyjętej odpowiedzialności za prawidłowe świadczenie jej usługi. Dodatkowo, TTP powinna określić zakres roszczeń, które mogą być zaakceptowane w odniesieniu do naruszeń bezpieczeństwa.

Odpowiedzialność zarówno TTP, jak i użytkownika powinna być jasno określona w każdej formalnej umowie zawartej między TTP i użytkownikiem. Większość przypadków odpowiedzialności powinna być częścią umowy, niektóre powinny być zdefiniowane co najmniej w ramach powszechnie stosowanej praktyki biznesowej, inne powinny mieścić się w standardowym poziomi jakości usługi.

W ustawie o podpisie elektronicznym odpowiedzialność została określona w **Art. 11. który stanowi że:**

1. Podmiot świadczący usługi certyfikacyjne odpowiada wobec odbiorców usług certyfikacyjnych, z zastrzeżeniem ust. 2 i 3, za wszelkie szkody spowodowane niewykonaniem lub nienależytym wykonaniem swych obowiązków w zakresie świadczonych usług, chyba że niewykonanie lub nienależyte wykonanie tych obowiązków jest następstwem okoliczności, za które podmiot świadczący usługi certyfikacyjne nie ponosi odpowiedzialności i którym nie mógł zapobiec mimo dołożenia należytej staranności.
2. Podmiot świadczący usługi certyfikacyjne nie odpowiada wobec odbiorców usług certyfikacyjnych za szkody wynikające z użycia certyfikatu poza zakresem określonym w polityce certyfikacji, która została wskazana w certyfikacie, w tym w szczególności za szkody wynikające z przekroczenia najwyższej wartości granicznej transakcji, jeżeli wartość ta została ujawniona w certyfikacie.
3. Podmiot świadczący usługi certyfikacyjne nie odpowiada wobec odbiorców usług certyfikacyjnych za szkodę wynikłą z nieprawdziwości danych zawartych w certyfikacie, wpisanych na wniosek osoby składającej podpis elektroniczny.
4. Podmiot świadczący usługi certyfikacyjne, który udzielił gwarancji za certyfikat zgodnie z art. 4 pkt 4, odpowiada wobec odbiorców usług certyfikacyjnych za wszelkie szkody spowodowane użyciem tego certyfikatu, chyba że szkoda wynikła z użycia certyfikatu poza zakresem określonym w polityce certyfikacji, która została wskazana w tym certyfikacie.

10. Podstawowe kategorie usług Zaufanej Trzeciej Strony

Zasadniczą usługą jest zarządzanie certyfikatami obejmujące wydawanie certyfikatów, ich zawieszanie i unieważnianie oraz archiwacja. Jednakże TTP świadczą szereg usług uzupełniających zarządzanie certyfikatami i dodatkowych takich jak usługa oznaczania czasem, usługa notariatu elektronicznego.

Zarządzanie certyfikatami / kluczami

Zgodnie ze standardem dotyczącym zarządzania kluczami ISO/IEC 11770-1, zarządzanie kluczami opiera się na podstawowych usługach generowania,

rejestrowania, certyfikacji, dystrybucji, instalacji, przechowywania, klucza tworzącego klucze pochodne, archiwizowania, unieważniania, usuwania z rejestru i niszczenia. Inne usługi związane z bezpieczeństwem, które mogą być stosowane obejmują usługi kontroli dostępu, audytowanie, uwierzytelnienie, kryptograficzne i oznaczania czasu.

Usługa generowania kluczy

Ta usługa jest uruchamiana w celu wygenerowania kluczy, w bezpieczny sposób, dla danego algorytmu kryptograficznego. Zasadniczą właściwością tej usługi jest wytworzenie tajnych, nieprzewidywalnych liczb o określonych właściwościach. Proces generowania liczb losowych obejmuje różne elementy takie, jak generowanie liczb losowych, walidacja procesu generowania liczb losowych, generowanie parametrów domeny, walidacja parametrów domeny, generowanie pary kluczy i walidacja klucza publicznego.

Certyfikacja kluczy publicznych

W tej usłudze TTP występuje jako akredytowany urząd certyfikacji (centrum certyfikacji), który tworzy certyfikat klucza. Urząd certyfikacji podpisuje klucze publiczne, aby potwierdzić ich ważność i autentyczność w zaufanej infrastrukturze kluczy publicznych. Podmioty wykorzystujące certyfikaty musza ufać temu samemu urzędowi certyfikacji lub mieć co najmniej jeden wspólny urząd w hierarchii certyfikacji. Certyfikowane klucze mogą być generowane przez usługę generowania TTP albo przez właściciela klucza. Urząd certyfikacji musi mieć pewność, że żądający jest w posiadaniu odpowiadającego kluczowi publicznemu klucza prywatnego. Certyfikaty dla kluczy publicznych zostały omówione szczegółowo w ISO/IEC 11770-1, Załącznik D.

Certyfikat jest wydawany z ograniczonym czasem ważności najczęściej certyfikaty wydawane są na okres jednego roku.

Zakres odpowiedzialności CA jest bardzo szeroki a w szczególności obejmuje:

a) zapewnienie technicznych i organizacyjnych możliwości szybkiego i niezawodnego wydawania, zawieszania i unieważniania certyfikatów oraz określenia czasu dokonania tych czynności
b) zidentyfikowanie podmiotów, których informacja o kluczu publicznym jest przedstawiona do certyfikacji;
c) zapewnienie środków przeciwdziałających fałszerstwom certyfikatów i innych danych poświadczanych elektronicznie przez te podmioty, w szczególności przez ochronę urządzeń i danych wykorzystywanych przy świadczeniu usług certyfikacyjnych,
d) zawarcie umowy ubezpieczenia odpowiedzialności cywilnej za szkody wyrządzone odbiorcom usług certyfikacyjnych,

e) poinformowanie osoby ubiegającej się o certyfikat, przed zawarciem z nią umowy, o warunkach uzyskania i używania certyfikatu, w tym o wszelkich ograniczeniach jego użycia,
f) uzyskanie od osoby, której wydano certyfikat zgody na publikację certyfikatu
g) zapewnienie jakości pary asymetrycznych kluczy używanych do tworzenia certyfikatów kluczy publicznych,
h) zabezpieczenie procesu certyfikacji oraz klucza prywatnego używanego do podpisywania informacji o kluczach publicznych,
i) zarządzanie danymi wewnętrznymi systemu, które mają być umieszczone w informacji o kluczu publicznym takimi, jak numer seryjny certyfikatu klucza publicznego, identyfikacja urzędu certyfikacji itp.,
j) przyporządkowanie i sprawdzenie okresów ważności,
k) powiadomienie podmiotu zidentyfikowanego za pomocą informacji o kluczu publicznym, że certyfikat klucza publicznego został wydany; środki używane do przekazu tego powiadomienia powinny być niezależne od metody używanej do przekazania informacji o kluczu publicznym do CA,
l) zapewnienie, że wszystkie informacje zawarte w certyfikacie spełniają wymagania stosowanej polityki certyfikacji, np. przez zagwarantowanie, że dwóm różnym podmiotom nie przyporządkowano tego samego identyfikatora, tak aby mogli być właściwie rozróżniani,
m) publikowanie danych, które umożliwią weryfikację, w tym również w sposób elektroniczny, autentyczności i ważności certyfikatów oraz innych danych poświadczanych elektronicznie przez ten podmiot oraz zapewnienie nieodpłatnego dostępu do tych danych odbiorcom usług certyfikacyjnych, oraz
n) rejestrowanie wszystkich kroków zrealizowanych w procesie generowania certyfikatu klucza publicznego.

Jeden CA może certyfikować informacje o kluczu publicznym innego CA i wydać mu certyfikat klucza publicznego. W tym przypadku zachodzi tzw wzajemna certyfikacja lub kroscertyfikacja

Unieważnienie certyfikatów

Celem tej usługi jest zapewnienie bezpiecznego wycofania z użytkowania certyfikatu w przypadku, gdy został ujawniony klucz prywatny lub zachodzi podejrzenie naruszenia jego ochrony. Lista unieważnionych certyfikatów musi być niezwłocznie publikowana, ustawa o podpisie elektronicznym nakłada na podmioty świadczące usługi certyfikacyjne obowiązek publikowania informacji o zawieszeniu bądź unieważnieniu w ciągu jednej godziny od unieważnienia lub zawieszenia certyfikatu . Unieważnienia może żądać właściciel certyfikatu lub inna upoważniona osoba albo jeśli istnieje jakiekolwiek podejrzenie, że ochrona klucza została naruszona inny zaufany podmiot w tym również podmiot, który wydał ten certyfikat , Zgodnie z ISO/IEC 11770-1, Załącznik D, każdy rekord listy

unieważnionych/zawieszonych certyfikatów powinien zawierać czas unieważnienia, czas żądania unieważnienia i czas znanego lub podejrzewanego naruszenia. TTP jest odpowiedzialna jedynie za unieważnianie kluczy swoich klientów.

Centra certyfikacji świadczą również wiele innych usług takich jak znakowanie dokumentów wiarygodnym czasem, poświadczanie dokumentów (notariat elektroniczny)

Usługa oznaczania czasem (ang. Time Stamping Service TSS) opatruje pieczęcią cyfrowy dokument przez kryptograficzne powiązanie wiarygodnego czasu z tym dokumentem (zwykle ze skrótem tego dokumentu, zwanym „streszczeniem wiadomości" lub „odciskiem wiadomości"2), i w ten sposób tworzy jego cechę umożliwiającą wykrycie każdej modyfikacji, takiej jak antydatowanie, uniknięcie ataków powtórzeniowych lub innych fałszerstw.

Usługa oznaczania czasu związana jest z autentycznością zegara, który został użyty, zatem usługa oznaczania czasu, winna używać zegara charakteryzującego się wysoką niezawodnością, dostępnością i wiarygodnością.

Ustawa o podpisie elektronicznym w szczególności stanowi:
Art7 ust 2.: *Znakowanie czasem przez kwalifikowany podmiot świadczący usługi certyfikacyjne wywołuje w szczególności* ***skutki prawne daty pewnej*** *w rozumieniu przepisów Kodeksu cywilnego* Art. 6. ust 1*Bezpieczny podpis elektroniczny weryfikowany przy pomocy ważnego kwalifikowanego certyfikatu* ***stanowi dowód*** *tego, że został on złożony przez osobę określoną w tym certyfikacie, jako składającą podpis elektroniczny.*

Opcjonalnie, TTP (ang. Trusted Third Party) świadcząca usługi oznaczania czasu powinna rejestrować wszystkie znaczniki elektroniczne w porządku chronologicznym w nieusuwalnym archiwum. Ponadto, może być realizowana usługa weryfikowania znacznika czasu.

Urząd oznaczania czasu (Time Stamp Authority TSA) jest zaufaną trzecią stroną, która tworzy tokeny znaczników czasu w celu poświadczenia istnienia dokumentu w momencie, w którym wygenerowany został ten znacznik. Tokeny znaczników czasu są opisane w ISO/IEC 13888-1.

TSA tworzy „dowód istnienia" dla danej wiadomości w danym momencie czasu. TSA może być także stosowany w przypadku, gdy wymagane jest odniesienie do zaufanego czasu oraz gdy lokalny zegar nie jest wiarygodny dla wszystkich podmiotów. Rolą urzędu oznaczania czasu jest opatrywanie znacznikiem czasowym skrótu wiadomości, w celu stworzenia poświadczenia, że w danym momencie czasu dokument istniał. Przykładowo, znacznik czasowy może być użyty do:

---

[2] *message digest, message imprint*

a) zweryfikowania faktu, że podpis cyfrowy został dołączony do dokumentu przed unieważnieniem certyfikatu klucza publicznego co oznacza, że certyfikat może być użyty do weryfikacji tego podpisu,
b) wskazania czasu złożenia dokumentu w przypadku, gdy jest to termin krytyczny lub nieprzekraczalny,
c) wskazania czasu transakcji.

Notariat elektroniczny nie wyręcza klasycznych notariuszy, choć w obrocie elektronicznym spełnia podobne funkcje. Publiczne usługi notarialne są usługami wysokiego poziomu, które wykorzystują wiele usług podstawowych takich, jak oznaczanie czasem, certyfikacja, usługa publikowania (zwana też usługą katalogu) certyfikatów, list unieważnionych certyfikatów, polityk i regulaminów itp., cyfrowa archiwizacja i niezaprzeczalność. Notariat elektroniczny polega na tym, że dokument jest przekazywany do Zaufanej Trzeciej Strony (ang. Trusted Third Party - TTP) która poświadcza lub certyfikuje ten dokument za pomocą poświadczenia elektronicznego (cyfrowego podpisu TTP) lub innych środków. Częścią tej usługi może być usługa publikacji, w której informacje takie, jak poświadczone uprzednio dokumenty mogą być uzyskiwane z bazy danych lub katalogu. Umożliwia to rozstrzyganie problemów, jakie mogą zaistnieć w wyniku przekazywania informacji w postaci elektronicznej.

Usługa publicznego notariatu może atestować i certyfikować określone klasy dokumentów, np. to, że dokument istniał w danym momencie w przeszłości w celu nadania mu wiarygodności i autentyczności. Taka usługa może być wykorzystana do mediacji w sporze między podmiotami i może być autoryzowana przez jakąś władzę.

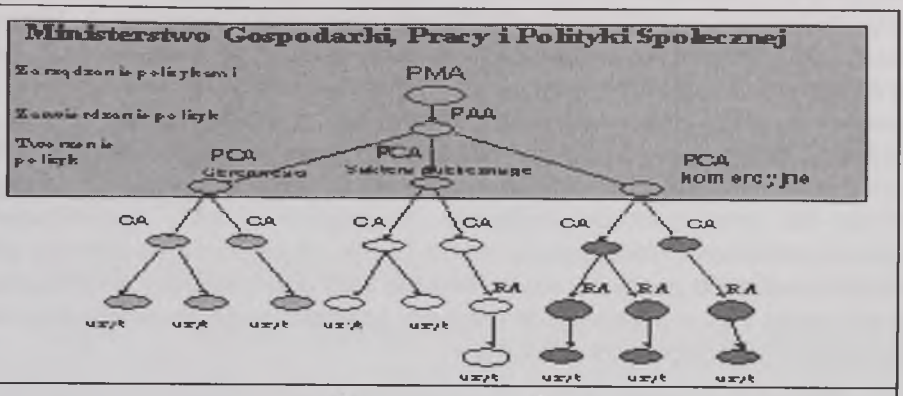
W chwili obecnej istnieje wiele złożonych problemów w obszarze elektronicznego urzędu notarialnego i odpowiedzialności związanej z wykonaniem tych czynności. Nie została rozstrzygnięta kwestia uprawnień do wykonania pewnych czynności a kwestionowanych przez środowisko notariuszy, którym stosowna ustawa nadaje pewne uprawnienia na wyłączność.

11. Hierarchia i środowisko PKI

Podmioty świadczące usługi certyfikacyjne na terenie kraju tworzą zwykle strukturę hierarchiczną. W zależności od stopnia rozwoju struktura ta jest bardziej rozwinięta lub spłaszczona.

Najbardziej rozwinięte struktury są nawet kilkupoziomowe. Użytkownik końcowy kontaktuje się bezpośrednio z Urzędami rejestracyjnymi (ang. Registration Autority RA), które ustalają zakres usług, ustanawiają związek pomiędzy tożsamością użytkowników i ich kluczem publicznym, podpisują umowy o świadczenie usług, przekazują wniosek certyfikacyjny do Centrum certyfikacji i udostępniają wystawiony certyfikat użytkownikowi często na karcie elektronicznej wraz z czytnikiem. Odbierają również zgłoszenia o unieważnienie certyfikatów , weryfikują je i przesyłają do CA.

Drugi z kolei poziom to Centra certyfikacji lub Urzędy Certyfikacji (ang. Certification Authority CA), które mogą spełniać również funkcje RA ale zasadnicza ich działalność to wystawianie certyfikatów zarządzają nimi oraz świadczą inne usługi TTP.



Hierarchia i Środowisko PKI

Trzeci z kolei poziom to Urząd Polityki Certyfikacji (Policy Certification Authority PCA). Jako urząd nadrzędny PCA ustanawia polityki certyfikacji i główne zasady dla całej podległej struktury. Akredytuje podległe jednostki a jednocześnie pełni role CA dla tej struktury czyli wydaje certyfikaty podrzędnym CA, zarządza nimi, zapobiega dublowaniu identyfikatorów wyróżniających DN (distinguish name) itd. Zwykle PCA obsługuje jakąś część sektora publicznego lub specjalistyczny segment rynku. Czwartym poziomem jest Urząd zatwierdzania polityk (Policy Approval Authority PAA) pełni on tą samą rolę wobec PCA jak PCA wobec CA. Zwykle jest to jeden taki urząd w kraju lub stanie.

Piątym poziomem jest Organ Zarządzający Politykami Certyfikacji (ang. Policy Management Authority PMA). Podmiot ten nie posiada żadnej struktury technicznej nie wydaje więc certyfikatów I nie zarządza nimi. Zajmuje się jedynie koordynacją polityk i zasad funkcjonowania podmiotów infrastruktury klucza publicznego i współpracą z podobnym poziomem zagranicznych PKI.

W naszej rzeczywistości struktura ta jest bardzo spłaszczona wszyscy dostawcy usług certyfikacyjnych podlegają Ministerstwu Gospodarki Pracy i Polityki Społecznej, które na podstawie aktów prawnych, norm i standardów określa szeroko rozumiane wymagania (funkcjonalne, bezpieczeństwa, środowiskowe, prawne itp.) dla całej struktury, prowadzi kontrolę tych podmiotów i udziela akredytacji tym podmiotom. Struktura ta nie umożliwia swobodnego wyboru przez CA jednostek akredytujących co stanowi jeden z zasadniczych wymogów Dyrektywy Parlamentu Wspólnoty Europejskiej

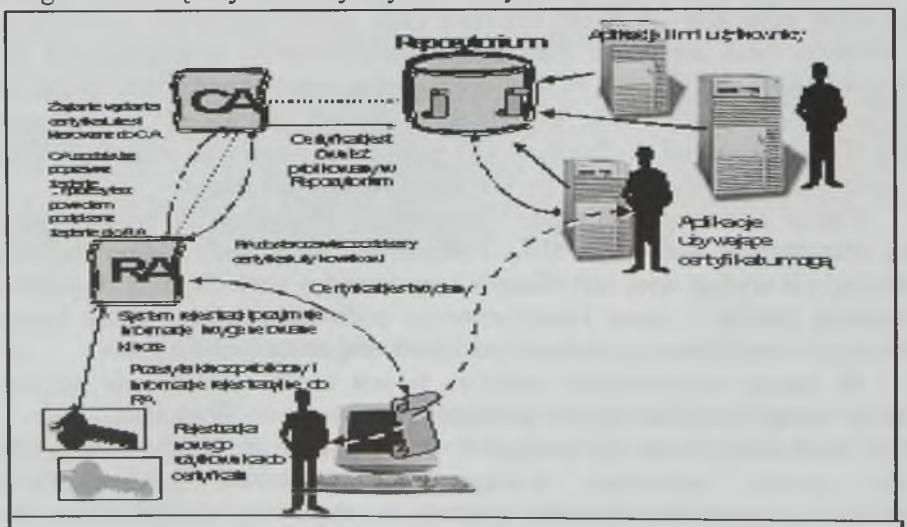
Współdziałanie i powiązania podmiotów infrastruktury klucza publicznego PKI

Współdziałanie i powiązania podmiotów zachodzą w ramach 4 reguł , które można wydzielić w architekturze PKI: ·akredytacji, certyfikacji, publikacji

oraz archiwizacji. **Akredytacja** ma na celu zapewnienie, że wszystkie podmioty występujące w PKI działają według jednolitych zasad i przygotowane są do ponoszenia odpowiedzialności finansowej w zakresie obowiązującym w ramach PKI. **Certyfikacja** ma na celu zapewnienie poprawnego i efektywnego realizowania podstawowych funkcji PKI - wydawania i unieważniania certyfikatów kluczy publicznych. **Publikacja** służy rozpowszechnianiu informacji takich jak certyfikaty, informacje o unieważnionych certyfikatach, listy akredytowanych urzędów certyfikacji, polityki i regulaminy certyfikacji itp. Zasadniczą cechą tego podsystemu musi być jego dostępność. Podsystem publikacji może być podzielony na część otwartą, dostępną dla wszystkich użytkowników i podmiotów działających w PKI oraz część zamkniętą, dostępną jedynie dla upoważnionych podmiotów działających w PKI. Archiwizacja•ma zapewnić możliwość rozstrzygania sporów (a więc dostarczać dowodów dla usługi niezaprzeczalności) pomiędzy użytkownikami PKI oraz pomiędzy użytkownikami i podmiotami PKI w perspektywie czasowej przekraczającej przewidywany okres działalności poszczególnych podmiotów.

12. Jak uzyskać Certyfikat

Jak już wspomniano aby móc stosować podpis elektroniczny i korzystać z usług z nim związanych należy uzyskać certyfikat.



Schemat działania PKI

W tym celu należy udać się do Punktu/urzędu rejestracji (ang Registration Authority RA) czyli jednej z placówek Centrum Certyfikacji (np. PCCE Sigillum), którego certyfikat chcemy uzyskać. W tym punkcie uzgadniamy zakres usług i rodzaj certyfikatu (do czego chcemy go używać) na podstawie przedstawionych dokumentów weryfikowana jest nasza tożsamość i generowana jest para kluczy.

Klucz publiczny i nasze niezbędne dane identyfikacyjne stanowią podstawę wystawienia wniosku o wydanie certyfikatu.

Wniosek ten wędruje do Centrum Certyfikacji (ang. Certification Authority -CA) wykorzystując odpowiednie zabezpieczenia, sprawdzana jest jego poprawność i wystawiany jest certyfikat, który wraz z danymi osobowymi jest umieszczony na karcie elektronicznej, która jest personalizowana (umieszczanie nazwiska i identyfikatora) po czym karta dostarczana jest użytkownikowi (sposób personalizacji i dostarczenia karty użytkownikowi może być różny w zależności od technologii stosowanej przez CA). Od tego czasu można używać certyfikatu do tych celów jakie zostały określone w dokumencie zwanym polityka certyfikacji np. do zabezpieczenia transakcji elektronicznych.

13. Repozytorium -Centralny katalog

Ustawa o podpisie elektronicznym nakłada obowiązek publikowania certyfikatów, polityk certyfikacji, regulaminów a zwłaszcza list unieważnionych certyfikatów (ang Certification Revocation List- CRL). Istotnym elementem PKI jest więc repozytorium czasem zwane katalogiem centralnym lub usługą katalogową. Często repozytorium tworzone jest w oparciu o protokół LDAP (Lightweight Directory Access Protocol) lub X.500. Do repozytorium użytkownik sięga najczęściej aby sprawdzić czy certyfikat, przy pomocy którego chce zweryfikować podpis elektroniczny jest nadal ważnym certyfikatem. Z repozytorium można też pobrać certyfikat osoby bądź podmiotu, z którym chcemy wymieniać informację. Należy jednak pamiętać, że podmioty muszą wyrazić zgodę na umieszczenie ich certyfikatu w repozytorium, tak więc może się zdarzyć, że w repozytorium możemy nie znaleźć certyfikatu podmiotu mimo iż jesteśmy przekonani, że podmiot ten posiada certyfikat.

14. Obszary zastosowań usług kryptograficznych, certyfikatów i podpisu elektronicznego,

Po okresie fascynacji w zakresie elektronicznego biznesu i jego lawinowym rozwoju, nastąpiło nagłe zahamowanie tego procesu a robiące kokosowe interesy firmy internetowe bankrutowały bądź traciły znaczną część swej klienteli. Było to spowodowane tym, że słabo zabezpieczone transakcje stały się łatwym łupem dla wielu oszustów a informacje o często popełnianych przestępstwach odstraszyły entuzjastów tego rodzaju transakcji od robienia dalszych niepewnych interesów. Zachwiane zaufanie będzie trudne do odzyskania.

Omówiony powyżej podpis elektroniczny będący podstawą **niezaprzeczalności,** czyli niemożności wyparcia się udziału w transakcji, **uwierzytelnienia podmiotów biorących udział w transakcji,** co oznacza, że podmioty mają pewność, że właściciel certyfikatu biorący udział w transakcji jest naprawdę tym za którego się podaje i **integralności tej transakcji** czyli niemożliwości zmiany treści określającej przedmiot transakcji, podnosi

bezpieczeństwo transakcji do poziomu co najmniej równego z transakcjami tradycyjnymi w których strony wymieniają się odręcznie podpisanym zleceniem, protokołem odbioru, umową i fakturą.

Dyrektywa 99/93/EC w sprawie wspólnotowej struktury dla podpisów elektronicznych mówi: „(19) *Podpisy elektroniczne będą stosowane w sektorze publicznym jako środek komunikowania się jednostek administracji Wspólnoty ze sobą oraz miedzy tymi jednostkami a obywatelami i przedsiębiorcami, przykładowo, przy zamówieniach publicznych, podatkach, ubezpieczeniach społecznych, w systemach ochrony zdrowia i sprawiedliwości*"

Transakcję elektroniczną można również ustrzec przed wścibskim okiem nieupoważnionych osób dokonując szyfrowania jej treści, czyli zapewniając poufność transakcji.
Wykorzystanie certyfikatów i podpisu elektronicznego wydaje się już dzisiaj wręcz nieograniczone, zwłaszcza tam gdzie występuje potrzeba przekazywania informacji. Nie sposób wymienić wszystkich zastosowań jednym tchem wymienia się zwykle

Zabezpieczoną pocztę: korespondencja wewnętrzna pomiędzy pracownikami,, filiami, oddziałami, korespondencja z kooperantami – oferty, zamówienia, umowy, przesyłanie dokumentacji medycznej – karty pacjentów, wyniki badań, recepty, rezerwację biletów, biura maklerskie biznes, gra na giełdzie. Transakcje B2B B2C, Bankowość internetowa, internetowe zakupy, sprzedaż, reklama, usługi doradcze, ubezpieczeniowe, turystyczne, leasingowe .Zarządzanie łańcuchem dostaw. Zabezpieczanie sieci lokalnych, Zdalny dostęp, Protokoły SSL i SET. Ustawa stwarza możliwość komunikowanie się z organami administracji publicznej, zobowiązując je do dostosowania swojej infrastruktury informacyjnej w ciągu czterech lat od wejścia w życie ustawy do kontaktu z obywatelami i podmiotami gospodarczymi drogą elektroniczną, można więc będzie wiele spraw w urzędach załatwiać bez wystawiania w kolejkach i stresującego kontaktu z urzędnikami. Lista zdaje się nie mieć końca.

Zakres zastosowania usług kryptografii klucza publicznego będzie jednak różny w zależności od rodzaju podmiotu, który będzie go stosował.. I tak w środowisku korporacji oprócz zastosowania podpisu elektronicznego do prowadzenia transakcji zewnętrznych niezbędnym będzie zbudowanie wewnętrznego PKI do celów rozliczeń wewnętrznych (wiele jednostek posiada znaczną niezależność finansową) i wzajemnego uwierzytelnienia. Korporacje będą również w odrębny sposób współpracować z Centrami Certyfikacji w zakresie ubiegania się o certyfikaty (np. banki, które posiadają tysiące klientów o jednoznacznie ustalonej tożsamości mogą skorzystać z uproszczonej procedury składania wniosków certyfikacyjnych) jak również sposobami rozliczeń z usług świadczonych przez CC.

Inaczej będzie przebiegała współpraca z małymi firmami, które często powierzają prowadzenie księgowości i rozliczeń wyspecjalizowanym podmiotom takim jak doradcy podatkowi. W tym przypadku podmioty te najprawdopodobniej będą również pośredniczyć w udostępnieniu firmom certyfikatów jak również wielu usług związanych z ich wykorzystaniem. Pośredniczenie będzie polegało na tym, że podmioty świadczące usługi małym firmom będą w porozumieniu z centrami certyfikacyjnymi tworzyły punkty rejestracji, sprawdzając tożsamość ubiegających się o certyfikat, pomagając im w wyborze rodzaju certyfikatu i złożeniu wniosku o wystawienie certyfikatu. Po uzyskaniu certyfikatu współpraca obu zainteresowanych będzie mogła odbywać się na zasadach powyżej opisanych, czyli w pełni zabezpieczonej wymiany informacji. Stosując podpis elektroniczny można nie tylko znacznie podnieść poziom bezpieczeństwa elektronicznej wymiany informacji ale można również obniżyć koszty i znacznie rozszerzyć zarówno zakres usług jak i zasięg terytorialny na którym te usługi są świadczone, gdyż wszystkie usługi (nawet podpisywanie umów) będą mogły być realizowane drogą elektroniczną bez konieczności bezpośredniego kontaktowania się.

15. Korzyści z stosowania certyfikatów i usług kryptograficznych

- ✓ Zmniejszenie strat związanych z manipulacjami transakcjami, utratą poufności oraz uszczerbkiem dobrego imienia
- ✓ Zmniejszenie kosztów administracyjno- zarządczych o ok. 30% ,
- ✓ Oszczędność czasu (dużo szybsze przesyłanie korespondencji, brak konieczności drukowania, podpisywania, skanowania)
- ✓ Uproszczenie podpisywania dokumentów wielostronicowych
- ✓ Łatwiejsze katalogowanie i archiwizowanie dokumentów
- ✓ Ułatwienie i zabezpieczenie korespondencji zbiorowej
- ✓ Zwiększenie dochodów przy mniejszym kapitale
- ✓ Niższe koszty prowadzenia działalności gospodarczej (zakupy, zmniejszenie inwentarza)
- ✓ Szybszy i bardziej powszechny dostęp do rynku
- ✓ Bardziej niezawodna obsługa konsumenta
- ✓ Wzrost biznesu (nowa sprzedaż) i elastyczności
- ✓ Użytkownicy będą uwolnieni od kajdan jednej jednostki lub lokalizacji

16. Bibliografia

1. Public Key Infrastructure Study Final Report National Institute of Standards and Technology
2. EG 201 057 V1.1.2 (1997-07) Telecommunications Security; Trusted Third Parties (TTP); Requirements for TTP services ETSI Guide

3. Public Key Infrastructure Study Final Report National Institute of Standards and Technology
4. Elementarz kryptologii - dr inż. Krzysztof Gaj mgr inż. Karol Górski mgr inż. Anna Zugaj
5. ISO/IEC TR 14516 | ITU-T X.842 - Wytyczne dla użytkowania i zarządzania usługami Zaufanej Trzeciej Strony
6. PN-ISO/IEC 13888-1do3:1999 - Technika informatyczna. Techniki zabezpieczeń. Niezaprzeczalność
7. USTAWAz dnia 18 września 2001 r. o podpisie elektronicznym
8. The Essential Role of Trusted Third Parties in Electronic Commerce © 1996 A. Michael Froomkin . Version 1.02 Oct. 14, 1996.
9. DIRECTIVE 1999/93/EC DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 December 1999 on a Community framework for electronic signatures
10. American National Standard for Financial Services ANS X9.79-1:2001 Part 1: PKI Practices and Policy Framework
11. Request for Comments: RFC 2527, Certificate Policy and Certification Practices Framework, S. Chokhani, W.Ford, March 1999
12. Request for Comments: 2459 - Internet X.509 Public Key Infrastructure Certificate and CRL Profile
13. Request for Comments: 1422 Privacy Enhancement for Internet Electronic Mail

FRANCISZEK WOŁOWSKI
Kierownik "Sigillum" Polskiego Centrum Certyfikacji Elektronicznej
Polska Wytwórnia Papierów Wartościowych"
tel (+48 22) 5302752 fax: (+48 22) 5032 571
E-mail: f.wolowski@pwpw.pl

ZARZĄDZANIE CIĄGŁOŚCIĄ DZIAŁANIA

Janusz ZAWIŁA-NIEDŹWIECKI

- Niezawodność układu a ciągłość działania systemu
- Bezpieczeństwo systemu a jego ciągłość działania
- Niepewność – ryzyko – zagrożenie – zakłócenie – ciągłość działania
- Czynniki oceny skali zakłócenia
- Strategie postępowania z zagrożeniami
- Lista zagrożeń – Mapa zakłóceń
- Polityki postępowania z zakłóceniami
- Organizacja zarządzania ciągłością działania

Tak jak niezawodność, rozumiana jako stopień zdolności obiektu do spełniania stawianych mu wymagań, jest kluczowa dla oceny jakości układu czy wyrobu technicznego, tak w ocenach jakości systemów działania istotnym elementem tej oceny jest zdolność systemu do utrzymywania ciągłego działania. Tego typu dążenie w ocenie jakości wynika przede wszystkim z orientacji systemu działania na szeroko rozumianego klienta jako odbiorcę efektu działania systemu (produktu lub usługi), a składa się na również szeroko rozumianą kulturę organizacji.

Od strony organizacyjnej problematyka zarządzania ciągłością działania zbiega się z zarządzaniem jakością, zarządzaniem bezpieczeństwem i tego typu uzupełniającymi (wspierającymi) procesami, w znaczeniu analizy procesowej, działalności danej organizacji. Rozważanie integracji zarządzania tymi zagadnieniami, również w sensie wspólnych struktur organizacyjnych zarządzania, jest racjonalne, a pośrednio jest też rekomendowane przez normy ISO.

Ciągłość działania można rozpatrywać jako jeden z postulatów doskonałości teoretycznego systemu idealnego. Postulaty te to:

- Optymalny organizacyjnie
- Optymalny kosztowo
- Dokładnie powtarzalny co do jakości efektu działania
- W pełni skuteczny
- Bezpieczny
- Nieprzerwany

Praktyczne próby realizacji poszczególnych postulatów doskonałości oczywiście mogą i przeważnie pozostają we wzajemnej sprzeczności np. typowe, dla zapewnienia nieprzerwanego działania systemu, przedsięwzięcia zwiększania redundancji łączności lub wprowadzania środków bezpieczeństwa są sprzeczne z postulatami optymalnej organizacji i optymalnych kosztów działania. Postulat doskonałości nieprzerwanego działania systemu czyli inaczej jego ciągłości działania, jest w praktyce nieosiągalny, choć może służyć wyznaczaniu celu

optymalizacji działania systemu w podejściu prognostycznym do projektowania lub reengineeringu.

Fundamentem rozumowania nad mechanizmem powodującym niemożność uzyskania stanu pełnej ciągłości działania systemu, jest niepewność jako immanentna cecha rzeczywistości otaczającej człowieka i wytworzoną przezeń cywilizację, a wynikająca z wielkiej złożoności i zmienności obiektów i zjawisk w naturze oraz zależności zachodzących między nimi. Filozoficzna zasada niepewności stwierdza, że zjawiska w rzeczywistości otaczającej człowieka są zawsze niepewne, a skoro tak to należy je traktować jako przypadkowe i jeśli można odnaleźć w przyrodzie pewne prawidłowości odnośnie danego zjawiska, to tylko oparte na prawdopodobieństwie i statystycznym prawie wielkich liczb [1, str.276 i następne]. Ta filozoficzna definicja została wywiedziona z oraz zweryfikowana przez współczesne dokonania fizyki, zwłaszcza konfrontację klasycznej teorii mechaniki newtonowskiej z mechaniką kwantową i atomową.

Niepewność, która nieustannie dotyka człowieka wcale nie oznacza, że podejmuje on ryzyko. Aby tak się stało, musi bowiem zaistnieć potrzeba działania, które może mieć charakter ryzykowny, co jest konsekwencją potrzeby podjęcia decyzji.

NIEPEWNOŚĆ	
Potrzeba decyzji i działania	Brak potrzeby decyzji i działania
Powstanie ryzyka	Brak ryzyka

Ryzyko zachodzi gdy podjęte działanie lub decyzja może być traktowane jako próba w eksperymencie podzielnym, tj. gdy wynik może być określony za pomocą jednego z trzech rodzajów prawdopodobieństwa: matematycznego, statystycznego lub szacunkowego, z których każdy opiera się na obiektywnej wiedzy. Wiedza ta musi usprawiedliwić przekonanie, że to co zaszło w przeszłości powtórzy się w przyszłości. Istotą ryzyka jest prawdopodobieństwo, które zakłada wiedzę, a ta wyklucza niepewność [2, str.30]. W sensie praktycznym ryzyko jako kategoria pojęciowa ogólna polega na tym, że podejmowana decyzja i będące jej konsekwencją działanie systemu może napotykać na konkretne utrudnienia i przeszkody właściwe dla ogólnej niepewności w obszarze i środowisku działania systemu, a antycypacyjnie postrzegane jako zagrożenia. A więc zagrożenie:

- Jest formą realizowania się potencjalnego dotąd ryzyka
- Ma postać i cechy właściwe mierzalne obiektywnie
- Ma źródło i przyczyny
- Cechuje specyficzny mechanizm realizowania się

- Oddziałuje na system działania w sposób mierzalny subiektywnie z perspektywy systemu działania, a stopień wpływu jest zależny od podatności systemu działania i jego środowiska.

Zagrożenia, które rozważamy w praktyce biznesowej, dzielą się na następujące kategorie:

- Katastrofy naturalne
- Terroryzm
- Destrukcja fizycznego środowiska pracy
- Destrukcja funkcjonalnego środowiska pracy
- Destrukcja technicznego środowiska pracy
- Destrukcja informatycznego środowiska pracy.

Gdy dane zagrożenie oddziałuje na system działania lub jego środowisko, a system staje się podatny na to oddziaływanie, mamy do czynienia z zakłóceniem, które:

- Jest skutkiem interakcji zagrożenia z systemem działania lub środowiskiem tego systemu
- Skutkuje istotnymi zmianami w obszarze działania systemu
- Nie poddaje się ocenie obiektywnej, zaś subiektywna dokonywana jest z perspektywy systemu działania.

Zgodnie już choćby ze zdrowym biznesowym rozsądkiem, ale także np. z normą ISO17799, zarządzając działaniem systemu należy tworzyć rozwiązania skutecznie zapewniające zachowanie ciągłości tego działania. Przez analogię do organizmów żywych rozwiązania takie mają stanowić

o zdolności do homeostazy tj o cesze systemu działania polegającej na uruchamianiu własnego wewnętrznego mechanizmu przeciwdziałania przez system zakłóceniu celem przywrócenia stanu sprzed pojawienia się tego zakłócenia. Skuteczność rozwiązań antycypujących zakłócenia i ich adekwatność do faktycznych zdarzeń powinna się plasować powyżej progu minimalnej akceptacji przez decydentów, którzy oceny dokonują zwykle w świetle dwu kryteriów:

- Prestiżu organizacji i stopnia jego podważenia w wyniku zawieszenia lub ograniczenia działania systemu
- Kosztów rozwiązań zabezpieczających oraz kosztów strat i przywracania działania naruszonego zakłóceniem.

Racjonalnie pojmowana homeostaza systemu działania świadomie prowadzi do okresowego ograniczenia jakości działania systemu do poziomu zawczasu ustalonego w świetle wyznaczników takich jak:

- Utrata nie usatysfakcjonowanego lub poszkodowanego klienta
- Benchmarking wobec konkurentów lub najlepszych praktyk rynkowych
- Solidne standardy współpracy z partnerami i klientami tzw. „Service Level Agreement".

Ograniczenie jakości nie powinno trwać dłużej aniżeli czas potrzebny na usunięcie przyczyn i skutków zakłócenia, przy czym niekiedy te pierwsze mogą ustąpić samoistnie, jeśli taki jest charakter zakłócenia.

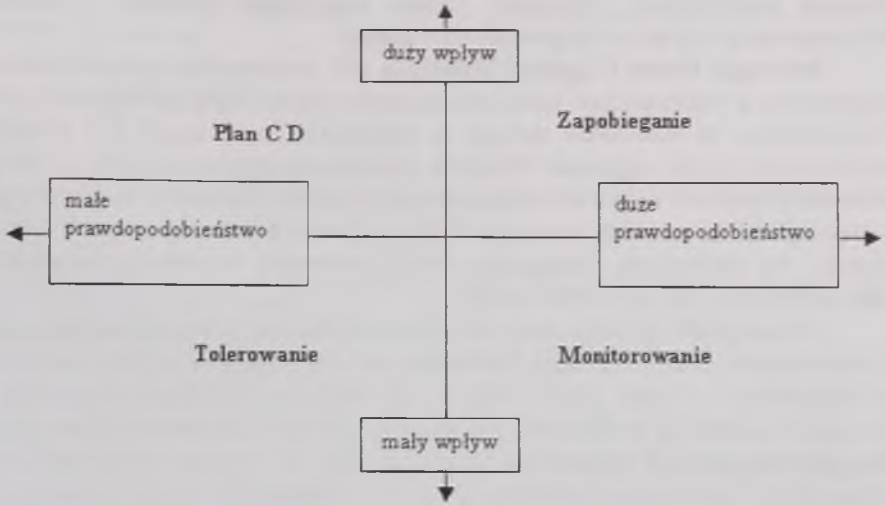
Percepcja zakłóceń jako zjawisk naruszania ciągłości działania uwzględnia i opiera się na dwóch zasadniczych czynnikach oceny ich istotności:

- Prawdopodobieństwie lub częstości występowania zakłócenia
- Wpływie (destrukcyjnym lub nie) zakłócenia na ciągłość działania.

Czynniki oceny nie oznaczają miar, te bowiem odpowiadają naturze zjawiska każdego konkretnego zakłócenia z osobna. Ocena powinna być indywidualna i dokonana z punktu widzenia danej organizacji.

Modelowe podejście do reagowania na zagrożenia/zakłócenia może odbywać się na cztery sposoby określane mianem strategii reagowania (patrz rys 1). Klasyfikacja poszczególnych przypadków zagrożeń do poszczególnych strategii jest nie tylko indywidualna oraz subiektywna, ale i tymczasowa. Wskazana jest okresowa, w miarę częsta weryfikacja takiej oceny uwzględniająca rozwój organizacji i jej systemu działania oraz narastanie wiedzy o wpływie poszczególnych zagrożeń i prawdopodobieństwie ich wystąpienia w formie zakłóceń. Przydział do każdej z kategorii powinien uwzględniać kryteria ekonomiczne i prestiżu.

Strategie reagowania na zagrożenia



Rys 1. Strategie reagowania na zagrożenia

Tolerowanie oznacza pogodzenie się z przejściowymi niedogodnościami. **Monitorowanie** oznacza, że wiedza o zakłóceniu jest dostateczna do uruchomienia mechanizmu kompensacji. **Zapobieganie** oznacza działania (i wydatki) w celu zapobieżenia negatywnym skutkom zakłócenia. **Plan Ciągłości Działania** jest zestawem scenariuszy przewidywanego materializowania się zagrożeń oraz działań zaplanowanych na taką okoliczność.

Strategia Tolerowania (T) powinna być przypisywana do postępowania z zakłóceniami
w swej naturze zewnętrznymi wobec danej organizacji, a tylko wtórnie jej dotyczącymi, w szczególności nie inwazyjnymi, a zwłaszcza nie destrukcyjnymi. Przykład: Kolporter prasy – przeczekanie mgły porannej i późniejsze rozwiezienie gazet.

Strategia Monitorowania (M) powinna być przypisywana do postępowania z zakłóceniami w swej naturze drobnymi, choć częstymi (przez co należy zakładać ich incydentalnie większy wpływ przez kumulację zdarzeń w krótkim czasie), ale wyraźnie nie destrukcyjnymi. Z tej strategii ma wynikać obowiązek szczegółowego rozwiązania przez posunięcia organizacyjne oraz nawet drobiazgowe regulacje wewnętrzne reakcji na wszelkie typowe zakłócenia. Istotą jest nikły lub wręcz żaden wzrost kosztu z tytułu rozwiązań reagowania, bowiem mają one przede wszystkim organizacyjny charakter. Przykład: Nieobecności chorobowe pracowników – obowiązek zawiadamiania zawczasu zakładu pracy oraz opracowane zasady organizowania zastępstw.

Strategia Zapobiegania (Z) powinna być przypisywana do postępowania z zakłóceniami istotnymi, destrukcyjnymi i potencjalnie często występującymi. Jest

to strategia prewencji,, jej naturalnymi konsekwencjami są inwestycje i rozwiązania ograniczające ryzyko zagrożenia. Typowe posunięcia to dublowanie rozwiązań technicznych. Przykład: Częste wyłączenia zasilania – instalacja podtrzymywaczy napięcia lub generatorów energii.

Strategia Planu Ciągłości Działania (P) powinna być przypisywana do postępowania z zakłóceniami istotnymi, destrukcyjnymi, lecz potencjalnie rzadko występującymi, co uzasadnia decyzje o rezygnacji ze Strategii Z i świadome podejmowanie ryzyka zagrożeń. Przykład: Giełda – światowe statystyki mówią, że zawieszenie notowań z powodu niesprawności systemu komputerowego zdarza się nie częściej niż raz na 3 lata i trwa nie dłużej niż jeden dzień, uzasadnione jest więc poleganie na scenariuszu zastępczego funkcjonowania w trakcie usuwania tak rzadko zdarzającej się, poważnej awarii.

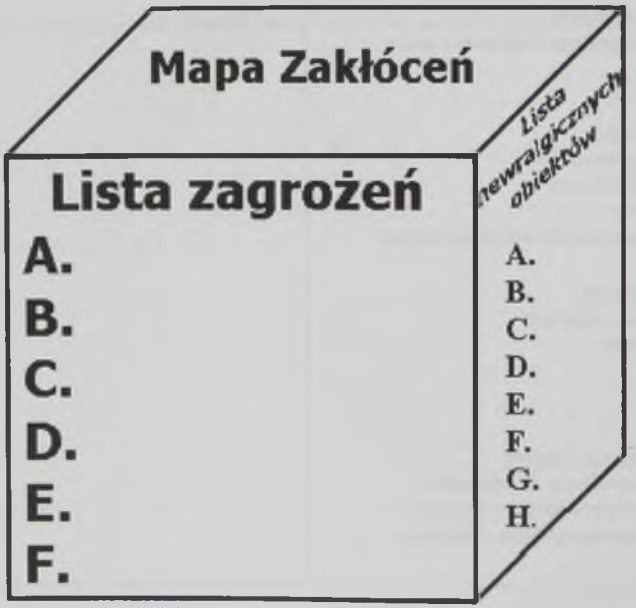
Poszczególne strategie, jako modelowe wytyczne postępowania, przypisuje się zakłóceniom, których **Mapa Zakłóceń** jest opracowywana jako zestawienie par: zagrożenie – obiekt (patrz tabl. 1 i 2 oraz rys. 2). **Lista Zagrożeń** jest sporządzana zgodnie z podziałem na kategorie zagrożeń przedstawione wcześniej. **Lista Newralgicznych Obiektów** tworzona jest w wyniku nałożenia obrazu podstawowych procesów biznesowych [3] i związanych z nimi przepływów informacji na plan przestrzenny całej infrastruktury budowlanej i technicznej organizacji. Modelowe strategie są następnie rozwijane do postaci **Polityk Postępowania z Zakłóceniami.**

Tablica 1. lista zagrożeń

Grupy / Zagrożenia	Ocena
A. Katastrofy naturalne • Trzęsienie ziemi • Skażenie środowiska • Powódź • Huragan • Wyładowania atmosferyczne	
B. Terroryzm • Szantaż • Zamach	
C. Zakłócenia fizycznego środowiska pracy • Brak dostępu do siedziby • Uszkodzenie budynku • Za niska / wysoka temperatura powietrza • Za duża wilgotność powietrza • Pożar • Zalanie	
D. Zakłócenie funkcjonalnego środowiska pracy • Strajk • Sabotaż • Niedostępność pracowników • Wypadek	
E. Zakłócenie technicznego środowiska pracy • Wyczerpanie zapasów mat. • Brak zasilania • Awaria klimatyzacji	
F. Zakłócenie informatycznego środowiska pracy Infrastruktura techniczna: • Serwerów • Stacji roboczych • Urządzeń pomocniczych • Urządzeń sieciowych • Okablowania • Brak połączenia z sieciami zewnętrznymi Oprogramowanie: • Wygaśnięcie licencji • Nieautoryzowane usunięcie • Wadliwe działanie Wirusy: Itd. Dane: • Utrata lub zniszczenie danych • Nieautoryzowany dostęp do informacji • Nieautoryzowane powielanie informacji • Nieautoryzowana modyfikacja informacji	

Tablica 2. lista newralgicznych obiektów

Kategorie	Obiekty	Np.
A. Budynki		własny biurowiec
B. Obiekty przemysłowe, Techniczne		hala fabryczna, kotłownia, ośrodek komputerowy
C. Ośrodki Biurowe		powierzchnia biurowa wynajęta w obcym budynku
D. Zewnętrzne Urządzenia Techniczne		wolnostojący generator napięcia zewnętrzny
E. Wewnętrzne Urządzenia Techniczne		klimatyzator, generator wewnętrzny
F. Infrastruktura Informatyczna		
G. Zewnętrzne Urządzenia Telekomunikacyjne		antena satelitarna na dachu, router, który znajduje się u klienta
H. Usługi Obce		telekomunikacja



Rys. 2.

Polityka Tolerowania (T) powinna określać podstawowe zasady przystępowania organizacji do stanu pogodzenia z zaistniałym zakłóceniem, badania przesłanek jego utrzymywania się, stwierdzania jego ustąpienia oraz powrotu do rutynowego funkcjonowania. Dokumentowi **Polityki T** powinny towarzyszyć procedury/instrukcje szczegółowo określające konieczne działania komórek organizacji w sytuacjach zakłóceń zakwalifikowanych do poddania ich tej polityce. Przykład: mimo, że reakcja organizacji na zakłócenie może krańcowo polegać na zawieszeniu wypełniania statutowych funkcji, to być może należy poinformować o tym partnerów handlowych lub opinię publiczną, skierować pracowników do prac zastępczych nie poddających się działaniu zakłócenia, uruchomić rozwiązania śledzące stopień intensywności zakłócenia. W momencie zaś ustąpienia zakłócenia należy dokonać weryfikacji czy jest możliwe podjęcie zawieszonych dotąd czynności/funkcji.

Polityka Monitorowania (M) powinna określać podstawowe zasady reagowania organizacji na zakłócenia, co do których świadomość ich zaistnienia w połączeniu z istniejącymi regułami zachowań (ewentualnie spisanymi w postaci procedur i instrukcji) powinna w dostatecznym stopniu uruchamiać organizacyjne mechanizmy kompensacji zakłócenia. Dokumentowi **Polityki M** powinny towarzyszyć procedury/instrukcje szczegółowo określające konieczne działania komórek organizacji w sytuacjach zakłóceń zakwalifikowanych do poddania ich tej polityce. Przykład: w banku, pracowników bezpośredniej obsługi klientów obowiązuje procedura uprzedzania o nieobecnościach spowodowanych np. chorobą, zaś określona liczba pracowników zaplecza jest przeszkolona do obsługi klientów na zasadzie zastępstw, przy czym każdego dnia określona ich liczba ma być gotowa do podjęcia takiej zastępczej pracy na wypadek absencji nadzwyczajnej, o której pracownik nie uprzedził zawczasu.

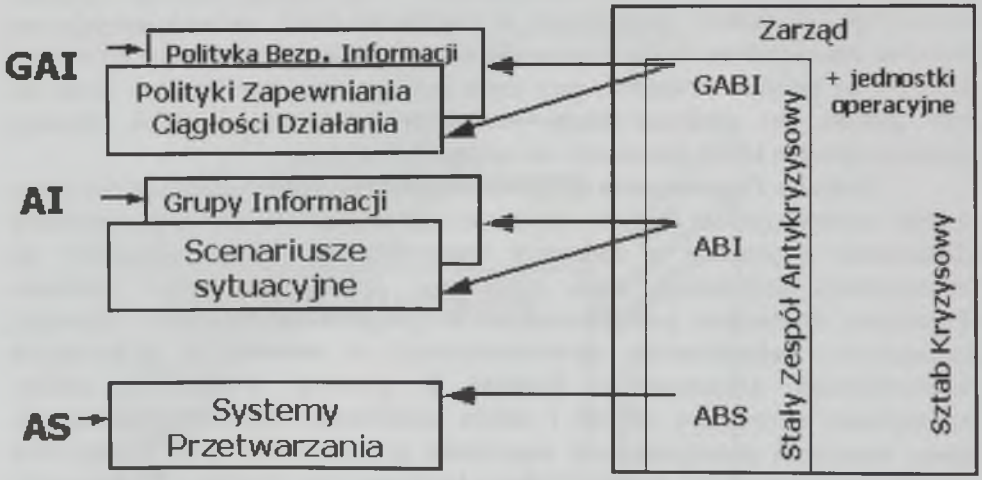
Polityka Zapobiegania (Z) powinna określać plany organizacji dotyczące działań prewencyjnych, które w odniesieniu do szczególnie istotnych elementów działalności organizacji, a zwłaszcza szczególnie wrażliwych elementów jej infrastruktury technicznej, mają zniwelować destrukcyjny wpływ zakłóceń. Typowymi działaniami podejmowanymi w tym celu są: tworzenie rozwiązań zapasowych, nadmiarowych, zwielokrotnionych w stosunku do przeciętnych zapotrzebowań. Dokumentowi **Polityki Z** powinny towarzyszyć analizy szczegółowo określające stopień i zakres wrażliwości rozwiązań istniejących, plany rozwiązań zmniejszających zagrożenia, procedury/instrukcje szczegółowo określające organizację i zasady działania bieżącego oraz specjalnych interwencji specjalistycznych zespołów do zwalczania specyficznych zagrożeń (pożar, atak hakerski, awaria informatyczna). Przykład: zapasowy ośrodek komputerowy, dublowane linie komunikacyjne prowadzone fizycznie różnymi drogami i/lub z wykorzystaniem odmiennych mediów transmisji. Także dyżury specjalnych ekip interwencyjnych o stosownych kwalifikacjach.

Równocześnie należy podkreślić, że każda z par obiekt – zagrożenie ujęta w **Polityce Z**, a więc w planie działań prewencyjnych, o ile polegają one na inwestycjach zmniejszających zagrożenie, to do czasu ich zakończenia powinny

być ujęte także w jednej z pozostałych **Polityk** (zaleca się aby w Polityce P) celem zapewnienia stosownej reakcji na zagrożenie.

Polityka Planu Ciągłości Działania (P) powinna określać plany organizacji dotyczące działań koniecznych w przypadku zmaterializowania zagrożenia w postaci konkretnego zakłócenia. Plany powinny obejmować rozwiązania organizacyjne dotyczące prowadzenia samej **Polityki** oraz scenariusze przypadków zakłóceń i zakładanych wobec nich działań, mających na celu zapewnienie kontynuacji przynajmniej podstawowej aktywności biznesowej organizacji. Ponadto **Polityka P** powinna określać zasady reagowania *ad hoc* na zdarzenia, których niestety nie udało się przewidzieć w scenariuszach (w ogóle lub co do skali). Dokumentowi **Polityki P** powinny towarzyszyć procedury/instrukcje szczegółowo określające organizację służb prowadzących plany ciągłości działania, podstawowe reguły komunikowania się w warunkach awarii, zasady reagowania na typowe zagrożenia, scenariusze przewidywanych rozległych zakłóceń i reagowania na nie, zasady uwzględniania w nowych wersjach planów awaryjnych doświadczeń ze zwalczania świeżo zaszłych zakłóceń.

Zasadniczą ideę co do struktury organizacyjnej służb odpowiadających za bieżące zarządzanie problematyką przygotowywania **Polityk Postępowania z Zakłóceniami** oraz organizację sztabu kryzysowego powoływanego na wypadek awarii przedstawia rys. 3.



Rys. 3. Zintegrowane zarządzanie bezpieczeństwem informacji i ciągłością działania

Gdzie, zgodnie z [4]:

GAI - Główny Administrator Informacji

GABI - Główny Administrator Bezpieczeństwa Informacji

AI - Administrator Grupy Informacji

ABI - Administrator Bezpieczeństwa Grupy Informacji

AS - Administrator Systemu

ABS - Administrator Bezpieczeństwa Systemu

Podsumowanie

jako streszczenie wytycznych normy ISO 17799 (rozdz.11)

- W działalności organizacji należy się liczyć z możliwością zaistnienia wydarzeń krytycznych, które uniemożliwią normalne kontynuowanie tej działalności.
- Niezależnie od charakteru przyczyn tych wydarzeń, w ramach formalnego lub postrzeganego
  w kategoriach biznesowych obowiązku dołożenia należytej staranności w wypełnianiu swych zadań, organizacja powinna dążyć do choćby ograniczonego kontynuowania działalności.
- Staranie takie powinno opierać się na uprzednio opracowanym, konsekwentnie doskonalonym
  i testowanym „Planie Ciągłości Działania" (PCD).
- Zapewnianie ciągłości działania to przewidywanie potencjalnych scenariuszy zakłóceń oraz rozdzielne projektowanie:
  * rozwiązań zapobiegających samym zagrożeniom,
  * rozwiązań służących jak najszybszemu usuwaniu skutków zakłóceń ,
  * rozwiązań kontynuowania ograniczonej działalności w warunkach krytycznych (PCD).
- Podejście do problemu ciągłości działania powinno być racjonalne, a więc przede wszystkim obliczone na zapewnienie równowagi pomiędzy oczekiwanym stopniem pewności zachowania ciągłości działania a kosztami jego uzyskania. Konieczne więc jest przyjęcie założenia stopniowego rezygnowania z kolejnych elementów normalnej działalności stosownie do zidentyfikowanych rozmiarów sytuacji krytycznej na wzór wojskowego planu wycofywania się na z góry upatrzone pozycje.
- PCD powinien być na tyle elastyczny, aby umożliwiał adaptatywne reagowanie na zakłócenia odbiegające od przewidywań będących podstawą planu.
- Konieczne jest zdefiniowanie procesowej istoty działalności danej organizacji jako minimum czynności, które można jeszcze uznać za wypełnianie przez nią jej obowiązków. Niemożność kontynuowania takiego minimum czynności jest podstawą decyzji o rezygnacji z korzystania z PCD i skupieniu się tylko na usuwaniu skutków zakłóceń.

- W ramach przygotowywania PCD rozważa się w pierwszym rzędzie aspekty biznesowe, prawne i organizacyjne, bo one decydują o koniecznym zakresie rozwiązań technicznych.
- Analiza biznesowa dotyczyć może kwestii prestiżu firmy, a na pewno swoistego bilansu ryzyka oraz środków finansowych przeznaczonych na jego ograniczenie. Rozsądnym jest potraktowanie PCD jako długofalowego projektu, w którym założone cele biznesowe będą osiągane stopniowo, kolejnymi przybliżeniami (wersjami PCD).
- Analiza prawna jest szczególnie ważna przy tworzeniu założeń PCD, bowiem pozwala zdefiniować zakres odpowiedzialności firmy za poszczególne obszary jej działalności, wskazać obszary newralgiczne oraz dobrać pozatechniczne formy zabezpieczeń.
- Analiza organizacyjna pozwala wyodrębnić kadrę osób właściwą dla posługiwania się PCD w warunkach krytycznych, stworzyć im odpowiedni zakres autonomii decyzyjnej w takiej sytuacji, a w warunkach codziennych umożliwić przygotowywanie się do takiej trudnej roli.
- Żaden z elementów analizy, a podobnie i projektowanie rozwiązań technicznych, nie jest etapem zamkniętym. Doskonalenie PCD polega na stałym ponawianiu takich analiz i projektowania w odniesieniu do zmian w działalności organizacji, rozwoju rozwiązań PCD oraz wniosków z wystąpienia faktycznych zagrożeń.

Literatura

1. Tatarkiewicz W. „Historia filozofii" tom 3, PWN, Warszawa 1995
2. Samecki W. „Ryzyko i niepewność w działalności przedsiębiorstwa przemysłowego", PWE, Warszawa 1967
3. Rummler G, Brache A. „Podnoszenie efektywności organizacji", PWE, Warszawa 2000
4. „Metodyka zapewniania ciągłości działania TISM-BCP", European Network Security Institute, Warszawa, 2003; www.ensi.net,

dr inż. Janusz Zawiła-Niedźwiecki
Politechnika Warszawska
Instytut Organizacji Systemów Produkcyjnych
Warszawa, ul.Narbutta 84
jzawila@poczta.onet.pl

WYKORZYSTANIE INFRASTRUKTURY KLUCZA PUBLICZNEGO DO UWIERZYTELNIANIA W SYSTEMACH INFORMATYCZNYCH

Wojciech WILEŃSKI, Leszek ZDAWSKI

Streszczenie: artykuł ma celu przedstawienie istoty infrastruktury klucza publicznego oraz jej zastosowania w uwierzytelniania dostępu do zasobów informatycznych. Ostatni rozdział zawiera opis rzeczywistego zastosowania przedstawionych funkcjonalności na przykładzie zintegrowanego systemu uwierzytelniania i podpisu elektronicznego proCertum ADS firmy Unizeto Sp. zo.o.

1. Istota podpisu elektronicznego

W dniu 16 sierpnia 2002 zaczęła obowiązywać w Polsce Ustawa o podpisie elektronicznym[1]. Ustawa ta zrównuje podpis elektroniczny z podpisem własnoręcznym[2] i określa między innymi warunki stosowania podpisu elektronicznego oraz skutki prawne jego stosowania.

Co to jest podpis elektroniczny? Zgodnie z definicją zawartą w przytoczonej ustawie są to cyt.: ... *dane w postaci elektronicznej, które wraz z innymi danymi, do których zostały dołączone lub z którymi są logicznie powiązane, służą do identyfikacji osoby składającej podpis elektroniczny*".

W praktyce podpis elektroniczny realizowany jest w oparciu o kryptografię asymetryczną. Kryptografia ta, krótko mówiąc jest sposobem szyfrowania wiadomości opartym o dwa klucze[3]. Klucze te stanowią parę i są tak dobrane, że to, co zaszyfrujemy jednym kluczem możemy odszyfrować **tylko** drugim kluczem. Ponadto, klucze te charakteryzują się jeszcze dodatkowo tym, że poznanie jednego klucza nie umożliwia wyliczenia drugiego.

Jeżeli użytkownik Internetu dysponuje parą wyżej opisanych kluczy i chce, aby treść wysyłanej do niego korespondencji była utajniona to wystarczy, że udostępni nadawcom wysyłanych do niego wiadomości jeden z kluczy. Wówczas mogą oni szyfrować wysyłane do niego wiadomości, a odszyfrować je może tylko właściciel drugiego klucza. Ten pierwszy klucz, powszechnie rozpowszechniany, nazwany jest kluczem publicznym, a drugi, tajny, kluczem prywatnym.

Dostępne oprogramowanie i sprzęt np. karty kryptograficzne zapewniają pełne bezpieczeństwo przy generowaniu kluczy tzn. właściciel klucza prywatnego ma pewność, że nikt poza nim nie zna tego klucza i nikt nie jest w stanie go odtworzyć.

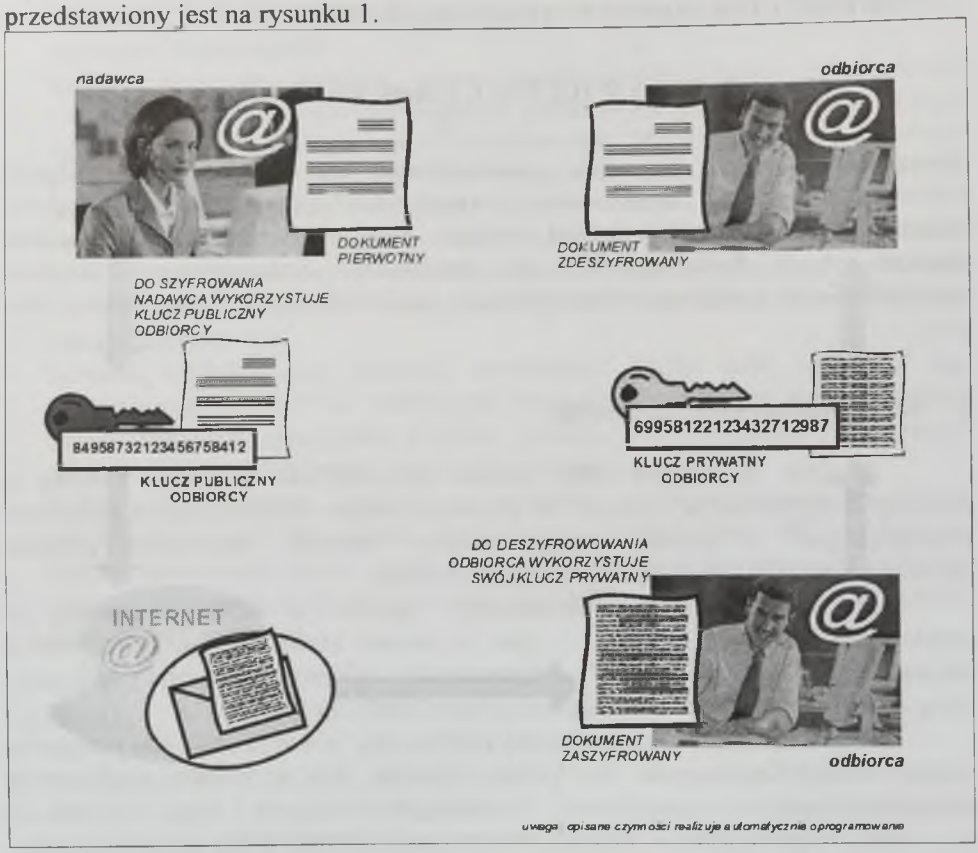
---

[1] Dziennik Ustaw nr130/2001 z dnia 15 listopada 2001 r.

[2] Tamże, Art.5 ust.2

[3] klucz – składnik kryptosystemu określający sposób szyfrowania wiadomości

Proces szyfrowania wiadomości przy użyciu klucza publicznego przedstawiony jest na rysunku 1.



Rys. 1 Szyfrowanie wiadomości przy użyciu klucza publicznego

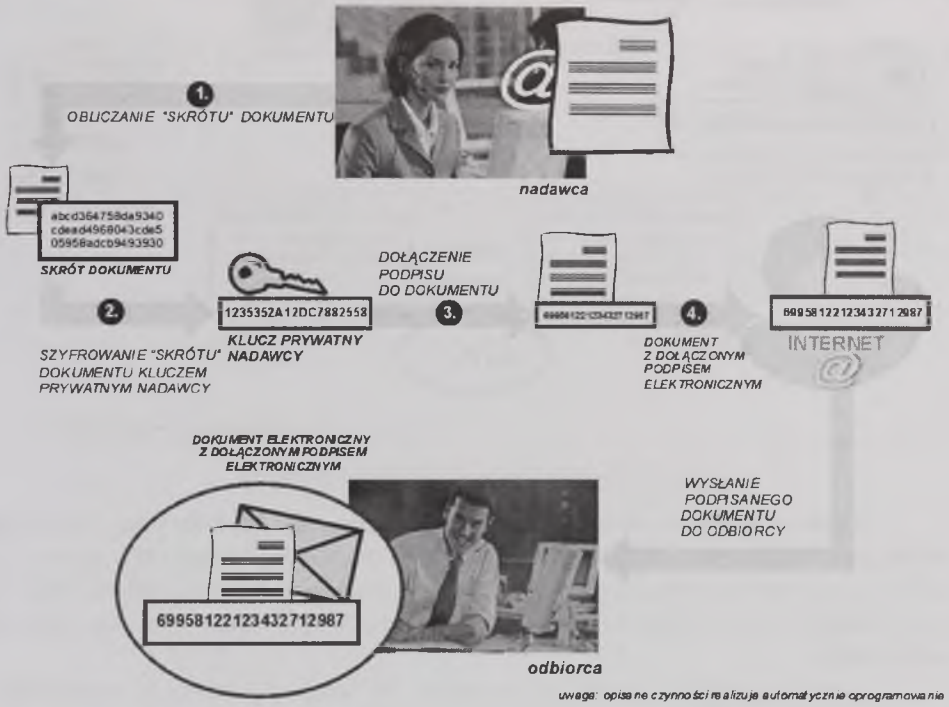
Wyżej opisane właściwości kryptografii asymetrycznej zostały wykorzystane również w innym celu, a mianowicie do potwierdzania tożsamości autora wiadomości. Idea takiego zastosowania kryptografii asymetrycznej jest bardzo prosta. Jeżeli otrzymamy zaszyfrowany tekst, który można odszyfrować kluczem publicznym nadawcy to mamy pewność, że został on zaszyfrowany kluczem prywatnym nadawcy. Klucz prywatny jest tajny i znany tylko nadawcy, w związku z tym nie ma wątpliwości, że ta wiadomość pochodzi od niego.

Opisane powyżej zastosowanie klucza publicznego i prywatnego jest podstawą funkcjonowania podpisu elektronicznego. Składanie podpisu elektronicznego pod dokumentem przebiega w następujący sposób[4]:

[4] C. Adams, St. Lloyd, *Podpis elektroniczny klucz publiczny,* Wydawnictwo Robomatic, Wrocław 2002, s.16

- za pomocą funkcji mieszającej (ang. cryptographic hash function) wyliczany jest skrót wiadomości; skrót ten charakteryzuje się tym, że ma stałą długość, jest niepowtarzalny tzn. dla dwóch nawet nieznacznie różniących się wiadomości jest inny oraz tym, że znając skrót praktycznie jest niemożliwe odtworzenie wiadomości.
- wyliczony skrót jest szyfrowany kluczem prywatnym autora wiadomości.

Tak wygenerowany podpis elektroniczny jest unikalny tzn. jest ściśle związany z podpisaną wiadomością, dla różnych wiadomości jest inny i w odróżnieniu od podpisu własnoręcznego nie istnieje w oderwaniu od wiadomości. Proces składania podpisu elektronicznego ilustruje rysunek 2.



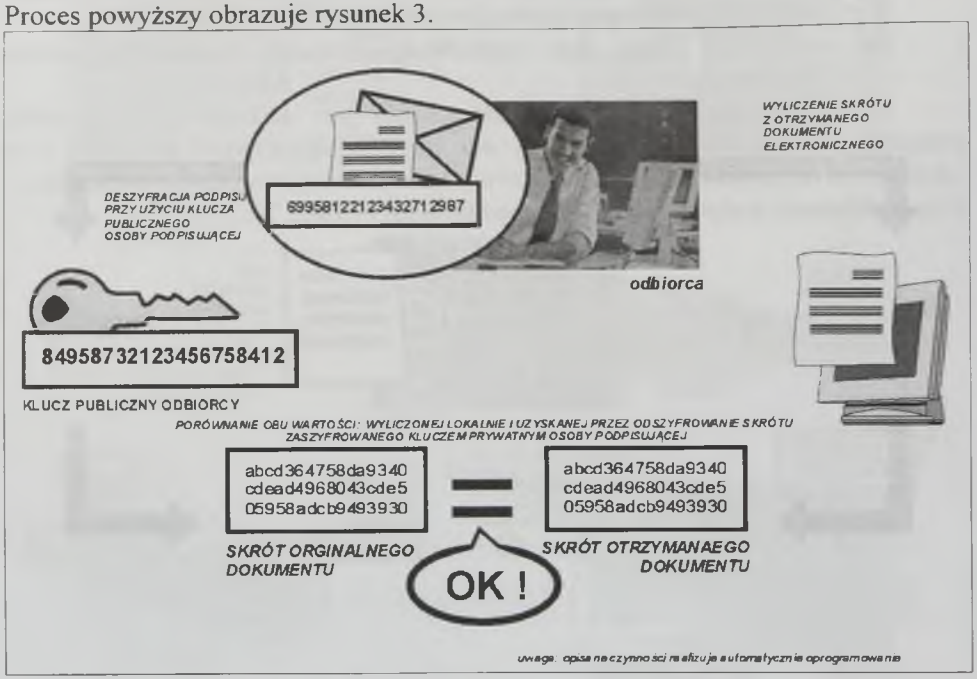
Rys. 2 Proces składania podpisu elektronicznego

Odbiorca podpisanej wiadomości, dysponując kluczem publicznym nadawcy, ma możliwość upewnienia się, kto podpisał wiadomość oraz sprawdzenia czy wiadomość ta od momentu podpisania do momentu odczytania nie została zmieniona. Proces ten zwany weryfikacją podpisu odbiorca przeprowadza następująco:

- za pomocą funkcji mieszającej wylicza skrót wiadomości
- za pomocą klucza publicznego nadawcy odszyfrowuje jego podpis (zaszyfrowany skrót)

- porównuje otrzymane wartości; jeżeli są równe to znaczy, że wiadomość została podpisana przez nadawcę oraz nie została zmieniona od momentu podpisania – sprawdzona została integralność.

Proces powyższy obrazuje rysunek 3.



Rys. 3 Weryfikacja podpisu elektronicznego

Oparcie podpisu elektronicznego na kryptografii asymetrycznej pozwala dodatkowo szyfrować podpisaną wiadomość przed wysłaniem (używamy do szyfrowania klucza publicznego odbiorcy) oraz szyfrować archiwizowane dane. W tym ostatnim przypadku do szyfrowania danych używamy własnego klucza publicznego.

Należy pamiętać jednak, że ze względów bezpieczeństwa do szyfrowania powinno się używać innej pary kluczy niż do podpisywania, a w przypadku podpisu „kwalifikowanego" Ustawa o podpisie elektronicznym zakazuje szyfrowania kluczami używanymi do podpisywania[5].

2. Funkcje Infrastruktury Klucza Publicznego

W opisywanych w poprzednim rozdziale zasadach funkcjonowania podpisu elektronicznego założyliśmy, że odbiorca dysponuje kluczem publicznym

[5] Ustawa z dnia 18 września 2001 r. o podpisie elektronicznym, (Dz.U. 2001, nr 130, poz.1450)

nadawcy. Jest to możliwe, jeżeli uczestnicy korespondencji znają się i wcześniej wymienili swoje klucze publiczne. Ale co zrobić w przypadku powszechnego używania podpisu elektronicznego? Skąd odbiorca ma wziąć klucz publiczny nieznanego nadawcy i mieć pewność, kto jest jego właścicielem?

Klucze publiczne z ich właścicielami wiążą certyfikaty klucza publicznego. Zgodnie z definicją certyfikatu, podaną w ustawie o podpisie elektronicznym jest to, cyt.: *„... elektroniczne zaświadczenie, za pomocą którego dane służące do weryfikacji podpisu elektronicznego są przyporządkowane do osoby składającej podpis elektroniczny i które umożliwiają identyfikację tej osoby."* Certyfikat powinien zawierać między innymi imię i nazwisko lub pseudonim osoby składającej podpis, termin ważności certyfikatu i klucz publiczny.

Do szerokiego funkcjonowania podpisu elektronicznego potrzebna jest jeszcze instytucja, która wydaje certyfikaty, weryfikuje zawarte w nich dane i zarządza certyfikatami. Instytucją tą jest Urząd Certyfikacji zwany też Centrum Certyfikacji lub Zaufaną Trzecią Stroną (ang. Trusted Third Party). Urząd ten jest częścią infrastruktury wspierającej powszechne stosowanie podpisu elektronicznego. Infrastruktura ta, znana jako Infrastruktura Klucza Publicznego (ang. PKI-public key infrastructure) składa się ze sprzętu, oprogramowania, ludzi, polityki oraz procedur niezbędnych do tworzenia, posługiwania się, przechowywania, dystrybucji i unieważniania certyfikatów opartych na kryptografii klucza publicznego[6].

3. Usługi certyfikacyjne

Podstawowe usługi świadczone przez Urząd Certyfikacji to:

- wydawanie certyfikatów
- zawieszanie certyfikatów
- unieważnianie certyfikatów
- publikowanie list unieważnionych i zawieszonych certyfikatów – CRL (ang. Complete Certificate Revocation Lists)

Wydanie certyfikatu poprzedzone jest rejestracją subskrybenta – przyszłego właściciela certyfikatu. Rejestracja polega przede wszystkim na weryfikacji tożsamości i jest realizowana bezpośrednio przez Centrum Certyfikacji lub przez podległy mu Punkt Rejestracji.

Dodatkowe usługi, które mogą być świadczone przez Urząd Certyfikacji to:

---

[6] A. Kruk, P. Matusiewicz, J. Pejaś, A. Ruciński, W. Ślusarczyk, *Podpis elektroniczny – sposób działania, zastosowanie i korzyści,* Ministerswo Gospodarki, Pracy i Polityki Społecznej, Warszawa 2003, s. 26

- znakowanie czasem (ang. time stamping) – umożliwia oznaczenie wiarygodnym czasem każdej transakcji (aukcja, logowanie, uruchomienie aplikacji) w sieci,
- informacja w czasie rzeczywistym o statusie certyfikatu – OCSP (ang. Online Certificate Status Protocol), to usługa, która umożliwia zweryfikowanie statusu certyfikatu on-line,
- notariat elektroniczny (ang. notary authority) umożliwia notaryzację transakcji elektronicznych w celu potwierdzenia ich autentyczności,
- kurier elektroniczny (ang. delivery authority)

Usługa kuriera elektronicznego jest odpowiednikiem tradycyjnej przesyłki za zwrotnym potwierdzeniem odbioru. Usługa ta polega na wydawaniu poświadczeń nadania i odbioru przesyłek elektronicznych. Kurier elektroniczny archiwizuje daty wysyłania i odebrania elektronicznych przesyłek, ich skróty kryptograficzne oraz umożliwia monitorowanie statusu tych przesyłek.

4. Zastosowanie certyfikatów klucza publicznego przy uwierzytelniania dostępu

4.1. Porównanie cech różnych typów uwierzytelniania dostępu

Zastosowanie tradycyjnych systemów uwierzytelniania dostępu, typu: „login/hasło" jest wystarczającym rozwiązaniem dla małych i średnich przedsiębiorstw, których wszyscy pracownicy (czyt.: użytkownicy systemu) często spotykają się w siedzibie firmy. Jest możliwość zweryfikowania tożsamości użytkownika „na żywo" a ilość zdalnych użytkowników jest stała i jest to znana grupa osób.

Niestety poziom bezpieczeństwa tradycyjnych systemów uwierzytelniania znacznie spada w dużych przedsiębiorstwach, które posiadają placówki na terenie całego kraju, kontynentu oraz szeroką grupę użytkowników zdalnych. Często muszą pilnie przydzielić dostęp do zasobów informatycznych użytkownikom, którzy nie mogą szybko stawić się do placówki firmy w celu zweryfikowania tożsamości i przekazania sekretu[7]. Użytkownicy wykonują autoryzację do wielu usług i aplikacji korporacyjnych, co wymaga od nich pamiętania kilku (do kilkunastu!) zestawów: login, hasło. Mimo zastosowania najbardziej wyrafinowanej technologii w budowie bezpieczeństwa systemu najsłabszym ogniwem wciąż pozostaje człowiek. Konieczność posługiwania się, częściej lub rzadziej, kilkoma zestawami haseł powoduje, że użytkownicy wspomagają swoją pamięć w sposób „niedozwolony": zapisują poszczególne hasła w kalendarzach, na biurkach i monitorach (!), udostępniają współpracownikom. Tego typu zachowania

[7] Sekret – na potrzeby artykułu przyjmijmy definicję, iż sekretem jest para informacji: nazwa użytkownika (potocznie: login) oraz hasło, pewna cecha biometryczna lub certyfikat klucza publicznego.

z kolei są znakomitą pożywką dla crackerów wykorzystujących socjotechniki w celu dostania się do dobrze zabezpieczonego systemu. Skuteczność polityki bezpieczeństwa znacznie spada. Kolejnym wyzwaniem, które stawiają sobie nieetyczni użytkownicy to kradzież baz z nazwami i hasłami użytkowników.

Rozwiązaniem problemu zapamiętywania sekretów może być oparcie systemu uwierzytelniania o cechy biometryczne użytkownika (np. linie papilarne kciuka, barwa głosu czy tęczówka oka), których na pewno użytkownik nie zapomni, nie zostawi w domu i ma cały czas ze sobą. Zaczynają się komplikacje, gdy pojawi się blizna na palcu, angina czy choroba oka. Innym zagadnieniem do dyskusji może być pytanie: co zrobić, gdy użytkownik w wypadku straci rękę? Można odpowiedzieć: wykorzystajmy drugi kciuk. Rodzi się kolejne pytanie: jak często można w ten sposób „odnawiać" sekret, którym w tym wypadku jest cecha biometryczna?

System logowania oparty o certyfikaty klucza publicznego łączy w sobie zalety obu wspomnianych wcześniej typów uwierzytelniania. Uniemożliwia nawet kradzież baz z hasłami użytkowników, bo... one nie istnieją (klucz prywatny jest w posiadaniu użytkownika) tak więc nawet administrator systemu nie ma do nich dostępu. Ten typ uwierzytelniania niesie ze sobą szereg dodatkowych funkcjonalności:

1. Weryfikacja tożsamości osoby nie musi być przeprowadzana w placówce przedsiębiorstwa – jest realizowana przez zaufaną trzecią stronę,
2. Można przeprowadzić weryfikację prawdziwości wniosku elektronicznego o przydzielenie uprawnień do systemu bez stawiennictwa wnioskodawcy w placówce firmy,
3. Składanie podpisów elektroniczny weryfikowanych kwalifikowanym certyfikatem – pozwala to na szybkie zawieranie umów bez potrzeby spotykania się z kontrahentem. W oparciu o tę cechę oraz system pocztowy można oprzeć obieg dokumentów elektronicznych.
4. Autoryzację do zasobów informatycznych z wykorzystaniem posiadanego certyfikatu klucza publicznego, który jest niepowtarzalnym potwierdzeniem jego tożsamości.
5. Wykorzystanie nośnika pary kluczy z certyfikatem – karty kryptograficznej jako:
    a. Identyfikatora pracownika,
    b. tokena, w kontroli dostępu do pomieszczeń,
    c. tokena, w rejestracji czasu pracy,
    d. tokena, umożliwiającego sprzętową ochronę stacji roboczych
6. Wykorzystanie kryptografii (m.in. AES 256) do zapewnienia poufności plikom, przesyłkom poczty elektronicznej itp.

Porównanie cech poszczególnych typów uwierzytelniania przedstawia Tablica 1.

Tablica 1 Porównanie cech systemów uwierzytelniania

Sposób uwierzytelniania → Funkcjonalność ↓	Login/hasło	Biometryczna	Certyfikat
Skuteczność ochrony sekretu przed skopiowaniem	–	–/+	+
Bezpieczne przeprowadzenie procesu nadania uprawnień tylko on-line	–	–	+
Fizyczne powiązanie sekretu z użytkownikiem, tak aby wykluczyć zapomnienie przez użytkownika	–	+	–
Przeprowadzenie weryfikacji tożsamości przyszłego użytkownika bez jego stawiennictwa w firmie	–	–	+
Unieważnianie sekretu w wypadku jego kompromitacji	+	–	+
Wydanie nowego sekretu przy domniemaniu kompromitacji	+	–	+

Źródło: opracowanie własne

Legenda:

– oznacza brak wymienionej cechy
+ oznacza posiadanie wymienionej cechy

4.2. Proces nadawania uprawnień w oparciu o certyfikat użytkownika

W celu otrzymania uprawnień do zasobów informatycznych użytkownik powinien legitymować się swoim wirtualnym dowodem osobistym – certyfikatem klucza publicznego. Wówczas nie ma potrzeby rozbudowywania w przedsiębiorstwie komórki weryfikującej tożsamość przyszłych użytkowników.
Aby wejść w posiadanie własnego certyfikatu należy:

1. Udać się do najbliższego punktu rejestracji Zaufanej Trzeciej Strony – Centrum Certyfikacji, np. Centrum Certyfikacji Powszechne Unizeto CERTUM, w celu przedłożenia wniosku o certyfikat klucza publicznego (lub przesłać listem poleconym niezbędne dokumenty, których autentyczność została potwierdzona przez notariusza),
2. Przedłożyć dokumenty, dzięki którym operator centrum certyfikacji zweryfikuje tożsamość,
3. Centrum certyfikacji przekazuje użytkownikowi kartę kryptograficzną zawierającą cechy wymienione przez wnioskodawcę we wniosku.

Posiadając certyfikat kwalifikowany przyszły użytkownik systemu:

1. Wypełnia elektroniczny wniosek o nadanie uprawnień podpisując go elektronicznie.
2. Podpisany i zaszyfrowany (w celu zachowania poufności) wniosek jest wysyłany Kurierem Elektronicznym do administratora zasobów informatycznych,
3. Administrator deszyfruje wniosek oraz weryfikuje jego integralność,
4. Na podstawie informacji zawartych we wniosku administrator systemu dopisuje wnioskodawcę do bazy użytkowników.
5. Wnioskodawca staje się użytkownikiem systemu.

Wnioskodawca, teraz już użytkownik może korzystać z zasobów informatycznych uwierzytelniając się z wykorzystaniem posiadanej karty kryptograficznej, a ściślej: swojego wirtualnego dowodu osobistego - certyfikatu klucza publicznego.

Szeroka funkcjonalność posiadanej przez użytkownika karty kryptograficzne powoduje, iż staje się ona dla niego niezbędna aby mógł:

- pracować przy komputerze,
- poruszać się po firmie,
- uwierzytelniać się do zasobów informatycznych.

Dzięki temu sekret w postaci certyfikatu nabiera cech sekretu biometrycznego: niemal fizycznie związanego z użytkownikiem.

5. Zintegrowany system uwierzytelniania i podpisu elektronicznego na przykładzie proCertum ADS

Przykładem praktycznego zastosowania rozwiązań opisanych w poprzednim rozdziale jest proCertum ADS (ang. Authorization and Digital signature System). proCertum ADS jest rozwiązaniem, które pozwala na wdrożenie bezpiecznego systemu pojedynczego uwierzytelniania oraz funkcjonalności podpisu elektronicznego w istniejącym już oprogramowaniu i systemach informatycznych klienta.

System pozwala na spójne zarządzanie uprawnieniami dostępu do zasobów informatycznych w modelu pojedynczego logowania (ang. Single Sign On) w oparciu o technologię PKI. Dodatkowo proCertum ADS pozwala na wykorzystanie potencjału podpisu elektronicznego i osiągnięcie korzyści wynikających z eliminacji bądź znacznego ograniczenia liczby dokumentów papierowych.

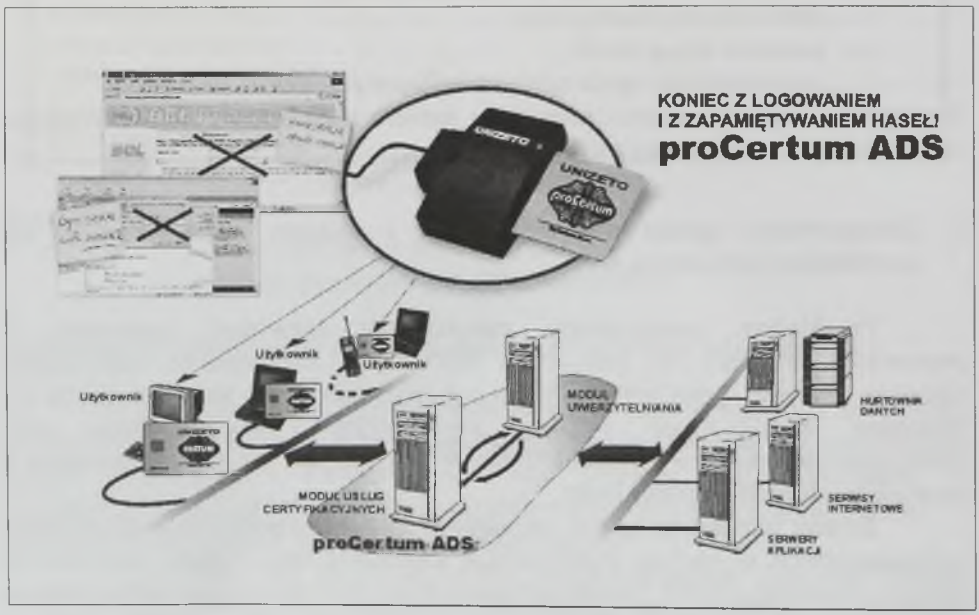
Zintegrowany System Uwierzytelniania i Podpisu Elektronicznego proCertum ADS zapewnia:

1. silne uwierzytelnienie użytkownika na najwyższym poziomie za pomocą certyfikatów klucza publicznego (X.509),
2. brak potrzeby pamiętania przez użytkownika wielu nazw użytkowników i wielu haseł do różnych systemów, aplikacji – wykorzystanie Single Sign On.
3. zaawansowany system niezaprzeczalności umożliwiający rozstrzyganie sporów – niezaprzeczalność czasu utworzenia archiwów

dokumentów elektronicznych (w tym: danych skompresowanych, oprogramowania), niezaprzeczalność czasu logowania (ochrona przed modyfikacjami logów systemowych przez osoby nieuprawnione),

4. uwierzytelnienie obiegu informacji oraz dokumentów elektronicznych,
5. sprzętową ochronę stacji roboczej użytkownika przed nieautoryzowanym dostępem osób trzecich,
6. korzystanie z podpisu elektronicznego w celu zapewnienia integralności stworzonych dokumentów, wysłanej poczty elektronicznej
7. szyfrowanie danych w celu zapewnienia poufności przesyłanych informacji, plików, wiadomości e-mail.

Zintegrowany System Uwierzytelniania i Podpisu Elektronicznego proCertum ADS, którego funkcjonalność przestawia rysunek 4, jest modułowym systemem zbudowanym z wykorzystaniem technologii PKI.



Rys. 4 Zintegrowany system podpisu elektronicznego i uwierzytelniania proCertum ADS
Źródło: opracowanie własne Unizeto Sp. z o.o.

W Systemie wykorzystywane są między innymi:

1. moduł usług certyfikacyjnych – proCertum PKI, modułowy system Infrastruktury Klucza Publicznego, realizujący usługi zarządzania certyfikatami klucza publicznego, znacznika czasu, notariatu elektronicznego, skarbca elektronicznego, kuriera elektronicznego oraz OCSP (weryfikacji statusu certyfikatu on-line).
2. moduł uwierzytelniania – produkty uznanych producentów umożliwiające bezpieczne wdrożenie systemu pojedynczego logowania, w tym między

innymi: Novell Secure Login ®, Novell Modular Authentication System, RSA Secure ID ®, Microsoft Smart Card Logon.

3. proCertum API – zestaw bibliotek, które pozwalają na implementację funkcjonalności infrastruktury klucza publicznego w oprogramowaniu i systemach informatycznych klienta (podpis elektroniczny, szyfrowanie danych, weryfikacja statusu certyfikatu, znakowanie wiarygodnym czasem, kurier elektroniczny).
4. proCertum Plug'in – implementacja funkcjonalności podpisu elektronicznego w standardowych aplikacjach biurowych klienta (m.in. Adobe Acrobat – od wer. 5.05, Microsoft Word – od wer. 2000, Microsoft Excel – od wer. 2000, Microsoft Project – od wer. 2000) umożliwiająca szerokie zastosowanie podpisu elektronicznego w firmie oraz stworzenie obiegu dokumentów elektronicznych w oparciu o pocztę elektroniczną Klienta,
5. UniLock – wykorzystanie karty kryptograficznej użytkownika do sprzętowej ochrony dostępu do stacji roboczej w systemach Windows NT 4.0, Windows 2000, Windows XP.

6. Literatura

1. Ustawa o podpisie elektronicznym z dnia 18 września 2001r. (Dz.U. Nr 130, poz. 1450, z dnia 15.11.2001r.)
2. C. Adams, St. Lloyd, *Podpis elektroniczny klucz publiczny,* Wydawnictwo Robomatic, Wrocław 2002
3. A. Kruk, P. Matusiewicz, J. Pejaś, A. Ruciński, W. Ślusarczyk, *Podpis elektroniczny – sposób działania, zastosowanie i korzyści,* Ministerswo Gospodarki, Pracy i Polityki Społecznej, Warszawa 2003

mgr Wojciech Wileński
Unizeto Sp. z o.o.
ul. Królowej Korony Polskiej 21
70-486 Szczecin, tel. (091) 4801 327
e-mail: wwilenski@unizeto.pl

mgr inż. Leszek Zdawski
Unizeto Sp. z o.o.
ul. Królowej Korony Polskiej 21
70-486 Szczecin, tel. (091) 4801 207
e-mail: lzdawski@unizeto.pl

ROZDZIAŁ 7

VARIA

GENETIC CLUSTERING AS AN EFFECTIVE METHOD OF APPROXIMATING ATTRACTORS OD A MULTIMODAL FUNCTION

Katarzyna ADAMSKA

Abstract. This work describes the strategy of genetic clustering which is a combination of genetic optimization and clustering (here using finite mixture models). It is an efficient tool for approximating central parts of attractors of a multimodal objective. The strategy can be desribed in a concise mathematical way, which enables precise estimation of its error based on several approximation theorems and the theory of SGA convergence. Presented computational tests confirm the efficiency of the approach.

Introduction

This paper concerns an approximating of basins of attraction of a multimodal function while optimizing the function with a genetic algorithm. That problem is interesting at least because of two reasons. Firstly, since a genetic algorithm rarely leads to an exact solutions, the approximation of a basin of attraction determines an area where a local optimization method can find easily an extremal point. As the second, describing a landscape explored by a genetic algorithm can provide curious information about the capabilities and limitations of the algorithm. It can be also interesting for many practical applications; for example it can be useful in chemical computation concurring chemical reactivity.

The approximation of basins of attraction will be performed with genetic clustering strategy, which means applicating some clustering technique to data obtained in genetic optimization. In this case clustering is done with the use of finite mixture model [2] and the parameters of the model are calculated with EM algorithm [1]. The approach gives an opportunity of formal analysis that enables estimating the error of the method, that is introduced in section 3. Tests presented in section 4 show that the strategy proves correct in practice.

1. The idea of genetic clustering

1.1. Finite mixture model and EM algorithm

As mentioned, *genetic clustering* means a technique of approximating central parts of attractors of an objective consisting in clustering a genetic population,

obtained as a result of some genetic optimization algorithm. In this work a finite mixture model is used for clustering. The main assumption of the model is that distribution of clustered data can be described with some density function ρ. It is also assumed that the density function ρ may be represented by the convex combination of some components

$$\rho(x) = \sum_{k=1}^{s} \gamma_k g_k(x; q_k) \tag{1}$$

where $g_k(x; q_k)$ stands for a component function, depending on input data x and set of specific parameters q_k. Each component should describe one cluster, which is indexed
with k. Coefficients γ_k in (1.1) are called *mixing proportions* and the function ρ is named a *mixture*. The number of clusters s must be predicted. The functional form of components $g_k(x; q_k)$ is assumed, but their specific parameters remain unknown, so the mixture depends on the parameter vector $q = (q_1, \ldots, q_k)$. In most practical works
components have a form of Gauss distribution, so a covariance matrix C_k and a mean vector m_k are characteristic parameters of a kth component, $q_k = (C_k, m_k)$.

The mixing proportions γ_k must be also evaluated. To perform that a *probability matrix* $\Gamma = [\gamma_{i,j}]_{i=1,\ldots,s}^{j=1,\ldots,n}$ is introduced. An element γ_{ij} of the matrix stands for the probability that j-th data element belongs to i-th cluster. A mixing proportion γ_k is computed from Γ as a normalized sum over the k-th row;

$$\gamma_k = \frac{\sum_{i=1}^{n} \gamma_{ki}}{n} \text{ for all } k=1,\ldots,s.$$

An elementary calculation based on the Bayes rule shows that sets of mixing proportions and component parameters are related with each other. Because of that fact, given one of these values sets, the second can be calculated. It constitutes the basis of the Expectation-Maximization Algorithm (EM) [1], which is the effective method of computing the mixture density function. One step of the EM algorithm consists of two substeps. The E-step is for calculating expected values of Γ entries. In the M-step component parameter vector q is worked out in such a way, that some auxiliary *likelihood function* [2, Sect. 3.2] is maximized. Starting the EM algorithm requires providing some initial evaluation of q or Γ. Then the iteration of EM steps begins. Assuming that initialization concerns Γ matrix the scheme of EM algorithm can be written as follows:

```
InitializeProbabilityMatrix;
Ltminus1 = ∞; t = 0; stop = 0;
do{
```

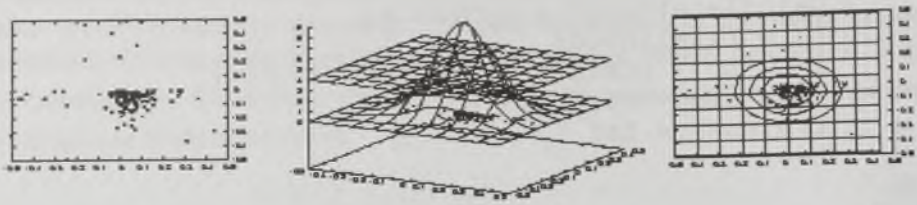
```
        CalculateGaussianParameters;
        CalculateProbabilityMatrix;
        Lt = Likelihood;
        if (|Lt - Ltminus1|< const ) stop++;
        else stop = 0;
        Ltminus1 = Lt;
        t++;
}
while (t < t_max) and (stop < 5)
```

The time and memory complexity of EM algorithm has been thoroughly analyzed in [3]. Denoting number of clusters with s, number of clustered data points with n and data dimension with p one can write the total complexity of one EM step as $8snp^3 + \frac{9}{2}snp^2 + \frac{13}{2}snp + \Theta(np)$. One should mention that among all parameters affecting the complexity n takes always the biggest value in practical tasks. Complexity depends linearly on n, which means that the method is quite effective.

The amount of memory required for basic data representation in EM algorithm is of order of $O\left(np + c\left(n + p^2\right)\right)$, where a constant coefficient depends on implementation.

1.2. Application of clustering to genetic optimization results

Now we are ready to refer the finite mixture model to genetic optimization and the problem of approximating basins of attraction. In that case, an input data set, which the model is constructed for, is constituted by a genetic sample. As a genetic sample we understand a multiset of phenotypes included in an admissible set, unambiguously corresponding to a population from a *genetic space* (for a comprehensive explanation see [4]). If the genetic algorithm is able to find areas close to extremal points of an objective function (i.e. an algorithm is *well-tuned*, see [4,5] for a formal definition), a genetic sample will be concentrated in these areas. Thus a mixture function ρ constructed for a genetic sample will have greater values in these parts of admissible domain which are close to extrema. Now, by cutting a mixture function for some threshold, we get level sets of the mixture. Those level sets will be called cluster extensions and in the next section there will be presented an argumentation why they can be treated as approximations of basins of attraction. The described process is illustrated on fig. .1.



(a) genetic sample (b) mixture density function (dotted plane indicates a cutting threshold) (c) cluster extension

Fig .1. The idea of genetic clustering with finite mixture model

2. Theoretical properties of genetic clustering

2.1 Measure density for SGA

Assume that a genetic population is a n-elemented multiset of elements from some r-sized genetic universe. A population can be represented by a frequency vector which i-th coordinate is a normalized contribution of an i-th genotype to a population. According to the theory of SGA introduced by Vose [6], evolution of a population in SGA can be described with a genetic operator (named also "*heuristics*") $G:\Lambda^{r-1} \to \Lambda^{r-1}$. $\Lambda^{r-1} \subset R^r$ denotes r-1-dimensional unit simplex, which contains frequency vectors for all populations. Each frequency vector constitutes a discrete measure in a genotype space [6, sect.3]. Using the fact that it is one-to-one coding between genotypes and phenotypes, the measure can be transported from a genotype space into a phenotype space D_r (refer to [7, 402] for details). It has been shown in [7] that under some mild geometrical assumptions concerning an admissible set D and D_r, a discrete measure μ on a phenotype space D_r can be extended to the particular measure on the whole D which has a density $\rho_\mu \in L^p(D)$. To distinguish between the measures let us take the following notation: if x stands for a population, let $\Theta(x)$ mean a corresponding discrete measure in a phenotype space, and $\rho_{\Theta(x)}$ a corresponding density on D. Note, that D_r is a discrete set included in D. The idea of constructing $\rho_{\Theta(x)}$ consists in dividing D onto hypercubes so that each point y from D_r is placed in a separate hypercube of the volume v_y. Size of a hypercube is dictated by the accuracy of coding. There is set $\rho_{\Theta(x)}|_{v_y} = \dfrac{\Theta(x)(y)}{meas(v_y \cap D)}$. As a result one gets a step-wise

measure density defined on a set of adjacent hypercubes intersected with D (refer for details to [7]).

Another theoretically important fact concerns the genetic operator G. Assume that G meets the following conditions:

- G has a finite set of fixed points (let us denote this set with K)
- $\forall x \in \Lambda^{r-1} \ \lim_{t\to\infty} G^t(x) \subset K$

fixed points of G correspond to populations that represent the best information about basins of attraction of an objective available for a given genetic algorithm (see [5], chapter 4)

Let z denote a fixed point of G and K a set of all fixed points of G. Then the following theorem about the convergence of measure densities can be written after [7]:

Theorem 2.1 $\forall \varepsilon > 0 \forall \eta > 0, \exists \widetilde{N} \in N, \exists W(\widetilde{N}) \in N, \exists z \in K$ such that $\forall n > \widetilde{N} \forall k > W(\widetilde{N})$

$$P\left\{\left\|\rho_{\Theta(x_n^k)} - \rho_{\Theta(z)}\right\|_{L^p} < \frac{\varepsilon}{c}\right\} > 1 - \eta$$

where x_n^k stands for a n-sized population obtained after k epochs, $p \in [1,\infty]$ and c is a coding-dependent constant.

2.2 Error of genetic clustering

Now let us return to the finite mixture model and the EM algorithm. As already mentioned in section 2 , each iteration of the EM algorithm results in a set of parameters unambiguously determining a mixture density function. Assume that for a given input data the EM algorithm is convergent. Thus one can construct a contractive function H leading from a space of mixture parameters onto itself, which describes a single step of the algorithm. According to known from the literature ([8]) theorem about an iterative function, the function H has a stationary point. A detailed description of construction of H and adaptation of the theorem for the EM algorithm is presented in [9]. Denote with ρ_i a mixture density which can be obtained in i-th step of the algorithm, and with ρ_A a density corresponding to a stationary point of the algorithm. As a conclusion from the mentioned theorem one gets the following estimation ([4]):

$$\|\rho_i - \rho_A\|_{L^p} \le \frac{b^i}{1-b}\chi \tag{2}$$

where $b \in (0,1)$ and χ is a norm-dependent constant. Therefore by putting together the theorem 2.1 and (2) one gets the following:

Proposition 2.2 $\forall \varepsilon > 0 \forall \eta > 0, \exists \widetilde{N} \in N, \exists W(\widetilde{N}) \in N, \exists z \in K$ such that $\forall n > \widetilde{N} \forall k > W(\widetilde{N})$

$$P\left\{\left\|\rho_i - \rho_{\Theta(z)}\right\|_{L^p} < \frac{\varepsilon}{c} + \frac{b^i}{1-b}\chi + \sigma_{GC}\right\} > 1 - \eta$$

where i means a number of EM iterations and $\sigma_{GC} = \left\|\rho_{\Theta\left(x_n^k\right)} - \rho_A\right\|_{L^p}$ (all the remaining symbols explained in the above text).

The estimated difference $\left\|\rho_i - \rho_{\Theta(z)}\right\|_{L^p}$ is important for the theory of genetic clustering, since it expresses the error of clustering. It is a difference between the best density available for genetic optimization and a mixture density calculated after i iterations of the EM algorithm. The first two quantities estimating the difference can be arbitrarily small; in the first component ε can be any positive number and the second
decreases with the increase of value i. Therefore the error of clustering depends on the quantity σ_{GC}, which is the difference between density induced by a a n-sized genetic population obtained after k epochs and some ideal measure density corresponding to stationary point of EM algorithm. Assume that fitting for $\rho_{\Theta\left(x_n^k\right)}$ and ρ_A is optimal; that means $\rho_{\Theta\left(x_n^k\right)}$ can be treated as an approximation spline function for ρ_A, consisting of zero degree polynomials. Then an approximation theorem from [10] can be applied to estimate σ_{GC}. It implies the following evaluation (see [11] for detailed considerations):

$$\left\|\rho_A - \rho_{\left(x_n^k\right)}\right\|_{L^p(D)} \le C(measD)^{\frac{1}{p}} d_V^{\max} \left|\rho_A\right|_{\infty,1,V}^{\max} \tag{3}$$

where $d_V^{\max} = \max_{V \in D} d$ is a maximal diameter of a ball included in V (hypercube of a domain D) and $\left|\rho_A\right|_{\infty,1,V}^{\max} = \max_{V \in D}\left|\rho_A\right|_{\infty,1,V}$ maximal density in the sense of seminorm in a Sobolev space. The constant C depends on the norm and geometrical properties of a partition of a domain.

Concluding, the approximation error depends on the measure of an admissible domain as well as on partition of a domain which constitutes a base for constructing genetic measure density. The smaller measures of subsets V are, the lower value of the quantity $d_V^{\max}$ is and thus, the approximation error is less, too. In particular, if the binary coding is more accurate (coordinates are coded with longer strings), the error will decrease.

3. Computational tests

This section presents some examples of practical performance of genetic clustering. The optimization stage has been accomplished with the use of the Hierarchical Genetic Strategy (HGS, see [12,13]). HGS is a very effective multi-populational parallel algorithm. It results in several populations, which occupy neighborhoods of extrema in an admissible domain.

3.1. Two-dimensional Rastrigin function

In the first example two-dimensional Rastrigin function was used, which was defined on the range [-0.5,0.5]x[-0.5,0.5] by the formula :

$$f_1(x,y) = 2 + x^2 + y^2 - \cos(40x) - \cos(40y)$$

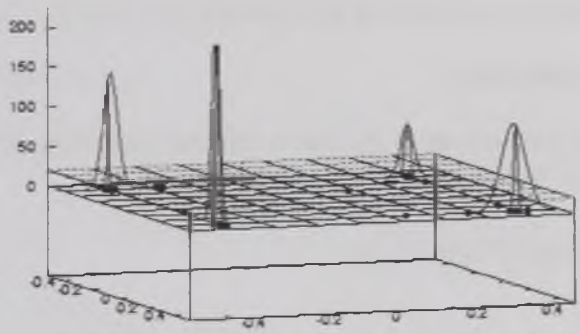


Fig. 2. Mixture density model for Rastrigin function

Figure 2 shows four populations computed with HGS for the Rastrigin function, which correspond to four global maxima in the considered domain, as well as four Gaussian peaks, obtained as a result of clustering with the use of mixture resolving approach. The surface drawn with dots marks the threshold for cutting density function to form the approximations of central parts of attractors. The approximations are shown on figure 3 as the projection of density mixture on the surface.

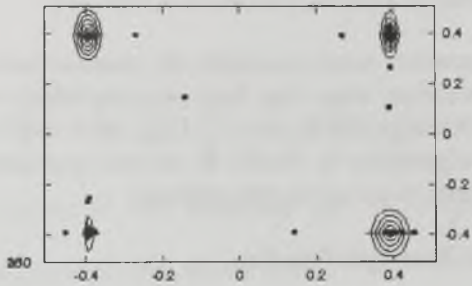


Fig .3 Approximations of attractors for Rastrigin function

This example is rather easy for clustering since populations are well separated and concentrated on relatively small areas. It shows that mixture resolving method works, enables approximating central parts of attractors (fig. 4.) and is not sensitive for some outside laying points.

3.2 Two-dimensional sinus

The next instance is a bit more difficult clustering task. The fitness function was two-dimensional sinus

$$f_2(x,y) = \sin(xy) + 1$$

optimized in a range [-3,3]x[-3,3].

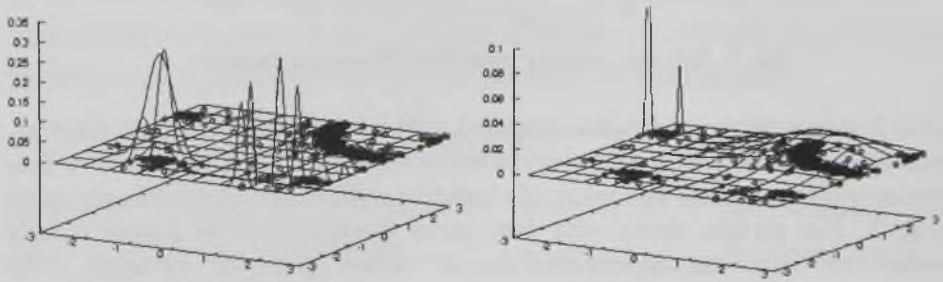


Fig. 4. Mixture density model for sinus function

In this case populations are not so well concentrated and one of them has a non-spherical shape. Since the Gaussian peaks determined by clustering have different height, results are presented on two separate figures for a reader's convenience. Clustering process has finished successfully. Gaussian peaks are located in proper parts of the domain. There is a quite big fraction of points that lie far from regions of high concentration of individuals, but they have not a significant influence on the

results. The non-spherical population has been described with a wide peak (fig. 5). The approximations of central parts of attractors are determined as a projection of calculated density on the surface drawn with a dotted line on figure 4.4. In spite of complicated shape of clusters the approximation of central parts of attractors seems to be quite satisfactory (fig. 5).

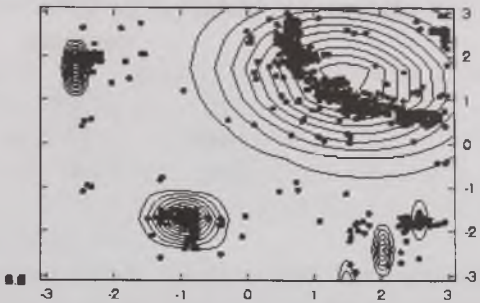


Fig. 5 Approximation of attractors for sinus function

3.3 Multidimensional data

There were performed calculations for ten-dimensional modified Rastrigin function

$$f_3(x) = 20 + 4x_1^2 + \cos(5\pi x_1) + \sum_{i=2}^{10}\left(x_i^2 + \cos(5\pi x_i)\right)$$

in a hypercube $[-0{,}5;0{,}5]^{10}$.

10-dimensional Rastrigin function has four global extrema in $[-0{,}5;0{,}5]^{10}$. HGS produces four populations on the highest level, which correspond to neighborhoods of the extrema. EM algorithm deals with the genetic sample in a few iterations (see table 1). Although genetic data seems to be a bit chaotic, four Gaussian peaks have been evaluated quickly and all they got the same mixing proportion 0.25. It means that each Gaussian corresponds to one population, thus as a consequence, to a neighborhood of one extremum.

3.4. Velocity of calculation versus size of data

This section includes time statistics for the examples presented before. Unfortunately, parameters, which have an influence on computational complexity, are not consistent within the examples and it is the result of the peculiarity of genetic calculation. Therefore data introduced here is just to show that EM algorithm is quite

fast. Computation was done with a processor Intel Celeron 633MHz, the last column of table contains processor time usage.

Table 1. Time statistics for computing mixture density function.

fitness function	dimension	number of clusters	number of points	number of EM iterations	number of initializa-tion iterations	time [msec]
Rastrigin	2	4	350	7	6	114
sinus	2	4	1100	22	28	1176
Rastrigin	10	4	200	8	9	161

The first conclusion from data in table1 is that number of EM iterations needed to evaluate mixture density function can be very small. It does not seem to depend on data dimension or number of clusters. Amount of EM iterations as well as number of initialization steps is remarkably higher for sinus function. The reason is non-Gaussian shape of clusters in that case.

General conclusion is, that EM algorithm is very effective computational tool, especially when shape of clusters corresponds to features of mixture components. The effectiveness for non-Gaussian-shaped clusters can be probably improved by implementation of other types of components (for instance Lorentzian).

4. Conclusions

- Genetic clustering performed with the use of finite mixture model and the EM algorithm in application to the problem of finding basins of attraction in genetic optimization enables wide opportunities of theoretical analysis. It also proves correct in practical tests. Moreover, it is practically very efficient because of its low memory and time complexity.
- If binary-coded genetic algorithm based on SGA is utilized as the optimization tool, one can estimate the error of the discussed method. The error in respect to the accuracy of genetic optimization has upper limit, which depends on the measure of an admissible set and the precision of coding.
- Computational tests performed for data optimized with HGS confirm correctness and high efficiency of the strategy.

References

1. Dempster A.P., Laird N.M., Rubin D.B.; Maximum likelihood from incomplete data via EM algorithm, J. Royal Statis. Soc., Ser. B, vol. 39, pp. 1-38, 1977
2. Duda R.O., Hart P.E. Pattern Classification and Scene Analysis, Wiley, New York 1973
3. Adamska-Piskorz K., *Clustering in genetic algorithms*, Jagiellonian University 2002 (Msc thesis)
4. Schaefer R., Adamska K., Jab³oński Z.J., *Clustering driven by the genetic sampling measure, Methods of Artificial Intelligence*, Proceedings of the Symposium on Methods of Artificial Intelligence AI-METH 2002, pp. 361-366, Gliwice, Poland, 2002
5. Schaefer R., *Introduction to the genetic global optimization* (Podstawy genetycznej optymalizacji globalnej), Jagiellonian University Press, Kraków 2002 (in Polish)
6. Vose M. D.: *The Simple Genetic Algorithm,* MIT Press, 1999
7. Schaefer R., Jab³oński Z.J.: O*n the convergence of sampling measures in the global genetic search*, Lecture Notes in Comp. Science, vol. 2328, pp. 593-600, Springer 2002
8. Stoer J.; *Introduction to numerical methods,* Sect. 5.2, PWN 1979 (in Polish)
9. Schaefer R. Adamska-Piskorz K., *Approximation of basins of attraction with mixture resolving method*, Proc. of Conf. on Evolutionary Algorithms and Global Optimization, 23-25 Sept. 2002, Kraków
10. Ciarlet P., *Finite Element Method for Elliptic Problems,* North-Holland Publishing Company, 1978
11. Schaefer R., Adamska K. On genetic clustering using finite mixture models - error estimation and practical tests, *Proc. of KAEiOG 2003*,
12. Ko³odziej J., Gwizda³a R., Wojtusiak J.; Hierarchical Genetic Strategy as a method of improving search efficiency, A*dvances in multi-agent systems*, pp.149-159,Jagiellonian University Press, 2001
13. Schaefer R., Ko³odziej J., Gwizda³a R., Wojtusiak J.; How simpletons can increase the community development - an attempt to hierarchcal genetic computation, *Proc. of 4th KAEGiOG*, pp.187-197, L¹dek Zdrój, 2000

mgr Katarzyna Adamska
Instytytut Informatyki Uniwersytetu Jagiellońskiego,
ul. Nawojki 11, 30072Kraków
adamska@ii.uj.edu.pl

SYMULACYJNE BADANIA PROTOKOŁÓW TRANSPORTOWYCH DLA BEZPRZEWODOWYCH SIECI LOKALNYCH WLAN

Andrzej GŁOWACZ, Michał MAJ

Streszczenie: Implementacja protokołów transportowych dla współczesnych sieci telekomunikacyjnych wymaga wcześniejszego przeprowadzenia szczegółowych badań symulacyjnych. Badania takie często pozwalają na wskazanie problemów w projektowanych mechanizmach protokołów. Obecne sieci heterogeniczne cechują się dużą liczbą charakteryzujących je parametrów, które zmieniają się bardzo dynamicznie i mają znaczny wpływ na wydajność. Uzasadnia to konieczność rozwoju narzędzi informatycznych pomocnych zarówno przy projektowaniu samych protokołów jak też pozwalających na testowanie rozwiązań sieci.

Artykuł przedstawia wyniki badań symulacyjnych nad wydajnością protokołów transportowych TCP i UDP. Dzięki użyciu złożonego modelu symulacyjnego stworzonego w ramach pracy wykryto i w pełni zidentyfikowano błędy w algorytmie szacowania pasma jednej z najnowszych wersji protokołu TCP Westwood.

Wprowadzenie

Sieci heterogeniczne łączące różne technologie dostępu do medium transmisyjnego charakteryzują się różnymi parametrami pozwalającymi opisać stan sieci. Parametry te zależą od konkretnego przypadku sieci i stąd wynika podstawowa trudność przy projektowaniu uniwersalnego protokołu transportowego, który mógłby działać w całej sieci. Dla przyszłych sieci takim rozwiązaniem może stać się protokół TCP (ang. *Transmission Control Protocol*), który obecnie jest podstawowym protokołem Internetu.

W pracy rozpatrujemy bezprzewodową sieć lokalną WLAN (ang. *Wireless Local Area Network*) używającą standardu IEEE 802.11 [9] jako przykład sieci heterogenicznej składającej się z dwóch podsieci: infrastruktury przewodowej oraz węzłów mobilnych stanowiących bezprzewodową część mobilną sieci. Główne wnioski z przeprowadzonych badań mogą być uogólnione dla dowolnej sieci heterogenicznej.

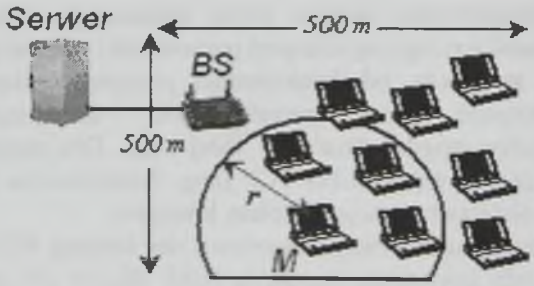
Technologie bezprzewodowe charakteryzują się stosunkowo wysoką stopą błędów transmisji, które mogą wynieść nawet do 10 %. Dzieje się tak dlatego, że kanał bezprzewodowy jest bardziej podatny na zakłócenia niż tradycyjny przewód, zwłaszcza w pewnych przedziałach pasma częstotliwości fal radiowych np. w paśmie 2,4 GHz. Stąd wynika trudność w rozróżnianiu błędów i strat pakietów wskutek przeciążeń sieci, zwłaszcza dla protokołu TCP, który został zaprojektowany dla sieci przewodowych. Osobnym zagadnieniem jest niska

żywotność baterii w urządzeniach bezprzewodowych. Ogranicza to wiele możliwości wprowadzania złożonych obliczeniowo metod w algorytmach sterowania przepływem. Ograniczone są także przepustowość sieci i efektywny zasięg transmisji. Pozostałymi problemami są uwarunkowania wynikające z topologii takie jak: brak jednoznacznej struktury sieci i mobilność użytkowników.

Spośród wielu metod poprawy wydajności protokołu [1] TCP na szczególną uwagę zasługują mechanizmy związane z inteligentnym ograniczaniem generowanego do sieci strumienia ruchu przez nadajnik. Nadajnik na podstawie informacji zwrotnej powinien być zdolny do właściwej oceny stanu sieci opisywanego przez dostępną przepustowość, opóźnienia transmisji, dostępność łącz i aktualną topologię. Ograniczenie prędkości transmisji pozwala na znacznie lepsze osiągi sieci niż w przypadku pasywnego doprowadzania do przeciążeń i spóźnionej reakcji w przypadku wcześniejszych wersji TCP [2]. Ciekawe rozwiązanie polegające na odfiltrowaniu składowej szybkozmiennej z ogólnej charakterystyki ruchu zostało zaprezentowane w wersji TCP Westwood [6].

1. Model symulacyjny

W ramach pracy stworzono złożony model symulacyjny pozwalający na skuteczne badanie wydajności protokołów transportowych. Model bazuje na symulatorze sieciowym *ns-2* [8] ogólnie dostępnym m. in. dla systemu Linux dzięki czemu uzyskano dobrą wydajność stworzonych narzędzi symulacyjnych.



Rys. 1. Ogólny schemat topologii badanej sieci WLAN

Na rysunku 1 przedstawiono ogólny schemat badanej sieci WLAN. Sieć taka może rozciągać się na znacznym obszarze, często nawet kilkuset metrów. Stąd należy przyjąć ograniczony zasięg transmisji węzłów mobilnych *M* wynoszący *r* [m] i uwzględnić wpływ doboru trasy na wydajność. W przypadku mniejszych sieci to ograniczenie oczywiście nie występuje i wszystkie stacje pozostają w zasięgu bezpośredniej transmisji. Sieć może tworzyć grupa samych węzłów mobilnych lub stacjonarnych wyposażonych w bezprzewodowe karty sieciowe obsługujące dany standard komunikacji. Taka sieć jest określana mianem *ad-hoc* i ma ograniczone możliwości. Pełny model stanowi połączenie sieci

bezprzewodowej z infrastrukturą przewodową. W takim przypadku jeden lub więcej węzłów pełni rolę stacji bazowej *BS* pośredniczącej w transmisji między obszarami. Możliwy jest wtedy m. in. pełny dostęp do Internetu dla użytkowników wyposażonych w laptopy np. poprzez serwer.

Na potrzeby badań stworzono trzy rodzaje skryptów: automatyzacji, symulacyjne i przetwarzania wyników [4]. Takie podejście jest optymalne uwzględniając własności symulatora *ns-2* i generowane przez niego pliki wynikowe zawierające opis zdarzeń w symulowanej sieci. Skrypty automatyzacji służą do zebrania danych wejściowych i uruchomienia wszystkich niezbędnych modułów. Skrypty symulacyjne języka *tcl* (ang. *Tool Command Language*) są zasadniczą częścią stworzonych narzędzi. Zawierają one zadane modele symulacyjne i wszelkie dane dla symulatora *ns-2*. Skrypty przetwarzania wyników służą do wyliczenia parametrów wyjściowych symulacji oraz prezentacji ich dla użytkownika. Integracja tych trzech modułów na bazie systemu operacyjnego Linux umożliwia wydajne i efektywne badanie sieci bezprzewodowej, zwłaszcza dla dużej liczby scenariuszy symulacyjnych [5].

2. Protokół TCP Westwood

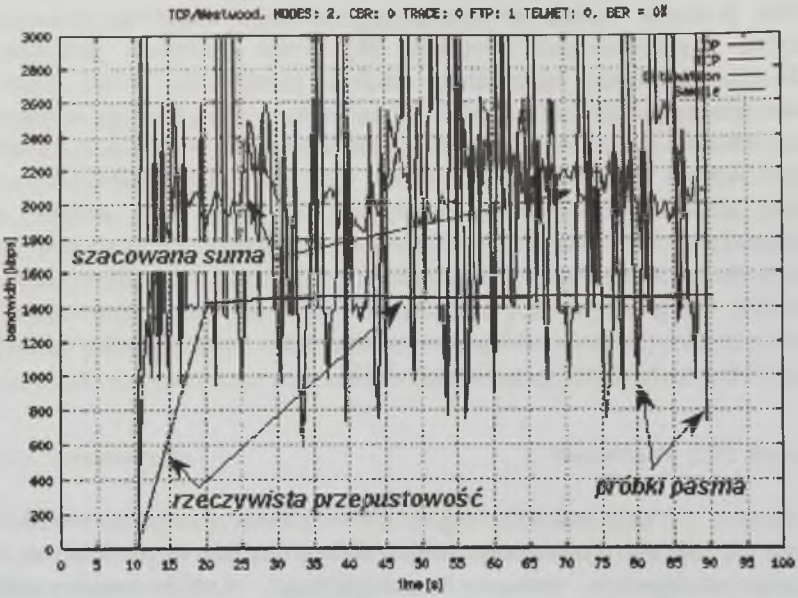
Jeden z najnowszych algorytmów eksperymentalnej wersji TCP Westwood [6] używa dyskretnego filtru dolnoprzepustowego Tustina w celu szacowania dostępnego pasma transmisyjnego. Odfiltrowanie składowej szybkozmiennej z ogólnej charakterystyki ruchu pozwala na bardzo dokładne szacowanie dla typowego łącza przewodowego.

Niestety, jak pokazują wyniki pracy [3, 4, 5], dla sieci bezprzewodowej pojawiają się niepożądane efekty przy obliczaniu szacowanej sumy przepustowości. Badania symulacyjne pokazują wyraźnie, że algorytm szacowania zastosowany w metodzie *Bandwidth Estimation* daje zbyt duże wartości wyjściowe brane następnie do sterowania mechanizmami zapobiegania przeciążeniom. W rezultacie wydajność protokołu jest znacznie niższa niż można by oczekiwać.

Powodem tego jest przyjęta zasada, że kolejne kroki adaptacyjne obliczane są w momencie nadejścia potwierdzeń protokołu. Dla badanej sieci bezprzewodowej potwierdzenia te mogą nadchodzić niesekwencyjnie i w małych odstępach czasu. Tak więc interwał czasowy przyjęty jako podstawa obliczania każdej próbki pasma może być mały rzędu milisekundy i w efekcie próbka obliczana jako stosunek potwierdzanej liczby danych do interwału staje się zbyt duża. Wliczenie dużej próbki do sumy z małą wagą nie powinno mieć dużego wpływu na szacowanie. Jak pokazują wyniki symulacji liczba błędnych próbek jest rzędu ponad 50 % i metoda nie może dać poprawnych rezultatów.

Pozostałymi odkrytymi problemami były niektóre kombinacje skumulowanych i opóźnionych potwierdzeń. Każda taka kombinacja musi zostać odpowiednio uwzględniona w algorytmie, tak aby np. nie wliczał on już uwzględnionych segmentów danych do badanej sumy. Symulacje ujawniły tę słabość metody, gdzie niektóre segmenty uwzględniane były z błędnym znakiem

i mnożnikiem, przez co następowały dodatkowe wahania estymacji i w znacznej liczbie badanych przypadków mogły doprowadzić do uniemożliwienia pracy agentowi nadawczemu TCP.



Rys. 2. Błędne szacowanie pasma TCP Westwood dla sieci bezprzewodowej

Na rysunku 2 przedstawiono przykład błędnego wyliczenia szacowanej przepustowości dla łącza bezprzewodowego pomiędzy dwoma węzłami mobilnymi. Rzeczywista przepustowość sieci jest znacznie niższa niż suma wyliczona przez algorytm Bandwidth Estimation protokołu TCP Westwood. Dzieje się tak dlatego, iż pobierane przez nadajnik próbki nie odpowiadają stanowi sieci. Interwały czasowe pomiędzy potwierdzeniami są zbyt małe w efekcie próbki pasma są zbyt duże.

Można się spodziewać, że ograniczenie się do interwałów czasowych pomiędzy potwierdzeniami daje niepełny obraz o stanie sieci i jest niewystarczające. Najnowsza propozycja *Adaptive Bandwidth Share Estimation* [10, 11] próbuje częściowo rozwiązać problem poprzez adaptację interwału, jednak dotąd nie przedstawiono działającego rozwiązania.

3. Wyniki symulacyjne

Przeprowadzone badania rozpoczęto od symulacji transmisji opartej o protokół UDP w sieci bezprzewodowej. Do symulacji jak źródło ruchu UDP wybrano transmisję wideokonferencyjną zgodną z H.263 generującą ruch o średniej przepustowości 65 kbps. Pierwsze symulacje ruchu skupiły się na badaniu wpływu wielkości segmentu UDP oraz wzrostu stopy błędów w kanale na

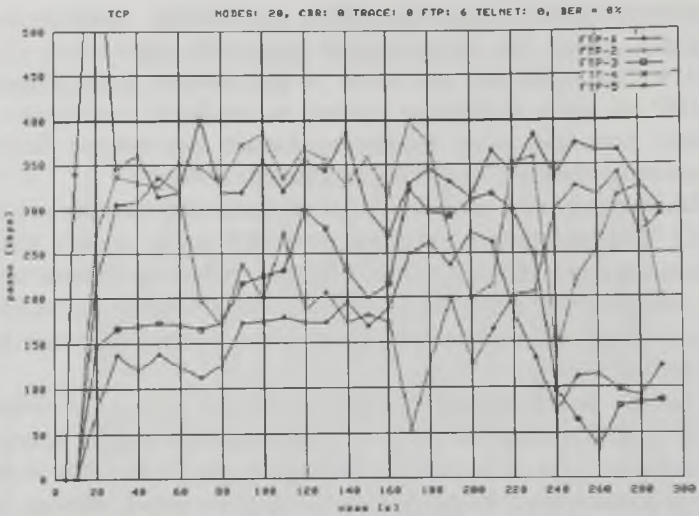
osiąganą przepustowość transmisji oraz na średnie opóźnienia. Zgodnie z oczekiwaniami, tylko dla najmniejszych badanych segmentów (256 bajtów) odnotowano wzrost opóźnień, natomiast w pozostałych przypadkach wielkość segmentu UDP nie miała wyraźnego wpływu na prędkość i opóźnienia transmisji. Podobnie przy wzroście stopy błędów w kanale zanotowano liniowy wzrost opóźnień transmisji i liniowy spadek łącznej przepustowości.

Badanie zachowania protokołu UDP w sieciach o odmiennych topologiach wykazało już tutaj podstawowe różnice pomiędzy małą, a dużą siecią *ad-hoc*: znacznie mniejsza przepustowość i dużo większe średnie opóźnienie transmisji dla dużej sieci, będące m.in. konsekwencją braku bezpośredniej łączności pomiędzy wszystkimi stacjami i konieczności przekazywania transmitowanych pakietów przez stacje pośredniczące.

Na koniec badań protokołu UDP wykonano symulacje równoczesnych połączeń TCP i UDP w sieci 802.11 w celu zaobserwowania jaki będzie wzajemny wpływ tych połączeń. Zgodnie z teorią CSMA/CA, przy braku strat w kanale, sesje TCP i UDP równomiernie obciążyły łącze bezprzewodowe, dzieląc się dostępną przepustowością w przybliżeniu po połowie.

Na podstawie wyników powyższych symulacji ustalono, iż sesje UDP dobrze nadają się do symulowania dodatkowego obciążenia w kanale transmisyjnym dla badań protokołu TCP. Dalsze symulacje skupiały się na obserwowaniu właściwości protokołu TCP, przy czym najczęściej jako aplikację generującą ruch TCP stosowany był transfer FTP. Badania rozpoczęto od określenia wpływu ruchu tła na transmisję TCP, porównując przypadki gdzie w kanale transmisyjnym jest: jedna sesja TCP; po jednej sesji TCP i UDP; dwie sesje TCP. Otrzymane wyniki wykazały, iż TCP ma tendencję to zabierania ponad połowy dostępnego pasma, gdy współdzieli kanał z połączeniem UDP. W przypadku dwóch jednoczesnych sesji TCP zaobserwowano znikomą różnicę pomiędzy średnimi wykorzystywanymi przepustowościami obydwu sesji TCP, co oznaczało równomierny podział dostępnej przepustowości. Obserwacje wpływu ruchu tła na opóźnienia transmisji TCP wskazały tutaj wyraźną tendencję wzrostu opóźnień przy zwiększeniu liczby jednoczesnych sesji TCP w kanale.

Badając ruch TCP dla różnej liczby jednoczesnych połączeń, otrzymano interesujący rezultat, wykazujący nieregularności sumarycznej przepustowości w czasie wszystkich jednoczesnych sesji TCP w kanale transmisyjnym. Nieregularności te występują tym częściej, im więcej jest jednoczesnych połączeń TCP, jednakże wraz ze wzrostem liczby tych połączeń, maleje również amplituda wahań sumarycznej przepustowości. Zjawisko to jest konsekwencją rywalizacji każdej z sesji TCP o zajęcie możliwie dużej przepustowości, a zarazem efektem interakcji mechanizmów zapobiegania przeciążeniom funkcjonujących w każdej sesji TCP niezależnie. Dobrą ilustracją tych obserwacji jest znajdujący się na rysunku 3. wykres zmian w czasie przepustowości zajmowanych przez poszczególne sesje FTP (oparte na TCP Reno – najszerzej spotykanej dziś implementacji TCP).



Rys. 3. Zmiany przepustowości w czasie dla poszczególnych sesji FTP w małej sieci *ad-hoc* z wykorzystaniem protokołu TCP Reno

Badania powyższe jako źródło ruchu TCP w sieci przyjmowały transmisję FTP. Aby się upewnić, iż rodzaj aplikacji generującej ruch w badanej sieci nie ma znacznego wpływu na osiągnięte wyniki, wykonano symulacje mające na celu zbadanie wpływu znacznej liczby sesji aplikacji Telnet na notowane wyniki. Jak wykazały symulacje, dodanie 20 jednoczesnych sesji Telnet nie powoduje wyraźnych zmian łącznej przepustowości osiąganej przez aplikacje TCP, ani nie zmienia także średnich opóźnień transmisji.

Kolejna seria przeprowadzonych symulacji miała na celu zbadanie wpływu liczby stacji oraz topologii sieci bezprzewodowej na efektywność transmisji TCP. Zwiększając liczbę stacji w małej sieci *ad-hoc* notowano niewielki, lecz zauważalny spadek łącznej przepustowości (cały czas nadawała taka sama liczba stacji, tylko ilość stacji w sieci się zwiększała). Co ciekawe, wraz ze wzrostem liczby stacji w sieci bezprzewodowej znacznie zwiększała się nieregularność sumarycznej przepustowości transmisji TCP. Tak samo wraz ze wzrostem liczby stacji rosła względna liczba pakietów, które transmitowane były poprzez stacje pośrednie (dla sieci z 64 stacjami było to ok. 10% pakietów), choć w małej sieci *ad-hoc* nie ma potrzeby przekazywania pakietów, gdyż wszystkie stacje pozostają we wzajemnym zasięgu. Wykonując analogiczne symulacje dla dużej sieci *ad-hoc* otrzymano całkiem odmienne wyniki: łączna przepustowość zmieniała się nieregularnie, bez związku z liczbą stacji w sieci. Analiza wyników wskazała na silną zależność łącznej przepustowości od liczby pakietów, które zostały przesłane z wykorzystaniem stacji pośrednich – im mniej pakietów było przekazywanych pośrednio, tym większa sumaryczna przepustowość. Zjawisko to wynika bezpośrednio z właściwości sieci bezprzewodowej. Natomiast obserwowana nieregularność liczby pakietów przekazanych z udziałem stacji pośrednich jest konsekwencją losowego dobierania podczas symulacji par stacji

nadawca - odbiorca. Jeśli losowo wybrane zostało więcej par znajdujących się we wzajemnym zasięgu, wówczas nie zachodziła konieczność przekazywania pakietów przez pośredników, co z kolei odciążało kanał radiowy umożliwiając transmisję kolejnych danych (średnia przepustowość liczona była jako stosunek przetransmitowanych danych do czasu symulacji). Podobnie jak w przypadku małej sieci, tak i tutaj dokonano także symulacji dla różnych rodzajów aplikacji generujących ruch, obserwując brak wpływu charakteru generowanego ruchu na osiągi całej sieci.

Interesująca okazała się tutaj symulacja przedstawionych powyżej zależności dla różnych implementacji protokołu TCP. Symulacjom poddano protokoły TCP Reno, New Reno, TCP SACK i TCP Vegas. O ile obserwowane średnie przepustowości dla tych protokołów okazały się bardzo zbliżone, to średnie opóźnienia transmisji znacznie się od siebie różniły. Najmniejsze opóźnienia okazały się mieć TCP Vegas i TCP New Reno, natomiast TCP SACK i TCP Reno odnotowały kilkukrotnie większe opóźnienia. Efekt ten spowodowany był liczbą segmentów odrzuconych na skutek przepełnienia kolejki stacji pośredniczących w transmisji.

Kolejne testy porównawcze odmiennych implementacji protokołów TCP wykonane zostały przy okazji badań wpływu stosowania opóźnionych potwierdzeń (*Delayed ACK*) na osiągi transmisyjne. Testy te przeprowadzono dla trzech topologii sieci. Zarówno w przypadku małej sieci *ad-hoc*, jak i sieci z infrastrukturą, stosowanie opóźnionych potwierdzeń zgodnie z założeniem troszkę podnosi przepustowość, lecz w przypadku dużej sieci *ad-hoc* pogarsza ono osiągi tej sieci (utrudnia ono wykrywanie strat pakietów i pogarsza efektywność retransmisji). Porównywane tutaj wersje TCP to Reno, New Reno i Vegas (SACK bazuje na opóźnionych potwierdzeniach – stąd nie wystąpił w tej serii badań) osiągnęły takie same rezultaty w zakresie średniej przepustowości. Analizując opóźnienia transmisji we wspomnianych warunkach zauważono, iż zarówno w małej jak i dużej sieci *ad-hoc* opóźnione ACK wprowadzają znaczny wzrost opóźnień transmisji, wyjątkiem jest tu sieć z infrastrukturą, gdzie opóźnienie jest mniejsze i w dodatku z opóźnionym ACK jeszcze nieznacznie spadło – efekt taki z pewnością powoduje Access Point, przez który przechodzi cały ruch w badanym przypadku. Analiza opóźnień transmisji dla wszystkich trzech badanych topologii ujawnia przewagę protokołu TCP Vegas, który za każdym razem osiągał najniższe średnie czasy opóźnień (kilkukrotnie mniejsze niż pozostałe wersje TCP). Jest to zapewne efektem zastosowania w tym protokole zaawansowanej techniki szacowania dostępnej przepustowości.

Kolejnym istotnym badaniem było określenie wpływu wielkości segmentu TCP na osiągi poszczególnych wersji protokołu TCP przy wzroście stopy błędów w kanale transmisyjnym. Testowano w tym celu cztery wersje TCP: Reno, New Reno, SACK i Vegas. Wszystkie te wersje wykazały się bardzo podobnymi osiągami: ze wzrostem segmentu notowany był wzrost przepustowości, a zwiększanie stopy błędów w kanale powodowało proporcjonalny spadek przepustowości. Opisany wzrost łącznej przepustowości wykorzystywanej przez połączenia wynika z faktu obniżenia narzutu wprowadzanego przez nagłówki

warstw niższych (IP i MAC) i strat czasu wynikających z konieczności znacznie częstszego przystępowania do rywalizacji o medium by przesłać tę samą ilość danych przy pomocy małych segmentów TCP. Analiza opóźnień transmisji w rozpatrywanych przypadkach ukazuje pewien wzrost opóźnień wraz ze wzrostem stopy błędów w kanale, lecz zjawisko to jest wyraźne tylko dla segmentów o małych rozmiarach. Obserwacja zbadanych tu opóźnień pod kątem porównania wersji TCP ponownie ukazuje przewagę TCP Vegas nad pozostałymi trzema wersjami, jednakże odwrotnie niż pozostałe wersje, Vegas notuje najniższe opóźnienia przy najmniejszych segmentach, a większe opóźnienia dla większego rozmiaru segmentu. Dzieje się tak ponieważ Vegas nie redukuje tak gwałtownie zajmowanej przepustowości i nie odczuwa negatywnych skutków nakładania się na siebie naraz konieczności retransmisji oraz jednoczesnego ograniczenia zajmowanej przepustowości. Możliwe jest to dzięki wbudowanemu nowemu mechanizmowi podejmowania decyzji o retransmisji, pozwalającemu znacznie szybciej zidentyfikować zagubione segmenty i je retransmitować, co z kolei bezpośrednio skutkuje dużo mniejszymi czasami opóźnień transmisji.

Obserwacje przepustowości kanału dla różnych wielkości segmentu TCP zaowocowała także spostrzeżeniem, iż dla większych segmentów ruch staje nierównomierny z dużą amplitudą skoków sumarycznej przepustowości wszystkich połączeń TCP. Dzieje się tak dla tego, iż przy bardzo dużym segmencie nawet małe zmiany wielkości okna transmisyjnego owocują znacznymi zmianami ilości danych transmitowanych bez potwierdzeń (czyli ruch bardziej przybiera charakter wybuchowy – ang. *BURST*). Także w przypadku większych segmentów TCP łączna przepustowość w kanale zwiększana była kosztem ruchu tła symulowanego przez UDP, dla którego rozmiar pakietu nie był zmieniany (zatem przy tej samej liczbie wysłanych pakietów do sieci, stacje korzystające z TCP wysłały więcej danych niż stacje nadające UDP).

Wykonano także badania wpływu liczby jednoczesnych sesji TCP na wydajność tego protokołu. Obserwując łączną przepustowość, można powiedzieć, iż w przypadku braku błędów w kanale liczba sesji nie ma wpływu na wykorzystanie dostępnego pasma. Jeżeli jednak w kanale transmisyjnym pojawią się błędy, to wraz ze wzrostem stopy błędów spada sumaryczna przepustowość, jednakże, jak należałoby się spodziewać, im więcej jednoczesnych sesji, tym wolniej spada przepustowość kanału przy wzroście stopy błędów. Generalnie, wzrost stopy błędów powoduje także wzrost opóźnień transmisji, lecz (jak zaobserwowano) czynnikiem znacznie bardziej zwiększającym opóźnienia jest liczba jednoczesnych sesji w kanale.

Kolejną istotną cechą badanych protokołów TCP jest podział dostępnej przepustowości pomiędzy konkurujące między sobą sesje TCP. Analiza wyników otrzymanych dla poszczególnych wersji TCP pozwoliła stwierdzić, iż rywalizując między sobą (w obrębie tej samej wersji TCP) bardzo dużą nieregularnością i nierównym podziałem przepustowości charakteryzowały się wersje Reno i New Reno. W przypadku TCP SACK podział przepustowości był trochę bardziej równomierny, lecz zdecydowanie najlepsze wyniki osiągnął TCP Vegas wykazując całkiem równomierny podział przepustowości pomiędzy poszczególne sesje, oraz

wykazał się niewielkimi nieregularnościami przebiegu przepustowości w czasie. Tak dobry rezultat TCP Vegas został osiągnięty dzięki nowym algorytmom przewidywania przeciążeń i ograniczania przepustowości. Równie interesujący okazał się rezultat symulacji mających na celu zbadanie interakcji sesji TCP pracujących w oparciu o odmienne implementacje tego protokołu. Jak pokazały symulacje, najlepszy wynik osiągnął TCP SACK z bardzo dobrymi średnimi przepustowościami i niewielkimi zmianami zajmowanej przepustowości w czasie. Dobry rezultat osiągnął także TCP New Reno zajmując także dość dużą przepustowość, natomiast słabo w rywalizacji o przepustowość poradził sobie TCP Vegas – w symulacjach zajmował on zazwyczaj najmniejszą część dostępnego pasma. Jak widać, jego skuteczne mechanizmy przewidywania przeciążeń sieci w przypadku rywalizacji o pasmo z innymi implementacjami TCP przegrywają, gdyż protokół ten przewidując natłok w sieci tym samym ustępuje przepustowości innym wersjom TCP.

4. Podsumowanie

Zaprezentowane tutaj symulacje ukazują efekty występujące w sieci 802.11, gdy transmisja oparta jest o protokoły TCP i UDP. Wyniki przeprowadzonych badań potwierdzają skuteczność mechanizmów przeciwdziałania przeciążeniom sieci, jednocześnie wskazując na problemy jakie powstają podczas funkcjonowania tych mechanizmów w sieciach bezprzewodowych, gdzie odmienne warunki transmisji znacznie obniżają efektywność ich działania. Badania nowych implementacji protokołu TCP dowodzą, iż sporo można w tej dziedzinie osiągnąć, jednakże nadal brak jest dopracowanych rozwiązań dla TCP, które spełniłyby wszystkie wymagania stawiane przez rozwój dzisiejszych sieci zarówno przewodowych jak i bezprzewodowych.

5. Przyszłe prace

Wyniki badań otrzymane w pracy pozwalają twierdzić, iż zarówno dla współczesnych sieci jak i dla sieci przyszłej generacji kluczową rolę w ich wydajności będą odgrywać protokoły transportowe. Z uwagi na rosnącą złożoność i różnorodność sieci opracowanie uniwersalnego protokołu lub zestawu protokołów dla celów sterowania przepływem staje się trudnym zadaniem dla projektantów. Algorytmy szacowania pasma mogą pozwolić na osiągnięcie celu, pod warunkiem dokładnego zbadania i odpowiedniego dostosowania tak do charakteru sieci jak i jej dynamicznie zmieniających się parametrów. Dalsze prace skierowane są na opracowanie nowych bardziej użytecznych mechanizmów sterowania dla protokołu TCP. Celowym wydaje się również rozwój narzędzi symulacyjnych, zwłaszcza wobec drogich alternatywnych rozwiązań komercyjnych i aktualnych problemach na tym polu w dziedzinie otwartych symulatorów sieciowych.

Literatura

1. Balakrishnan H., V. N. Padmanabhan, S. Seshan, R. H. Katz; „*A Comparison of Mechanisms for Improving TCP Performance over Wireless Links*"; IEEE/ACM Transactions on Networking 1996.
2. Brakmo L. S., S. W. O'Malley, L. L. Peterson; „*TCP Vegas: New Techniques for Congestion Detection and Avoidance*"; SIGCOMM 1994.
3. Chodorek R. R., A. Głowacz; „*Behavior of TCP Westwood in Wireless Network*". Advanced Technologies, Applications and Market Strategies for 3G and Beyond; Kraków 2002 r.
4. Głowacz A.: „*Zastosowania protokołów TCP i UDP w sieci bezprzewodowej*"; Referaty laureatów XXXIX Sesji Studenckich Kół Naukowych Pionu Hutniczego – AGH Kraków 2002 r.
5. Głowacz A., M. Maj; „*Badanie wydajności protokołów TCP i UDP w sieci bezprzewodowej*"; Praca magisterska. AGH Kraków 2002 r.
6. Mascolo S., C. Casetti, M. Gerla, M. Y. Sanadidi, R. Wang; „*TCP Westwood: Bandwidth Estimation for Enhanced Transport over Wireless Links*"; ACM Mobile Computing and Networking; Rzym 2001 r.
7. Mathis M., J. Mahdavi, S. Floyd, A. Romanov; RFC 2018. „*TCP Selective Acknowledgement Option*"; Network Working Group; październik 1996 r.
8. McCanne S., S. Floyd; Symulator ns wersja 2.1b9; kwiecień 2002 r. http://www.isi.edu/nsnam
9. Standard ANSI/IEEE 802.11 „*Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*"; 1999 Edition. http://standards.ieee.org
10. Wang R., M. Valla, M. Y. Sanadidi, M. Gerla; „*Adaptive Bandwidth Share Estimation in TCP Westwood*"; IEEE Globecom; Taipei 2002.
11. Wang R., M. Valla, M. Y. Sanadidi, M. Gerla; „*Using Adaptive Rate Estimation to Provide Enhanced and Robust Transport over Heterogeneous Network*"; 10th IEEE International Conference on Network Protocols, Paris 2002.
12. Wielka Internetowa Encyklopedia Multimedialna; kwiecień 2003 r. http://wiem.onet.pl

mgr inż. Andrzej GŁOWACZ
Studium Doktoranckie kierunku Informatyka
Wydziału Elektrotechniki, Automatyki, Informatyki i Elektroniki
Katedra Telekomunikacji Akademii Górniczo-Hutniczej
Al. Mickiewicza 30, 30-059 Kraków, tel. (0 12) 617 32 55
e-mail: glowacz@kt.agh.edu.pl

mgr inż. Michał MAJ
Polskie Radio Kraków S.A.
al. Słowackiego 22
30-007 Kraków
e-mail: mmaj@krak.eu.org

ZASTOSOWANIA METODY HAZOP W INŻYNIERII OPROGRAMOWANIA

Aleksander JARZĘBOWICZ

Streszczenie: Artykuł przedstawia HAZOP – metodę analizy modeli systemów oraz jej zastosowania w dziedzinie inżynierii oprogramowania do identyfikacji błędów obecnych w modelach. Omawiane są różne obszary stosowalności związane z różnymi klasami błędów: defektów modelowania, zdarzeń niebezpiecznych dla otoczenia systemu oraz zagrożeń związanych z zabezpieczeniem systemu. Artykuł prezentuje również dotychczasowy stan badań związanych z metodą oraz najbliższe perspektywy jej rozwoju.

Wstęp

Technologie informacyjne zyskują coraz większą popularność i zakres zastosowań w różnych dziedzinach życia. Oprogramowanie z jednej strony zastępuje istniejące rozwiązania, oparte na innych technologiach, a także pracę ludzi, z drugiej natomiast jest „z marszu" wprowadzane do zupełnie nowych dziedzin problemowych, które dopiero się kształtują. Konsekwencją tego jest fakt, że mamy do czynienia z coraz większym stopniem zależności osób, organizacji, a nawet społeczeństw od oprogramowania. Pociąga to za sobą pewne niebezpieczeństwa. Niewłaściwe; obarczone błędami lub po prostu niezgodne z potrzebami użytkownika działanie oprogramowania może skutkować poważnymi problemami np. brakiem możliwości prawidłowego funkcjonowania przedsiębiorstwa (w przypadku systemu wdrożonego w firmie) czy też zagrożeniem życia ludzkiego (w przypadku systemu używanego w samolocie lub szpitalu).

Oprogramowanie w przeciwieństwie do innych, bardziej tradycyjnych technologii jest odporne na tzw. błędy przypadkowe czyli związane z eksploatacją. Nie ulega zużyciu, odkształceniom ani uszkodzeniom mechanicznym. Jednakże jest z nim nieodłącznie związany inny rodzaj błędów – błędy projektowe, polegające na wytworzeniu produktu nie spełniającego wyspecyfikowanych wymagań, bądź też w ogóle niewłaściwym wyspecyfikowaniu wymagań. Z tego typu błędami mamy do czynienia już w najwcześniejszych etapach tworzenia oprogramowania i wynikających z nich produktach. Jednym z tych etapów jest faza analizy, której głównym produktem jest pewien model systemu (np. obiektowy). Niniejszy artykuł koncentruje się na identyfikacji i poprawie błędów na podstawie analizy obiektowego modelu systemu.

W używanym powyżej pojęciu błędu projektowego kryje się co najmniej kilka różnych rodzajów nieprawidłowości. Podstawowym rodzajem są defekty polegające na nieprawidłowym odwzorowaniu w modelu (albo innej reprezentacji systemu) rzeczywistości dziedziny problemowej lub wymagań użytkownika. W sytuacji gdy defekty te pozostaną nie wykryte, powodują wręcz z definicji, że

wytworzony system nie będzie działał poprawnie. W przypadku niektórych klas systemów istnieją jednak pewne dużo trudniejsze do wykrycia i wymagające odrębnej analizy rodzaje błędów. Chodzi tu po pierwsze o sytuacje, w których prawidłowo funkcjonujący system może ulec awarii i zagrozić swojemu otoczeniu na skutek np. nietypowego pobudzenia, zakłóceń lub awarii jednego z komponentów. Po drugie, poprawne działanie systemu może zostać (przy braku wystarczających zabezpieczeń) zakłócone na skutek złośliwego ataku.

W pierwszym przypadku mówi się o tzw. *hazardzie* czyli zdarzeniu niebezpiecznym, mogącym prowadzić do wypadku w otoczeniu systemu. Jest to pojęcie związane z systemowym atrybutem bezpieczeństwa (ang. *safety*) definiowanym jako poziom gwarancji, że funkcjonowanie systemu nie zagrozi jego otoczeniu. Identyfikacja i analiza hazardów jest konieczna w przypadku systemów wdrażanych w środowisku, w którym może dojść do sytuacji groźnych lub wręcz katastrofalnych. Jako przykłady można wymienić systemy związane z dziedzinami takimi jak: medycyna, transport (w tym lotniczy), energetyka, wojskowość i wiele innych.

W drugim natomiast przypadku, przedmiotem zainteresowania jest zabezpieczenie (ang. *security*) czyli atrybut systemu określający stopień odporności systemu na celowe ataki z zewnątrz. Wymagana jest analiza zagrożeń dotyczących zabezpieczenia (ang. *security threats*). Z koniecznością wykonania takiej analizy mamy do czynienia w przypadku systemów, dla których bardzo istotne jest uniemożliwienie dostępu użytkownikom nieuprawnionym do korzystania z systemu oraz ochrona zasobów systemu przed nieupoważnionym ujawnieniem lub manipulacją. W kategorii tej mieszczą się na przykład systemy związanych z elektronicznym biznesem, bankowością, administracją itp.

Oczywiście powyższe trzy kategorie nie wyczerpują wszystkich możliwych błędów projektowych, z jakimi można się zetknąć podczas prac nad systemami informatycznymi. Ich znaczenie jest jednak na tyle istotne, że warto poświęcić więcej uwagi problemowi ich wykrywania, i to na możliwie najwcześniejszym etapie rozwoju systemu. Z potrzebą taką najczęściej spotykamy się w przypadku zwykłych defektów modelowania, dotyczy to praktycznie każdego systemu. Z hazardami i zagrożeniami można się zetknąć jedynie w kontekście niektórych, niekiedy dość specyficznych klas systemów, jednak potencjalne konsekwencje ich wystąpienia nadają zagadnieniu ich wykrywalności duże znaczenie.

W przypadku wszystkich trzech rodzajów błędów, ich wykrywanie nie jest zwykle sprawą łatwą. Główną przyczyną jest rosnąca złożoność systemów, w efekcie której coraz trudniejsze staje się intelektualne panowanie nad całym systemem, a co za tym idzie również jego kolejnymi reprezentacjami opracowywanymi w cyklu wytwarzania. Dodatkowo, dla hazardów i zagrożeń nie wystarczy jedynie porównanie analizowanej reprezentacji systemu z wymaganiami systemowymi lub opisem dziedziny problemowej. Ich identyfikacja wymaga bardzo kreatywnego myślenia, rozpatrywania specyficznych sytuacji i kombinacji różnych zdarzeń. Wszystko to sprawia, że konieczne staje się usystematyzowanie procesu wykrywania błędów oraz zapewnienie niezbędnego wsparcia biorącym w

nim udział analitykom. Wsparcie takie polega na umiejętnym sterowaniu uwagą analityka i wyposażeniu go w odpowiednie wzorce, listy kontrolne, narzędzia itp. Jednym z możliwych rozwiązań jest wykorzystanie do tego celu metody HAZOP, a ściślej rzecz biorąc którejś z jej adaptacji ukierunkowanej na określony rodzaj błędów. W dalszej części artykułu zostanie pokazane, w jaki sposób możliwe jest wykrywanie poszczególnych omawianych powyżej rodzajów błędów za pomocą różnych wariantów metody HAZOP.

1. Metoda HAZOP

Metoda HAZOP (ang. *Hazard and Operability Study*) została opracowana z myślą o wykrywaniu błędów w modelach systemów na drodze systematycznego przeglądu modelu, koncentracji na jego wybranych fragmentach i sugerowaniu analitykowi potencjalnych nieprawidłowości.

Z uwagi na ciekawą historię metody, warto omówić jej genezę i kilka znanych zastosowań. Początki oryginalnej metody HAZOP sięgają lat 60. W brytyjskim przemyśle chemicznym i naftowym pojawiła się wówczas potrzeba opracowania techniki pozwalającej na identyfikację i analizę hazardów na podstawie modeli instalacji fabryk chemicznych. Instalacje takie, w dużym uproszczeniu, składają się z różnych elementów takich jak pompy, zbiorniki czy reaktory chemiczne oraz z połączeń (rur) pomiędzy nimi, którymi przepływają substancje chemiczne. Z doświadczeń dostępnych w tej dziedzinie przemysłowej wynikało, że na ogół wszystkie błędy prowadzące do awarii systemu manifestują się w połączeniach, w postaci odchyleń od prawidłowych (zgodnych z projektem) parametrów tych połączeń. Tak więc, przedmiotem zainteresowania metody stały się połączenia oraz ich parametry (atrybuty) takie jak: przepływ, temperatura, ciśnienie. W celu zapewnienia kompletności analizy zdecydowano się na wykorzystanie zbioru słów kluczowych (ang. *guidewords*), które w sposób ogólny sugerują wszelkie możliwe rodzaje odchyleń. Słowa kluczowe przyjmują postać krótkich fraz np.: NO, MORE, LESS, REVERSE, AS WELL AS, PART OF, OTHER THAN. Zestawienie atrybutu połączenia i słowa kluczowego daje konkretną interpretację odchylenia od prawidłowej wartości atrybutu.
Przykłady takich interpretacji:

- atrybut: *przepływ*, słowo kluczowe: *NO*, interpretacja: *brak przepływu w rurze*;
- atrybut: *temperatura*, słowo kluczowe: *LESS*, interpretacja: *temperatura substancji w rurze jest niższa niż zakładana w projekcie*;

Należy zauważyć że niektóre kombinacje atrybutów i słów kluczowych nie dają sensownych interpretacji. Analiza polegała na systematycznym rozpatrywaniu wszystkich połączeń i ich atrybutów oraz stosowaniu do poszczególnych atrybutów kolejnych słów kluczowych. Daje to pewne gwarancje, że żaden fragment systemu nie zostanie pominięty oraz ogranicza w danej chwili zakres rozważań do pojedynczej, konkretnej nieprawidłowości, jaka może

wystąpić w systemie. Poszczególne nieprawidłowości są rozpatrywane pod kątem możliwości ich wystąpienia oraz potencjalnych konsekwencji.

Metoda HAZOP odniosła sukces i do dziś jest szeroko stosowana w przemyśle chemicznym, a także w innych gałęziach przemysłu np. nuklearnej. Tymczasem wraz z rozwojem informatyki i coraz powszechniejszym stosowaniem oprogramowania w obszarach wymagających rozpatrzenia kwestii bezpieczeństwa, pojawiła się potrzeba opracowania podobnych technik analitycznych w dziedzinie informatycznej. W tej sytuacji podjęto próby zaadaptowania do potrzeb inżynierii oprogramowania sprawdzonych metod, stosowanych w innych dziedzinach inżynierskich, m.in. HAZOPu. Przystosowanie takie wymagało opracowania następujących zagadnień:

- wybór odpowiedniej reprezentacji systemu;
- wybór elementów reprezentacji stanowiących analogię połączeń i atrybutów chemicznego HAZOPu;
- dobór słów kluczowych;
- ustalenie interpretacji dla poszczególnych par atrybutów i słów kluczowych;

Prace w tym zakresie zostały wykonane w pierwszej połowie lat 90. Początkowo stosowano HAZOP jedynie do diagramów przepływu danych [1], z uwagi na podobieństwa pojęciowe pomiędzy przepływami substancji w instalacjach chemicznych i przepływami informacji w systemach informatycznych. Z czasem okazało się, że jako połączenie (podstawowe pojęcie metody, element podlegający analizie) można rozważać bardzo odmienne pojęciowo elementy różnego rodzaju reprezentacji systemów np. związek na diagramie ERD lub diagramie klas, czy też przejście na diagramie stanów. W efekcie opracowane zostały reguły adaptacji metody HAZOP dla kilku reprezentacji systemów i związanych z nimi konkretnych notacji modelowania ([2],[3],[4]). Wszystkie te doświadczenia zostały zebrane w standardzie 00-58 [5] brytyjskiego Ministerstwa Obrony dotyczącym stosowania metody HAZOP do systemów zawierających oprogramowanie. Standard definiuje uniwersalny zestaw słów kluczowych (patrz Tablica 1), kryteria prowadzenia analizy dla różnych rodzajów reprezentacji systemów oraz model procesu prowadzenia analiz w ramach m.in. spotkań grupowych.

Fakt stosowania metody w różnych dziedzinach inżynierskich, do różnych rodzajów diagramów i notacji oraz z ukierunkowaniem na różne cele (defekty, hazardy, zagrożenia), sprawia że trudno niekiedy dopatrzyć się cech wspólnych i dostrzec, że chodzi o tę samą metodę. Wśród wspólnych, najbardziej charakterystycznych mechanizmów metody, które zarazem decydują o jej skuteczności znajdują się:

- koncentracja na pewnych dobrze określonych elementach modelu (*połączenia*) i wyróżnienie ich możliwie atomowych, niepodzielnych fragmentów lub własności (*atrybuty*);
- wykorzystanie zestawu *słów kluczowych* sugerujących różne rodzaje nieprawidłowości;

- zestawianie słów kluczowych z atrybutami elementów i opracowanie *interpretacji* takich par jako sugestii pewnych błędów projektowych, które mogą wiązać się z użyciem w modelu elementu danego typu;
- opracowanie list kontrolnych zawierających dla wszystkich rozpatrywanych elementów sugestie błędów, które mogą ich dotyczyć;
- systematyczny przegląd modelu z wykorzystaniem list kontrolnych i podejmowanie dla każdej sugestii błędu decyzji, czy rzeczywiście jest ona prawdziwa dla aktualnie analizowanego elementu modelu;
- stosowanie do rejestracji wyników charakterystycznych struktur danych zawierających informacje o analizowanym elemencie, sugestii błędu oraz decyzji analityka/analityków (tzw. tabele HAZOP);

Tablica 1

Słowo kluczowe	**Ogólna interpretacja**
NO	Kompletny brak zamierzonego rezultatu, przy jednoczesnym braku innych efektów
MORE	Zbyt wysoka wartość parametru ilościowego
LESS	Zbyt niska wartość parametru ilościowego
AS WELL AS	Oprócz zamierzonego rezultatu otrzymano inny, dodatkowy
PART OF	Zamierzony rezultat został uzyskany tylko częściowo
REVERSE	Otrzymano logiczne przeciwieństwo zamierzonego rezultatu
OTHER THAN	Brak zamierzonego rezultatu, ale otrzymano inny, odmienny
EARLY	Zdarzenie zachodzi wcześniej niż zamierzono
LATE	Zdarzenie zachodzi później niż zamierzono
BEFORE	Zdarzenie zachodzi przed innym zdarzeniem, w stosunku do którego miało być późniejsze
AFTER	Zdarzenie zachodzi po innym zdarzeniu, w stosunku do którego miało być wcześniejsze

2. Obszary zastosowań metody

Zarówno standard 00-58, jak i pozostałe cytowane w poprzednim rozdziale prace odnoszą się do wykorzystania metody HAZOP do wykrywania hazardów, a więc do zastosowań w dziedzinie bezpieczeństwa systemów. W ostatnich latach pojawiły się jednak pomysły ukierunkowania analizy na inne rodzaje błędów projektowych: zagrożeń związanych z zabezpieczeniem systemu oraz zwykłych defektów modelowania. Te nowe zastosowania wynikają w sposób dość naturalny

z głównej „idei" HAZOPu i nie stanowią zbyt daleko idącej modyfikacji metody. Różnice dotyczą doboru analizowanych elementów i ich atrybutów oraz słów kluczowych, a także (i to przede wszystkim) sposobu tworzenia interpretacji dla par atrybut-słowo kluczowe. Mając na uwadze określoną kategorię błędów projektowych, które mają być wykrywane, można tak dobrać interpretacje poszczególnych par, aby sugerowały właśnie taki rodzaj błędów. Kwestie te stanowią główną różnicę pomiędzy poszczególnymi wariantami metody, chociaż nie jedyną. W każdym z trzech omawianych zastosowań spotykane są odmienne rozwiązania dotyczące organizacji procesu analizy (np. podział pracy - indywidualna lub zespołowa, zakres rozważań nad pojedynczą sugestią błędu).

Poniżej, w kolejnych podrozdziałach przedstawiane są szczegóły zastosowań metody HAZOP w trzech omawianych wcześniej obszarach problemowych.

2.1. Wykrywanie defektów

Analiza ukierunkowana jest na wykrywanie defektów polegających na błędnym odwzorowaniu modelowanej rzeczywistości w danej reprezentacji systemu. Przykładem takiej reprezentacji mogą być diagramy klas wchodzące w skład notacji UML. Podstawowymi elementami podlegającymi w ich przypadku analizie są związki obecne na tych diagramach, w szczególności te najczęściej używane: powiązanie (ang. *association*) i uogólnienie (ang. *generalization*). Jako atrybuty przyjmowane są przyporządkowane im w UML charakterystyki (nazwa, liczność itp.) a także „tożsamość" tych związków – sam fakt ich istnienia. Wykorzystywany jest uniwersalny zestaw słów kluczowych pokazany w Tablicy 1. Interpretacje nadawane zestawieniom atrybut-słowo kluczowe ustalane są pod kątem wychwycenia typowych defektów popełnianych przy wykorzystaniu danego atrybutu w modelowaniu. Przykładowe interpretacje przedstawiają Tablice 2-4. Atrybutem w Tablicy 2 jest związek uogólnienia (jego „tożsamość"), natomiast atrybutami Tablic 3 i 4 są charakterystyki związku typu powiązanie, odpowiednio: nazwa (ang. *name*) i agregacja (ang. *aggregation*).

Tablica 2

Atrybut: Generalization	
Słowo kluczowe	**Interpretacja**
AS WELL AS	Klasy nie pozostają ze sobą w związku dziedziczenia, cały związek jest zbędny i powinien być usunięty
MORE	Zbyt wiele klas dziedziczących, niektóre są zbędne
LESS	Zbyt mało klas dziedziczących, brak niektórych potrzebnych klas
OTHER THAN	Zdefiniowany związek jest niepoprawny, powinien być zdefiniowany inny związek niż uogólnienie

Tablica 3

Atrybut: Association.name	
Słowo kluczowe	**Interpretacja**
NO	Związek nie ma nazwy mimo, że powinien być nazwany
REVERSE	Błędny kierunek odczytywania nazwy związku, w którym znajdują się klasy
OTHER THAN	Nazwa związku jest nieprawidłowa i nie oddaje jego znaczenia; należy użyć innej nazwy
AS WELL AS	Nazwa związku jest zbyt ogólna, należy ją uszczegółowić
PART OF	Nazwa oddaje jedynie częściowo istotę związku, należy ją uogólnić

Tablica 4

Atrybut: Association.aggregation	
Słowo kluczowe	**Interpretacja**
NO	Nie zaznaczono agregacji, która powinna wystąpić
AS WELL AS	Zaznaczono agregację, która w rzeczywistości nie ma miejsca
OTHER THAN	Zaznaczono niewłaściwy typ agregacji tzn. silną zamiast słabej lub odwrotnie
MORE	Agregacja obejmuje zbyt wiele klas składowych, niektóre nie powinny wchodzić w jej skład
LESS	Agregacja nie obejmuje niektórych klas (obecnych na diagramie lub też nie), które powinny wchodzić w jej skład

W przypadku błędów projektowych będących defektami modelowania, analiza polega jedynie na ich wykryciu i udokumentowaniu. Nie dochodzi się ich przyczyn ani nie zastanawia nad konsekwencjami, jakie mogły by mieć gdyby pozostały nie wykryte.

Zalecane jest przyjęcie procesu dwufazowego. W fazie pierwszej niezależny (niezaangażowany w tworzenie modelu) analityk wykrywa i rejestruje defekty. W fazie drugiej wyniki pracy analityka trafiają do autora modelu, który dokonuje ich weryfikacji. Weryfikacja polega na ocenie czy defekt opisany przez analityka rzeczywiście ma miejsce, czy też np. analityk nie zrozumiał jakiegoś rozwiązania. Potwierdzone defekty powinny następnie zostać poprawione, najlepiej również przez autora modelu. Więcej informacji o tym zastosowaniu metody można znaleźć w [6].

2.2. Analiza hazardów

Jest to najstarsza i najbardziej znana dziedzina zastosowań metody HAZOP. Analizie podlegają modele opisujące dynamiczne zachowanie systemu np. znane z UML diagramy stanów lub diagramy interakcji. Przedmiotem zainteresowania analizy są elementy reprezentujące dynamiczne zmiany zachodzące w systemie np. przejście (ang. *transition*) w przypadku diagramu stanów lub komunikat (ang. *message*) w przypadku diagramu interakcji. Za atrybuty przyjmowane są wybrane charakterystyki tych elementów.

W dziedzinie analizy hazardów, interpretacje tworzone przez odnoszenie słów kluczowych do atrybutów mają za zadanie sugerować różnego rodzaju zakłócenia dotyczące zmian zachodzących podczas pracy systemu. W Tablicach 5 i 6 znajdują się zaczerpnięte z [7] przykłady interpretacji dla zdarzenia uruchamiającego przejście (ang. *trigger*) oraz dla komunikatu. Inne zestawy interpretacji zostały zaproponowane w pracy [8].

Tablica 5

Atrybut: Transition.trigger	
Słowo kluczowe	**Interpretacja**
NO	Zdarzenie nie zachodzi
AS WELL AS	Oprócz oczekiwanego zdarzenia zachodzi inne
OTHER THAN	Zamiast oczekiwanego zdarzenia zachodzi inne
EARLY	Zdarzenie zachodzi wcześniej niż jest spodziewane
LATE	Zdarzenie zachodzi później niż jest spodziewane
BEFORE	Zdarzenie zachodzi przed zdarzeniem/akcją które powinno je poprzedzać
AFTER	Zdarzenie zachodzi po zdarzeniu/akcji które powinno zajść po nim

W ramach analizy hazardów dla każdej tego typu sugestii wystąpienia zakłócenia należy odpowiedzieć na następujące pytania:

- Czy w ogóle możliwe jest wystąpienie takiego zakłócenia?
- Jakie mogą być tego przyczyny?
- Jakie może mieć ono konsekwencje dla całego systemu?

Jeśli wykryte zostanie zakłócenie, które może zagrozić bezpieczeństwu systemu (czyli hazard), odpowiedzi na powyższe pytania stanowią punkt wyjścia do zmiany projektu systemu w taki sposób aby wyeliminować przyczyny hazardu lub ograniczyć jego konsekwencje. Wszystko to stanowi zagadnienie dużo bardziej skomplikowane niż wykrywanie defektów modelowania i wymaga zaangażowania większej liczby osób. Analiza przeprowadzana jest w ramach spotkania grupowego

z udziałem m.in. projektantów systemu, użytkowników oraz ekspertów dziedzinowych.

Tablica 6

Atrybut: Message	
Słowo kluczowe	**Interpretacja**
NO	Komunikat nie zostaje przekazany
AS WELL AS	Komunikat zostaje powtórzony
OTHER THAN	Przekazany komunikat jest niepoprawny
EARLY	Komunikat zostaje przekazany zbyt wcześnie
LATE	Komunikat zostaje przekazany zbyt późno
BEFORE	Komunikat zostaje przekazany przed innym, który powinien go poprzedzać
AFTER	Komunikat zostaje przekazany po innym, który powinien być przekazany po nim

2.3. Analiza zagrożeń

W tej dziedzinie zastosowań do metody wprowadzono pewne modyfikacje, które na pierwszy rzut oka wydają się odbiegać w znaczny sposób od pierwowzoru. W rzeczywistości jednak wszystkie główne założenia metody HAZOP pozostają aktualne.

Pierwsza różnica polega na tym, że analizowanymi elementami nie są połączenia, lecz wchodzące w skład rozważanego systemu komponenty, które wymagają odpowiedniego zabezpieczenia. Jest to pojęcie bardzo szerokie. Do komponentów takich można zaliczyć zarówno rekord w bazie danych lub wiadomość e-mail, jak i firewall lub sieć komunikacyjną. Druga różnica dotyczy atrybutów, które są takie same dla każdego komponentu. Atrybutami tymi są trzy aspekty składające się na pojęcie zabezpieczenia: poufność, integralność i dostępność. Zrezygnowano również z tradycyjnego zestawu słów kluczowych. Zamiast tego wyróżniono dwa odrębne zestawy: pierwszy z nich sugeruje intencjonalność zaistniałego zagrożenia (celowe lub przypadkowe), drugi natomiast wskazuje przyczynę/sprawcę zagrożenia (np. wirus, haker, usterka techniczna).

Przykładowe kombinacje atrybutów i słów kluczowych przedstawione są w Tablicy 7, opracowanej na podstawie [9]. Tablica posiada dość interesującą konstrukcję. Interpretacje można otrzymać po prostu czytając kolejne pola poszczególnych wierszy.

Tablica 7

Słowo kluczowe I	utrata	Atrybut	Komponent	spowodowana przez	Słowo kluczowe II
celowa	utrata	integralności	programu pocztowego	spowodowana przez	wirusa
przypadkowa	utrata	dostępności	serwera	spowodowana przez	usterkę techniczną
celowa	utrata	poufności	rekordu danych	spowodowana przez	hakera

Wszystkie zagrożenia otrzymane w wyniku takich kombinacji muszą zostać przeanalizowane pod kątem możliwych przyczyn i konsekwencji, podobnie jak ma to miejsce w przypadku hazardów. Wyniki analizy mogą stanowić przesłanki do wprowadzenia zmian w projekcie systemu np. wprowadzeniu dodatkowych zabezpieczeń.

3. Zakończenie

Przedstawione w niniejszym artykule przykłady zastosowań metody HAZOP świadczą o jej potencjalnie dużej użyteczności w różnych dziedzinach inżynierii oprogramowania. Na szczególną uwagę zasługuje wykorzystanie metody do wykrywania defektów modelowania, które może okazać się pomocne w przypadku praktycznie każdego projektu informatycznego. Skuteczność tego typu analiz została już potwierdzona eksperymentalnie.

W chwili obecnej, w kilku ośrodkach naukowych trwają prace nad jak najpełniejszym przystosowaniem metody do notacji UML, za pomocą której tworzona jest obecnie ogromna większość modeli systemów. Jest to o tyle istotne, że standard 00-58 i inne źródła wiedzy o metodzie odwołują się do starszych notacji i metodyk, które zostały już w powszechnym użyciu zastąpione przez UML. W rezultacie pewna część podawanych w tych dokumentach informacji straciła swoją wartość i nie może mieć zastosowania w aktualnie realizowanych przedsięwzięciach informatycznych.

W najbliższym czasie przewidywany jest rozwój „młodszych" obszarów zastosowań metody (wykrywania defektów oraz analizy zagrożeń), dla których nie opracowano jeszcze kompletnych rozwiązań. Praca badawcza autora artykułu związana jest głównie z zagadnieniem zastosowania HAZOPu do wykrywania defektów i obejmuje zarówno działania związane z adaptacją metody, jak również prace nad wytworzeniem narzędzia informatycznego wspomagającego prowadzenie analiz.

Literatura

1. Earthy J. V.; *„Hazard and operability studies as an approach to software safety assessment"*; I.E.E. Computing and Control Division Colloquium on Hazard Analysis, Institution of Electrical Engineers, Digest No.: 1992/198; 1992
2. Chudleigh M.; *„Hazard Analysis Using HAZOP: A Case Study"*; materiały konferencyjne SAFECOMP'93; 1993
3. McDermid J., Pumfrey D.; *„Development of hazard analysis to aid software design"*; materiały konferencyjne COMPASS `94; 1994
4. Fencott P. C., Hebbron B. D.; *„The Application Of Hazop Studies To Integrated Requirements Models For Control Systems"*; materiały konferencyjne SAFECOMP'94; 1994
5. UK Ministry of Defense; *„HAZOP Studies on System Containing Programmable Electronics"*; Defense Standard 00-58; 2000
6. Górski J., Jarzębowicz A.; *„Wykrywanie anomalii w modelach obiektowych za pomocą metody UML-HAZOP"*; w mat. IV Krajowej Konferencji Inżynierii Oprogramowania; 2002
7. Jarzębowicz A.; *„Wspomaganie analizy systemu informatycznego metodą dewiacji przepływów"*; praca magisterska, Wydział Elektroniki, Telekomunikacji i Informatyki, Politechnika Gdańska; 2002
8. Lano K., Clark D., Androutsopoulos K.; *„Safety and security analysis of object-oriented models"*; materiały konferencyjne SAFECOMP'2002; 2002
9. Winther R., Johansen O-A., Gran B. A.; *„Security assessments of safety critical systems using HAZOPs"*; materiały konferencyjne SAFECOMP'2000; 2002

mgr inż. Aleksander Jarzębowicz
Katedra Zastosowań Informatyki
Politechnika Gdańska, ul. Narutowicza 11/12, 80-952 Gdańsk
tel. (58) 347-24-75
e-mail: olek@eti.pg.gda.pl

BIBLIOGRAFIA PUBLIKACJI PRACOWNIKÓW POLITECHNIKI CZĘSTOCHOWSKIEJ W ZINTEGROWANYM SYSTEMIE APIS-NB

Lidia SZCZYGŁOWSKA

Oprogramowanie APIS-NB umożliwiło stworzenie pod WWW bazy danych publikacji pracowników uczelni. System pracuje na platformie Windows 2000, wraz z zainstalowanym serwerem WWW – Apache. APIS-NB umożliwia wyszukiwanie rekordów poprzez strony WWW, posiada narzędzie administratorskie dla bibliotekarzy, oraz system archiwizacji zbiorów. Przy tworzeniu bazy wykorzystano po konwersji danych dotychczasową bazę publikacji opartą o oprogramowanie Micro ISIS. Zachowano funkcje poprzedniego oprogramowania wraz ze wszystkimi wprowadzonymi danymi.

Aplikacja ta została zaprojektowana w podobnej konwencji funkcjonalnej i graficznej jak główne katalogi Biblioteki - książek i czasopism - tak by maksymalnie ułatwić użytkownikowi dostęp do informacji i wyeliminować konieczność korzystania z tzw. „pomocy".

Baza danych - "Bibliografia publikacji pracowników Politechniki Częstochowskiej" -zawiera informację na temat publikacji pracowników Politechniki Częstochowskiej od 1988 roku. Umożliwia czytelnikom wyszukiwanie informacji natomiast pracownikom Biblioteki - po uprzednim zalogowaniu się do systemu - wprowadzanie, modyfikację, usuwanie rekordów oraz tworzenie zestawień statystycznych - raportów. Domyślnym użytkownikiem systemu nie wymagającym logowania do bazy jest "czytelnik".

Baza dostępna jest w Internecie ze strony Biblioteki Głównej pod adresem www.bg.pcz.pl - zalecaną przeglądarką jest Internet Explorer.

APIS-NB jest systemem zintegrowanym. Administrator systemu tworzy konta dla nowych użytkowników, przydziela ich do odpowiednich grup posiadających różne poziomy dostępu do danych. Zmian można dokonywać w trakcie eksploatacji systemu w zależności od potrzeb. Obecnie w systemie możemy wyróżnić trzy poziomy dostępu:

1. dostęp z poziomu użytkownika – przeglądanie bazy
2. dostęp z poziomu pracownika biblioteki – tworzenie bazy danych (edycja, dodawanie rekordów, generowanie raportów)
3. funkcje administratorskie (edycja kont użytkowników, ustawianie parametrów systemu)

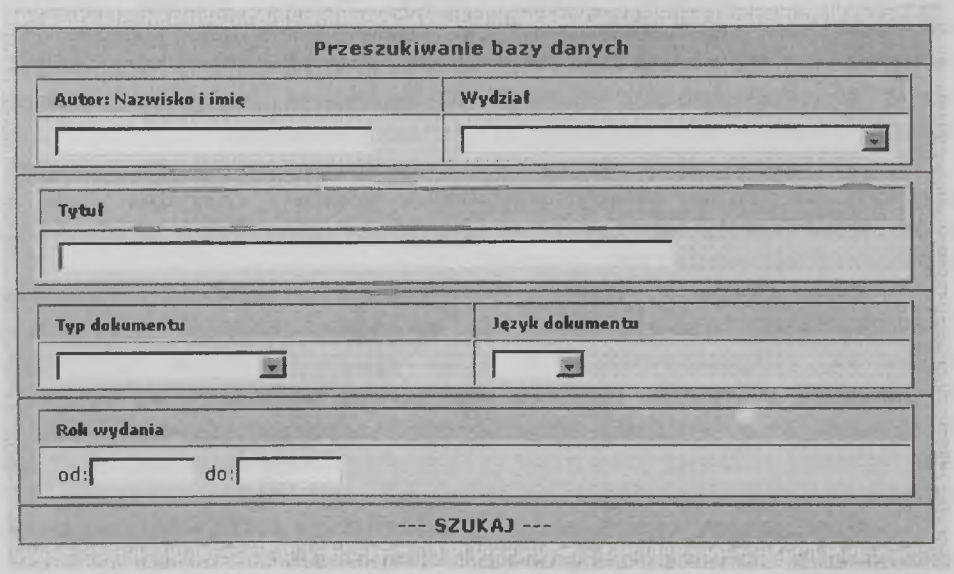
1. Przeglądanie bazy

Korzystanie z funkcji przeglądania bazy nie wymaga logowania do systemu – domyślnym użytkownikiem jest „czytelnik" – może on przeglądać

rekordy wykorzystując w tym celu „katalog alfabetyczny" lub formularz wyszukiwawczy („wyszukiwanie w bazie").

Przeszukiwanie bazy za pomocą formularza wyszukiwawczego

Opcja „WYSZUKIWANIE W BAZIE" umożliwia odnalezienie interesującego nas opisu publikacji z wykorzystaniem formularza. W celu odnalezienia opisu powinniśmy wypełnić przynajmniej jedno pole formularza. Wyszukiwanie następuje jako złożenie reguł podanych w polach formularza (tytuł = „dane dotyczące tytułu" i (AND) autor = „dane dotyczące autora"). Domyślnie puste pole w formularzu oznacza dowolną wartość w bazie.

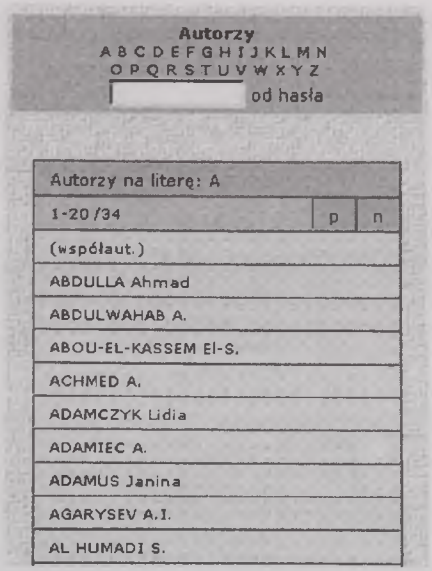


Rys. 1. Formularz wyszukiwawczy

Opis pól formularza:

Autor: Nazwisko i imię – należy tu podać dane autora którego opisy poszukujemy. Może to być:

a) pełne nazwisko + imię – pełne dane autora wyszukane w indeksie autorów wyświetlanym z prawej strony ekranu – możliwe jest wówczas uściślenie wyników o wydział na jakim dany autor był – wybór wydziału.



Rys. 2. Indeks autorów

b) pełne nazwisko – otrzymujemy opis publikacji autora lub autorów o danym nazwisku - uściślenie wyników o wydział na jakim dany autor był jest niemożliwe – nie przyniesie rezultatu.

Wydział: wybieramy z listy wydział dla którego przeszukujemy bazę publikacji. W przypadku wypełnienia pola „autor" należy podać pełne dane autora – wyszukane w indeksie autorów, wyświetlanym z prawej strony ekranu.

Tytuł: podajemy fragment tytułu, cały tytuł nie jest obowiązkowy; możemy podać fragmenty słów tytułu w dowolnej kolejności; spacja oznacza logiczny AND, „+" logiczny OR

Wyszukiwanie możemy zawęzić do danego typu dokumentu, języka publikacji, roku wydania, oraz - po zalogowaniu - do zakresu numerów systemowych. Służą do tego odpowiednio pola: **typ dokumentu, język dokumentu, rok wydania, nr systemowy** (pole „nr systemowy" dostępne jest po zalogowaniu się do bazy).

Pusta zawartość pola „rok wydania od" oznacza wpisanie roku 1988 (data rozpoczęcia katalogowania). Pusta zawartość pola „rok wydania do" oznacza wpisanie bieżącego roku.

Po wypełnieniu formularza należy kliknąć napis „--- SZUKAJ ---" znajdujący się w jego dolnej części. W wyniku otrzymujemy listę opisów publikacji – spis tytułów (tytuł + autorzy).

W celu otrzymania opisu publikacji - po wypełnieniu formularza (przynajmniej jedno pole) – należy kliknąć w jej tytuł. W ten sposób otrzymujemy – w zależności od praw aktualnego użytkownika - opis publikacji (czytelnik) bądź formularz umożliwiający modyfikację rekordu (użytkownicy z prawami 0,1)

Bibliografia publikacji pracowników Politechniki Częstochowskiej

WYSZUKIWANIE W BAZIE | ALFABETYCZNY SPIS TYTUŁÓW
logowanie do systemu | wyloguj się
aktualny użytkownik: czytelnik

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z | od hasła

Opis rekordu (nr systemowy: 008321)	
Autorzy	JOŃCZYK Anna - Wydział Inżynierii Proces.,Mater.i Fizyki Stos. KUŹNICKA-KUBAT Urszula - Wydział Inżynierii Proces.,Mater.i Fizyki Stos. STANIEWSKA Ewa - Wydział Inżynierii Proces.,Mater.i Fizyki Stos. WASZKIELEWICZ Wiesław - Wydział Inżynierii Proces.,Mater.i Fizyki Stos.
Tytuł	Analiza rynku jako podstawowy element wyboru strategii zaopatrzenia w hutnictwie.
Typ dokumentu	Referat
Rok wydania	2001
Nazwa czasop./patent./raport	W:II Międzynarodowa Sesja Naukowa.Nowe technologie i osiągnięcia w metalurgii i inżynierii materiałowej.Częstochowa 2001
Liczba stron	s.391-394
Materiały ilustrowane	tab.bibliogr.poz.2.
	powrót

Rys. 3. Opis publikacji dla czytelnika

2. Obsługi bazy danych dla pracowników

Korzystanie z funkcji dla pracowników Biblioteki wymaga zalogowania się do systemu, a więc posiadania konta w systemie (**login, hasło**). Administracją kontami (dodawaniem, usuwaniem) zajmuje się **administrator systemu**.

W celu zalogowania się do systemu należy kliknąć napis „logowanie do systemu" znajdujący się w górnej części ekranu (nagłówku). W formularzu „Logowanie do bazy danych" należy podać swój login i hasło otrzymane od administratora bazy, następnie kliknąć klawisz „WEJDŹ" lub nacisnąć „Enter" z klawiatury.

Logowanie do bazy danych		
Login		
Hasło		
	WEJDŹ	

Rys. 4. Formularz logowania do systemu

Dodawanie rekordów

Po zalogowaniu się do systemu – w górnej części ekranu pojawi się opcja „dodaj rekord". Po kliknięciu napisu automatycznie tworzony jest nowy rekord w bazie o następnym wolnym numerze systemowym bazy – z tytułem w postaci „!!!

nowy rekord !!!". Porzucenie edycji rekordu w w tym momencie pozostawia w bazie rekord o tytule „!!! nowy rekord !!!" – możemy go łatwo odszukać pod literą „A" w „alfabetycznym spisie tytułów" lub wyszukując za pomocą formularza – „wyszukiwanie w bazie".

W celu zapisania nowego rekordu - po wypełnieniu formularza – należy kliknąć napis „zapisz opis" – otrzymamy wówczas komunikat o zapisie nowego rekordu do bazy oraz jego opis (nr systemowy, opis bibliograficzny).

W celu porzucenia edycji nowego rekordu należy kliknąć napis „usuń opis" – otrzymamy komunikat o usunięciu rekordu z bazy (jest to konieczne, gdyż po kliknięciu „dodaj rekord" – automatycznie tworzony jest pusty rekord z tytułem „!!! Nowy rekord !!!").

Modyfikacja rekordów, usuwanie

Po zalogowaniu się do systemu – każde „wejście" w opis rekordu – kliknięcie tytułu opisu - powoduje wyświetlenie formularza umożliwiającego modyfikację wybranego opisu.

W celu zachowanie dokonanych zmian należy kliknąć napis „uaktualnij opis" –otrzymamy wówczas komunikat o modyfikacji istniejącego rekordu bazy oraz jego uaktualniony opis bibliograficzny.

W celu zapisania rekordu jako nowy należy kliknąć napis „zapisz jako nowy" – otrzymamy wówczas komunikat o zapisie nowego rekordu do bazy oraz jego opis (nr systemowy, opis bibliograficzny) – będzie on wierną kopią istniejącego rekordu (oczywiście z uwzględnieniem ewentualnych zmian w polach formularza) z nowym numerem systemowym – stary rekord pozostanie baz zmian.

W celu usunięcie rekordu z bazy należy odszukać rekord, wejść do jego edycji – klikając w jego tytuł, następnie należy kliknąć napis „usuń opis" znajdujący się w dolnej części formularza. Otrzymamy komunikat o usunięciu rekordu z bazy (usunięty numer systemowy)

Dodawanie, usuwanie, modyfikacja kodów wydziałów, języków, typów dokumentów, źródeł opisów możliwe jest po uprzednim zalogowaniu się do sytemu. W przypadku braku jakiegoś pola w listach wyborów należy uzupełnić tablicę kodów danego pola. W tym celu klikamy napis **„administracja bazą"** znajdujący się w górnej części ekranu. Otrzymujemy w ten sposób tablicę kodów wydziałów gotową do edycji.

Bibliografia publikacji pracowników Politechniki Częstochowskiej

WYSZUKIWANIE W BAZIE | ALFABETYCZNY SPIS TYTUŁÓW | RAPORTY
logowanie do systemu | administracja bazą | dodaj rekord | wyloguj się
aktualny użytkownik: **Politechnika Czestochowska**

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z | od hasła

wydziały | typy dokumentów | języki publikacji | źródła opisów

Kody wydziałów

Kod	Opis
WM	Wydział Inżynierii Proces.,Mater.i Fizyki Stos.
WBM	Wydział Inżynierii Mechanicznej i Informatyki
WE	Wydział Elektryczny
WZ	Wydział Zarządzania
WBud	Wydział Budownictwa
WIiOŚ	Wydział Inżynierii i Ochrony Środowiska

Rys. 5. Administracja bazą – tablica kodów wydziałów

Wybór tablicy kodów

W celu zmiany tablicy kodów na inną tj. typy dokumentów, języki publikacji lub „źródła opisów" klikamy odpowiedni napis – np. dla typów jest to „typy dokumentów" – znajdujący się pod nagłówkiem strony. Pamiętać musimy jednak o każdorazowym zapisywaniu zmian po edycji tablicy – w tym celu klikamy przycisk „ZAPISZ ZMIANY" znajdujący się w dolnej części tablicy kodów.

wydziały | typy dokumentów | języki publikacji | źródła opisów

Typy dokumentów

Kod	Opis
AR	Artykuł w czasopiśmie
RE	Referat
KS	Książka
ST	Skrypt
MN	Monografia

Rys. 6. Wybór tablicy kodów typów dokumentów

Edycja tablicy kodów, zatwierdzenie zmian.

Procedura edycji kodów jest identyczna dla każdej z tablic. Wpisywany w tablicy „kod" powinien być unikatowy, „opis" może być dowolny – pamiętać jednak musimy, iż to opis kodu widoczny jest w trakcie wypełniania formularza wyszukiwania lub edycji rekordu. Zaleca się, aby wszystkie używane do katalogowania kody były ujęte w tablicach – ma to znaczenie dla zestawień

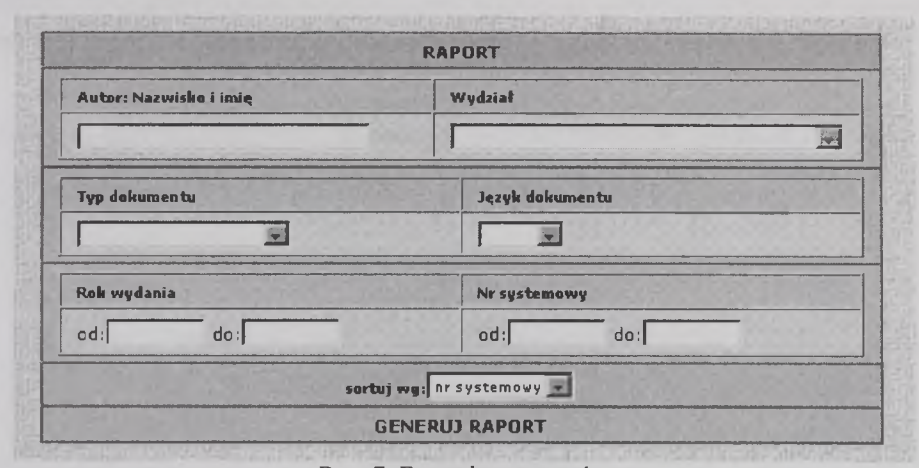
statystycznych. **Każdy rekord powinien być opisany przynajmniej jednym wydziałem, typem dokumentu, językiem publikacji.**

W przypadku braku pól na wprowadzenie nowego kodu i opisu należy zapamiętać istniejącą tablicę – w tym celu klikamy przycisk „ZAPISZ ZMIANY" znajdujący się w dolnej części tablicy kodów. Po tej operacji dodana zostanie następna linia umożliwiająca wpisanie następnych kodów. Czynność tą należy powtarzać aż do wpisania wszystkich kodów

Konieczne jest każdorazowe zapisywanie zmian po edycji tablicy – w tym celu klikamy przycisk „ZAPISZ ZMIANY" znajdujący się w dolnej części tablicy kodów.

Raporty

Korzystanie z raportów wymaga zalogowania się do systemu. Po zalogowaniu w górnej części ekranu pojawi się dodatkowa opcja „Raporty". Po kliknięciu w napis „Raporty" przechodzimy do formularza umożliwiającego sprecyzowanie typu raportu.



Rys. 7. Formularz raportów

Możemy podać autora, wydział dla którego ma być dany raport, typ dokumentów, język dokumentów, zakres czasowy jaki raport ma obejmować oraz zakres numerów systemowych rekordów. Domyślnie puste pole w formularzu oznacza dowolną wartość w bazie. Opisy rekordów w raporcie możemy posortować według numeru systemowego bądź tytułu.

Opis pól formularza:

Autor: Nazwisko i imię – należy tu podać dane autora dla którego generujemy raport. Może to być:

c) pełne nazwisko + imię – pełne dane autora wyszukane w indeksie autorów wyświetlanym z prawej strony ekranu – możliwe jest wówczas uściślenie wyników o wydział na jakim dany autor był – wybór wydziału.

d) pełne nazwisko – otrzymujemy raport dotyczący opisów autora lub autorów o danym nazwisku - uściślenie wyników o wydział na jakim dany autor był jest niemożliwe – nie przyniesie rezultatu.

W wyniku otrzymujemy dane statystyczne dotyczące publikacji danego autora oraz listę opisów pogrupowaną według lat publikacji.

Wydział: wybieramy z listy wydział dla którego chcemy otrzymać raport. W wyniku otrzymujemy dane statystyczne dotyczące publikacji w których brali udział autorzy z zadanego wydziału oraz listę opisów pogrupowaną według lat publikacji.

Otrzymane w raporcie wyniki możemy zawęzić do danego typu dokumentu, języka publikacji, roku wydania, numeru systemowego.

Pusta zawartość pola „rok wydania od" oznacza wpisanie roku 1988 (data rozpoczęcia katalogowania). Pusta zawartość pola „rok wydania do" oznacza wpisanie bieżącego roku.

Po wypełnieniu formularza raportu należy kliknąć napis „GENERUJ RAPORT" znajdujący się w jego dolnej części. W wyniku otrzymujemy raport z aktualną datą. Całość możemy zaznaczyć oraz wkleić do programu Word 97/2000 i dodatkowo przekształcić według własnych potrzeb lub wydrukować w otrzymanej postaci.

PUBLIKACJE PRACOWNIKÓW POLITECHNIKI CZĘSTOCHOWSKIEJ

/dostępne w bazie komputerowej BIBLIO/

Stan na dzień: **17.11.2001**

Autor	
Wydział	Wydział Budownictwa
Typ dokumentu	wszystkie
Język dokumentu	wszystkie
Rok publikacji	od: 1988 do: 2001

Dane statystyczne

Liczba wszystkich publikacji: 675
język obcy: 0, język polski : 0

Typy dokumentów

Typ dokumentu	Język polski	Język obcy	Razem
Artykuł w czasopiśmie	0	0	347
Referat	0	0	278
Książka	0	0	49
Skrypt	0	0	0

Rys. 8. Fragment raportu dla Wydziału Budownictwa

3. Funkcje administratorskie

Administracja kontami użytkowników bazy danych

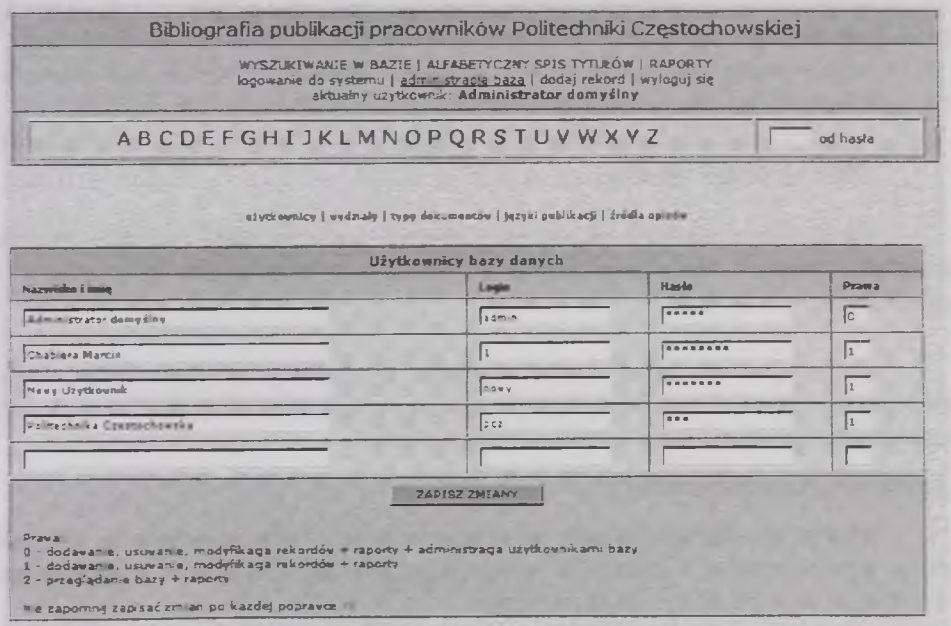
Pierwsze logowanie

Administracja kontami użytkowników bazy danych, wymaga zalogowania się do systemu jako administrator, polega na dodawaniu, usuwaniu lub modyfikacji istniejących kont. Domyślne konto administratorskie (administrator domyślny) to:

Login: admin
Hasło: admin

Po pierwszym zalogowaniu należy wymazać konto: „administrator domyślny" (admin) lub zmienić hasło na mniej oczywiste.

Proces dodawania, usuwania rekordów polega na edycji tablicy, której stan wyświetlany jest po kliknięciu napisu „administracja bazą" położonego w nagłówku strony lub napisu „użytkownicy" po wejściu do edycji innych kodów (wydziały, typy dokumentów ...). Po każdej zmianie zawartości tablicy użytkowników należy zapamiętać zmiany wciskając przycisk „ZAPISZ ZMIANY".



Rys. 9. Tablica użytkowników

Dodawanie, modyfikacja kont użytkowników

Dodanie nowego konta polega na wypełnieniu pól: nazwisko i imię, login, hasło, prawa i kliknięciu przycisku „ZAPISZ ZMIANY" znajdującego się w dolnej części tablicy użytkowników. W przypadku wyczerpania wszystkich pól umożliwiających wprowadzenie nowego konta należy zapamiętać tablicę użytkowników klikając przycisk „ZAPISZ ZMIANY" – pojawi się wówczas nowa linia tablicy umożliwiająca wprowadzenie następnego konta. Czynność tą należy powtórzyć, aż do momentu wpisania wszystkich nowych użytkowników.

Opisy pól formularza:

Nazwisko i imię	- dane użytkownika,
Login	- nazwa konta użytkownika,
Hasło	- hasło użytkownika, – powinno być wpisane i zapamiętane przez nowego użytkownika w trakcie tworzenia nowego konta przez administratora
Prawa	- prawa jakie użytkownik posiada w bazie - liczba 0,1 lub 2

Znaczenie kodów praw użytkowników:

0 - administrator bazy, może dodawać usuwać, modyfikować rekordy, tworzyć raporty,
administrować kontami użytkowników bazy, przeglądać bazę, ustawia parametry bazy

1 - pracownik Biblioteki, może dodawać usuwać, modyfikować rekordy, tworzyć raporty, przeglądać bazę

2 - pracownik Biblioteki lub upoważniona osoba może jedynie przeglądać bazę i tworzyć raporty

Domyślnym użytkownikiem bazy danych jest **czytelnik**, może on jedynie przeglądać bazę – wpis tego konta do tablicy użytkowników jest zbędne i niewskazane.

Przynajmniej jedno konto wpisane do tablicy użytkowników musi mieć prawa typu 0. Musi istnieć przynajmniej jeden administrator bazy, który posiada prawa edycji kont innych użytkowników.

Usuwanie kont użytkowników

Usunięcie konta polega na wyczyszczeniu odpowiadających mu pól tablicy użytkowników i kliknięciu przycisku „ZAPISZ ZMIANY" znajdującego się w dolnej części tablicy użytkowników.

Przewidziane sytuacje wyjątkowe

Użytkownik zapomniał hasło: użytkownik powinien udać się do admnistratora systemu w celu ponownego wpisania hasła do tablicy użytkowników.

Administrator zapomniał hasło: jest to wyjątkowa sytuacja, należy się wówczas zalogować do

bazy wpisując:

Login: tryb awaryjny bazy
Hasło: tryb awaryjny bazy

Tryb awaryjny pozwala na edycję tablicy użytkowników.

Do tablicy użytkowników nie wpisano konta o prawach 0 – brak konta administracyjnego: w takiej sytuacji po kliknięciu przycisku „ZAPISZ ZMIANY" do tablicy kont użytkowników zostanie automatycznie dodane konto: administrator domyślny:

Login: admin
Hasło: admin

Po zalogowaniu należy wymazać konto: „administrator domyślny" (admin) lub zmienić hasło na mniej oczywiste.

Do zadań administratora należy także robienie kopii danych, co ułatwia zastosowanie mechanizmu archiwizacji zbiorów wbudowanego w system.

Baza bibliografii publikacji pracowników uzupełniana jest systematycznie przez pracowników Biblioteki w oparciu o wiarygodne źródła informacji (Bibliografia Zawartości Czasopism, Current Contens, Przewodnik Bibliograficzny itp.), publikacje wpływające do Biblioteki oraz prace przedłożone przez autorów. Raporty przedstawiane są Władzom Uczelni w okresie sprawozdawczym, a także na życzenie. Pracownicy uczelni zobowiązani są do dostarczania oryginału lub wiarygodnego potwierdzenia opublikowania pracy, w ich interesie leży też sprawdzanie zgodności wykazu ze stanem faktycznym.

mgr inż. Lidia Szczygłowska
Kustosz, Z-ca Dyrektora
Dyrektora Biblioteki Głównej
ds. Komputeryzacji
Politechnika Częstochowska
42-200 Częstochowa

SPAM - AGRESYWNA FORMA PRZEKAZU INFORMACJI

Agnieszka ULFIK, Stefan NOWAK

Internet jest złożonym tworem. Zbudowany jest nie tylko z maszyn, komputerów –jego istotę stanowią ludzie: realna społeczność internetowa. Niestety nie zawsze ludzie Ci stosują się do przyjętych zasad i reguł. Każdy użytkownik komputera wie co to wirus i co może on zrobić z naszymi danymi, a nawet ze sprzętem. To zagrożenie jest znane od dawna i istniało jeszcze przed epoką Internetu. Dzisiejsze wirusy często potrafią same ewoluować, zmieniać się i oczywiście powielać. Obrona antywirusowa pochłaniała i pochłania olbrzymie nakłady finansowe. W ostatnich latach pojawiła się jednak nowa plaga. Jest nią spam.

Samo słowo „spam" ma długą i burzliwą historię. Pojawiło się jako marka sprzedawanej w USA puszki z mielonką. Intensywna reklama radiowa zrobiła swoje i w czasie II wojny światowej był to produkt bardzo popularny. Jego nazwa trafiła więc do skeczu grupy Monty Pythona, gdzie zastosowana została w specyficzny sposób („Well, we have spam, tomato & spam, egg & spam, egg, bacon & spam... Spam spam spam spam, spam spam spam spam, lovely spam, wonderful spam..."). W efekcie spam stał się swego rodzaju słowem-wytrychem, przybierającym różne znaczenia i kojarzącym się z nadmiarem, powtarzaniem itp.

W roku 1994 firma Cartner & Siegel z Phinix w USA wprowadziła „mielonkę" do internetu. Firma ta przeprowadziła pierwszą w historii akcję zasypywania grup dyskusyjnych Usenetu reklamami. Reakcja internautów była bardzo spontaniczna – tysiące maili z protestami. Była to pierwsza akcja bombardowania pocztowego (mail-bombing). W efekcie listy wciąż zapełniające dyski skłoniły providera, u którego firma Cartner & Siegel miała konto pocztowe, do zablokowania tego konta. Prawnicy uzyskali odblokowanie konta po podpisaniu zobowiązania, że nie będą więcej dopuszczać się takich czynów. Ponieważ firma ta nie dotrzymała zobowiązania w lutym 1995 roku została całkowicie odcięta od sieci. W międzyczasie firma ta zdołała wydać świetnie sprzedającą się książkę o marketingu w Internecie oraz założyć firmę oferującą organizowanie takich akcji.

1. Definicja spamu

Sieć zmienia się bardzo szybko, pojawiają się kolejne usługi, a z nimi nowe możliwości dla spamerów, ale i nowe metody walki z nimi. W rezultacie określenie „spam" inaczej zwany również „junk e-mail" stale nabiera szerszego znaczenia i trudno jest przewidzieć, czy kiedyś proces tworzenia się definicji tego słowa dobiegnie końca. Uznanie czegoś za spam zawsze będzie dyskusyjne. Kiedyś za spam uważano wiadomość wysłaną na kilka grup dyskusyjnych. Obecnie definicja obejmuje także niechciane listy elektroniczne – rozsyłane na przeróżne sposoby.

Poniższy tekst przedstawia definicję spamu i jest tłumaczeniem Łukasza Kozackiego[1] [7] standardowej definicji spamu według Mail Abuse Prevention System:

Elektroniczna wiadomość jest spamem, jeżeli

1) treść i kontekst wiadomości są niezależne od tożsamości odbiorcy, ponieważ ta sama treść może być skierowana do wielu innych potencjalnych odbiorców,

oraz

2) jej odbiorca nie wyraził uprzedniej, możliwej do weryfikacji, zamierzonej, wyraźnej i zawsze odwoływalnej zgody na otrzymanie tej wiadomości,

oraz

3) treść wiadomości daje odbiorcy podstawę do przypuszczeń, iż nadawca wskutek jej wysłania może odnieść korzyści nieproporcjonalne w stosunku do korzyści odbiorcy wynikających z jej odebrania.

Uwagi

i Trywialne lub zautomatyzowane metody personifikacji treści, takie jak „Szanowny Panie Jones, zauważyliśmy że jest Pan właścicielem domeny JONES.COM ..." w żadnym wypadku nie czynią wiadomości zależną od tożsamości jej odbiorcy.

ii Zaniedbanie zaznaczenia opcji typu „proszę nie przesyłać mi mailem treści reklamowych / marketingowych" w trakcie wypełniania formularza na stronie WWW w żadnym razie nie oznacza wyraźnej zgody na otrzymywanie takich treści. Jedynie w wypadku, gdy domyślną opcją jest „nie-wysyłanie żadnych maili", zaznaczenie przez użytkownika ewentualnej zgody na taką wysyłkę oraz gdy sens tej zgody jest jasno określony przed podjęciem jakichkolwiek działań, które mogłyby zmienić wynik zaznaczenia takich opcji, można tak wyrażoną zgodę uznać za zamierzoną.

iii Ewentualność nienależnej korzyści, jaką może odnieść nadawca w stosunku do odbiorcy oraz zależność treści wiadomości od osoby i tożsamości odbiorcy są autorytatywnie ustalane przez odbiorcę tej wiadomości i nie podlegają jakiejkolwiek dyskusji lub reinterpretacji ze strony nadawcy.

iv Wiadomość mailowa o charakterze innym niż osobisty zawsze w sposób nieproporcjonalny obciąża swymi kosztami odbiorcę i uważa się, iż przynosi nieproporcjonalną korzyść nadawcy, o ile odbiorca w weryfikowalny sposób nie zamówił otrzymania tej wiadomości lub z własnej woli nie zachce zaakceptować danej wiadomości w drodze wyjątku.

v Wiadomość nie musi mieć treści obraźliwej lub komercyjnej aby podlegać pod definicję spamu. Treść wiadomości jest bez znaczenia, poza możliwością określenia na jej podstawie zależności od tożsamości odbiorcy, jego zgody oraz możliwych dla niego korzyści.

Dodatkowe zastrzeżenia pochodzące z Definition of „spam" w/g UCEB SPAM blocking blackhole list

---

[1] *Definicja autorstwa Łukasza Kozickiego została zaczerpnięta ze strony Nospam-PL.NET*

vi *Wydobycie listy adresów mailowych z jakiejkolwiek publicznie dostępnej bazy danych, np. bazy whois (jak RIPE NCC, ARIN, APNIC) lub od rejestratora domen internetowych również nie sprawiają, że wiadomość staje się zależna od tożsamości jej odbiorcy.*

vii *Listy mailingowe typu Opt-Out są nie do przyjęcia, ponieważ brak danego adresu w dowolnej liście opt-out nie może być traktowany jak zaproszenie do wysyłanie na ten adres niezamawianych e-maili.*

viii *Jeśli wiadomość zawiera sfałszowane nagłówki sesji SMPT, może być uważana za spam niezależnie od jakichkolwiek innych cech zawartości.*

2. Rodzaje spamu

Wśród spamu możemy spotkać wiadomości o różnej treści i o różnym charakterze. Poniżej wyszczególnione zostały podstawowe rodzaje mailowych śmieci.

2.1. Reklamy

Spam marketingowy jest najstarszym i najczęściej spotykanym rodzajem spamu. Maile reklamowe to jeden z rodzajów reklamy bezpośredniej. W 1999 roku firma IMT Strategies przeprowadziła badania postrzegania i skuteczności tego rodzaju reklamy. Wynika z nich jasno, że wiadomości wysyłane za zgodą ich adresatów są bardzo efektywną formą marketingu. Z drugiej jednak strony wśród wniosków znalazło się stwierdzenie, że wysyłanie reklam bez zgody zainteresowanych powoduje spadek skuteczności tej akcji, a potencjalni klienci nastawiają się negatywnie do przedsiębiorstw ogłaszających się w ten sposób. Badanie to zostało zlecone przez właścicieli największych baz adresów e-mail. Sporo firm tych przytoczonych wniosków nie zna lub nie uważa ich za prawdziwe, ponieważ spam reklamowy nadal istnieje.

W przypadku reklam trzeba jednak dodać, że istnieje jeden wyjątek. Przy darmowych kontach pocztowych często w regulaminie istnieje zapis o zgodzie na otrzymywanie reklam.

2.2. Hoax

Angielskie słowo „hoax" dosłownie oznacza żart, figiel, kawał. Polska nazwa hoaxów to „fałszywy alarm". Są to wiadomości, które używając różnego typu treści mają na celu wywołanie jak największej paniki a co za tym idzie - ruchu w sieci. Wiadomości te są zazwyczaj wysyłane przez tzw. spamerów naiwnych. Istnieją trzy podstawowe typy hoaxów:

- łańcuszki szczęścia (np. *Jest to mail szczęścia, natychmiast po jego otrzymaniu prześlij go 7 osobom a będziesz szczęśliwy, bogaty, zdrowy, zostaniesz prezydentem itp., w przeciwnym razie stracisz wszystkich przyjaciół, zostaniesz okradziony, przejedzie Cię tramwaj itp.*),

- ostrzeżenia o wirusach (np. *Po Internecie krąży groźny wirus o nazwie ..., który jest w stanie skasować wszystkie twoje dane, spalić płytę główną, wysadzić monitor w powietrze. Natychmiast prześlij tą wiadomość wszystkim których znasz aby zdążyli zabezpieczyć się przed tym wirusem*),
- prośba o pomoc (np. *10 letni Piotruś umiera. 4 miesiące temu wykryto u niego białaczkę. Aby zdobyć pieniądze potrzebne na przeszczep szpiku potrzebuje twojej pomocy. Prześlij ta wiadomość do wszystkich swoich znajomych, a każdy wysłany e-mail przyniesie zysk który zasili konto chorego dziecka*).

Takie listy to w 99,99% przypadków fałszywki, plotka rozpowszechniana przez kogoś dla uciechy. Głupi dowcip, który bardzo wielu ludziom zabiera czas i pieniądze. Wiadomości dotyczące wirusów same w sobie są swego rodzaju wirusami, działającymi dzięki ludziom, a nie programom.

Hoaxy naprawdę szkodzą - wysyłane w ogromnych ilościach (wystarczy pomnożyć sobie np. 50 czy 150 adresów z własnej skrzynki kontaktowej przez ilość maili, jaka może zostać wysłana w 2, 3-cim „pokoleniu"...) - blokują sieć, niepotrzebnie przyczyniają się do wzrostu kosztów funkcjonowania Internetu, a przede wszystkim zabierają cenny czas setkom i tysiącom ludzi - i powodują zupełnie niepotrzebne stresy.

W dodatku niektóre firmy cynicznie zbierają w ten sposób adresy mailowe - wiedząc, że większość osób wysyłających dalej takie fałszywki nie usuwa z nich emaili poprzednich adresatów. Taki mail po kilku tygodniach czy miesiącach wraca do nadawcy, i może on uzupełnić swoją bazę danych adresów do spamowania o kolejną porcję.

A tego typu fałszywki nie wykryje w dodatku skaner antywirusowy. Ponadto, utrudniają one rozpoznanie rzeczywistych ostrzeżeń. Prawdziwe ostrzeżenia o wirusach mogą być wysłane przez osoby, którą naprawdę jakiś wirus dotknął - np. ktoś wykrył go u siebie po zanotowaniu dziwnego zachowania komputera.

Jeśli jednak masz pewne wątpliwości, co do informacji na temat groźnego wirusa, możesz sprawdzić czy wiadomość nie zawiera prawdy. Bazy wirusowe dostępne na stronach producentów oprogramowania antywirusowego, są uaktualniane co kilka godzin, a czasami nawet częściej. Niektóre serwisy posiadają również informacje na temat fałszywych alarmów.

Jeśli nie jesteś pewien czy wiadomość, którą właśnie otrzymałeś jest hoaxem poniżej znajdziesz kilka wskazówek jak wykrywać fałszywki. Zazwyczaj jest to stosunkowo proste, wystarczy przyjrzeć się wiadomości pod kątem niżej spisanych regułek, jakim zwykle podlegają hoax. Oto one:

- każdy hoax zawiera tekst w rodzaju *„prześlij wszystkim których znasz", „wyślij na wszystkie adresy ze swojej skrzynki", „zawiadom znajomych najszybciej jak to możliwe"* itp.,
- zawsze autor takiego listu stara się ze wszystkich sił zachęcić odbiorców do rozesłania dalej tego listu - używając argumentów bazujących na ludzkiej psychice. Często powołuje się na szansę pomocy innym, czasem miłosierdzie,

albo obiecuje w zamian powodzenie i szczęście, a do tego ze wszystkich sił stara się przekonać, że wiadomość jest prawdziwa,
- prawie zawsze autor takich pogłosek powołuje się na autorytety typu Microsoft, IBM, AOL, Symantec, McAfee, tymczasem wystarczy zajrzeć na witrynę Symantec, by dowiedzieć się, że oni nigdy nie rozsyłają podobnych informacji,
- często też posługują się technicznymi lub pseudo-technicznymi argumentami albo komputerowym żargonem – *„będąc w Twojej skrzynce zaraza adresy"*, *„dostaliśmy wiadomość od kogoś, kogo książka adresowa* (a tym samym nasza) *została zainfekowana wirusem"* - co jest, wirus zaraża książki adresowe a nie komputery, a w dodatku zarażając jedną książkę automatycznie zaraża wszystkie inne (?),
- ostrzeżenia o wirusach nigdy nie są wysyłane przez duże firmy do przypadkowych ludzi ani nie trafiają do Usenetu. Między innymi dlatego, że mogłyby zostać potraktowane jak spam, a to sprowadziłoby na firmę poważne kłopoty. Jeśli w mailu jest adnotacja typu *„przed chwilą McAfee ogłosił istnienie groźnego wirusa"*, *„Microsoft poinformował o istnieniu nowego groźnego wirusa"* - to jest to kłamstwo,
- w liście takim jest opis horroru jaki nas może spotkać jeśli nie zareagujemy, nie ma natomiast żadnych szczegółów dotyczących warunków w jakich wirus się aktywuje – *„wirus otwarty w programie pocztowym powoduje masakrę ! Kasuje całe dyski"*,
- w fałszywce jest zazwyczaj sporo wykrzykników, wielkich liter - i ogólnie wygląda on na pisany w panice. W biuletynach wysyłanych przez jakąkolwiek firmę antywirusową wykrzykników prawie nie ma, za to są poprawnie sformatowane od strony typograficznej – tekst po prostu wygodnie się czyta
- biuletyny na temat aktualnych wirusów są wysyłane do osób które tego sobie zażyczyły, wpisując się na odpowiednią listę na stronie www - ale wyglądają zupełnie inaczej niż fałszywka. Nigdy nie ma w nich zachęty do rozsyłania tej wiadomości dalej,
- ilość fałszywych pogłosek najczęściej nasila się kilka tygodni po atakach prawdziwych wirusów.

2.3. Nigeryjski szwindel albo „Makumba 419"

Są to wiadomości w stylu: *Czy chciałbyś dostać w prezencie 12.000.000 $?*

Oczywiście, taka oferta jest zwykłym oszustwem. Mechanizm tego oszustwa jest bardzo stary, znacznie starszy od Internetu. Już na początku ubiegłego wieku pojawiały się tradycyjne listy o treści proponującej o pośrednictwo w przekazywaniu pieniędzy celem wykupienia rosyjskiego arystokraty z hiszpańskiego więzienia.

Oszustwo w formie elektronicznej pojawiło się w Internecie pod koniec lat '80. Jedna z jego nazw, „419" określa numer paragrafu Nigeryjskiego Kodeksu

Karnego, dotyczący tego rodzaju wyłudzeń. Inne nazwy to Nigerian Scam (Nigeryjski szwindel), Advance Fee Fraud (Oszustwo z kosztami wstępnymi), 419 Fraud (Oszustwo cztery-jeden-dziewięć), The Nigerian Connection (Nigeryjski łącznik).

Zasada działania oszustów w każdym wypadku pozostaje identyczna: należy ofiarę skusić niebotycznie wysoką prowizją za udział w niezbyt legalnym przedsięwzięciu (przy czym pozwala to łatwo tłumaczyć wszelkie machinacje dotyczące przepływu dokumentów), a następnie wyciągać od ofiary oszustwa stopniowo coraz większe kwoty „na pokrycie kosztów operacji" - na łapówki, koszty transferu, wyprania pieniędzy, uzyskanie dokumentów umożliwiających transfer. Następnie pojawia się tysiąc innych komplikacji, a wszystkie bardzo, coraz bardziej kosztowne. Oczywiście, o pokrycie wszystkie kosztów proszona jest ofiara oszustwa, bo osoba proponująca całe przedsięwzięcie „nie ma aktualnie dostępu do zdeponowanych funduszy i potrzebuje pomocy, by ten dostęp uzyskać". Oszuści liczą na to, że ich ofiara, oczekując gratyfikacji wielkości kilku milionów dolarów, zgodzi się na wydanie kilku czy kilkudziesięciu tysięcy dolarów celem pokrycia kosztów wstępnych. Mało tego - oszuści posuwają się nawet do tworzenia fałszywych stron www afrykańskich agend rządowych lub banków. Przysyłają ofierze nazwę użytkownika i hasło, pozwalając sprawdzić, ze pieniądze już zostały przelane na utworzone na nazwisko ofiary konto w tym „banku" (i teraz potrzebne są jedynie pieniądze na pokrycie kosztów zagranicznego transferu).

I oczywiście - tacy naiwni się znajdują: defraudują pieniądze firm w których pracują, sprzedają swoje domy czy samochody. Nie otrzymują potem nic w zamian - ich „kontrahenci" rozpływają się gdzieś w Afryce, a próby dotarcia do nich osobiście mogą się skończyć nawet śmiercią oszukanego desperata. Oszustwo to aktualnie wyciąga od naiwnych w samych Stanach Zjednoczonych około 1 miliona dolarów dziennie - i tak już od lat, a zjawisko to coraz bardziej się nasila. Dla Nigerii jest to trzecia co do wielkości przychodów „gałąź przemysłu"! W wielu krajach świata obecnie każdy, lub prawie każdy list z Nigerii, a także listy z Południowej Afryki, Zimbabwe, Angoli, Sierra Leone czy Wybrzeża Kości Słoniowej, jest uważany za wstęp do wyłudzenia pieniędzy.

Od 1999 w Nigerii działa agenda Amerykańskiego Secret Service, ale mimo, że dokonała ona setek aresztowań nielegalny biznes „kręci się" coraz mocniej i obejmuje coraz więcej państw świata. Niestety, okazuje się, że rząd Nigerii nie chce współpracować w zwalczaniu tego procederu - nikt nie został przez nigeryjski sąd skazany z paragraf 419 kodeksu karnego, żadna z ofiar nie dostała też odszkodowania.

2.4. Wiadomości łamiące netykietę

Brak wiedzy na temat ogólnie przyjętych zasad przestrzeganych przez internautów też może być odbierane jako spamowanie. Początkujący użytkownicy Internetu często nic nie wiedzą na temat netykiety (zwyczajów i reguł panujących w Internecie) i w skutek swej niewiedzy stają się spamerami. Osoby takie nieświadomie utrudniają życie innym użytkownikom sieci ale zazwyczaj sytuacja

taka jest przejściowa i jeśli spamer przeczyta netykietę lub ktoś zwróci mu uwagę – poprawia się i przestaje być huliganem.

Netykieta dotyczy zarówno formy jak i treści wysyłanych listów. Generalnie należy unikać „szumu informacyjnego". Poniżej podane są wskazówki którymi należy się kierować aby nie być posądzonym o łamanie netykiety:

- unikanie nadmiernie długich i rozbudowanych sygnatur (podpisów pod listami umieszczanych w sposób zautomatyzowany) czyli nie dłuższych niż pięć wierszy tekstu,
- umiejętne cytowanie; jeśli odpowiadamy na czyjś list i chcemy go zacytować to fragment cytowany nie powinien być dłuższy niż nasza odpowiedź (czasami warto zacytować jedynie fragmenty listu na który odpowiadamy),
- unikanie dołączania dużych plików do wiadomości dla grup dyskusyjnych; do transferu dużych plików wymyślono inne sposoby przesyłania niż poczta (np. FTP), zamiast przesyłać plik umieśćmy w liście link do miejsca w którym jest on dostępny lub jeśli nie ma takiej możliwości należy wysłać plik osobom bezpośrednio zainteresowanym,
- błędne używanie DUŻYCH LITER; duże litery traktowane są jako krzyczenie lub „mówienie" podniesionym tonem głosu więc może być traktowane jako nietakt;
- treść odpowiadająca tematowi; nie należy, zwłaszcza podczas wypowiadania się na grupach dyskusyjnych, pisać listów na inny temat niż ten jaki jest zaproponowany,
- temat odpowiadający treści; pole Subject powinno zawierać informacje dotyczącą zawartości, nie będąc ani zbyt zwięzłe ani zbyt „przegadane"
- unikanie zbędnych dyskusji na grupach dyskusyjnych czyli tzw. flame war czyli wojny ogniowej z osobami z którymi się nie zgadzamy,
- bezwzględne dochowanie tajemnicy korespondencji; nie należy publikować treści wiadomości skierowanych do nas, lub tych w których posiadanie weszliśmy przypadkowo.

2.5. Pozostałe rodzaje spamu

Istnieje jeszcze wiele innych odmian spamu. Jedną z nich są SMS-y przesyłane na telefony komórkowe, a zawierające treści spełniające definicje spamu. Inny rodzaj to wiadomości które widzimy używając komunikatorów internetowych. Ostatnio pojawił się w Internecie nowy rodzaj spamu - to podobne do komunikatów systemowych okienka, „wyskakujące" na monitorach użytkowników systemów Windows NT, XP i 2000. Są to wiadomości wyświetlane przez usługę Windows Messengera (nie mylić z MSN Messenger'em). W normalnych warunkach usługa ta jest wykorzystywana np. przez administratora serwera do komunikacji z użytkownikami.

3. Koszt spamu

Spam choć wygląda niewinnie i na pierwszy rzut oka poza tym, że jest denerwujący nie robi wiele szkody. Każda wiadomość to jednak bity wędrujące po trasach routingu, miejsce zajęte na dyskach providerów, czas potrzebny na odfiltrowanie i usunięcie „śmieci", a to już są bardzo konkretne pieniądze.

Ważną cechą spamu jest fakt, że płaci za nią nie spamer, ale odbiorca. To właśnie odbiorca płacąc za łącze ponosi główny koszt niechcianych śmieci.

Firma Ferris Research opublikowała raport dotyczący kosztów spamu. Liczby wskazane w raporcie, to strata 8,9 miliardów (8.900.000.000) USD rocznie przez amerykańskie firmy, 2,5 miliarda USD w przypadku firm europejskich i dodatkowo strata 500 milionów USD przez providerów internetu łącznie w USA i Europie. Ponadto 4,4 miliarda USD kosztuje użytkowników internetu czas tracący na kasowanie spamu, a 3,7 miliardów USD wynoszą koszty dodatkowej mocy serwerów, koniecznej do uporania się ze spamem oraz koszty ponoszone dodatkowo na dzierżawę łącz internetowych.

4. Zabezpieczenia, sposoby radzenia sobie z problemem

Istnieje wiele metod radzenia sobie z problemem junk e-mail. Poniżej przedstawione są najczęściej stosowane sposoby radzenia sobie z tą plagą.

4.1. Obrona pozycyjna

Na niechcianą korespondencję odbiorcy reagują różnie. Podejście określane jako „siła spokoju" polega na ograniczeniu się do zwyczajnego kasowania bez czytania. Następnym krokiem jest zautomatyzowanie tej operacji. Większość programów pocztowych dysponuje konfigurowalnymi filtrami pozwalającymi automatycznie usunąć maile spełniające określone warunki. Kluczem do sukcesu jest odnalezienie w podsyłanych nam śmieciach wspólnych elementów. Pole manewru w tym względzie zależy od używanego oprogramowania, ale zawsze można sprawdzać nagłówki, a zwłaszcza dane w polach From:, To: i Subject:. Jeśli jakieś słowo się powtarza, to zablokowanie wszystkich listów je zawierających może w znacznej mierze rozwiązać nasze problemy.

4.2. Administrator na wagę złota

O ile w przypadku kont darmowych antyspamowa pomoc administratorów jest zwykle bardzo ograniczona, o tyle użytkownicy kont komercyjnych (nawet tych najtańszych) liczyć już mogą na więcej. Często serwery mają własne filtry antyspamowe, o których aktualizację dba administrator z własnej inicjatywy. Wspomaga się istniejącymi w Sieci bazami spamerów (np. www.polspam.org) czy informacjami ze specjalnych grup dyskusyjnych przeznaczonych do zgłaszania źródeł spamu (np. pl.news.mordplik).

Jeśli mimo wszystko coś się w większych ilościach prześlizgnie do naszej skrzynki, można poprosić administratora o pomoc. Stosowne adresy e-mail, kontaktowe numery telefonów itd. znajdziemy na stronie WWW firmy prowadzącej nasze konto. Ważne jest, aby zwrócić się bezpośrednio do administratora, gdyż inni pracownicy biurowi, mówiąc delikatnie, mogą być słabo zorientowani w sytuacji.

Mimo zdenerwowania, zapewne uzasadnionego atakiem spamera na nasze konto, wskazane jest zachowanie spokoju i powstrzymanie się od wylewania żalów oraz niecenzuralnych sformułowań. Trzeba załatwić to profesjonalnie! W mailu piszemy krótko o zaistniałej sytuacji (ile i jakiego spamu dostaliśmy, jakie interwencje były już podjęte itp.). Dalej wklejamy kilka pełnych nagłówków spamerskich listów lub dołączamy je w całości (jako załącznik), aby administrator mógł to przeanalizować. Na koniec konkretnie przedstawiamy, jakich działań oczekujemy (zablokowanie wszystkiego od tego nadawcy, wszystkiego o podobnej tematyce, „wycinanie" załączników itd.). Gotowe – czekamy na efekty pracy zarządcy serwera i odpowiedź, która po fachowo przygotowanej prośbie powinna być szybka i skuteczna.

4.3. Atak odwetowy

Jeśli jakiś spamer „zasypuje" nas wyjątkowo perfidnie, spróbować możemy interwencji bezpośrednio u niego. Wbrew pozorom może to przynieść bardzo pozytywne skutki, zwłaszcza kiedy listy pochodzą od firmy, która otwarcie podaje swoje dane adresowe. Piszemy grzecznie, acz zdecydowanie, że nie życzymy sobie więcej takiej korespondencji. Taka interwencja często pomaga. Możemy nawet zostać przeproszeni!

Gorzej sprawa ma się z anonimowym spamem zagranicznym. List często napisany jest bardzo grzecznie (po angielsku), a na końcu znajduje się nawet instrukcja opisująca wypisanie się z listy dystrybucyjnej. Czasem podjęcie opisanych tam działań przynosi pożądany skutek i spam od tego nadawcy już nigdy się nie pojawia, ale zazwyczaj jest zupełnie odwrotnie! Zdarza się, że taki uprzejmy spam jest tylko testem świeżo zdobytego adresu. Jeśli zaatakowany użytkownik odpowie, to będzie oznaczało, że zna angielski, zna się jako tako na komputerach itd. To już bardzo konkretne informacje, które pozwalają tak sprawdzony adres drogo sprzedać innym spamerom! Na takie maile lepiej nie odpowiadać w sposób zalecany przez nadawcę. Lepiej przyjrzeć się nagłówkom listu i nawiązać kontakt z administratorem serwera, z którego natręt korzysta.

Ponieważ wpisy w polach From: czy Return-Path: mogą być sfałszowane, pozostają nam pola Received: pozostawione przez kolejne serwery, które pośredniczyły w przesyłaniu danego listu. Z oczywistych względów interesuje nas ten pierwszy, od którego spam rozpoczął swoją wędrówkę. Pojawiająca się w nim informacja o domenie nadawcy także może nie być autentyczna, ale numer IP będzie już wartościową informacją.

Trzeba tylko ustalić, do kogo należy maszyna o danym numerze. Pomoże nam w tym usługa whois, dająca dostęp do bazy zarejestrowanych domen. Skorzystać można z niej za pomocą uniksowej komendy whois, strony WWW instytucji rejestrującej albo graficznego narzędzia, typu VisualRoute. W uzyskanych tą drogą informacjach powinniśmy znaleźć dane kontaktowe administratora serwera. Gdyby ich brakowało, można spróbować standardowych adresów typu „postmaster" lub „abuse" w danej domenie. Treść skargi przygotowujemy podobnie jak w przypadku listu do „naszego admina" (dołączmy pełne nagłówki!), lecz tym razem prosimy o zdyscyplinowanie użytkownika, który wysyła spam. Zazwyczaj już kilka takich skarg od różnych internautów skłania admina do zainteresowania się uciążliwym klientem i podjęcia stosownych działań (od upomnień do usunięcia konta włącznie).

4.4. Z naszego podwórka

Jeśli wytropiony winowajca korzystał z łączy znanego nam operatora telekomunikacyjnego, łatwo możemy pozyskać jeszcze jednego sojusznika. Wbrew pozorom np. nasz krajowy monopolista traktuje poważnie sprawy internetowych nadużyć – w tym spamu. Na specjalnie utworzonej stronie możemy znaleźć precyzyjne informacje na temat przyjętych zasad zgłaszania „zdarzeń" oraz statystykę podjętych przez TP S.A. działań. Funkcjonuje też konto pocztowe (abuse@tpnet.pl), na które przesłać możemy nasze uwagi.

Nie zniechęcajmy się przy tym pozorną anonimowością użytkowników numeru dostępowego 0-20 21 22. Mimo braku indywidualnego loginu i hasła oraz dynamicznie przydzielanego numeru IP możliwe jest bezproblemowe zidentyfikowanie spamera za pomocą systemu billingowego centrali telefonicznej. Pierwszą reakcją ze strony TP SA jest zwykle „ostrzeżenie" – doręczony na domowy adres list polecony, przedstawiający dokładnie, „co, gdzie i kiedy. Taki list potrafi naprawdę zdziałać cuda. Na „opornych" są inne środki: odłączenie telefonu czy zgłoszenie do organów ścigania (stosowane rzadko, ale w 100% skuteczne).

4.5. Profilaktyka

Zgodnie z zasadą „lepiej zapobiegać, niż leczyć" powinniśmy przede wszystkim unikać działań sprzyjających „zarażeniu". Podstawową sprawą będzie chronienie własnego adresu przed wpadnięciem w ręce spamera. W tym celu zakładamy dodatkowy alias dla naszego konta (lub dodatkową skrzynkę), przeznaczony wyłącznie do stosowania w „ryzykownych kontaktach". Jeśli rozprzestrzeni się on zbytnio wśród spamerów, zmieniamy go. Wszyscy znajomi znający nasz inny, stały adres na tym nie ucierpią, a spamerom zafundujemy wielki powrót wysyłanego przez nich śmiecia.

Aktywni użytkownicy grup dyskusyjnych użyć mogą wybiegu polegającego na dodaniu kilku znaków do swojego adresu. Człowiek chcący do nas napisać zorientuje się, że np. tekst *„wytnij_to"* nie jest częścią adresu

i bez problemu domyśli się prawidłowej nazwy naszej skrzynki. Spamerski automat przeszukujący grupy w poszukiwaniu potencjalnej ofiary da się jednak nabrać i skieruje swój atak na nieistniejące konto.

Każdy sprytny internauta opracowywać będzie prawdopodobnie też inne metody walki ze spamem. I słusznie, bo problem jest poważny. Pamiętajmy tylko, aby nasze antyspamerskie akcje nie szkodziły innym, uczciwym użytkownikom Sieci.

4.6. Spamerzy przed sąd

Błądzącymi po Sieci megabajtami natrętnej korespondencji, reklam i nikomu niepotrzebnych informacji zainteresowało się już wiele firm oraz organizacji. Powstały specjalne serwisy pomagające sporządzić skargę na spamera.

Temat zajmują się też parlamenty niektórych krajów. Powstają już przepisy dające usługodawcom internetowym kompetencje w ograniczaniu spamerom dostępu do Sieci, a nawet pozwalające na wymierzenie spamerowi wysokiej kary pieniężnej. Na razie przodują w tym USA[2] oraz Japonia. W większości krajów Unii Europejskiej również istnieją mniej lub bardziej skuteczne uregulowania prawne.

Dnia 10.03.2003 weszła w życie ustawa z dnia 18 lipca 2002r. o świadczeniu usług drogą elektroniczną (Dz.U. Nr 144 z 2002 r., poz. 1204). Z ustawą tą wiązano wielkie nadzieje. Ustawa ta dotyczy wszelkiego typu usług świadczonych drogą elektroniczną. W przypadku uregulowań dotyczących poczty elektronicznej w wielu miejscach ustawa zawiera normy ogólne dopuszczające dowolność interpretacji. W związku z powyższym należy się spodziewać nowelizacji niniejszej ustawy i jeszcze skuteczniejszego narzędzia do walki ze spamem.

5. Literatura:

1. http://echelon.pl/pub/spam-faq.html
2. http://nospam-pl.net
3. http://www1.gazeta.pl
4. http://www.arch.ipsec.pl
5. http://www.polspam.org
6. http://www.spam.com
7. http://www.spam.abuse.net
8. http://www.wsp.krakow.pl/papers
9. Marcin Nowak, Wielkie sprzątanie, Chip nr1 / 2003, str. 130
10. Tomasz Hrycuniak, Łapać spamera, Chip nr4 / 2002, str. 134
11. Sebastian Kuniszewski, Groźne inaczej..., Chip nr10 / 2001, str. 160

---

[2] Niektóre stany nie posiadają uregulowań prawnych np. na Florydzie nie obowiązują ograniczenia dotyczące wysyłania niechcianej poczty (http://www1.gazeta.pl/wyborcza/1093892,34471,1463548.html?as=2&ias=3)

Agnieszka Ulfik
Instytut Ekonometrii i Informatyki
Wydział Zarządzania
Politechnika Częstochowska
Ul. Armii Krajowej 19b
42-200 Częstochowa
ulfik@zim.pcz.czest.pl

Stefan Nowak
Instytut Ekonometrii i Informatyki
Wydział Zarządzania
Politechnika Częstochowska
Ul. Armii Krajowej 19b
42-200 Częstochowa
snowak@zim.pcz.czest.pl

ROZWIĄZANIA FIRMY CITRIX URZECZYWISTNIENIE KONCEPCJI WIRTUALNEGO STANOWISKA PRACY INTERNET JAKO INFRASTRUKTURA. WYŻSZY POZIOM DZIAŁALNOŚCI PRZEDSIĘBIORSTWA

„Od ponad dwóch lat korzystamy z Citrix® MetaFrame™ dla zapewnienia pracownikom naszego banku, rozrzuconym w 200 lokalizacjach, szybkiego dostępu do ponad 60 najważniejszych aplikacji. W ubiegłym roku czas przestoju systemu wyniósł zero. Opierając się na tych wynikach oraz na wstępnych testach wersji Citrix MetaFrame XP™, oczekujemy z niecierpliwością nowego produktu, który zapewni nam jeszcze większe możliwości zarządzania i skalowania. Technologia firmy Citrix umożliwia naszemu zespołowi informatyków osiągnięcie więcej przy użyciu mniejszych środków. Dzięki temu możemy zapewnić nieprzerwany rozwój naszej podstawowej działalności bankowej".

— Eric Wedemeyer, zastępca wiceprezesa, Mercantile Safe Deposit & Trust

Koncepcja wirtualnego stanowiska pracy szybko ewoluuje. Początkowo ograniczała się jedynie do przekazywania informacji za pośrednictwem stron WWW, a obecnie obejmuje wszechstronne wykorzystanie Internetu, intranetów i ekstranetów w charakterze infrastruktury do pracy zespołowej, wykonywania transakcji i zarządzania wiedzą. Istotnym wymaganiem stawianym przez te nowe środowiska zorientowane na WWW jest zapewnienie bezpiecznego wdrażania aplikacji o krytycznym znaczeniu.

Bez względu na to, czy infrastruktura obejmuje jedno czy więcej przedsiębiorstw, działy informatyki odpowiedzialne za komputeryzację działalności stoją dziś przed tymi samymi problemami. Należy wdrożyć przewidywalne, niezawodne i bezpieczne wirtualne stanowisko pracy, z którego użytkownik może korzystać w dowolnym miejscu i w dowolnym czasie. Ponadto należy zapewnić dostęp do wielu różnorodnych aplikacji i informacji — i to nie tylko ze stacjonarnych komputerów osobistych, lecz także z najrozmaitszych urządzeń przenośnych i wyspecjalizowanych, takich jak notebooki, kioski, portale, urządzenia bezprzewodowe i wiele innych.

Mamy do czynienia z nową epoką w technologii, która stawia przed administratorami przytłaczające wyzwanie w postaci ogromnie skomplikowanego środowiska informatycznego. Narzędzia, urządzenia i technologie często nie dają się dobrze zintegrować i nie chcą harmonijnie współpracować. Niedopasowane do siebie zasoby uniemożliwiają współużytkowanie najważniejszych informacji w czasie rzeczywistym, co naraża przedsiębiorstwo na stratę czasu i pieniędzy oraz pogorszenie wydajności.

Z tych powodów tysiące instytucji i przedsiębiorstw zaczyna wykorzystywać rozwiązania firmy Citrix, światowego lidera na rynku oprogramowania i usług do obsługi aplikacji. Systemy Citrix umożliwiają firmom

zagwarantowanie użytkownikom spójnego interfejsu oraz stałego dostępu do informacji i aplikacji, wpływając tym samym na zapewnienie efektywności i wydajności na poziomie niezbędnym w dzisiejszym wymagającym świecie biznesu.

Rozwiązania firmy Citrix. Możliwość natychmiastowego wdrożenia wirtualnego środowiska pracy

Jednak teraz — dzięki sprawdzonym rozwiązaniom Citrix — przedsiębiorstwa i instytucje na całym świecie mogą sprostać wyzwaniu, jakie stawia potrzeba wdrożenia wirtualnego środowiska pracy. Rodzina produktów Citrix obejmuje rozwiązania do obsługi aplikacji i portali. Dzięki tym rozwiązaniom przedsiębiorstwa i instytucje mogą wykorzystać Internet do dostarczenia aplikacji o krytycznym znaczeniu.

Nasze podejście jest oparte na najbardziej zaawansowanych technologiach na świecie — architekturze Citrix ICA® (*Independent Computing Architecture* — niezależna architektura komputerowa) i Citrix MetaFrame. Jest to jedyne rozwiązanie tego rodzaju, które zapewnia bezpieczny dostęp za pośrednictwem Internetu do aplikacji na platformach Windows®, UNIX® i Java™ — praktycznie z dowolnego urządzenia i za pośrednictwem dowolnego połączenia. Zapewniona jest przy tym niezrównana łatwość zarządzania i skalowania.

Citrix jest ekspertem w dziedzinie rozwiązań do obsługi aplikacji i portali. Od ponad 10 lat dostarczamy rozwiązania charakteryzujące się skalowalnością, niezawodnością, łatwością zarządzania i bezpieczeństwem. Naszymi klientami są przedsiębiorstwa na całym świecie. Nie proponujemy zatem Państwu żadnych eksperymentów. I nie musicie Państwo szukać gdzie indziej. Rozwiązania Citrix są niezrównane pod względem przewidywalnej wydajności, jakości interfejsu użytkownika oraz niskich kosztów eksploatacji. W sytuacji coraz większego zróżnicowania aplikacji, urządzeń, systemów operacyjnych i połączeń, coraz większej komplikacji środowisk informatycznych oraz coraz bardziej bezwzględnej konkurencji, możemy pomóc przedsiębiorstwom przejść na nowy poziom komputeryzacji — poziom bazujący na Internecie. Zapewniamy korzyści, których nie znajdą Państwo nigdzie indziej — korzyści zarówno dla działu informatyki, jak i użytkowników.

Rozwiązanie Citrix umożliwia informatykom zarządzanie wszystkimi aplikacjami z jednego stanowiska. Takie podejście zapewnia niezrównaną prostotę i skuteczność zarządzania nawet w najbardziej skomplikowanych środowiskach informatycznych.

Citrix ICA. Przetwarzanie na serwerze

Architektura Citrix ICA umożliwia przeniesienie przetwarzania z poszczególnych urządzeń-klientów na serwer, dzięki czemu można uzyskać prostsze, scentralizowane zarządzanie aplikacjami przy ich instalacji oraz wyższy

poziom pomocy technicznej i zabezpieczeń — a wszystko to przy nadzwyczaj niskich wymaganiach na przepustowość połączeń.

Zaawansowana kontrola administracyjna. Niezrównana łatwość zarządzania i skalowania

„Technologia Citrix to ważny element, dzięki któremu firma Chevron może prowadzić działalność w dowolnym czasie i w dowolnym miejscu. Citrix MetaFrame XP i Citrix NFuse™ stanowią podstawę naszego ogólnoświatowego, jednolitego środowiska operacyjnego NetGIL (*Global Information Link* — globalne łącze informacyjne). Środowisko bazuje na serwerach i umożliwia nam sprostanie specyficznym dla branży energetycznej problemom z dostępem do informacji. Na przykład, pracownik na szybie naftowym daleko w Zatoce Perskiej może uzyskać dostęp do aplikacji z prostego urządzenia, które w razie uszkodzenia po prostu się wymienia bez wzywania pomocy technicznej. Dla zapewnienia obsługi NetGIL, eksploatujemy ponad 40 serwerów Windows 2000. Citrix MetaFrame XP ułatwia nam pomyślną realizację tej obejmującej całe nasze przedsiębiorstwo inicjatywy".

— David Gallimore, starszy analityk systemowy, Chevron

Wdrożenie we współczesnym rozszerzonym przedsiębiorstwie środowiska informatycznego charakteryzującego się przewidywalnością, niezawodnością, bezpieczeństwem i ekonomicznością wymaga zapewnienia pełnych możliwości zarządzania i skalowania. Nasza sztandarowa rodzina rozwiązań wykorzystujących Citrix MetaFrame XP daje przedsiębiorstwom i instytucjom możliwości zarządzania bardzo zróżnicowanymi zasobami, w tym zespołami serwerów, aplikacjami i użytkownikami. Możliwości te obejmują scentralizowane zarządzanie licencjami, drukarkami i siecią. Dla zapewnienia jeszcze lepszej kontroli i skalowalności, rozwiązania Citrix do obsługi aplikacji zapewniają zgodność z czołowymi platformami zarządzania oferowanymi przez inne firmy. Zapewniona jest także obsługa Active Directory i ODBC oraz kompatybilność z bazami danych Microsoft® SQL, Oracle® i Access.

Aplikacje mogą być instalowane w ciągu minut

Dzięki funkcjom pakietowania i dostarczania aplikacji, zapewnianym przez rozwiązanie Citrix, administratorzy systemów informatycznych mogą centralnie konfigurować proces powielania oprogramowania na wiele serwerów Citrix MetaFrame i zarządzać tym procesem. Oznacza to, że zadanie to może być wykonane w ciągu minut, a nie jak dotychczas w ciągu wielu dni lub nawet tygodni. Wystarczy, by administrator utworzył pakiet aplikacji, aktualizacji lub plików (kompatybilnych z systemem Windows), po czym wskazał docelowe serwery Citrix MetaFrame XP i określił harmonogram dostarczania aplikacji na te serwery w obrębie całego rozszerzonego przedsiębiorstwa. Instalowanie aplikacji i

ich weryfikacja odbywa się całkowicie automatycznie. Dzięki temu nie tylko można wyeliminować błędy instalacji, lecz także zaoszczędzić czas i pieniądze oraz zlikwidować kłopoty zarówno administratorów, jak i użytkowników.

Zarządzanie zasobami dla zapewnienia dostępności przez całą dobę, siedem dni w tygodniu

Możliwości monitorowania i analizowania zapewniane przez system Citrix pozwalają administratorom środowisk informatycznych na śledzenie systemów, aplikacji i użytkowników w czasie rzeczywistym w celu zapewnienia nieprzerwanej dostępności systemu. Monitorowanie wykorzystania aplikacji umożliwia administratorowi ocenę wpływu aplikacji na zasoby systemu, takie jak procesory, pamięć i licencje. Możliwość wszechstronnej wizualizacji i analizowania aktualnego wykorzystania systemu i aplikacji pozwala administratorowi na przewidywanie wymagań systemu oraz na wprowadzanie korekt na długo przed pojawieniem się opóźnień odczuwalnych przez użytkowników.

Konwergencja organizacyjna. Stworzenie sieciowego środowiska pracy

Przyjęcie Internetu jako platformy dla działalności przedsiębiorstw jest siłą stymulującą koncepcję „uniwersalnej sieci", w której sieci lokalne, sieci rozległe, Internet, intranety i ekstranety są ściśle zintegrowane, umożliwiając przedsiębiorstwu lub instytucji optymalizację komunikacji.

Jednak obecnie istnieje tylko niewiele aplikacji o krytycznym znaczeniu, zaprojektowanych z myślą o dostarczaniu ich przez Internet. Wiele przedsiębiorstw i instytucji przerabia istniejące aplikacje, przystosowując je do pracy w sieci przy użyciu takich metod jak HTML, skrypty w Javie i różne środki niestandardowe. Podejście takie wymaga znacznych nakładów czasu i pieniędzy. Aby zapewnić szybkie i ekonomiczne wykorzystanie sieci, potrzebne jest rozwiązanie umożliwiające przedsiębiorstwom i instytucjom integrację i publikowanie nowych lub istniejących aplikacji przy użyciu dowolnej standardowej przeglądarki internetowej. Wymagane jest przy tym, aby taka integracja i publikowanie możliwe były niezwłocznie. Oto dlaczego klienci zwracają się ku rozwiązaniu Citrix NFuse — naszemu oprogramowaniu dla portali aplikacji.

Pełne wykorzystanie sieci

Citrix NFuse współpracuje bezproblemowo z rozwiązaniami firmy Citrix do zarządzania aplikacjami i obsługi aplikacji. Dzięki temu przedsiębiorstwa mogą grupować aplikacje i zawartość informacyjną w jednym widoku portalu. Użytkownicy mają wówczas dostęp do tych aplikacji i treści za pośrednictwem dowolnej standardowej przeglądarki internetowej — przy użyciu dowolnego

urządzenia i dowolnego połączenia oraz z dowolnej lokalizacji. Zawartość portalu może być zindywidualizowana, dzięki czemu każdy użytkownik otrzymuje te aplikacje i informacje, których potrzebuje do jak najefektywniejszej pracy. Citrix NFuse to najszybsza droga do poprawy wydajności oraz zapewnienia prawdziwej mobilności użytkowników.

Dla zagwarantowania wysokiego poziomu bezpieczeństwa w przedsiębiorstwach używających Internetu jako głównej infrastruktury, udostępniamy rozwiązanie Citrix Extranet™, które ściśle współpracuje z naszymi rozwiązaniami do obsługi aplikacji, zapewniając kompletną wirtualną sieć prywatną (VPN). Citrix Extranet stosuje najnowsze technologie uwierzytelniania, kontroli dostępu i szyfrowania, dzięki czemu gwarantuje bezpieczne dostarczanie informacji i aplikacji o krytycznym znaczeniu za pośrednictwem Internetu do użytkowników na całym świecie.

Największa elastyczność. Uniwersalny dostęp z każdego urządzenia, każdego miejsca, każdej sieci

Tylko Citrix MetaFrame — seria rozwiązań do obsługi aplikacji wykorzystująca architekturę Citrix ICA — umożliwia przezroczysty dla użytkownika dostęp do dowolnej aplikacji z komputerów z systemem Windows, komputerów z systemem Macintosh, stacji roboczych z systemem UNIX lub Linux®, a także urządzeń bezprzewodowych z przeglądarką WWW, notebooków, przystawek internetowych do telewizorów, kieszonkowych urządzeń bezprzewodowych i innych wyspecjalizowanych urządzeń sieciowych. Citrix MetaFrame zapewnia bezproblemowy dostęp do wszystkich aplikacji za pośrednictwem Internetu, sieci lokalnej, sieci rozległej, łączy telefonicznych lub łączy bezprzewodowych. Z punktu widzenia interfejsu użytkownika i działania, aplikacje takie nie różnią się od aplikacji pracujących na lokalnym komputerze osobistym.

Citrix MetaFrame zapewnia pełną gamę funkcji pozwalających na elastyczny dostęp do aplikacji na żądanie, w tym pełną integrację okien aplikacji lokalnych i zdalnych, dzięki czemu praca użytkownika z aplikacjami zdalnymi jest równie wygodna, jak z lokalnymi. Citrix zapewnia również jednolity mechanizm łączenia się z opublikowanymi aplikacjami na wypadek przerwania łączności z serwerem. Szyfrowanie strumienia danych od klienta ICA do serwera Citrix MetaFrame zapewnia wysoki poziom bezpieczeństwa. Niezrównana elastyczność tego rozwiązania pozwala na zmniejszenie liczby wąskich gardeł spowodowanych niedostatkami sprzętu oraz na zmniejszenie liczby koniecznych modernizacji sieci. Dzięki temu system informatyczny działa sprawniej, użytkownicy mogą przystąpić do pracy szybciej, zaś całe przedsiębiorstwo lub instytucja staje się bardziej wydajne.

Nasze kompleksowe rozwiązanie gwarancją Państwa sukcesu

System wdrażania aplikacji Citrix jest podstawą naszego sukcesu jako firmy. Jednak Citrix dostarcza nie tylko doskonałe produkty, lecz także kompleksowe rozwiązanie, zaprojektowane tak, by zapewnić klientom i partnerom firmy Citrix wygodne korzystanie z systemów informatycznych. Citrix Solution Network™ zapewnia pomoc techniczną na najwyższym światowym poziomie, a także szkolenie, konsultacje i elastyczne programy licencji — wszystko, co jest niezbędne do pomyślnego wdrożenia systemu informatycznego w dzisiejszym, dynamicznym środowisku.

Dział pomocy technicznej firmy Citrix jest przygotowany do udzielania wsparcia klientom na całym świecie i gwarantuje, że każdy zawsze otrzyma skuteczne rozwiązanie. Przodujemy pod względem najnowszych technologii dzięki naszemu ogólnoświatowemu sojuszowi Citrix Business Alliance™, który obejmuje ponad 600 czołowych producentów urządzeń oraz czołowych firm dostarczających oprogramowanie, sprzęt i rozwiązania telekomunikacyjne. Oferujemy szkolenia, certyfikację, narzędzia do obsługi technicznej, konsultacje oraz dostosowywane poziomy usług pomocy technicznej, dzięki czemu zapewniamy zakres usług pozwalający Państwu na maksymalne wykorzystanie inwestycji w informatykę. Możemy śmiało twierdzić, że rozwiązania Citrix wykraczają poza dostarczanie aplikacji, a stanowią kompleksowe wirtualne stanowisko pracy — które mogą Państwo wdrożyć natychmiast.

Dzięki rozwiązaniom firmy Citrix, przedsiębiorstwa i instytucje mogą już dzisiaj wdrożyć cyfrowe stanowisko pracy, wymagane przez współczesny, wysoce konkurencyjny świat biznesu.

Więcej informacji na temat korzyści, jakie dają rozwiązania firmy Citrix do obsługi aplikacji, można uzyskać pod adresem www.citrix.com.

O firmie Citrix

Citrix Systems, Inc. (Nasdaq:CTXS) jest światowym liderem w dziedzinie infrastruktury dostępowej dla przedsiębiorstw, instytucji rządowych oraz edukacyjnych. Najbardziej zaufana marka w dziedzinie infrastruktury dostępowej dla przedsiębiorstw, Citrix® MetaFrame® Access Suite, umożliwia prosty i bezpieczny dostęp na żądanie w przedsiębiorstwach, z dowolnego miejsca, w dowolnym czasie, przy pomocy niemal każdego urządzenia i połączenia sieciowego. Niemal 50 milionów ludzi w ponad 120,000 przedsiębiorstwach na świecie wykorzystuje technologię firmy Citrix każdego dnia. Do grona klientów firmy Citrix należy 100% firm z listy *Fortune* 100, 95% z listy *Fortune* 500 oraz 95% z listy *Financial Times European* 100. Siedziba główna firmy mieści się w Fort Lauderdale, w stanie Floryda (USA). Citrix posiada biura w 22 krajach oraz ponad 7,000 partnerów w ponad 100 krajach. Więcej informacji znajduje się pod adresem: http://www.citrix.com.

Citrix®, MetaFrame® oraz Citrix Solutions Network™ są zarejestrowanymi znakami towarowymi lub znakami towarowymi Citrix Systems, Inc. w Stanach Zjednoczonych oraz innych krajach. Wszystkie inne znaki oraz zarejestrowane znaki towarowe stanowią własność odpowiednich osób lub instytucji.

O firmie Pretor sp. z o.o.

Firma PRETOR sp. z o.o. rozpoczęła działalność 19 października 1991 roku. Właścicielami firmy są absolwentami Politechniki Gdańskiej wydziału Elektroniki.

W obecnej chwili PRETOR sp. z o.o. proponuje kompleksowe rozwiązania informatyczne dla dużych przedsiębiorstw, firm, instytucji i uczelni.

Pretor wytwarza serwery i komputery **PHOBOS** - produkowane zgodnie z normą ISO 9001:2000 ze ściśle określonych elementów, wyselekcjonowanych w procesie szczegółowych testów. Zaletą firmy jest duża elastyczność oferty i jej dostosowanie do zamówień klienta. Pretor proponuje swoim klientom profesjonalne rozwiązania serwerowych platform sieciowych opartych na systemie **MICROSOFT WINDOWS 2000 SERVER**. Oferujemy usługi związane z konserwacją i utrzymaniem prawidłowej pracy systemów. Projektujemy i wykonujemy kompleksowe rozwiązania rozległych sieci Internet, Intranet, WAN i sieci LAN oparte na elementach sieciowych firmy CISCO, HP, 3COM. Od 2000 roku działając w sieci **CITRIX SOLUTION NETWORK** oferuje rozwiązania zcentralizowanej pracy terminalowej na bazie aplikacji **CITRIX MetaFrame XP**. Zapewniamy także swoim klientom kompleksową opiekę nad bazą sprzętu i oprogramowania firmy w ramach usługi SKOS – IT (System Kompleksowej Obsługi Serwisowej - IT). Obszarem działania firmy są również instalacje teleinformatyczne w obiektach przemysłowo-handlowych. W zakres wchodzą tu systemy sygnalizacji włamania i napadu, kontrola dostępu, system śluz bankowych, telewizja przemysłowa, system sygnalizacji pożaru, systemy gaszące, systemy nagłośnienia, centrale telefoniczne, systemy telewizji kablowej, wizualizacja systemów nadzoru.

Dane teleadresowe firmy PRETOR: siedziba: ul. Rakoczego 31, 80-169 Gdańsk
tel. (058) 303-31-33, fax (058) 303-21-64
pretor@pretor.com.pl , www.pretor.com.pl

CZY ISTNIEJE PRAWO INTERNETOWE?

Piotr WAGLOWSKI

Transgraniczność internetu może doprowadzić do stopniowego likwidowania jednostek państwowych, jakie znamy teraz, na rzecz społeczności globalnej, międzynarodowej. Może się jednak zrealizować również zupełnie inny wariant rozwoju Sieci. Wariant, w którym ugruntują się wirtualne granice państw, a w tych granicach będzie obowiązywać niepodzielnie prawo... Powstaje tylko pytanie: czy konieczne jest definiowania nowej gałęzi prawa? Prawa internetowego?

Moda na prawo internetowe

Od pewnego czasu panuje moda na prawo internetowe. Przyznam, że sam się dałem porwać krucjacie na rzecz tego prawa[1]. Moja krucjata zmierzała do uświadamiania społeczności internetowej, iż system norm, jakim rządzą się różne aspekty funkcjonowania w Internecie nie powinien się opierać wyłącznie na netykiecie - dokumencie spisanym w zamierzchłych czasach jako regulamin postępowania w jednej z sieci akademickich... Starałem się wykazać, że normy, jakimi rządzi się aktywność w Internecie powinny opierać się na prawie państwowym lub międzynarodowym, w zależności od tego, jakiej aktywności dotyczy, gdzie ma miejsce zdarzenie i kto jest jego uczestnikiem.

Mimo iż - cytując jednego z dyskutantów w Usenecie - "dla wielu zasiadających w ławach sejmowych, internet, to ikonka na pulpicie": ostatnio ustawodawca polski sporo energii poświęcił na uchwalanie kolejnych ustaw mających doniosłość dla działalności w Internecie. Na uwagę zasługuje zwłaszcza wielość normatywnych definicji i terminów używanych w tych ustawach.

Kilka terminów i definicji normatywnych

Kluczowa dla internetu ustawa z 21 lipca 2000 r. prawo telekomunikacyjne nie wspomina o internecie. Z tej ustawy możemy dowiedzieć się, iż eksploatacja sieci telekomunikacyjnych to faktyczne i bezpośrednie dysponowanie całością funkcji urządzeń i sieci zapewniających telekomunikację; interoperacyjność sieci to zdolność sieci telekomunikacyjnych do efektywnej współpracy w celu zapewnienia wzajemnego dostępu użytkowników do usług świadczonych w tych sieciach; połączenie sieci telekomunikacyjnych to fizyczne i funkcjonalne połączenie sieci telekomunikacyjnych, eksploatowanych przez tego samego operatora lub przez

---

[1] W efekcie napisałem artykuł "2001 Odyseja prawno-internetowa. Podsumowanie wydarzeń związanych z prawem internetowym", dostępny w serwisie VaGla.pl Prawo i Internet.

różnych operatorów, w celu zapewniania telekomunikacji użytkownikom tych sieci lub świadczenia usług telekomunikacyjnych zarówno przez operatorów łączonych sieci, jak i przez inne podmioty mające dostęp do sieci; sieć telekomunikacyjna to urządzenia telekomunikacyjne i linie telekomunikacyjne, zestawione i połączone w sposób umożliwiający przekaz sygnałów pomiędzy określonymi zakończeniami sieci, za pomocą przewodów, fal radiowych bądź optycznych lub innych środków wykorzystujących energię elektromagnetyczną; sieć publiczna to sieć telekomunikacyjna, nie będąc siecią wewnętrzną, służąca do świadczenia usług telekomunikacyjnych; sieć wewnętrzna - sieć telekomunikacyjna eksploatowana przez podmiot wyłącznie dla własnych potrzeb lub założona w budynkach niemieszkalnych usytuowanych na terenie jednej nieruchomości gruntowej; użytkownik końcowy - podmiot korzystający z usług telekomunikacyjnych lub żądający ich świadczenia albo udostępnienia takich usług, dla zaspokojenia własnych potrzeb; zakończenie sieci to punkt sieci telekomunikacyjnej przeznaczony do zapewniania użytkownikowi dostępu do sieci...

W innych przepisach ustawodawca postanowił jednak wymienić internet, tak jak w przypadku ustawy z dnia 26 lipca 2001 r. o zmianie ustawy o zamówieniach publicznych (Dz.U. z 2001 r. Nr 113, poz. 1208): Artykuł 29 ww. ustawy stanowi: "Zamawiający zaprasza do udziału w przetargu nieograniczonym, zamieszczając ogłoszenie w miejscu publicznie dostępnym w swojej siedzibie, na stronie internetowej oraz za pośrednictwem Prezesa Urzędu w Biuletynie Zamówień Publicznych". W kolejnych ustępach czytamy "Jeżeli wartość zamówienia nie przekracza wyrażonej w złotych równowartości kwoty 30 000 EURO, ogłoszenie o przetargu nie ograniczonym zamieszczane jest w siedzibie zamawiającego, w miejscu publicznie dostępnym oraz na stronie internetowej, co najmniej na 7 dni przed terminem składania ofert". I dalej: "Jeżeli wartość zamówienia przekracza równowartość kwoty 30 000 EURO zamawiający obowiązany jest zamieścić ogłoszenie o przetargu nieograniczonym w swojej siedzibie oraz na stronie internetowej, od dnia publikacji ogłoszenia w Biuletynie Zamówień Publicznych do dnia składania ofert". Super. A co to jest strona internetowa?

Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 17 maja 2002 r. w sprawie Biuletynu Informacji Publicznej (Dz. U. Nr 67 z 5 czerwca 2002 r., poz. 619) definiuje kolejne pojęcia. Adres URL - adres wskazujący lokalizację strony głównej lub podmiotowej strony Biuletynu w sieci teleinformatycznej (adres strony www). Czy termin "strona internetowa" przewidziana w ustawie o zamówieniach publicznych ma ten sam zakres przedmiotowy co "strona www" w Rozporządzeniu? Zastanawia mnie umieszczenie w ustawie angielskojęzycznych skrótów. URL to skrót od Uniform Resource Locators zdefiniowany m.in w RFC[2] o numerze 1738[3]. WWW to angielskojęzyczny skrót od World Wide Web. Na mocy § 6 Rozporządzenia: Informacje publiczne udostępniane w Biuletynie nie mogą zawierać

---

[2] Request for Comments

[3] dostępnym m.in. pod adresem http://www.w3.org/Addressing/rfc1738.txt

niewyjaśnionych skrótów. Samo rozporządzenie takimi, niewyjaśnionymi skrótami operuje.

Dostęp do informacji publicznych zawartych w Biuletynie jest możliwy poprzez stronę główną Biuletynu posiadającą adres URL - www.bip.gov.pl (§ 3 Rozporządzenia). Biorąc pod uwagę powyższy zapis - wydaje się, że "strona www" oraz "strona internetowa" ma ten sam zakres znaczeniowy. Ale w myśl zasad interpretacyjnych: nie można różnym terminom stosowanym przez ustawodawcę przypisywać tego samego znaczenia.

Z rozporządzenia możemy się dowiedzieć nieco o teorii tworzenia stron internetowych. Wiemy już, że strona www może mieć stronę główną. W tym akcie normatywnym zdefiniowano dodatkowo takie terminy jak: menu przedmiotowe - wykaz grup tematycznych, w które są zestawione informacje publiczne udostępniane w Biuletynie, na podstawie ustawy lub ustaw szczególnych, mechanizmy baz danych - elementy oprogramowania serwera pozwalające na gromadzenie, porządkowanie, aktualizowanie i udostępnianie informacji publicznych w Biuletynie Informacji Publicznej, moduł wyszukujący - element oprogramowania serwera umożliwiający znalezienie stron Biuletynu zawierających wyrażenie poszukiwane przez odwiedzającego Biuletyn, moduł administracyjny - element oprogramowania serwera umożliwiający dostęp (po podaniu identyfikatora i hasła dostępu) do Biuletynu w celu dokonania niezbędnych zmian w treści informacji, moduł bezpieczeństwa - programowe lub sprzętowo-programowe rozwiązanie techniczne pozwalające na odseparowanie informacji publicznych zawartych na podmiotowych stronach Biuletynu od sieci Internet, w sposób uniemożliwiający niekontrolowaną penetrację lub eksplorację tych informacji albo ich zniszczenie lub modyfikację przez osoby nieuprawnione. Czy chodzi o firewall? Wiemy czym jest serwer. To komputer, w którym są zgromadzone informacje publiczne udostępniane w Biuletynie.

Rozporządzenie daje również normatywne podstawy dla branży reklamowej i marketingowej, działającej w Internecie. Dzięki rozporządzeniu wiemy, że odwiedzający Biuletyn to osoba poszukującą informacji publicznej w Biuletynie. Termin ten zapewne odpowiada angielskojęzycznemu "unique user"... Daje również podstawy normatywne do prowadzenia statystyk oglądalności serwisów internetowych. Zdefiniowano bowiem termin odwiedziny Biuletynu jako wywołanie strony głównej lub podmiotowej strony Biuletynu... Strona główna Biuletynu zawiera informację o liczbie odwiedzin podmiotowych stron Biuletynu i poszczególnych grup tematycznych w menu przedmiotowym. Biuletyn również powinien zawierać Logo - termin ten wyjaśniono jako "znak graficzny".

Wiemy co to jest adres elektroniczny. Pojęcie to definiuje art. 2 ust. 1 ustawy z dnia 18 lipca 2002 r. o świadczeniu usług drogą elektroniczną. Adres elektroniczny to oznaczenie systemu teleinformatycznego umożliwiające porozumiewanie się za pomocą środków komunikacji elektronicznej, w szczególności poczty elektronicznej. Oznaczenie to będzie miało zatem znaczenie również w przypadku doręczania podań, które jest przewidziane w KPA. Wedle KPA (art. 63. § 1.) podania (żądania, wyjaśnienia, odwołania, zażalenia) mogą być wnoszone pisemnie, telegraficznie lub za pomocą dalekopisu, telefaksu, poczty

elektronicznej, a także ustnie do protokołu. Wedle innych zaś przepisów - za pomocą poczty elektronicznej można zawrzeć umowę z konsumentem: "umowy zawierane z konsumentem bez jednoczesnej obecności obu stron, przy wykorzystaniu środków porozumiewania się na odległość, w szczególności drukowanego lub elektronicznego formularza zamówienia niezaadresowanego lub zaadresowanego, listu seryjnego w postaci drukowanej lub elektronicznej, reklamy prasowej z wydrukowanym formularzem zamówienia, reklamy w postaci elektronicznej, katalogu, telefonu, telefaksu, radia, telewizji, automatycznego urządzenia wywołującego, wizjofonu, wideotekstu, poczty elektronicznej lub innych środków komunikacji elektronicznej w rozumieniu ustawy z dnia 18 lipca 2002 r. o świadczeniu usług drogą elektroniczną (Dz.U. z 2002 r. Nr 144, poz. 1204), są umowami na odległość, jeżeli kontrahentem konsumenta jest przedsiębiorca, który w taki sposób zorganizował swoją działalność" (art. 6. 1. ustawy z dnia 2 marca 2000 r. o ochronie niektórych praw konsumentów oraz o odpowiedzialności za szkodę wyrządzoną przez produkt niebezpieczny). W rozumieniu tych przepisów poczta elektroniczna będzie więc jednym ze środków porozumiewania się na odległość.

Wspomniana wyżej ustawa o świadczeniu usług drogą elektroniczną definiuje system teleinformatyczny - zespół współpracujących ze sobą urządzeń informatycznych i oprogramowania, zapewniający przetwarzanie i przechowywanie, a także wysyłanie i odbieranie danych poprzez sieci telekomunikacyjne za pomocą właściwego dla danego rodzaju sieci urządzenia końcowego w rozumieniu ustawy z dnia 21 lipca 2000 r. - Prawo telekomunikacyjne. Do tej ostatniej ustawy odwołuje się również w przy definiowaniu pojęcia świadczenia usługi drogą elektroniczną. Jest to wykonanie usługi, które następuje przez wysyłanie i odbieranie danych za pomocą systemów teleinformatycznych, na indywidualne żądanie usługobiorcy, bez jednoczesnej obecności stron, przy czym dane te są transmitowane za pośrednictwem sieci publicznych. Wiemy również co to są wspomniane wyżej środki komunikacji elektronicznej. To rozwiązania techniczne, w tym urządzenia teleinformatyczne i współpracujące z nimi narzędzia programowe, umożliwiające indywidualne porozumiewanie się na odległość przy wykorzystaniu transmisji danych między systemami teleinformatycznymi, a w szczególności pocztę elektroniczną. Wiemy, że ustawy o świadczeniu usług drogą elektroniczną nie stosuje się do używania poczty elektronicznej lub innego równorzędnego środka komunikacji elektronicznej między osobami fizycznymi, w celach osobistych niezwiązanych z prowadzoną przez te osoby, chociażby ubocznie, działalnością zarobkową lub wykonywanym przez nie zawodem (art. 3 ust. 2). Jest też zalążek regulacji dotyczącej spamu: Zakazane jest przesyłanie niezamówionej informacji handlowej skierowanej do oznaczonego odbiorcy za pomocą środków komunikacji elektronicznej, w szczególności poczty elektronicznej (art. 10 ust.1). Poczta elektroniczna jest to jednak mniej uprawniona forma komunikacji, gdyż, jak czytamy w ustawie z dnia 10 czerwca 1994 r. o zamówieniach publicznych (art. 21 ust. 3): "Oświadczenia lub zawiadomienia przekazane za pomocą teleksu, poczty elektronicznej lub telefaksu uważa się za złożone w terminie, jeżeli ich treść

dotarła do adresata przed upływem terminu i została niezwłocznie potwierdzona na piśmie przez przekazującego".

Czy termin poczta elektroniczna da się porównać do korespondencji? Do sejmu wpłynął rządowy projekt ustawy o usługach pocztowych (druk sejmowy nr 743). Wedle tego projektu usługę pocztową stanowi miedzy innymi wykonywane w obrocie krajowym lub zagranicznym zarobkowe przyjmowanie i doręczanie przesyłek zawierających korespondencje, lub druki bezadresowe, których treść jest przesyłana do placówki operatora lub doręczana adesatowi za pomocą technik stosowanych w telekomunikacji. Wedle projektu ustawy - korespondencja to informacja utrwalona na dowolnym nośniku, w tym utrwalona pismem wypukłym, z wyłączeniem książek, katalogów, gazet i czasopism lub innych publikacji.

W ustawie z dnia 5 lipca 2002 r. o ochronie niektórych usług świadczonych drogą elektroniczną opartych lub polegających na dostępie warunkowym ustawodawca zawarł kolejne ważne dla internetu definicje: dostęp warunkowy - wszelkie środki oraz przedsięwzięcia techniczne, warunkujące korzystanie z usług chronionych przez indywidualnego usługobiorcę, urządzenia dostępu warunkowego - sprzęt lub oprogramowanie, które zostały zaprojektowane lub przystosowane w celu umożliwienia korzystania z usług chronionych, usługi oparte na dostępie warunkowym - usługi, z których korzystanie jest uzależnione od uprzedniego nabycia przez usługobiorcę urządzenia dostępu warunkowego lub uzyskania indywidualnego upoważnienia dostępu do danej usługi, usługi polegające na dostępie warunkowym - usługi, których przedmiotem jest umożliwienie korzystania z urządzeń dostępu warunkowego, przekaz informacji handlowej - usługi, których przedmiotem jest przekazywanie, w dowolnej formie, informacji służących promocji działalności gospodarczej, w szczególności reklama, marketing bezpośredni, sponsorowanie, promocja sprzedaży i inne akcje promocyjne, urządzenia niedozwolone - sprzęt lub oprogramowanie, które zostały zaprojektowane lub przystosowane w celu umożliwienia korzystania z usług chronionych bez uprzedniego upoważnienia usługodawcy. Usługami chronionymi są: 1) rozpowszechnianie i rozprowadzanie programów telewizyjnych i radiowych w rozumieniu przepisów o radiofonii i telewizji, 2) inne usługi świadczone na indywidualne żądanie usługobiorcy drogą elektroniczną bez jednoczesnej obecności stron - jeśli są świadczone za wynagrodzeniem i oparte na dostępie warunkowym. Przepisy cytowanej ustawy mają olbrzymie znaczenie dla wszystkich serwisów internetowych, które oferują odpłatnie dostęp do usług lub treści, a które na przykład zabezpieczone są za pomocą hasła...

A skoro jesteśmy przy pieniądzach... Ustawa z dnia 12 września 2002 r. o elektronicznych instrumentach płatniczych nie wspomina o Internecie, zawiera zaś definicję elektronicznego instrumentu płatniczego. Elektroniczny instrument płatniczy to każdy instrument płatniczy, w tym z dostępem do środków pieniężnych na odległość, umożliwiający posiadaczowi dokonywanie operacji przy użyciu elektronicznych nośników informacji lub elektroniczną identyfikację posiadacza niezbędną do dokonania operacji, w szczególności kartę płatniczą lub instrument pieniądza elektronicznego" (art. 2 ust. 4 ustawy).

Interesujących definicji dostarcza ustawa z dnia 29 sierpnia 1997 roku o ochronie danych osobowych (Dz.U. 1997 Nr 133 poz. 883 z póź zm.) W rozumieniu ustawy za dane osobowe uważa się wszelkie informacje dotyczące zidentyfikowanej lub możliwej do zidentyfikowania osoby fizycznej i dalej: Osobą możliwą do zidentyfikowania jest osoba, której tożsamość można określić bezpośrednio lub pośrednio, w szczególności przez powołanie się na numer identyfikacyjny albo jeden lub kilka specyficznych czynników określających jej cechy fizyczne, fizjologiczne, umysłowe, ekonomiczne, kulturowe lub społeczne. Informacji nie uważa się za umożliwiającą określenie tożsamości osoby, jeżeli wymagałoby to nadmiernych kosztów, czasu lub działań. Czy dotyczy to adresu elektronicznego lub poczty elektronicznej? System informatyczny w rozumieniu cytowanej ustawy, to zespół współpracujących ze sobą urządzeń, programów, procedur przetwarzania informacji i narzędzi programowych zastosowanych w celu przetwarzania danych. Ustawa definiuje również zabezpieczenie danych w systemie informatycznym. Jest to wdrożenie i eksploatacja stosownych środków technicznych i organizacyjnych zapewniających ochronę danych przed ich nieuprawnionym przetwarzaniem.

Kolejną regulacją, która ma kluczowe znaczenie dla internetu jest Kodeks karny. Kodeks karny wykorzystuje terminy, które, rozpatrując konkretne zdarzenia w Internecie, należy brać pod uwagę: Kto bez uprawnienia uzyskuje informację dla niego nie przeznaczoną, otwierając zamknięte pismo, podłączając się do przewodu służącego do przekazywania informacji lub przełamując elektroniczne, magnetyczne albo inne szczególne jej zabezpieczenie, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 2 (267 § 1 kk). Tej samej karze podlega, kto w celu uzyskania informacji, do której nie jest uprawniony, zakłada lub posługuje się urządzeniem podsłuchowym, wizualnym albo innym urządzeniem specjalnym (§ 2)... Kto, nie będąc do tego uprawnionym, niszczy, uszkadza, usuwa lub zmienia zapis istotnej informacji albo w inny sposób udaremnia lub znacznie utrudnia osobie uprawnionej zapoznanie się z nią, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 2 (Art. 268. § 1). I dalej: Jeżeli czyn określony w § 1 dotyczy zapisu na komputerowym nośniku informacji, sprawca podlega karze pozbawienia wolności do lat 3... Artykuł 269 kk przewiduje, że kto, na komputerowym nośniku informacji, niszczy, uszkadza, usuwa lub zmienia zapis o szczególnym znaczeniu dla obronności kraju, bezpieczeństwa w komunikacji, funkcjonowania administracji rządowej, innego organu państwowego lub administracji samorządowej albo zakłóca lub uniemożliwia automatyczne gromadzenie lub przekazywanie takich informacji, podlega karze pozbawienia wolności od 6 miesięcy do lat 8. Tej samej karze podlega, kto dopuszcza się czynu określonego w § 1, niszcząc albo wymieniając nośnik informacji lub niszcząc albo uszkadzając urządzenie służące automatycznemu przetwarzaniu, gromadzeniu lub przesyłaniu informacji.

W rozumieniu ustawy o ochronie baz danych termin baza danych to zbiór danych lub jakichkolwiek materiałów i elementów zgromadzonych według określonej systematyki lub metody, indywidualnie dostępnych w jakikolwiek sposób, w tym środkami elektronicznymi, wymagający istotnego, co do jakości lub ilości, nakładu inwestycyjnego w celu sporządzenia, weryfikacji lub prezentacji

jego zawartości. Ustawa definiuje również termin pobieranie danych jako stałe lub czasowe przejęcie lub przeniesienie całości lub istotnej, co do jakości lub ilości, części zawartości baz danych na inny nośnik, bez względu na sposób lub formę tego przejęcia lub przeniesienia z zastrzeżeniem wynikającym z art. 3 ustawy. W ustawie znalazł się również termin wtórne wykorzystanie, który oznacza publiczne udostępnienie bazy danych w dowolnej formie, a w szczególności przez rozpowszechnienie, bezpośrednie przekazywanie lub najem z takim samym zastrzeżeniem, że wypożyczenie baz danych nie stanowi pobierania danych lub wtórnego ich wykorzystania.

Jednym z ciekawszych dokumentem jest Załącznik do rozporządzenia Ministra Spraw Wewnętrznych i Administracji z dnia 3 lipca 2002 r. (wydanego na podstawie art. 28 ust 5 ustawy z dnia 6 lipca 2001 r. o gromadzeniu, przetwarzaniu i przekazywaniu informacji kryminalnych) "Wzór strukturalny elektronicznej karty zapytania", w którym jest po prostu fragment kodu XML, wraz ze znacznnikami tego języka oraz napisem "W3C Recommendation for XML version 1.0 Second Edition 6.10". Zresztą takich aktów normatywnych jest więcej: Określenie wzoru rejestru transakcji, sposobu jego prowadzenia oraz trybu dostarczania danych z rejestru Generalnemu Inspektorowi Informacji Finansowej (Rozporządzenie Ministra Finansów z dnia 21 września 2001 r.) opisuje normatywnie strukturę i zasady budowy plików w formacie XML!

Wykazując wielość przepisów, w których polski ustawodawca przewidział wykorzystywanie sieci komputerowych, chciałbym wskazać jeszcze jeden: "Państwowa Komisja Wyborcza może określić zasady i tryb wcześniejszego przekazywania danych z protokołu za pośrednictwem sieci telekomunikacyjnej użytku publicznego lub systemu elektronicznego przesyłania danych". Jest to zapis art. 126 ustawy z dnia 16 lipca 1998 roku ordynacja wyborcza do rad gmin, rad powiatów i sejmików województw. Jak widać internet znalazł się również w ustawach systemowych.

Oczywiście przytoczone powyżej definicje i terminy nie są pełnym wyliczeniem użytych w aktach normatywnych pojęć dotyczących internetu. Stanowią jedynie przykłady.

Prawo internetowe

Zdaniem jednego z moich adwersarzy: "Jeśli mamy prawo drogowe, wodne czy morskie, to możemy tez mieć internetowe. Czyli zasady regulujące działalnością związaną z Internetem". Jego zdaniem szkoda, ze takiego "kodeksu sieciowego" nie mamy.

Ja zaś tak nie uważam. Mamy szereg przepisów (powyżej zasygnalizowałem tylko niektóre) i powstają coraz to nowe. Chociaż to fikcja, prawo jako system jest z założenia spójne. Regulacje prawne dotyczące różnych sfer aktywności mogą dotyczyć również innych sfer. Przykładowo: jeśli publikacja internetowa spełnia kryteria jakimi ustawa prawo prasowe określa prasę - wówczas taka publikacja internetowa będzie podlegała pod prawo prasowe. Podobnie jeśli chodzi o

odszkodowania: art. 415 KC nie mówi ani o prasie, ani o włamaniu do domu: "Ktokolwiek uczynił szkodę drugiemu zobowiązany jest do jej naprawienia". Przepis ten nie wspomina o Internecie, ale może mieć zastosowanie do sytuacji faktycznej polegającej na przykład na opublikowaniu materiałów kompromitujących w Internecie, albo na usunięciu plików z komputera. Idąc dalej: jeśli coś spełnia kryterium utworu - jest chronione przez prawo autorskie niezależnie czy zostało opublikowane drukiem czy na stronie internetowej, w Usenecie czy na IRCu. W takich przypadkach nie ma potrzeby wskazywać internetu jako odrębnego bytu prawnego.

Na tle tego wszystkiego może powstać problem z udowodnieniem różnych faktów. Nie postulujmy jednak tworzenia prawa internetowego, a powstanie regulacji dotyczącej specjalnych zasad zbierania i zabezpieczania elektronicznych materiałów dowodowych, tylko wcześniej powinna powstać spójna normatywna koncepcja dokumentu elektronicznego. Przykładem regulacji dotyczącej dokumentu elektronicznego i możliwości zastosowania go w postępowaniu dowodowym jest zapis art. 34 ustawy z 13 października 1998 r. o systemie ubezpieczeń społecznych: "Informacje zawarte na koncie ubezpieczonego i koncie płatnika składek prowadzonych w formie elektronicznej, które przekazane zostały w postaci dokumentu pisemnego albo elektronicznego, są środkiem dowodowym w postępowaniu administracyjnym i sądowym z zakresu ubezpieczeń społecznych"[4]. Również Kodeks cywilny, po zmianach wynikających z wprowadzenia podpisu elektronicznego przewiduje, że oświadczenie woli złożone w postaci elektronicznej opatrzone bezpiecznym podpisem elektronicznym weryfikowanym przy pomocy ważnego kwalifikowanego certyfikatu jest równoważne formie pisemnej (art. 78. § 2).

Owszem są potrzebne pewne regulacje, ale nie jako kodeks, niekoniecznie jako dedykowana ustawa. Można by postulować zmianę ustawy prawo prasowe, zmianę ustaw dotyczących nadawania programów radiowych i telewizyjnych i jeszcze paru sfer i stworzyć jedno spójne prawo mediów. Taka ustawa nie będzie jednak wstanie regulować wszystkich sfer aktywności w internecie. "Pojęcie 'cyberprzestrzeń' to pułapka dla naiwnych! Tak naprawdę Internet to realna aktywność bardzo wielu ludzi, i tylko ten realny świat się liczy". To jedno ze zdań, jakie padło w internetowej dyskusji na temat "prawa internetowego", toczonej w Usenecie. Należy pamiętać, że poszczególne dziedziny prawne, takie jak prawo do prywatności, wolność słowa, prawo autorskie, nieuczciwa konkurencja, ochrona

---

[4] *nota bene* ustawa posługuję się m.in. określeniem "teletransmisja danych w formie dokumentu elektronicznego z aktualnego programu informatycznego udostępnionego przez Zakład". W celu przekazania dokumentów ubezpieczeniowych w formie elektronicznej do ZUS płatnik musi korzystać z aktualnej wersji programu udostępnianej przez Zakład Ubezpieczeń Społecznych oraz z Internetu przy użyciu komputera zgodnego z wymogami Zakładu. Dokumenty są szyfrowane i przekazywane poprzez konto poczty elektronicznej utworzone przez ZUS dla każdego płatnika (Rozporządzenie Ministra pracy i polityki społecznej z dnia 3 lipca 2001 roku określające zasady przekazywania dokumentów ubezpieczeniowych w formie elektronicznej do ZUS).

znaków towarowych, ochrona nieletnich i wiele, wiele innych, rozwijały się dużo wcześniej, przed powstaniem internetu[5].

Odpowiedzią na wyzwania technologii, dzięki której dzieła, znajdujące się na takich nośnikach jak płyty CD lub bazy danych w systemach komputerowych, mogą być łatwo przekazywane za pomocą Internetu, jest jedna ze zmian wprowadzonych w do ustawy o prawie autorskim i prawach pokrewnych. Powstaje nowe pole eksploatacji: udostępnienie utworu w taki sposób, aby każdy mógł mieć do niego dostęp w miejscu i czasie przez siebie wybranym. Nie ma tu ani słowa o Internecie, a przepis ten będzie się nadawał do ochrony utworów bezprawnie zamieszczanych w globalnej Sieci. Zamiast stosować do internetu odrębny reżim prawny - należy konstruować odpowiednie klauzule generalne, które będą obejmowały również tę sferę aktywności ludzkiej. Niektóre, istniejące i stosowane w "realnym" świecie regulacje również będą miały zastosowanie do internetu. Pozostaje jedynie problem umiejętności dokonania stosownej subsumpcji stanu faktycznego z ustaloną normą i wydanie orzeczenia.

Ponad to, wobec dynamicznego rozwoju techniki komunikowania się, internet cały czas się zmienia. W jego skład wchodzą coraz to nowe technologie, by podać jakiś przykład: m-commerce. Nie da się stworzyć jednolitego (i aktualnego w chwili wydania) prawa, które całościowo regulowałoby problematykę internetu, a co za tym idzie sama normatywna definicja internetu byłaby trudna do skonstruowania.

Dlatego uważam, że odrębnej gałęzi "prawa internetowego" nie należy tworzyć. Nie oznacza to, że aktywność w Internecie nie podlega prawnym regulacjom. Jeśli już ustawodawca postanowił uhonorować internet, definiując pewne terminy - można jedynie postulować, by w poszczególnych aktach normatywnych, poszczególne terminy były definiowane w taki sam sposób. By wymiar sprawiedliwości łatwiej mógł dokonywać interpretacji przepisów oraz stosowania obowiązującego prawa.

Piotr Waglowski
Prawnik, absolwent Wydziału Prawa i Administracji Uniwersytetu Warszawskiego, autor serwisu VaGla.pl Prawo i Internet, Członek Zarządu Internet Society Poland. Jest słuchaczem studiów doktoranckich prowadzonych w Instytucie Nauk Prawnych Polskiej Akademi Nauk, pełni funkcję Głównego Konsultanta Polskiej Izby informatyki i Telekomunikacji.

[5] W 1999 roku Internet obchodził trzydziestolecie swoich narodzin; W 2001 roku internet obchodził 10 lecie powstania w Polsce.

MIND MAPPING APPLICATIONS IN PROJECT MANAGEMENT
ZASTOSOWANIE MIND MAPPINGU W ZARZĄDZANIU PROJEKTAMI

Mariusz PASZKO, Anna OSZCZAK, Roman ZIĘBA

Streszczenie: Mind Mapping jest nową i ekscytującą metodą używania i usprawnienia pamięci, koncentracji i kreatywności poprzez planowanie i tworzenie struktury na wszystkich poziomach z wykorzystaniem kluczowych pojęć. Celem jej jest przyspieszenie zdolności uczenia się, zapamiętywania i zapisywania informacji. Jako proste i bardzo efektywne narzędzie wspomaga pracę kierownika projektu. Poprzez nieliniowe organizowanie informacji pozwala wizualizować naturalny proces twórczego myślenia i pozwala panować nad złożoną naturą współczesnych projektów. Obecnie jest ona stosowana z wielkim sukcesem przez międzynarodowe firmy, renomowane uniwersytety i wyróżniających się artystów.

Summary: Mind Mapping is a new exciting way of using and improving memory, concentration, and creativity through planning and structuring thought on all levels, in order to accelerate the ability to learn, remember, and record information. Thanks to that Mind mapping is one of the simplest yet most powerful tools that a project manager can use. As this creative technique uses a non-linear way of organizing information it enables you to capture the natural flow of your ideas and allows control over the whole complexity of nowadays projects. Currently it is used with great success by multinational organizations, leading universities, and outstanding artists.

„Przeszłość to zawsze prolog."
William Shakespeare

W dzisiejszym świecie wiedza staje się podstawowym zasobem przedsiębiorstwa oraz integruje wszystkie pozostałe zasoby. Wiedza przedsiębiorstwa jest ściśle powiązana z jego zasobami ludzkimi i warunkuje zdolność organizacji do efektywnego działania. Związana jest zawsze z ludźmi i powstaje na styku wiedzy poprzedniej. Bez wiedzy „starej", „złej" nie można stworzyć nowej i lepszej wiedzy. Jednak przedsiębiorstwa często napotykają problemy związane z wykorzystaniem, kontrolowaniem, gromadzeniem i tworzeniem nowej wiedzy, której nośnikiem są pracownicy oraz otaczający świat.

To potęga umysłu pracowników, ich szybkość uczenia się i dostosowywania się do nowych sytuacji i realiów, efektywność komunikowania się i przekazywania posiadanej wiedzy mają decydujące znaczenie w kreowaniu i podnoszeniu wartości firmy, dostosowywaniu się przedsiębiorstwa do zmieniającego się otoczenia i umiejętnego korzystania z nadarzających się okazji. Ona tworzy przewagę konkurencyjną, którą trudno jest pokonać i skopiować. Trudność polega jednak na tym, że każdy z pracowników musi nauczyć się jak korzystać z posiadanych informacji, jak je przekazać w sposób komunikatywny i prosty, by stały się wiedzą, tzn. pozwolą innym uczyć się i wykorzystywać je w swojej pracy. Dlatego konieczne jest odejście od postrzegania mózgu ludzkiego jedynie jako możliwej do zaprogramowania maszyny, na rzecz traktowania go również jako dynamicznego i uporządkowanego systemu nerwowego o ogromnych możliwościach zapamiętywania, przetwarzania informacji i kreowania nowej wiedzy. Ważne jest także zrozumienie, że wiedza może istnieć nie tylko w umyśle człowieka, lecz także w zespole. W praktyce oznacza to, że żaden człowiek nie posiada całej wiedzy na temat rozwiązania danego problemu, lecz jako członek zespołu jest w stanie przyczynić się do rozwiązania zadania.

Jednakże, szacuje się, ze 95% ludzi jest niezadowolonych lub ma zastrzeżenia do swojego „umysłowego funkcjonowania". Narzekamy głownie na brak koncentracji i kreatywności, trudności w zapamiętywaniu i uczeniu się. Konieczne zatem jest wprowadzenie takich technik i metodologii pracy, które poprzez oparcie się na naturalnych procesach przyswajania wiedzy i pracy mózgu pozwolą i wzmocnią procesy tworzenia, wykorzystania i dzielenia się wiedzą. Nie można jednak nigdy zapominać o pozytywnym wpływie odpoczynku oraz czynników zewnętrznych na stan i funkcjonowanie naszego umysłu.

Mind Mapping – twórcy i założenia

Pod koniec lat 60 zagadnieniem procesu zapamiętywania, uczenia się, postrzegania i przyswajania informacji zainteresował się angielski uczony Tony Buzan. Opracował on metodę pozwalającą wykorzystać potęgę umysłu i nazwał ją *Mind Mapping*. Pozwala ona zrozumieć jak pracuje nasz mózg, nauczyć się jak się uczyć i zapamiętywać, jak myśleć kreatywnie. Opiera się na wynikach badań dotyczących fizjologicznej, psychologicznej i społecznej strony funkcjonowania mózgu.

W latach 80' amerykański profesor Roger Sperry, uhonorowany Nagrodą Nobla w 1981 roku w dziedzinie medycyny za wkład w naukę o funkcjonowaniu kory mózgowej, dowiódł, że mózg każdego człowieka składa się z dwóch części, działających na wyższych poziomach fizjologicznych, z których każda charakteryzuje się innymi funkcjami mentalnymi:

- Lewa odpowiada za: słowa, logikę, cyfry, zdania, linearność, analizę, tworzenie list
- Prawa zaś za: rytm, zmysł przestrzeni, syntezę, wyobraźnię, marzenia, kolory, wymiary.

Proces uczenia się i zapamiętywania przebiega efektywnie i szybko, gdy obie półkule „zmuszone" są do pracy. Ponadto, najlepiej zapamiętujemy informacje, które docierają do nas na początku i końcu, są powiązane z czymś, co nas interesuje, co znamy lub jest w jakiś sposób wyróżnione. Umysł człowieka dąży także do zobaczenia ustrukturalizowanej całości. Nie lubi rzeczy nie dokończonych i nie uporządkowanych. Ponadto, myśli słowami kluczowymi, skojarzeniami, hasłami i obrazami.

Pomimo, iż fakty dotyczące pracy mózgu odkryto i udowodniono dopiero w latach osiemdziesiątych i dziewięćdziesiątych, to dowodzą one niezbicie skuteczności wcześniej powstałej metodologii *Mind Mappingu*.

Mind Mapping zmusza niejako do pracy obie półkule mózgu wykorzystując powiązania, wyróżnienie, strukturę, kolor i tworzenie obrazu całości. Jest to usturkturalizowana prezentacja idei, czy zagadnienia. W centrum umieszczone jest kluczowe słowo lub słowa, od niego promieniście rozchodzą się idee powiązane ze słowem – hasłem.

Metoda korzysta z szerokiego zakresu wizualnych pomocy, jak: kolor, symbol, obraz, wymiar, linia, forma, faktura, tło, czy też dźwięk lub animacja, gdy korzystamy ze specjalnego oprogramowania.

Każda mind mapa jest unikalna i jako niepowtarzalna całość zapisywana jest przez pamięć wzrokową. Różnorodność ułatwia zapamiętywanie i odtwarzanie mapy. Obrazy i symbole są bardziej wyraziste, bogatsze w treść i precyzyjniejsze niż słowa. Uważa się, ze jeden obraz wart jest więcej niż 1000 słów. W rezultacie, rysunki wyzwalają i wspomagają kreatywne myślenie, ułatwiają szybkie zapamiętanie wielu informacji.

Wciąż jednak, w szkole, na studiach czy w pracy dominują słowa pisane linearnie. Współcześnie dąży się, wbrew wynikom badań naukowych, aby to słowa, a nie obrazy stały się głównym przekazem informacji. Nie doceniamy mądrości, jaka zostawiły nam ludzie z czasów prehistorycznych – o ich życiu wiemy, nie z zapisów słownych, ale z rysunków i malowideł ściennych.

Jednakże, od kilkunastu lat techniki Mind Mappingu są znane i z powodzeniem stosowane w Stanach Zjednoczonych, czy krajach Europy Zachodniej. Również w Polsce zajęcia i szkolenia z tego zakresu prowadzone są już od 12 lat. Na co dzień, metodologia *Mind Mappingu* wykorzystywana jest m.in. w firmach takich jak: IBM, BP, Boeing, Electronic Data Systems, Hewlett-Packard, General Motors oraz wielu innych (szersza lista referencyjna na końcu referatu).

Zastosowanie

Mind Mapping generalnie znajduje zastosowanie w trzech sferach:

- biznesie
- edukacji
- życiu prywatnym

Świat biznesu wykorzystuje *Mind Mapping* do tworzenia projektów i biznes planów, przygotowania prezentacji, umów handlowych, raportów, czy zarządzania czasem poprzez plany dnia, tygodnia, miesiąca. Jest to także genialne narzędzie do przeprowadzenia burzy mózgów, prowadzenia negocjacji, robienia notatek oraz rozwiązywania problemów. W sposób jasny i przejrzysty pozwala przedstawić strukturę organizacji, problemu, czy strony internetowej. Jest niezastąpione przy organizacji spotkań - przedstawienie harmonogramu, kluczowych tematów i problemów oraz zapis przebiegu i podsumowanie spotkania oraz wkładu każdego uczestnika. Dzięki usprawnieniu i uproszczeniu komunikacji międzypersonalnej poprzez skonkretyzowanie problemów, jasne przypisanie obowiązków i odpowiedzialności, ujednolicenie stosowanych pojęć stanowi doskonałą i efektywną metodologię zarządzania zasobami ludzkimi.

W edukacji technika ta znajduje także szerokie zastosowanie. Wykład przygotowany w formie mind mapy jest o wiele bardziej czytelny niż długi monolog, czy tekst. Główna idea jest jasno zdefiniowana i umieszczona w centrum, a zależności między podrzędnymi ideami wskazane są poprzez odległość od środka oraz linie je łączące. Forma mind mapy pozwala również na proste dodawanie nowych pomysłów bez niepotrzebnego wymazywania, czy wykreślania poprzednio zapisanych informacji.

Mind mapa przydatna jest także przy nauce języków obcych, robienia notatek z lektur, wykładów i prezentacji, a także przy wygłaszaniu przemówień – plan, główne punkty i ich struktura i kolejność, dostępność dużej ilości informacji.

Mind Mapping pomaga studentom i profesorom zapamiętywać i przywoływać informacje, upraszcza pracę z tekstem, czy pisanie zaliczeń. Ułatwia także przyswajanie wiedzy skracając czas potrzebny do nauki poprzez eliminacje ponownego czytania, konsolidowania wiedzy z różnych źródeł, czy szukania kluczowych problemów. Eliminuje niepotrzebne słowa i tworzy relacje i powiązania miedzy kluczowi ideami. Mind mapa może prezentować ogromną ilość informacji na jednej stronie poprze odpowiednie użycie rysunków, kluczowych słów, powiązań miedzy nimi i stworzenie całości rozwiązań tematu, czy problemu. *Mind Mapping* wymaga jednak od użytkownika uważnego czytania i przemyślenia procesu tworzenia mapy, zmusza do bycia skoncentrowanym i precyzyjnego analizowania zagadnienia.

Mind Mapping oferuje nowe interesujące sposoby używania i usprawnienia pamięci, koncentracji, czy kreatywności. Dzięki swoim zaletom powoduje wzrost produktywności i oszczędza czas, co jest tak ważne dla rozwoju każdej organizacji.

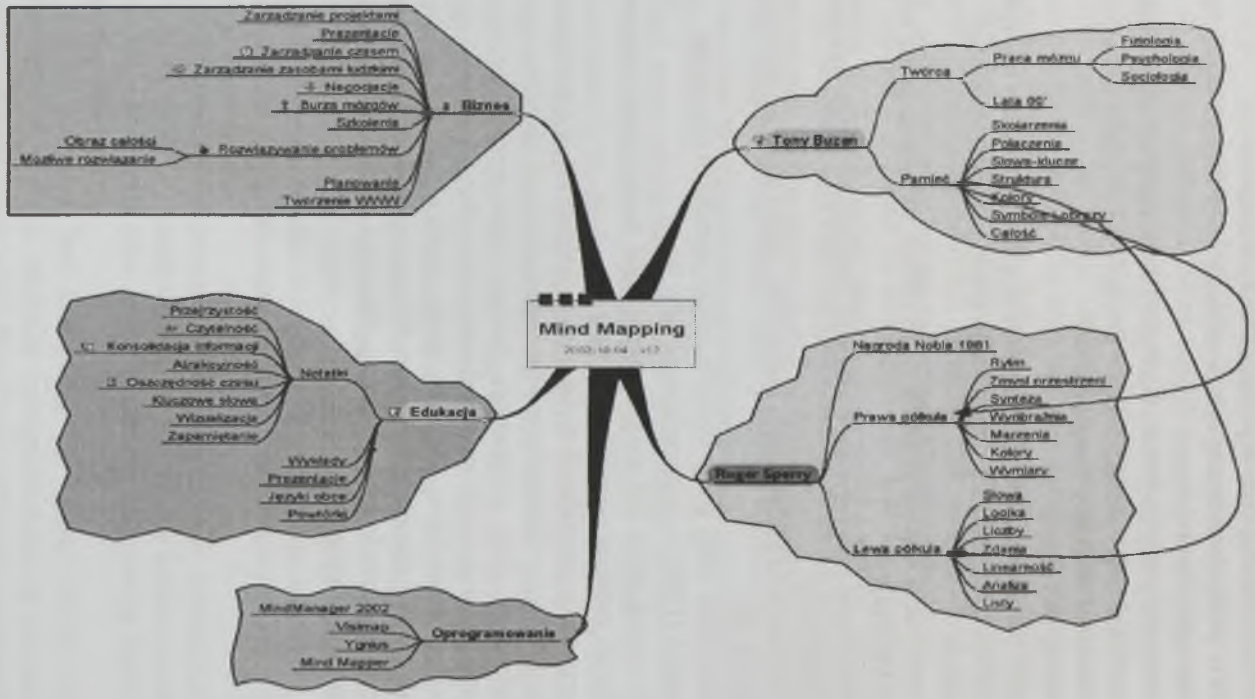
***Mind Mapping* i komputer**

Jako, że człowiek jest kreatywną, ale często leniwą istotą, a świat komputerów stale się rozwija, specjalne oprogramowanie do tworzenia mind map zostało zaprojektowane przez wiele firm. Poza walorami papierowej wersji mind mapy, komputerowa mind mapa stwarza o wiele więcej możliwości. Jednym z najbardziej funkcjonalnych programów wspierających metody *Mind Mappingu* jest MindManager 2002 firmy Mindjet. Program, oprócz możliwości

poprawiania i tworzenia od nowa mind mapy, podświetlania kluczowych haseł, czy dodawania komentarzy ukrytych w plikach, pozwala także na:

- Eksport i pełną dwustronną synchronizację z MS Project-bardzo przydatne przy zarządzaniu projektami
- Eksport mind mapy do MS PowerPoint i stworzenie szybkiej prezentacji
- Eksport mind mapy do MS Word – przygotowanie dokumentacji z zachowaniem struktury i nagłówków
- Eksport do MS Excel
- Eksport i pełną dwustronną synchronizację z MS Outlook- tworzenie planów dnia, miesiąca itp., dodawanie zadań i kontrolowanie ich wykonania
- Zapewnia pełną kompatybilność z MS Office, otrzymuje pełne wsparcie ze strony Microsoftu
- Zainstalowanie programu na większości palmtopów i jego sprawne funkcjonowania
- Wykorzystanie kilkudziesięciu szablonów: plany dnia, zarządzanie projektami itp.
- Wykorzystanie całej gamy kolorów, zdjęć, symboli i obrazków
- Eksport do HTML – tworzenie stron WWW
- Równoczesną pracę 100 osób na jednaj mind mapie dzięki serwerowi konferencyjnemu w wersji Enterprise
- Wykorzystanie specjalnych trybów wyświetlania mapy i poziomów szczegółowości wykorzystanie własnego modułu prezentacyjnego

MindManager jest programem bardzo elastycznym, intuicyjnym w obsłudze, a mapy generowane przez wbudowane algorytmy są najczęściej zgodne z oczekiwaniami użytkowników. Odciąża także użytkownika od wielu błahych problemów, jak choćby planowanie miejsca na nowe gałęzie, poprawianie błędów. Pozwala także na tworzenie relacji miedzy gałęziami, dodawanie komentarzy do gałęzi, formatowanie czcionek itp.



Rys. 1.Mind Mapping

Czym jest zarządzanie projektami?

Zarządzanie projektami (Project Management) to zbiór zasad, metod i technik zarządzania stosowanych w projekcie dla osiągnięciu zadowolenia klienta[1], które rozumie się jako zrealizowanie określonego celu w założonym czasie i w ramach przydzielonego budżetu.

Ogólny zbiór zasad PM został skodyfikowany przez Project Management Institute w aktualizowanym PMBOK (Project Mgmt Body of Knowledge). Dla potrzeb tego artykułu posłużono się metodyką zawartą w PMBOK.

Niektórzy rozszerzają domenę PM przydając jej atrybuty dziedziny naukowej czy nowej filozofii zarządzania (zarządzanie przez projekty). Pozostaje na pewno PM szeroką, interdyscyplinarną wiedzą, łączącą wiele dziedzin od psychologii po statystykę. Szczególny nacisk położony jest w niej na jasny opis celu projektu przez przyjęcie dokładnych kryteriów jego osiągnięcia. Kierownik projektu pracuje zawsze w ramach dynamicznego trójkąta zależności: czasu, budżetu i zakresu. Wymagania odnośnie terminów często bywają bezlitosne. We współczesnym, zmiennym otoczeniu biznesu szybkość odpowiedzi na potrzeby klienta, nacisk na innowacyjność i efektywne zarządzanie kosztami decydują o sukcesie bądź porażce w walce o pozycję na rynku. To wymusza na współczesnych organizacjach wprowadzanie zmian: coraz częstszych i na coraz większą skalę. Dla rosnącej grupy firm i organizacji strategią odpowiedzi na te nowe wyzwania staje się stosowanie zasad zarządzana projektami.

Doświadczenia z projektów różnej skali, realizowanych w różnych środowiskach oraz rosnąca złożoność i skala współczesnych przedsięwzięć zrodziły potrzebę usystematyzowania wiedzy dotyczącej zarządzania projektami..

Metodyki i narzędzia projektowe

Jakkolwiek współczesny Project Management jest dyscypliną stosunkową młodą, liczącą nieledwie dwie dekady, to część organizacji wypracowała już własne, kompleksowe metodyki projektowe, przystosowane najczęściej do konkretnych środowisk, w jakich realizują one własne przedsięwzięcia. Własną metodyką pochwalić się może IBM, Oracle czy Microsoft. Z Wysp Brytyjskich wywodzi się metodyka PRINCE2, a ogólny zbiór zasad PM, jak to już wcześniej zostało wspomniane, skodyfikował Project Management Institute w aktualizowanym PMBOK (Project Mgmt Body of Knowledge)

Nie ma potrzeby porównywać wymienionych metod ani oceniać ich wartości. Podlegają one ciągłym modyfikacjom wraz z zachodzącymi zmianami w świecie ekonomii, gospodarki i życia społecznego oraz nabywaniem nowych

---

[1] W referacie posłużono się metodyką projektową oraz definicjami zaczerpniętymi z *Project Management Body of Knowledge*, publikacji Project Management Institute, ©PMBOK 2000.

doświadczeń przez zespoły projektowe. Szczególnie wyraźne jest to na przykładzie młodego i niezwykle dynamicznie rozwijającego się rynku usług i produktów informatycznych, gdzie w krótkich odstępach czasu zachodzą kolejne rewolucje technologiczne. Równolegle do prac nad metodykami projektowymi trwały również poszukiwania najlepszych dróg propagowania zgromadzonej i zagregowanej wiedzy.

W ten sposób rozwinęła się kolejna dyscyplina zarządzania tzw. *Knowledge Management*. Jednym ze strategicznych nośników wiedzy stała się technologia informatyczna - szczególnie pomocna wszędzie tam, gdzie w pracach projektowych wykorzystywane są różnorodne aplikacje komputerowe.

Wszystkie metodyki, – mimo że obrosły własnymi terminologiami, standardami i szablonami dokumentów, wywodzą się z jednego zbioru kardynalnych zasad zarządzania projektami. Wszak podstawowymi narzędziami pracy kierownika projektu pozostaje ołówek i gumka.

Kierownik projektu ma dziś do dyspozycji szereg specjalistycznych narzędzi ułatwiających pracę w chaosie, jaki stanowi projekt z jego złożoną strukturą, zmiennymi wymaganiami, burzliwym otoczeniem i nierzadko konfliktami wewnątrz organizacji.

Zarządzanie czasem (harmonogramem) ułatwia PERT, metoda ścieżki krytycznej czy wykresy Gantta. Techniki estymacji i analiza wartości uzyskanej (Earned Value) wspierają kierownika projektu w zarządzaniu kosztami projektu (budżetem). Z kolei zarządzanie zakresem (specyfikacje) odbywa się z pomocą technik definiowania projektu oraz procedur zarządzania zmianami. Powstało szereg aplikacji komputerowych wspomagających zarządzanie projektami, wśród nich największą popularność zdobył program Microsoft Project.

Specyfika współczesnych projektów

Mimo stosowania coraz bardziej zaawansowanych technik i narzędzi wspomagających prace w projekcie ryzyko niepowodzenia pozostaje wysokie, o czym stale donoszą alarmujące statystyki. Dzieje się tak za sprawą niezwykłej złożoności współcześnie realizowanych projektów oraz dużej zmienności środowiska, w jakim są przeprowadzane. Mimo stosowania zaawansowanych technik zarządzania ryzykiem nie sposób przewidzieć obecnie wszystkich sytuacji, które mogą zagrozić uzyskaniu zaplanowanego rezultatu w założonym czasie i budżecie. Współczesny kierownik projektu wie doskonale, że o sukcesie projektu nie decyduje sama precyzyjna estymacja pracy, racjonalny harmonogram i dostęp do zasobów.

Ważną cechą współcześnie realizowanych projektów jest to, że prowadzone są przez zespoły powoływane jedynie na czas ich trwania spośród różnych pionów funkcjonalnych organizacji. To rodzi konflikty kompetencyjne i problemy w pozyskiwaniu zasobów do prac w projekcie.

Kierownicy projektów wymieniają najczęściej następujące przyczyny niepowodzeń projektów:

1. Brak rzeczywistego wsparcia ze strony sponsorów
2. Brak szybkiej reakcji udziałowców projektu na zagrożenia
3. Niewłaściwie sporządzona specyfikacja wymagań (definicja projektu)
4. Nierealistyczne wymagania klienta
5. Nieefektywna komunikacja w projekcie

Jak wynika z powyższej listy, przyczyny niepowodzeń projektów najczęściej nie są natury technicznej, ale wynikają z błędów w ocenie sytuacji i braku efektywnej komunikacji, czyli stanowią tzw. „czynnik ludzki".

O sukcesie projektu decyduje zatem skuteczne zarządzanie przepływem informacji, dostrzeganie na czas i właściwa ocena powstających zagrożeń i szans oraz szybkie komunikowanie idei pomiędzy udziałowcami projektu, zespołem projektowym i klientem, odbiorcą projektu. Niezbędną cechą kierownika projektu jest, zatem kreatywność, jednak rozumiana nie tyle jako wymyślanie nowych idei, lecz kreatywność jako umiejętność łączenia faktów, analizowania ich powiązań i tworzenia na ich bazie nowych rozwiązań, lepiej odpowiadających na bieżące potrzeby w projekcie.

Podsumowując można przypisać projektom następujące cechy:

1. Są bytami wysoce złożonymi
2. Przebiegają w zmiennym środowisku
3. Wymagają twórczego wysiłku
4. Przebiegają w wielu wymiarach

Dobór ludzi do projektu i metodyki doboru

Obecnie, kapitał, technologia, wyposażenie techniczne i strategia, adekwatne do celów projektu, nie gwarantują powodzenia realizacji projektu. Elementem decydującym staje się czynnik ludzki. Stworzenie odpowiedniego zespołu projektowego to jeden z ważniejszych etapów PM.

„Nie wszystkie grupy to od razu zespoły projektowe. Zespół to grupa ludzi związanych ze sobą i zaangażowanych do osiągnięcia wspólnych celów i wyprodukowania przedmiotów odbioru wysokiej jakości....nie do administracji i zarządzania."

Vijay Verma 1997

Tworzenie zespołu jest procesem transformacji grupy osób, które mogą posiadać różne zainteresowania, wykształcenie i doświadczenie w zintegrowany i efektywny zespół roboczy. Najpierw jednak należy wyselekcjonować odpowiednie osoby i przydzielić im zadania i obowiązki w taki sposób, aby naturalne predyspozycje człowieka odpowiadały konkretnym zadaniom w projekcie.

Jest wiele znanych metod doboru i selekcji pracowników, jak np. Predictive Index, OPQ, Meurs, metoda Thomas International. Wszystkie one określają potencjał i optymalne dopasowanie zasobów ludzkich do wymagań danego stanowiska pracy, czy zadania w projekcie.

Metoda Thomas International jest efektem wieloletniej pracy psychologów. Dowiedziono, że człowiek w zależności od środowiska pracy – antagonistycznego lub sprzyjającego, może przyjmować różne postawy: aktywną lub pasywną (tzw. teoria Marstona). W 1956 r. amerykański psycholog Thomas Hendricson oparłszy się na ww. teorii, określił, iż postawy te są uzależnione od cech osobowości i predyspozycji danej osoby. Początkowo metodę Thomasa wykorzystywał Pentagon, jednak od lat 70 zaczyna być wykorzystywana na całym świecie.

Metoda składa się z profesjogramu, czyli analizy profilu etatu: wymagane zachowanie członka zespołu i cechy osobowości, oraz z analizy profilu osobowego opisującego predyspozycje zawodowe ze szczególnym uwzględnieniem silnych stron badanej osoby. Charakter ludzi analizowany jest w czterech płaszczyznach: dominacji, komunikatywności, stabilizacji i adaptacji. Porównanie dwóch źródeł informacji: profesjogramu i analizy profilu osobowościowego może być podstawą do decyzji kadrowych.

Najczęstsze zastosowania metody Thomasa to:

- Analiza Profilu Osobowego
- Rozmowy okresowe
- Czynniki motywujące
- Rekrutacja i wybór
- Dokształcanie i doskonalenie pracowników
- Dobór pracowników w zespole
- Profil kierownictwa przedsiębiorstwa lub projektu
- Indywidualny styl kierowania i zarządzania
- Zachowanie w sytuacjach pod presją
- Konfliktowość w stosunkach międzyludzkich
- Wydajność
- Reorganizacja bądź restrukturyzacja

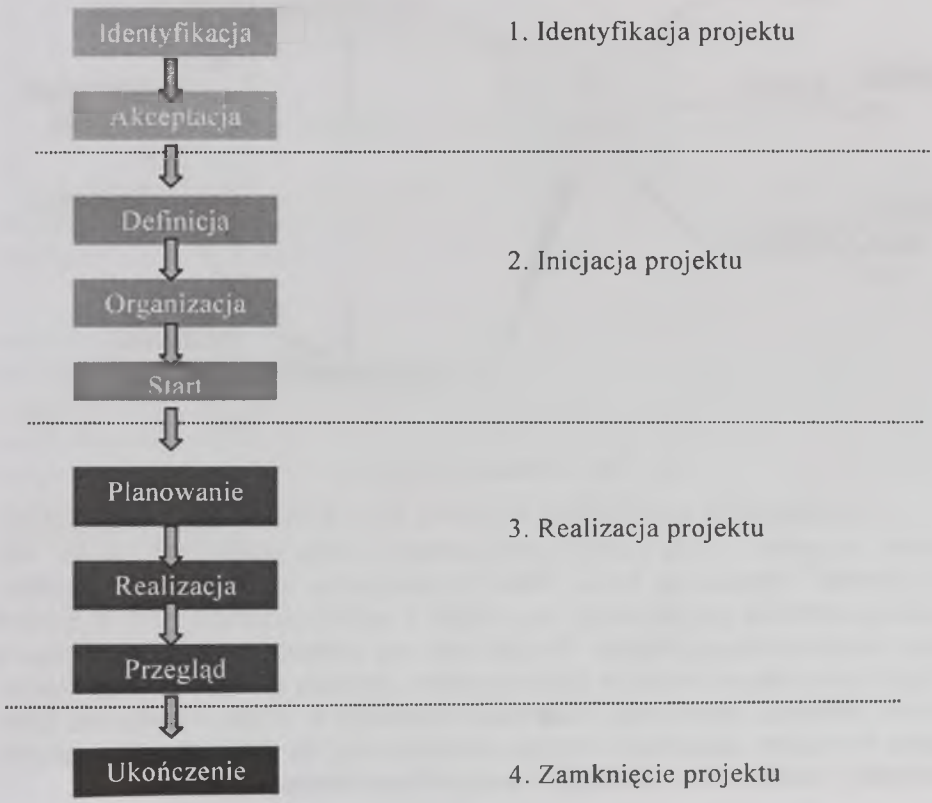
Projekt jako żywy organizm

Rzeczywistość projektowa prawie nigdy nie odpowiada przyjętemu planowi. Konieczna jest stała i bieżąca kontrola oraz aktualizacja planu projektu w stosunku do planu bazowego, zdefiniowanego w fazie inicjacji projektu. Efektywne zarządzanie sytuacjami wyjątkowymi (ryzykiem, zmianami, problemami i błędami) zapewnia właściwą reakcję na spodziewane lub już zaistniałe zaburzenia. Regularne spotkania zespołu projektowego i raporty z przebiegu prac, jak również wewnętrzne i zewnętrzne kontrole projektu dostarczają zweryfikowanej wiedzy o rzeczywistych nakładach i postępach prac, stanie budżetu, jakości produktów projektu oraz bieżącym stanie analizy sytuacji wyjątkowych. Projekt należy zatem traktować jak żywy organizm, który musi szybko reagować na bodźce płynące ze swojego otoczenia, w którym informacje krążą niczym decydujące o życiu płyny ustrojowe.

Podejście linearne a drzewo powiązań

Projekt z definicji stanowi sytuację wyjątkową, wydzieloną w czasie, mającą swój początek i koniec. Stanowi zatem w pewnym sensie przeciwieństwo procesu z jego rutynową, powtarzalną sekwencją czynności. Jakkolwiek mówi się o procesach w ramach poszczególnych etapów cyklu życia projektu (procesy inicjacji, procesy planowania, etc.) postrzeganie projektu w kategoriach procesów rodzi ryzyko popadnięcia w rutynę i automatycznego naginania rzeczywistości projektowej do raz wypracowanego schematu. Logiczna i „sprawdzona" iteracja czynności jest oczywiście ułatwieniem i stanowi podstawę większości metodyk projektowych, niemniej stanowi duże ograniczenie i utrudnienie w całościowym postrzeganiu pełnej złożoności projektu.

Lepszym podejściem byłoby postrzeganie holistyczne projektu jako jednego dużego obrazu, na którym rysuje się cała złożoność elementów przedsięwzięcia, ich wzajemnych powiązań i to we wszystkich wymiarach rzeczywistości projektowej. Punktem centralnym tego obrazu powinien być cel projektu - jasno określony i dokładnie rozumiany przez wszystkich jego uczestników.

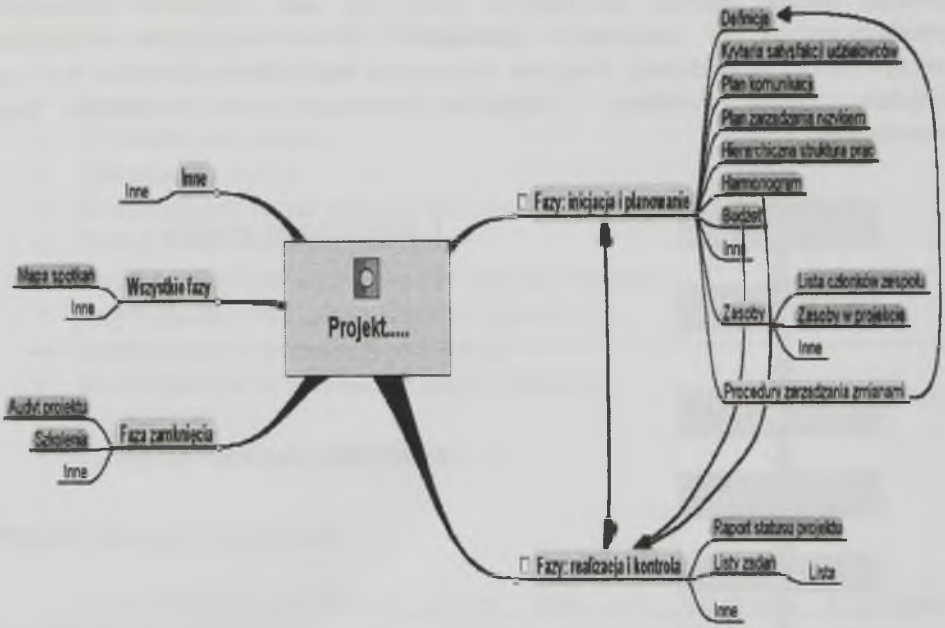


Rys.2. Przykład linearnego myślenia w projekcie: Projekt jako iteracja faz

Mind Mappig a zarządzanie projektami

Mind Mapping jest prostym i jednocześnie bardzo efektywnym narzędziem wspomagającym pracę kierownika projektu. Ta kreatywna technika pozwala organizować informację w sposób nieliniowy, zapewniając swobodny i naturalny przepływ idei i umożliwiając łatwe nimi zarządzanie oraz ich zapis.

Mind mapa pozwala prezentować strukturę problemu za pomocą prostego obrazu, na którym organizacja myśli jest czytelna i pełna.
„Gdy masz przed oczami cały obraz zagadnienia, zaczynasz spontanicznie tworzyć własne powiązania, które wcześniej nie przyszłyby ci do głowy" stwierdza Hobart Swan, pracownik Mindjet LLC, firmy tworzącej oprogramowanie Mind Manager 2002 do tworzenia mind map.



Rys.3. Mind mapa projektu

W mind mapie wykorzystuje się słowa, linie, symbole, skojarzenia, kolory, kształty, wymiary i całą możliwą różnorodność form graficznych po to, aby maksymalnie stymulować mózg. Cała informacja na temat definicji projektu, przebieg spotkania projektowego czy raport z audytu organizuje się w postaci mapy powiązanych rozgałęzień. Projekt staje się centralnym punktem obrazu, z którego rozchodzą się konary w postaci planów, struktury prac, kosztów, zasobów. Drzewo pozwala organizować wszystkie informacje w układ rozgałęzień, które można wyróżniać, grupować i łączyć, skupiając się na detalach najmniejszych elementów i jednocześnie nie tracąc z oczu pełnego obrazu projektu.

Korzyści z *Mind Mappingu* w zarządzaniu projektami

Mind mapping ma bardzo szerokie zastosowanie w wielu obszarach zarządzania projektami. Oprócz tego, że jest niezastąpionym narzędziem do budowania struktury projektu, stymulatorem kreatywności podczas sesji burzy mózgów, systemem prowadzenia spotkań czy sposobem na formatowanie dokumentów projektowych, *Mind Mapping* skutecznie wspomaga proces uczenia się i zapamiętywania, przesuwając znacząco krzywą uczenia się. Przyjęcie systemu *Mind Mappingu* w prowadzeniu projektu pozwala jednocześnie łatwiej zapamiętywać gromadzone informacje, jak również efektywniej je wykorzystywać.

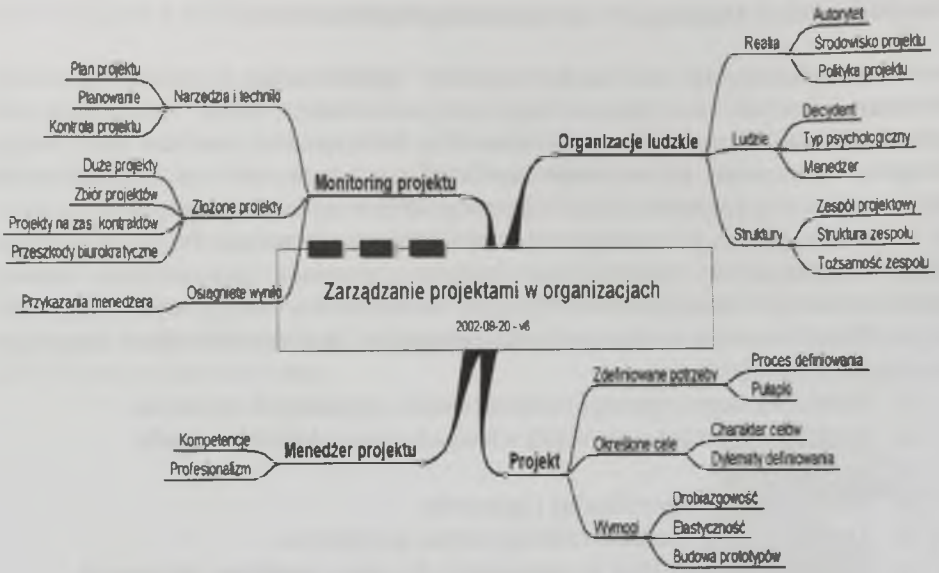
Podsumowując można wskazać następujące zastosowania *Mind Mappingu* w zarządzaniu projektami:

1. Właściwa ocena sytuacji i dokonywanie optymalnych wyborów
2. Szybkie organizowanie myśli własnych oraz członków zespołu projektowego
3. Kreatywność indywidualna i grupowa
4. Definiowanie, analiza i rozwiązywanie problemów
5. Redefiniowanie celów w odniesieniu do czasu, zasobów, zmiennych wymagań
6. Wspomaganie pamięci i procesu uczenia się
7. Zapewnienie efektywnej komunikacji w zespole projektowym.

Przykłady zastosowań *Mind Mappingu* w zarządzaniu projektami

Mind mapping stosowany jest z powodzeniem jako narzędzie wspomagające zarządzanie projektami przez wiele firm na całym świecie. Dość wymienić British Petroleum, Boeing, Xerox, Oracle.

Ciekawym przykładem zastosowania *Mind Mappingu* był projekt w zakładach lotniczych Boeing. Kierownik projektu Mike Stanley zmienił gruby podręcznik procedur jakości na półtorametrową mind mapę. Stosowany jako integralne narzędzie w programie identyfikowania projektów naprawy jakości mind *mapping* pozwolił 10-krotnie obniżyć zakładane koszty wprowadzenia programu. W ciągu miesiąca dzięki technice *Mind Mappingu* zidentyfikowano ponad 500 projektów naprawy jakości, co przyczyniło się do oszczędności rzędu 10 milionów dolarów.



Rys. 4. Wykorzystanie Mind Mappingu w zarządzaniu projektami

Ograniczenia *Mind Mappingu* w zarządzaniu projektami

Podstawowym zarzutem przeciwników *Mind Mappingu* w zarządzaniu projektami jest to, że jako metoda nieliniowa nie zapewnia wystarczającej kontroli czasowej nad zadaniami. Inny zarzut to zbytnie uproszczenie analizy projektu. Nie stanowiąc odpowiedzi na wszystkie potrzeby kierownika projektu jest *Mind Mapping* raczej uzupełnieniem pozostałych, bardziej specjalistycznych narzędzi, jakie ma do dyspozycji kierownik projektu. Stosując *Mind Mapping* należy także uniknąć błędu odstąpienia od stosowanej metodyki. Umiejętne włączenie technik *Mind Mappingu* do zarządzania projektem oraz przyjęcie tej metody przez wszystkich uczestników projektu jest warunkiem wykorzystania wszystkich jego zalet.

Podsumowanie

Mind Mapping, metoda twórczego stymulowania umysłu, opracowana przez Tonego Buzana, jest efektywną techniką organizowania informacji, zarządzania czasem i zasobami ludzkimi. Dobrze skonstruowana mind mapa może zawierać na jednej dużej kartce papieru wszystkie istotne informacje z kilkudziesięciu stron tekstu. Pozwala to na błyskawiczny dostęp do nich oraz ich przejrzystą prezentację.

Jest skutecznym i sprawdzonym narzędziem wspomagającym zarządzanie projektami. Najlepiej sprawdza się tam, gdzie potrzebna jest wielostronna analiza zagadnienia i kreatywne myślenie całego zespołu. Wspomaga proces komunikowania złożonych treści oraz ułatwia zapamiętywanie informacji. Mimo pewnych ograniczeń stosowanie *Mind Mappingu* w projektach dowiodło jego dużej wartości jako narzędzia umożliwiającego pełne zobrazowanie całej złożoności projektu we wszystkich jego wymiarach.

Lista referencyjna – firmy wykorzystujące metody Mind Mappingu

- Na świecie:
  AEG, Accenture, Alcatel SEL, Allianz, American Express, AT&T, Autodesk, Bank of America Corp., Bank One Corp., BASF, Bayer AG, Bayerische Landesbank, Bertelsmann, BMW, Boeing, British Airways, Boehringer Ingelheim Pharma KG, Caterpillar, Cap Gemini Ernst & Young, Coca-Cola, Compaq Computer, DaimlerChrysler, Dell Computer, DeTeCSM GmbH, Deutsche Bahn, Deutsche Bank, Deutsche Post, Deutsche Telekom, Dresdner Bank, E.ON, Eastman Kodak, Ford Motor, FedEx, Ford Motor Company, Fraunhofer Gesellschaft, General Electric, General Motors, Gillette, Goldman Sachs Group, Heidelberger Druckmaschinen AG, Henkel, Hewlett-Packard, Honeywell, Intel, IBM, Ingram Micro Inc., Itellium, Johnson & Johnson, KarstadtQuelle, Lufthansa Group, MAN, Merryll Lynch, Metro, Munich Re Group, Microsoft, Motorola Inc., NASA, Nike, Nokia Corporation, Otto Versand, Oracle, Procter and Gamble, Preussag, ProAlpha AG, Robert Bosch, Roche Diagnostics GmbH, Rödl und Partner, RWE DEA AG, SchwanStabilo, Siemens, Sonax, SAP, Sun Microsystems, ThyssenKrupp, Unilever, Visa InternationalTecis Ecom AG, Vitra AG, Volkswagen, Walt Disney...
- W Polsce:
  Towarzystwo Ubezpieczeniowe Commercial Union, Towarzystwo Ubezpieczeniowe Credit Suisse, Towarzystwo Ubezpieczeniowe Allianz Polska, PriceWoterhouseCoopers, Beskidzki Fundusz Inwestycyjny, Specjalistyczna Przychodnia Lekarska dla Pracowników Wojska, Nestle Polska, Polskie Koleje Państwowe, Leoni Kabel, Pilkington Automotive, Masterfoods Polska, Citicorp, jobpilot Polska, Hochtief, Uniwersytet Jagielloński, Politechnika Koszalińska, Politechnika Świętokrzyska, Akademia Ekonomiczna w Poznaniu, Uniwersytet Marii Curie-Skłodowskiej w Lublinie, Politechnika Wrocławska, Bertelsmann Media, Martin&Jacob ....

Literatura

1. Buzan, T. *The Mind Map book* A Plume Book 1996
2. Buzan, T. *Use Both Sides of Your Brain*
3. Buzan, T. *Make the Most of Your Mind*
4. *http://www.mindjet.com*
5. *http://www.maps.jcu.edu.au/netshare/learn/mindmap/*
6. *http://www.mind-map.com/*
7. *http://www.egov.com*
8. Pinto, J.K. *Successful Information System Implementation. The Human Side*
9. Millet, K. *Project Management. A Managerial Approach*
10. *The Project Management Body of Knowledge*, Project Management Institute
11. Choo, G. *Map Your Mind*
12. Kurtis, J. *Właściwy człowiek na właściwym miejscu* PRACAiTY 15/2001
13. Wągrzyn, M. *Thomas pomoże znaleźć właściwą osobę* GAZETA PRAWNA 78/2001

Mariusz Paszko
Anna Oszczak
Roman Zięba
Brand Integration Technologies
Ul. Szpacza 2, 04-238 Warszawa, tel.: +48 22 6114393,faks: +48 22 6114392
e-mail: brand@brand.com.pl, www.brand.com.pl

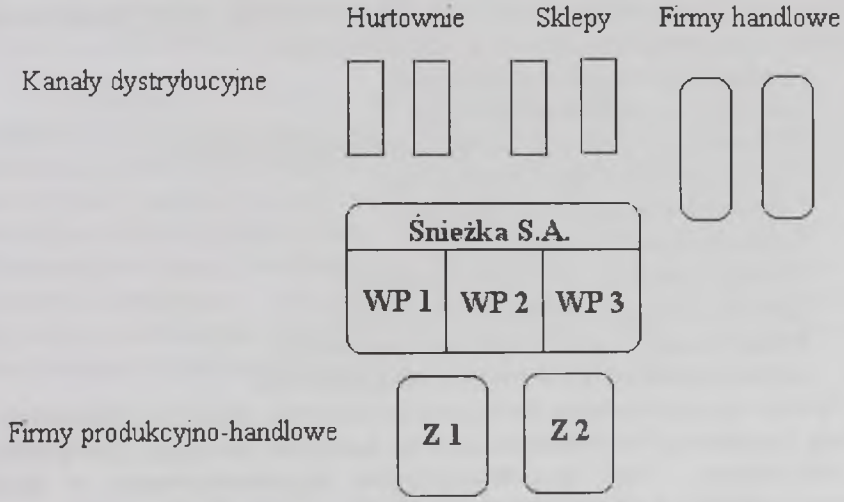
WDROŻENIE SYSTEMU IFS W FABRYCE FARB I LAKIERÓW „ŚNIEŻKA" S.A.

TOMASZ PAWLUS

Streszczenie: Artykuł ten przedstawia wdrożenie systemu IFS Applications dla 200 użytkowników w zakładach chemicznych. Ich specyfiką jest krótki czas produkcji oraz duża liczba różnorodnych zamówień klientów, a co za tym idzie szybki przepływ informacji i towarów oraz ścisłe dopasowanie produkcji do bieżącego popytu generowanego przez kanały dystrybucyjne. Prezentowany przykład pokazuje, iż możliwe jest szybkie i efektywne wdrożenie, przy zachowaniu określonego budżetu i limitu czasowego.

1. Śnieżka S.A.

Fabryka Farb i Lakierów „Śnieżka" S.A. jest firmą wielooddziałową. Zatrudnia ok. 500 pracowników. Posiada 3 oddziały produkcyjne w różnych lokalizacjach oraz udziały większościowe w kilku firmach produkcyjnych i dystrybucyjnych.
Sieć dystrybucyjną stanowią w głównej mierze odbiorcy hurtowi, przy czym hurtownicy o statusie głównego odbiorcy wykonują ok. 80% sprzedaży.



Rys.1.Organizacja firmy

Firma powstała w połowie lat osiemdziesiątych, początkowo zatrudniając łącznie z właścicielami 15 osób. W roku 1991 w wyniku rozwoju powstały dwa dodatkowe oddziały produkcyjne. Obecna nazwa spółki została przyjęta w roku 1998 przekształcając się ze spółki cywilnej w spółkę akcyjną. Rokrocznie produkcja ulegała zwielokrotnieniu dzięki czemu w roku 2002 wartość sprzedaży przekroczyła 250 mln złotych stawiając ŚNIEŻKĘ na drugiej pozycji wśród producentów farb i lakierów w Polsce.

Za rozwojem produkcji nie nadążały przemiany organizacyjne. Zarząd firmy często odczuwał brak szczegółowych i zarazem aktualnych informacji, koniecznych do podejmowania właściwych decyzji. Przygotowanie takich informacji nastręczało wielu problemów i pochłaniało cenny czas.
Rodziła się również potrzeba uruchomienia kanału internetowego do przyjmowania zamówień. Wynikało to z dużej liczby obszernych zamówień klientów (nawet do 200 pozycji na zamówieniu). Błędy na zamówieniach, jak również przy wydaniach z magazynu wyrobów gotowych generowały poważne straty. Problemem stawało się bieżące opanowanie wszystkich zdarzeń zachodzących w firmie i spółkach zależnych, jak również opracowywanie wyników oraz sprawozdań finansowych.
Właściciele firmy w momencie podpisywania kontraktu na wdrożenie systemu IFS Applications byli przekonani co do konieczności wprowadzenia szeroko zakrojonych zmian. Sytuacja ta w znacznym stopniu ułatwiała podejmowanie wielu trudnych decyzji w trakcie procesu wdrożeniowego.

2. Założenia przedwdrożeniowe.

Myśląc o wdrożeniu systemu zintegrowanego zarząd firmy jak i pracownicy zakładu posiadali pewne oczekiwania i wymagania. Jednak sposób ich prezentacji był raczej ogólny i nie określał precyzyjnie stanu oczekiwanego. Oczekiwania te można sklasyfikować w kilka obszarów :

1. Usprawnienie przepływu informacji;
2. Ułatwienie podejmowania decyzji;
3. Minimalizacja błędnych wydań na magazynie wyrobów;
4. Usprawnienie procesów planowania;
5. Zwiększenie skuteczności działania;
6. Ułatwienie (mechanizacja) pracy;
7. Obsługa i kontrola spółek powiązanych.
8. Kontrola kosztów produkcji i sprzedaży;
9. Porządkowanie i systematyzacja działalności;
10. Zachowanie dotychczasowych funkcjonalności;

W celu uporządkowania powyższych oczekiwań, przed przystąpieniem do wdrożenia opracowano we własnym zakresie dokument opisujący stan procesów przed wdrożeniem. Treść tego opracowania usystematyzowana w postaci tabelarycznej (Tablica 1.) pozwoliła na określenie wszystkich głównych procesów i podprocesów, definiując ich cechy oraz oczekiwane rezultaty. Uwagi zebrane w trakcie opracowywania tego dokumentu oraz zalecenia osób go współtworzących

stały się materiałem wyjściowym do wyznaczenia celów i miar tych celów. Precyzyjne określenie oczekiwań chroni nas przed niedomówieniami. Stanowi on obecnie punkt odniesienia do oceny wdrożenia.

Tablica I. Opis procesów gospodarczych.

Nazwa procesu Jednoznaczne i zwięzłe określenie procesu / podprocesu		
Cel Precyzyjnie określone cele, do których zmierza proces		
Właściciel Osoba decyzyjna, odpowiedzialna za prawidłowy przebieg procesu		
Dostawca Kto lub co dostarcza nam potrzebne informacje. Powiązanie z innymi procesami.	**Klient** Kto lub co jest odbiorcą naszej pracy. Powiązanie z innymi procesami.	
Start Co lub kto decyduje, że będziemy obsługiwać ten proces	**Koniec** Co lub kto decyduje, że obsługa procesu jest zakończona	
Wejście Określenie danych wejściowych. Jeśli pochodzą od różnych dostawców musimy to wyszczególnić	**Wyjście** Określenie danych wyjściowych. Co jest efektem zakończenia procesu.	
Miary Jeśli mówimy o efektach naszych działań, w tym miejscu, opisujemy miary jakimi się posługujemy przy ocenie procesu. Konkretne wartości wskaźników lub osiąganie wielkości.		
Wejście Opis jak wyżej	**Czynność** Kolejno wykonywane czynności lub odwołanie do podprocesu opisanego w podobny sposób.	**Wyjście** Opis jak wyżej
UWAGI W tym miejscu powinniśmy zebrać wszystkie nasze wnioski i zalecenia w stosunku do obsługi opisywanego procesu tak aby stanowiły wytyczne do poprawy tych procesów.		

Opracowana analiza sprecyzowała cechy charakterystyczne dla działalności firmy.Zagwarantowany umową czas dostawy do klienta wynosi 72 godziny, a krótki cykl produkcji umożliwia, w wielu wypadkach, na kilkakrotne odtworzenie zapasu magazynowego w ciągu doby. W tej sytuacji szybkość przepływu informacji ma kapitalne znaczenie na trafność podejmowanych decyzji. Uruchamianie zleceń produkcyjnych w oparciu o plan produkcji musi być na bieżąco korygowane, reagując na zmieniające się stany magazynowe oraz spływające zamówienia klientów. Stany minimalne zostały ustalone na poziomie umożliwiającym realizację zamówienia w 98% w ciągu doby.

3. Realizacja.

Zakres wdrożenia pokrywał praktycznie większość wymagań użytkowników. Kontrakt objął następujące moduły:

a. Dystrybucja: magazyn, zakupy, zamówienia klienta, fakturowanie

b. Produkcja: zlecenia produkcyjne, CRP, MRP, plan główny, koszty
c. Finanse: reguły księgowe, księga główna, ARK, księga należności, księga zobowiązań, środki trwałe, konsolidacja

Celowo zrezygnowano z modułu IFS Remonty, IFS Projektowanie oraz IFS Zarządzanie Zasobami koncentrując się na kluczowych z punktu widzenia firmy procesach – tzw. 'core business'. Natomiast wdrożenie IFS e-Business, jako funkcjonalności niezależnej od pozostałych odłożono na okres po wdrożeniu i uruchomieniu modułów bazowych.

Główne etapy wdrożenia:

a. przygotowanie kontraktu 04-05.2002
b. podpisanie kontraktu 05.2002
c. powołanie zespołu wdrożeniowego
d. wykonanie analizy przedwdrożeniowej przez dostawcę 06.2002
e. szkolenia od 06-08.2002
f. szkolenie z budżetowania 07.2002
g. logistyka magazynów wysokiego składowania - KK
h. krytyczne poprawki i uzupełnienia 09-12.2002
i. testowanie systemu 11.2002
j. uruchomienie systemu 01.2003
k. poprawki i uzupełnienia niekrytyczne 01-06.2003
l. ostateczne strojenie systemu – zakończenie wdrożenia prawdopodobnie 06.2003

Zdając sobie sprawę z możliwości wdrażanego systemu zintegrowanego równolegle prowadzono prace związane z **wprowadzeniem budżetowania**. Praktycznie bez ponoszenia dodatkowych kosztów stworzono narzędzie do kontroli budżetów poszczególnych komórek w rozbiciu na konkretne projekty. Daje to zarządowi potężne możliwości identyfikacji kosztów oraz ich planowania. Na konkretne efekty należy jeszcze poczekać aż pracownicy nauczą się w pełni wykorzystywać to narzędzie, ale już dzisiaj informacja zawarta w systemie jest wiele dokładniejsza niż przed wdrożeniem.

Zastosowanie **czytników kodów kreskowych**, pracujących on-line w trybie radiowym, od momentu spływu produkcji do przygotowania wysyłki, znacznie przyspieszyło obieg informacji. Szef produkcji dysponuje obecnie bieżącą informacją na temat aktualnego stanu zapasu magazynowego, który może bilansować ze złożonymi zamówieniami. Również magazynierzy mają bieżącą informację o ilości i lokalizacji towaru na magazynie, zaś sam system przeprowadza kontrolę wydawanego towaru. Ogranicza to ilość popełnianych błędów przy wydaniach. Oprogramowanie czytników 'radiowych' umożliwia:

- Przyjęcie surowców z KJ do magazynu
- Raportowanie operacji na zleceniu produkcyjnym
- Przyjęcie wyrobu ze zlecenia produkcyjnego
- Rezerwacja i wydanie na zamówienie klienta
- Zwrot z magazynu wysyłkowego
- Przesunięcie MM

- Inwentaryzacja
- Wydanie na zapotrzebowanie materiałowe

Wprowadzono również integrację z **kasami fiskalnymi** w obszarze sprzedaży detalicznej w sklepach - zarówno z towarem własnej produkcji, jak i w sklepie spożywczym reprezentującym całkowicie odmienny asortyment towarów.

W zakresie obsługi eksportu została wykonana **integracja z wykorzystywanym dotychczas programem HUZAR**.

Jak już wspomniałem, przyjmowanie zamówień klientów przed wdrożeniem systemu zintegrowanego nastręczało wielu problemów. Wprowadzanie długich, przekraczających często 100 pozycji zamówień rodzi błędy. Wynika to z faktu, iż materiałem źródłowym jest fax nie zawsze dobrej jakości oraz tempo pracy. Mimo weryfikacji zdarzają się błędy. W momencie, gdy zamówienie jest wypisywane bezpośrednio przez klienta sytuacja taka nie ma miejsca. Odpowiedzialność za poprawność zamówienia spoczywa po stronie klienta.

W tym celu wykorzystano opracowaną przez Śnieżkę i istniejącą wcześniej stronę WWW (o dużej oglądalności), gdzie umieszczona jest funkcjonalność sklepu interne-towego. Klient po zalogowaniu może wypełnić zamówienie lub przesłać go e-mailem (w odpowiednim formacie). Zamówienia takie odczytywane są kilka razy dziennie i przekształcane w sposób automatyczny na zamówienia systemu zintegrowanego.

Po jedenastu miesiącach od podpisania kontraktu z firmą IFS można powiedzieć, że większość celów została osiągnięta a zakończenie wdrożenia nastąpi w czerwcu 2003 roku. Uruchomienie systemu w spółkach zależnych oraz konsolidacja wyniku będą prowadzone w trakcie roku 2003, w znacznym zakresie samodzielnie przez firmę Śnieżka.

4. Pięć warunków dobrego wdrożenia.

Jasno sprecyzowane cele, zgodne ze strategią i misją zakładu. Wyartykułowane przez Zarząd, jak również przez bezpośrednich użytkowników, którzy najlepiej znają swoje potrzeby i wymagania. Określenie miar pozwalających na kontrolowanie postępu prac wskazuje konkretną wartość, do której zmierzamy. Zabezpiecza nas to przed ogólnikowymi dyskusjami, czy wdrożenie przyniosło zakładane korzyści, czy nie. Często zdarza się, że wiele czasu marnowane jest na przedkładanie swoich racji pomiędzy entuzjastami a oponentami wdrożenia. Podczas formułowania celów musimy zwracać uwagę na równoważenie potrzeb tak, aby pogodzić wszystkie interesy. W tym celu warto posłużyć się metodologią Balanced Scorecard.

Zdecydowanie zarządu firmy to otwartość na zmiany i elastyczność w podejmowaniu decyzji. Przyjmowanie nowej organizacji bardziej dostosowanej do rozwiązywania problemów (spłaszczenie struktury organizacyjnej). Szybkie rozstrzyganie problemów nie blokuje działania i wyzwala w pracownikach energię niezbędną do ciężkiej pracy, a nic tak nie zniechęca, jak brak szybkiej i właściwej

decyzji. Ze zdecydowania zarządu wypływać powinno także stworzenie systemu motywacyjnego mobilizującego do ciężkiej pracy, aby pracownik czuł się dowartościowany. Oby nie sprawdzało się stare amerykańskie powiedzenie, że *„pracownik pracuje na tyle, by go nie zwolnili, a pracodawca płaci tyle, aby pracownik nie odszedł."*

Komunikacja pomiędzy firmą wdrażającą a zespołem wdrażającym, jak i pomiędzy pracownikami firmy ma niebagatelne znaczenie do wzajemnego zrozumienia. Wypracowanie wspólnego języka i sposobów komunikowania się musi być kształtowana od samego początku wdrożenia. Każdy pracownik powinien znać dokładnie swój zakres obowiązków oraz wiedzieć jakie zadania są realizowane w zakładzie. Musi znać i rozumieć swoje miejsce w zachodzącym procesie zmian. Regularne spotkania informacyjne, szkolenia i warsztaty sprzyjają lepszemu zrozumieniu. W imię lepszej komunikacji nie można nikogo wyręczać, natomiast powinniśmy wzajemnie służyć sobie pomocą.

Kontrakt jest wzajemną gwarancją powodzenia całego przedsięwzięcia. Zarówno dostawcy jak i nabywcy zależy na powodzeniu przedsięwzięcia. Nie może tworzyć przywilejów, musi w jasny i precyzyjny sposób określać wymagania oraz regulować przewidywane sytuacje. Powinien również zawierać budżet i harmonogram wdrożenia. Zdając sobie sprawę z ciągłości procesu zmian, już na wstępie musimy określić warunki odbioru końcowego wdrożenia.

Kontrola realizacji to konsekwencja w działaniu. Nadzorowanie prac (harmonogramu), kontrola budżetu, ewentualna korekta harmonogramu czy budżetu oraz zarządzanie ryzykiem. Niezbędne jest również dokumentowanie prac, tak aby wiadomo było **kto, co, kiedy i jak** zostało zrobione.

5. Korzyści z szybkiego i kompleksowego wdrożenia.

Brak powtórzeń szkoleń – przeciągające się wdrożenie najczęściej stwarza sytuację, gdzie pracownicy po pewnym czasie od szkolenia zapominają zdobytą wiedzę. Szkolenia najczęściej realizowane są w jednym bloku na początku wdrożenia. Ma to swoje zalety dając możliwość poznania funkcjonalności sytemu na początku wdrożenia ale ma też swoje skutki uboczne gdy jakąś wiedzę odkładamy na potem, bo potem zazwyczaj nie pamiętamy i dochodzi do powtórzenia szkoleń w tym zakresie.

Ograniczenie zmian w oprogramowaniu – szybkie tempo wdrożenia powoduje najczęściej ograniczenie ilości zmian proponowanych przez użytkowników. Czując presję czasu starają się wykorzystywać tkwiące w systemie możliwości zdając sobie sprawę, że każda modyfikacja to dodatkowy czas i pieniądze. Ponadto mniejsza ilość modyfikacji sprzyja w przyszłości podnoszeniu wersji oprogramowania.

Szybszy zwrot nakładów – szybkie zakończenie wdrożenia to także szybsze efekty i zwrot poniesionych nakładów.

Większe możliwości systemu – w przypadku kompleksowego wdrożenia większej ilości modułów. Systemy zintegrowane przedstawiają sobą pewną

określoną koncepcję zarządzania przedsiębiorstwem. Wybieranie z całości fragmentów to tak jakbyśmy do swojego „malucha" chcieli wsadzić kierownicę i siedzenia z Mercedesa. Spotkałem się w literaturze, że niektórzy fachowcy od wdrożeń pragną zaliczyć jako wdrożenie systemu zintegrowanego przypadek uruchomienia trzech modułów.

Ograniczenie kosztów związanych z rozszerzeniem funkcjonalności – dzielenie wdrożenia na etapy odległe w czasie, to jest takie, gdzie celowo ograniczana funkcjonalność niektórych modułów nie ma moim zdaniem uzasadnienia. Powoduje to w przyszłości efekt powtórzenia wdrożenia modułów już eksploatowanych. Wówczas czas i pieniądze poświęcone na wdrożenie ulegają zwielokrotnieniu.

6. Napotkane trudności.

Wielu użytkowników końcowych wymagało dodatkowych szkoleń z podstawowego zakresu **obsługi komputera PC** mimo, że byli użytkownikami starych DOS-owskich aplikacji. Różnica w technologii stanowiła znaczącą barierę w początkowym okresie wdrożenia.

Ograniczony czas wdrożenia oraz jego znaczny zakres wymuszały wykonywanie zadań na **ostatni moment,** co powodowało powstawanie błędów, gdyż w wielu wypadkach brakowało czasu na rzetelne sprawdzenie i przetestowanie przygotowywanych danych czy rozwiązań, zarówno ze strony dostawcy, jak i użytkowników W efekcie strata czasu następowała na poprawie zaistniałych błędów. Część mniej krytycznych tematów przeszła na późniejsze terminy – dzięki takiemu podejściu nie opóźniono uruchomienia systemu, natomiast zamknięcie całego wdrożenia zostało przesunięte o około 3 miesiące.

We własnym zakresie dopasowano i uzupełniono bazę raportów ograniczając tym samym koszty wdrożenia. Tym samym zakład usamodzielnił się jeśli chodzi o modyfikację raportów pod często zmieniające się wymagania użytkowników.

7. Bibliografia

1. David K. Carr, kelvin J. Hard, William J. Trahant – *Zarządzanie procesem zmian* – PWN 1998.
2. Robert S. Kaplan, David P. Norton - *Strategiczna karta wyników* – PWN 2001.
3. Eugene H. Melan – *Process management : methods for improving products and service* – McGraw-Hill 1993

Tomasz Pawlus
Konsultant
tomek_ps@go2.pl

INDEKS AUTORÓW
Tom III

GOGOŁEK Waldemar; ; Państwowy Instytut Geologiczny Warszawa, ul. Rakowiecka 4; Tel. 0-22- 849 5351 **Artykuł 37, str. 59**	4	2
GOLA Elżbieta; Biuro Zarządu Banku, ING Bank Sląski S.A., 40-086 Katowice; ul. Sokolska 34 Tel.:(0-32)357-75-08 e-mail: elzbieta.gola@ingbank.pl **Artykuł 32, str. 5**	4	2
GÓRBIEL Krzysztof; Mgr; Szkoła Główna Handlowa w Warszawie; Katedra Informatyki Gospodarczej; Al. Niepodległości 162 e-mail: kgorbi@sgh.waw.pl **Artykuł 52, str. 235**	6	2
GRABARA Janusz; dr inż.; Instytut Ekonometrii i Informatyki; Politechnika Częstochowska; Ul. Armii Krajowej 19b, 42-200 Częstochowa e-mail: grabara@zim.pcz.czest.pl **Artykuł 39, str. 87**	5	2
HELT Piotr; Instytut Elektroenergetyki; Politechniki Warszawskiej; ul. Koszykowa 75, 00-662 Warszawa; Tel(0-22)660-73-14 e-mail:piotr.helt@ien.pw.edu.pl **Artykuł 22, str. 301**	3	1
ISKIERKA Iwona; Dr inż.; Instytut Informatyki i Ekonometrii Politechnika Częstochowska; ul. H. Dąbrowskiego 69, 42-200 Częstochowa; tel.: (0-34) 3 250 242 e-mail: iskierka@zim.pcz.czest.pl **Artykuł 41, str. 109**	5	2
JAGODZIŃSKI Mieczysław, dr inż.; IFS Poland, Senior Business Manager Al. Pokoju 78, 31-564 Kraków e-mail: m.jagodzinski@ifs.com.pl **Artykuł 12, str. 173**	2	1
JARZĘBOWICZ Aleksander; mgr inż.; Katedra Zastosowań Informatyki; Politechnika Gdańska, ul. Narutowicza 11/12, 80-952 Gdańsk; tel. (58) 347-24-75 e-mail: olek@eti.pg.gda.pl **Artykuł 62, str. 375**	7	2
JASIOROWSKI Krzysztof; Mgr inż.; FAMEG S.A. w Radomsku Kierownik dz. Informatyki, Project Managament MOVEX i SZJ LOAD, 97-500 Radomsko, ul. 11 Listopada 4a/56; Tel. 0-44/6822 312, 0-604 423 158 e-mail:krzysztofjasiorowski@wp.pl **Artykuł 13, str. 183**	2	1
Artykuł 23, str. 313	3	1
JASIOROWSKI Remigiusz; Mgr inż.; właściciel Firmy LOAD, 97-500 Radomsko; ul.11 Listopada 4a/56, tel. 0-602 132 374. e-mail:remekj@poczta.onet.pl **Artykuł 13, str. 183**	2	1
Artykuł 23, str. 313	3	1
JELONEK Dorota; dr; Wydział Zarządzania; Politechnika Częstochowska; Al. Armii Krajowej 19b; 42-200 Częstochowa tel. (0-34) 32-50-391, 32-50-388 **Artykuł 53, str. 249** e-mail: jelonek@zim.pcz.czest.pl	6	2

JONCZYK Monika; Instytut Informatyki Teoretycznej i Stosowanej, Politechnika Częstochowska; Ul. Dąbrowskiego 73, 42-200 Częstochowa e-mail: monikjon@wp.pl **Artykuł 45, str. 145**	5	2
KARDAS Michał; mgr inż.; Absolwent Wydziału Elektrycznego Politechniki Warszawskiej; ul. Nicejska 1 m.67 02-763 Warszawa e-mail: mkardas@ee.pw.edu.pl **Artykuł 35, str. 35**	4	2
KASPRZYK Artur; POTIS Software Development Tools ul. Jasielska 6a, 54-033 Wrocław e-mail: artur.kasprzyk@potis.com.pl **Artykuł 14, str. 193**	2	1
KATKOW Aleksander; Instytut Ekonometrii i Informatyki; Wydział Zarządzania; Politechnika Częstochowska; Ul. Armii Krajowej 19b 42-200 Częstochowa **Artykuł 42, str. 117**	5	2
KAZIMIERSKI Bogusław; Dr. Państwowy Instytut Geologiczny Warszawa, ul. Rakowiecka 4; Tel. 0-22- 849 5351 **Artykuł 37, str. 59**	4	2
KROLIKOWSKA Barbara; Dr. Instytut Informatyki w Zarządzaniu Uniwersytet Szczeciński; Ul. Mickiewicza 64; Tel.(091)444-19-15 e-mail: bkrol@uoo.univ.szczecin.pl **Artykuł 54, str. 257**	6	2
KRUPA Kazimierz; Uniwersytet Rzeszowski; ul. Rejtana 16 C 35 959 Rzeszów; tel. (0) 17 27 61 347 email: kkrupa@pf.pl **Artykuł 15, str. 209**	2	1
KUBIAK Dorota; Grupa Kęty S.A.; Kęty, ul.Kościuszki 111 e-mail: dkubiak@gk-kety.com.pl **Artykuł 24, str. 335**	3	1
KUJAWSKI Maciej; GramSoftware Sp. z o.o.;ul. Sabały 47 02-174 Warszawa; Tel. (22) 846 84 83 e-mail: m.kujawski@gramsoftware.pl **Artykuł 7, str. 103**	1	1
KULISIEWICZ Tomasz; Internet Obywatelski t.kulisiewicz@egov.pl Wrzosowa 18 05-830 Nadarzyn **Artykuł 10, str. 143**	1	1
KUROWSKI Jerzy; ZETO OLSZTYN Sp. z o.o.; ul. Pieniężnego 6/7; 10-005 OLSZTYN, tel. 089 – 5272113, kurowski@zeto.olsztyn.pl **Artykuł 55, str. 265**	6	2
KURZOK-DERDA Joanna; dr; Instytut Systemów Sterowania Chorzów ul. Długa 1/3 Tel.: (0-prefix-32) 247-28-20 e-mail: jkurzok@iss.pl **Artykuł 16, str. 219**	2	1
LASEK Mirosława ; Katedra Informatyki Gospodarczej i Analiz Ekonomicznych, Wydział Nauk Ekonomicznych, Uniwersytet Warszawski, ul. Długa 44/50, 00-241 Warszawa ; e-mail: mlasek@wne.uw.edu.pl, **Artykuł 33, str. 17**	4	2
LASOCKI Bogusław; mgr; Wyższa Szkoła Zarządzania Personelem ul. Hirszfelda 11; 02-776 Warszawa; tel. (22) 644 99 11, 643 06 83 e-mail: blasocki@e-sysinfo.pl; info@wszp.edu.pl **Artykuł 17, str. 231**	2	1

LIS Tomasz; mgr inż.; Instytut Ekonometrii i Informatyki; Politechnika Częstochowska; Wydział Zarządzania; Ul. Armii Krajowej 19b; 42-200 Częstochowa e-mail: tomlis1@wp.pl **Artykuł 25, str. 343** **Artykuł 34, str. 29**	 **3** **4**	 **1** **2**
ŁAPETA Jarosław; Mgr inż.; Politechnika Częstochowska Wydział Zarządzania; Instytut Ekonometrii i Informatyki Ul. Armii Krajowej 19B; 42-200 Częstochowa e-mail: jlapeta@zim.pcz.czest.pl **Artykuł 25, str. 343** **Artykuł 34, str. 29**	 **3** **4**	 **1** **2**
ŁATUSZYNSKA Małgorzata; dr; Uniwersytet Szczeciński Instytut Informatyki w Zarządzaniu e-mail: malgorzata.latuszynska@uoo.univ.szczecin.pl **Artykuł 43, str. 125**	 **5**	 **2**
MAJ Michał; mgr inż.; Polskie Radio Kraków S.A. al. Słowackiego 22; 30-007 Kraków e-mail: mmaj@krak.eu.org **Artykuł 61, str. 365**	 **7**	 **2**
MALINOWSKI Artur; Mgr inż.; Konsultant (MCSD, MCSE, MCDBA) Microsoft Services for the Enterprise Ul. Grzybowska 80/82, 00-844 Warszawa; Tel. +48 (22) 431 10 22 E-mail: arturmal@microsoft.com **Artykuł 26, str. 351**	 **3**	 **1**
MAŁACZEK Michał; Politechnika Warszawska; Instytut Sterowania i Elektroniki Przemysłowej e-mail: m.malaczek@iem.pw.edu.pl, m.malaczek@pzjudo.pl **Artykuł 18, str. 243**	 **2**	 **1**
MŁYNARSKI Stefan; Doc. Dr; Państwowy Instytut Geologiczny Warszawa, ul. Rakowiecka 4; Tel. 0-22- 849 5351 **Artykuł 37, str. 59**	 **4**	 **2**
NAFKHA Rafik; dr inż.; STOEN S.A.; ul. Wybrzeże Kościuszkowskie 41, 00-347 Warszawa e-mail: Rafik.Nafkha@stoen.pl **Artykuł 35, str. 35**	 **4**	 **2**
NAŁĘCKI Krzysztof; Centrum Komputerowe Politechniki Śląskiej 44-100 Gliwice, ul. Akademicka 5 **Artykuł 27, str. 369**	 **3**	 **1**
NIEMIEC Andrzej PRIM Sp. z o.o. Wrocław – tel. 0-71-328 3094; e-mail: prim@prim.com.pl **Artykuł 19, str. 257**	 **2**	 **1**
NOGA Agnieszka; mgr; Instytut Ekonometrii i Informatyki; Politechnika Częstochowska; Wydział Zarządzania; Ul. Armii Krajowej 19b; 42-200 Częstochowa **Artykuł 6, str. 95**	 **1**	 **1**
NOWAK Stefan; mgr inż.; Instytut Ekonometrii i Informatyki; Politechnika Częstochowska; Wydział Zarządzania; Ul. Armii Krajowej 19b; 42-200 Częstochowa e-mail: snowak@zim.pcz.czest.pl **Artykuł 49, str. 199** **Artykuł 64, str. 399**	 **5** **7**	 **2** **2**

NOWAKOWSKA Anna; Mgr; Wydział Zarządzania Politechniki Częstochowskiej; Instytut Ekonometrii i Informatyki al. Armii Krajowej 19B; 42-200 Częstochowa e- mail: nowakowskaanna@wp.pl **Artykuł 8, str. 111** **Artykuł 28, str. 385**	1 3	1 1
OGIELA Marek R.; Katedra Automatyki, Wydz. EAIiE, Akademia Górniczo-Hutnicza Al. Mickiewicza 30, 30-059 Kraków **Artykuł 1, str. 5**	1	1
OSZCZAK Anna; Brand Integration Technologies Ul. Szpacza 2, 04-238 Warszawa, tel.: +48 22 6114393, faks: +48 22 6114392 e-mail: brand@brand.com.pl, www.brand.com.pl **Artykuł 67, str. 429**	7	2
PASZKO Mariusz; Brand Integration Technologies Ul. Szpacza 2, 04-238 Warszawa, tel.: +48 22 6114393, faks: +48 22 6114392 e-mail: brand@brand.com.pl, www.brand.com.pl **Artykuł 67, str. 429**	7	2
PAWLUS Tomasz; konsultant, e-mail: tomek_ps@go2.pl **Artykuł 68, str. 445**	7	2
PELCZAR Małgorzata; mgr; Wydział Informatyki Politechniki Szczecińskiej; Instytut Sztucznej Inteligencji i Metod Matematycznych; Zakład Metod Matematycznych; ul. Żołnierska 49, 71-210 Szczecin tel. (+4891) 449 55 83 e-mail: mpelczar@wi.ps.pl **Artykuł 44, str. 135**	5	2
PĘCZKOWSKI Marek; Katedra Informatyki Gospodarczej i Analiz Ekonomicznych, Wydział Nauk Ekonomicznych, Uniwersytet Warszawski, ul. Długa 44/50, 00-241 Warszawa; e-mail: mpeczkowski@we.uw.edu.pl **Artykuł 33, str. 17**	4	2
PIECH Henryk; Politechnika Częstochowska; Instytut Matematyki i Informatyki; ul. Dąbrowskiego 73; tel. 604559300; e-mail: hpiech.adm.pcz.czest.pl, **Artykuł 4, str. 73**	1	1
PIOTROWSKI Paweł; dr inż.; Politechnika Warszawska, Instytut Elektroenergetyki; ul. Koszykowa 75, 00-662 Warszawa e-mail: pawel.piotrowski@ien.pw.edu.pl **Artykuł 35, str. 35**	4	2
PONDEL Maciej; Mgr inż.; Instytut Informatyki Ekonomicznej Akademia Ekonomiczna we Wrocławiu; ul. Komandorska 118/120, 53-345 Wrocław, tel/fax (071) 3680-516 e-mail: m-pondel@ae.wroc.pl **Artykuł 3, str. 61**	1	1
PRZENIOSŁO Stanisław; Doc. Dr; Państwowy Instytut Geologiczny; Warszawa, ul. Rakowiecka 4; Tel. 0-22- 849 5351 **Artykuł 37, str. 59**	4	2
PTAK Aleksandra; Politechnika Częstochowska; Wydział Zarządzania; Instytut Ekonometrii i Informatyki; ul. Armii Krajowej 19B; 42-200 Częstochowa **Artykuł 4, str. 73**	1	1

REKOSZ Dariusz; mgr; Kierownik Wydziału Informatyki PKE S.A. Elektrownia ŁAGISZA; 42-504 Będzin, ul. Pokoju 14 tel.: +32 267-13-31 e-mail: darekr@lagisza.pke.pl **Artykuł 29, str. 391**	3	1
REY Jan; Business Solution Architect; SAP Polska Sp. z o.o.; IBS Public Services; ul. Domaniewska 41; 02-672 Warszawa tel. +48-22-541-66-25; e-mail: jan.rey@z.pl **Artykuł 9, str. 121**	1	1
SAKOWSKI Andrzej; ERP-EKSPERT andrzej@saratoga.neostrada.pl **Artykuł 36, str. 47**	4	2
SENCZYNA Stefan, Dr inż.; Katedra Podstaw Systemów Technicznych, Wydział Organizacji i Zarządzania, Politechnika Śląska; 41-800 Zabrze, ul Roosvelta 26-28, tel. (0-32) 277 73 93 e-mail: sencz@polsl.gliwice.pl **Artykuł 30, str. 403**	3	1
SEWASTIANOW Paweł, Instytut Informatyki Teoretycznej i Stosowanej, Politechnika Częstochowska; ul. Dąbrowskiego 73, 42-200 Częstochowa e-mail sevast@k2.pcz.czest.pl **Artykuł 45, str. 145**	5	2
SKRODZKA Wioletta; Instytut Informatyki i Ekonometrii Zakład Matematyki Ekonomicznej; Politechnika Częstochowska Wydział Zarządzania; ul. Armii Krajowej 19B; 42-200 Częstochowa **Artykuł 46, str. 157** **Artykuł 47, str. 169**	5	2
SOŁTYSIAK Wioletta; Mgr inż.; Wyższa Szkoła Pedagogiczna w Częstochowie; e-mail: wsoltysiak@wsp.czest.pl **Artykuł 48, str. 185**	5	2
SZAMAŁEK Krzysztof; Dr hab.; Ministerstwo Srodowiska; Sekretarz Stanu – Gł. Geolog Kraju; 00-922 Warszawa, ul. Wawelska 52/54 tel. 0-22-57 92 337 **Artykuł 37, str. 59**	4	2
SZCZYGŁOWSKA Lidia; mgr inż.; Kustosz, Z-ca Dyrektora Dyrektora Biblioteki Głównej ds. Komputeryzacji ul. Armii Krajowej 36a; 42-200 Częstochowa **Artykuł 38, str. 77** **Artykuł 63, str. 387**	4 7	2 2
SZKIC-CZECH Ewa; Elektrownia „Opole" S.A., Brzezie k/Opola **Artykuł 31, str. 417**	3	1
SZOPA Janusz; dr hab. Prof. PCz; Instytut Ekonometrii i Informatyki; Politechnika Częstochowska; Wydział Zarządzania; Ul. Armii Krajowej 19b; 42-200 Częstochowa **Artykuł 49, str. 199**	5	2
SMIGIELSKA Anna; mgr; Politechnika Szczecińska; Wydział Informatyki; ul. Derdowskiego 18/24; 71-087 Szczecin; tel.: +48 607 93 48 92 e-mail: asmig1@tlen.pl **Artykuł 20, str. 273**	2	1

TADEUSIEWICZ Ryszard, Katedra Automatyki, Wydz. EAIiE, Akademia Górniczo-Hutnicza Al. Mickiewicza 30, 30-059 Kraków **Artykuł 1, str. 5**	1	1
ULFIK Agnieszka; mgr inż.; Instytut Ekonometrii i Informatyki; Wydział Zarządzania; Politechnika Częstochowska; ul. Armii Krajowej 19b; 42-200 Częstochowa **Artykuł 28, str. 385**	3	1
Artykuł 42, str. 117	5	2
Artykuł 49, str. 199	5	2
Artykuł 64, str. 399	7	2
WAGLOWSKI Piotr – www.vagla.pl **Artykuł 66, str. 419**	7	2
WIERZBICKI Marek, MOTTE, ul. Aleksandrowska 106/22, 91-224 ŁÓDŹ e-mail: motte@polbox.com, marek.wierzbicki@azymut.pl **Artykuł 56, str. 289**	6	2
WILENSKI Wojciech; mgr; Unizeto Sp. z o.o.; ul. Królowej Korony Polskiej 21; 70-486 Szczecin, tel. (091) 4801 327 e-mail: wwilenski@unizeto.pl **Artykuł 59, str. 341**	6	2
WŁODARCZYK Aneta; Politechnika Częstochowska Wydział Zarządzania; Katedra Ekonometrii i Statystyki; ul. Armii Krajowej 19B; 42-200 Częstochowa **Artykuł 50, str. 213**	5	2
WOŁOWSKI Franciszek; Kierownik "Sigillum" Polskiego Centrum Certyfikacji Elektronicznej; Polska Wytwórnia Papierów Wartościowych" tel (+48 22) 5302752 fax: (+48 22) 5032 571 e-mail: f.wolowski@pwpw.pl **Artykuł 57, str. 301**	6	2
ZAWADZKI Marek; IBM Lotus Polska Warszawa, ul. 1 Sierpnia 8 **Artykuł 21, str. 285**	3	1
ZAWIŁA-NIEDŹWIECKI Janusz; dr inż. Politechnika Warszawska Instytut Organizacji Systemów Produkcyjnych, Warszawa, ul.Narbutta 84 jzawila@poczta.onet.pl **Artykuł 58, str. 329**	6	2
ZDAWSKI Leszek; mgr inż.; Unizeto Sp. z o.o. ul. Królowej Korony Polskiej 21; 70-486 Szczecin, tel. (091) 4801 207 e-mail: lzdawski@unizeto.pl **Artykuł 59, str. 341**	6	2
ZIĘBA Roman; Brand Integration Technologies Ul. Szpacza 2, 04-238 Warszawa, tel.: +48 22 6114393, faks: +48 22 6114392 e-mail: brand@brand.com.pl, www.brand.com.pl **Artykuł 67, str. 429**	7	2

ISBN 83-204-2869-6

ISBN 83-204-2872-6