STUDIA INFORMATICA 2000
Volume 21 Number 1 (39)

Katarzyna KRAJEWSKA , Henryk KRAWCZYK
Politechnika Gdańska, Katedra Architektury Systemów Komputerowych

METODY WYSZUKIWANIA PROBLEMOWO ZORIENTOWANEJ INFORMACJI W INTERNECIE[1]

Streszczenie. Przedstawiono koncepcje wyszukiwania informacji w Internecie. Zaprezentowano własności różnego typu wyszukiwarek. Podano metodę inteligentnego wyszukiwania obrazów w bazach mutlimedialnych oraz oceniono przydatność tej metody dla zastosowań medycznych.

INTERNET SEARCH METHODS OF PROBLEM-ORIENTED INFORMATION

Summary. Concepts of information search methods are presented. Properties of problem-oriented browsers are described. Methods of intelligence search images in multimedia databases are given and their suitability for medical applications.

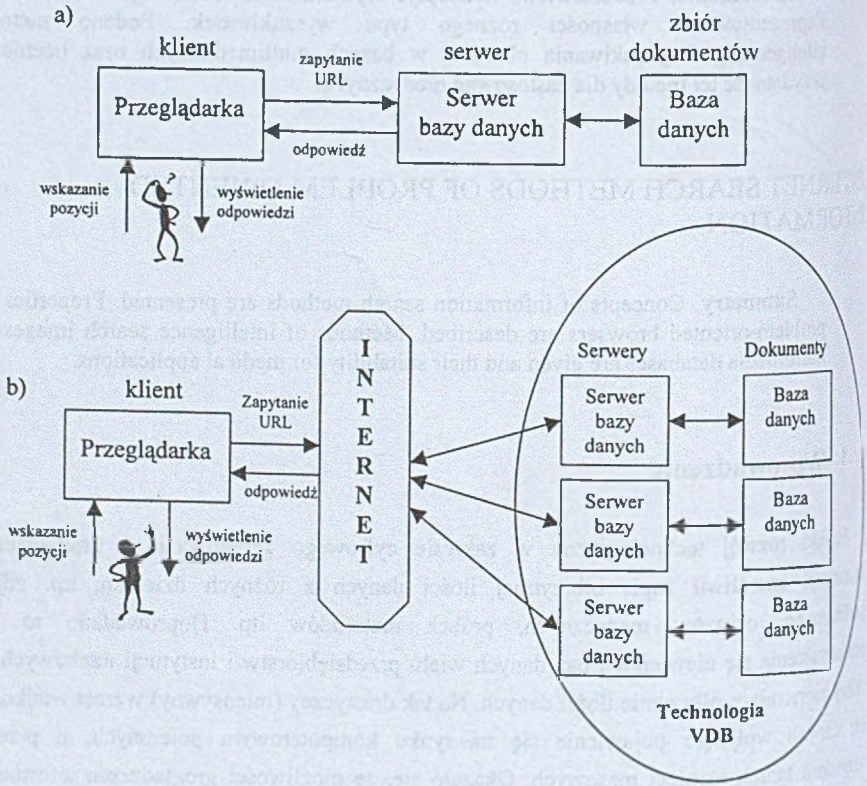
1. Wprowadzenie

Szybki rozwój technologiczny w zakresie cyfrowego generowania i gromadzenia informacji umożliwił zapis olbrzymiej ilości danych z różnych dziedzin, np. zdjęć satelitarnych, obrazów medycznych, próbek materiałów itp. Doprowadziło to do przekształcenia się niewielkich baz danych wielu przedsiębiorstw i instytucji naukowych w zbiory zawierające olbrzymie ilości danych. Na tak drastyczny (intensywny) wzrost wielkości baz danych wpłynęło pojawienie się na rynku komputerowym pojemnych, a przede wszystkim tanich pamięci masowych. Okazało się, że możliwości gromadzenia informacji przez człowieka są dużo większe od możliwości ich analizowania i przeszukiwania. Szczególnie, że w chwili obecnej duża część baz danych jest udostępniana poprzez sieć.

[1] Praca wykonana w ramach grantu KBN Nr 8T11C 00117

Problemem jest jednak to, że źródła informacji udostępniane w sieci różnią się sposobem organizacji i przedstawiania danych, używanym słownictwem, sposobami dostępu do danych, a także oferowanymi mechanizmami zapytań. Utrudnia to możliwość odczytu i łączenia danych z różnych źródeł. Dlatego też dąży się do stworzenia standardów sposobów przechowywania i prezentacji informacji. Przykładem tego są strony WWW, dostępne za pośrednictwem Internetu [6].

Główna strona organizacji jest nazywana witryną, zaś główna strona WWW danej osoby – stroną osobistą (domową). Strony posiadają odpowiedni format (HTML), dzięki czemu przeglądarki mogą je analizować i wyświetlać. Dokumenty HTML są identyfikowane za pomocą tzw. adresu, zwanego jednolitym adresem zasobu (URL). Za pomocą takiego adresu można wskazać określony dokument poprzez identyfikację podstawowych wskaźników: protokołu komunikacyjnego, nazwy komputera i adresu lokalnego dokumentu. Dodatkowo

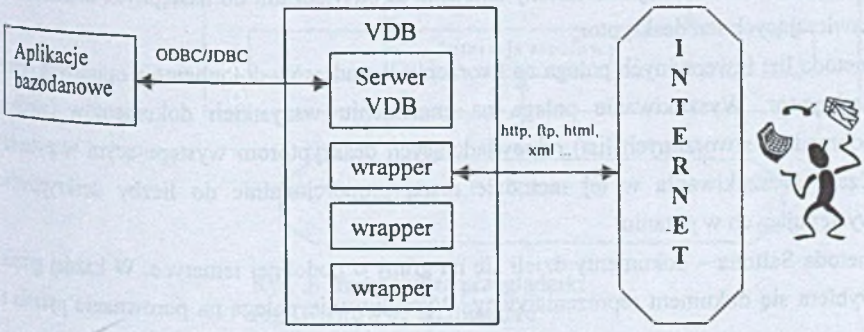


Rys. 1. Scentralizowany (a) i rozproszony (b) system informacyjny
Fig. 1. Information systems: centralized (a) and (b) distributed

wykorzystuje się odsyłacze, które (zawarte w danym dokumencie) są wskaźnikami do innych dokumentów. Na ogół przeglądarka umieszcza każdy pobierany plik w pamięci podręcznej na lokalnym dysku i przy każdym następnym odwołaniu sprawdza, czy wskazany dokument jest w pamięci podręcznej i dopiero jego brak uruchamia dostęp zdalny. Na rysunku 1 przedstawiono dwie podstawowe architektury systemu informacyjnego: scentralizowaną i rozproszoną. W pierwszym przypadku dokumenty są dostępne bezpośrednio i można zapewnić poprawność odsyłaczy między nimi. W drugim przypadku administrator każdego z serwerów może według własnego uznania zmieniać i usuwać przechowywane dokumenty, w konsekwencji czego odsyłacze nie muszą być poprawne.

Do poszukiwania informacji wykorzystuje się przeglądarki, takie jak AltaVista, Infoseek, czy Lycos. Na ogólne zapytanie otrzymuje się bardzo wiele dokumentów, które często dla poszukującego nie są przydatne. Podobnie wiele istotnych dokumentów może być pominiętych. W celu usprawnienia poszukiwań wykorzystuje się dodatkowe mechanizmy, np. operatory boolowskie, które jednak nie muszą zapewnić satysfakcjonujących rozwiązań. Najlepszym podejściem byłoby wykorzystanie języka naturalnego do formułowania zapytań, który następnie byłby transformowany na skuteczny język zapytań boolowskich.

W celu umożliwienia aplikacjom wyszukiwania informacji po całym Internecie proponuje się technologię wirtualnej bazy danych (VDB) [5]. VDB (*ang. Virtual Database*) zbiera informacje z danej dziedziny znajdujące się w całej sieci, indeksuje je w zależności od tytułów, kategorii i innych atrybutów. Programista ma dostęp do VDB poprzez standardowe zapytania SQL, zaś końcowy użytkownik może używać aplikacji z interfejsem graficznym, która buduje odpowiednie zapytanie. Aplikacje tego typu działają w sieci Internet, a ich interfejs jest napisany na ogół w formacie HTML. Dostęp do VDB jest możliwy poprzez VDB serwer (rysunek 2). Dla każdego zewnętrznego źródła danych VDB kreuje mechanizm owijacza (*ang. wrapper*), który jest interfejsem pomiędzy serwerem VDB a tym źródłem danych. Serwer VDB łączy więc poszczególne relacyjne bazy danych w jedną ujednoliconą



Rys. 2. Struktura wirtualnej bazy danych
Fig. 2. Structure of virtual database

wirtualną bazę. System zarządzający VDB, po otrzymaniu zapytania, dekomponuje je, określa części zapytań, które mogą być wysyłane do pojedynczych źródeł danych, po czym skleja otrzymane rezultaty. Technologia wirtualnej bazy danych umożliwia rozwój aplikacji, które charakteryzują się:

- dużą liczbą źródeł danych,
- autonomicznymi źródłami danych (bez scentralizowanej kontroli),
- źródłami danych o zróżnicowanej strukturze.

W pracy ograniczymy się do problemów wyszukiwania informacji specjalistycznej. Przedstawimy metody przeszukiwania dokumentów oraz architekturę wyszukiwarek różnego typu, jak również metody oceny skuteczności poszukiwań. Na bazie tego zasygnalizujemy architekturę wyszukiwarki, która może być wykorzystana do wynajdywania zdjęć medycznych o określonych własnościach. Przeglądarka taka może stanowić swego rodzaju system rekomendacji wspomagający pracę lekarza.

2. Inteligentne wyszukiwarki

Tradycyjne metody wyszukiwania informacji polegają na różnym sposobie przeglądania dokumentów:

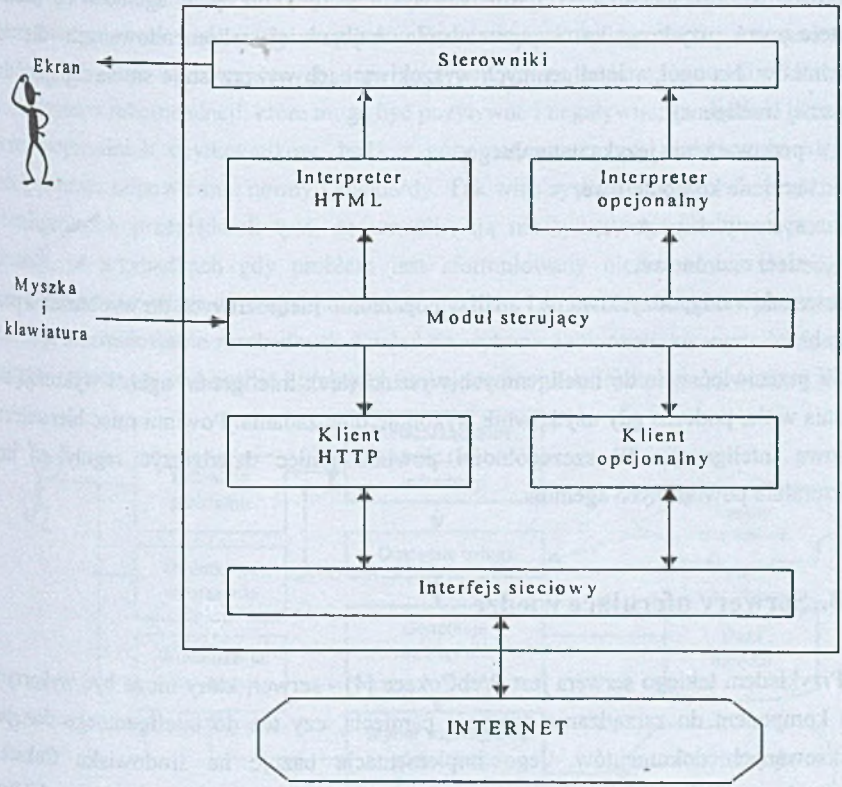
- metoda list prostych (klasyczna, modyfikacji z podziałem połówkowym, modyfikacji z grupowaniem) polega na przeglądaniu wszystkich dokumentów i sprawdzaniu, które z nich spełniają warunki zadane w pytaniu. Czas wyszukiwania w tej metodzie jest wprost proporcjonalny do ilości dokumentów, co jest wadą tej metody. Zaletami są: małe zapotrzebowanie na pamięć i proste algorytmy przetwarzania;
- metoda list łańcuchowych – dla dokumentów opisanych przez dowolny deskryptor tworzy się łańcuch adresów, którego pierwszym elementem jest adres jednego z tych dokumentów, zaś kolejne elementy łańcucha są odsyłaczami do następnych dokumentów zawierających ten deskryptor;
- metoda list inwersyjnych polega na tworzeniu list adresów dokumentów opisanych przez deskryptor. Wyszukiwanie polega na znalezieniu wszystkich dokumentów (według poprzednio stworzonych list) odpowiadających deskryptorom występującym w pytaniu. Czasy wyszukiwania w tej metodzie rosną proporcjonalnie do liczby deskryptorów występujących w pytaniu;
- metoda Saltona – dokumenty dzieli się na grupy o podobnej tematyce. W każdej grupie wybiera się dokument reprezentatywny. Wyszukiwanie polega na porównaniu pytania z

wszystkimi dokumentami reprezentatywnymi, a następnie z dokumentami z grup o dużym stopniu podobieństwa;

- metoda Chowa - dokumenty dzieli się na grupy, a adres grupy oblicza się poprzez wartość pewnego wyrażenia kombinatorycznego. Wadą jest założenie występowania stałej liczby deskryptorów w pytaniach;
- metoda Gosha – jest podobna do metody Chowa, z tym że adres grupy oblicza się poprzez rozwiązanie układu równań liniowych nad ciałem Galoisa.

Przeglądarki mają za zadanie pomoc użytkownikom w poruszaniu się po milionach stron WWW. Typowym sposobem wyszukiwania jest przeglądanie przy użyciu słów kluczowych. Schemat takiej przeglądarki podaje rysunek 3.

Zwykle wyszukiwarki pomagają użytkownikom znaleźć interesujące ich zasoby lub dane za pomocą słów kluczowych. Używają tylko minimalnej wiedzy o dziedzinie wyszukiwania i



Rys. 3. Architektura przeglądarki
Fig. 3. Browser architecture

przyswajają sobie bardzo ogólny model użytkownika. Niezdolność do poznawania użytkowników, dziedzin i słów kluczowych ogranicza ich zdolności do optymalizacji wyszukiwania poprzez unikanie niezgodnych tematycznie lub brakujących stron. Pojawia się też problem przeładowania informacji. Istnieją już miliony stron WWW i ich liczba rośnie w zawrotnym tempie. W celu ułatwienia wyszukiwania wykorzystuje się nowe rozwiązania dotyczące sztucznej inteligencji, jak np. wspomniane wcześniej inteligentne przeglądarki, inteligentni agenci czy serwery oferujące wiedzę. Przeglądarki uzupełnia się o dodatkowe funkcje, które umożliwiają leksykalny rozbiór zapytania i dopasowanie wyraźnych fraz rzeczownikowych i czasownikowych do dostępnych tezaurusów. Przykładem takiej przeglądarki jest FAQFinder [6], gdzie dla zapytania użytkownika określa się najbardziej prawdopodobną odpowiedź poprzez zbiór FAQ (*ang. Frequently Asked Questions*). Przykładami dwóch innych znanych prototypów inteligentnych przeglądarek są WebWatcher i Letizia [6]. Przeglądarka WebWatcher wykorzystuje koncepcje agentów do nauczenia zainteresowań użytkownika i opracowania najbardziej rekomendowanych dla niego hiperlinków. Na ogół w inteligentnych wyszukiwarkach wykorzystuje się następujące działy sztucznej inteligencji:

- przetwarzanie języka naturalnego,
- uczenie komputerowe,
- systemy ekspertowe,
- sieci neuronowe,

do przeszukiwania, odzyskiwania i analizy poprzednio niemożliwych do wyobrażenia pozycji danych.

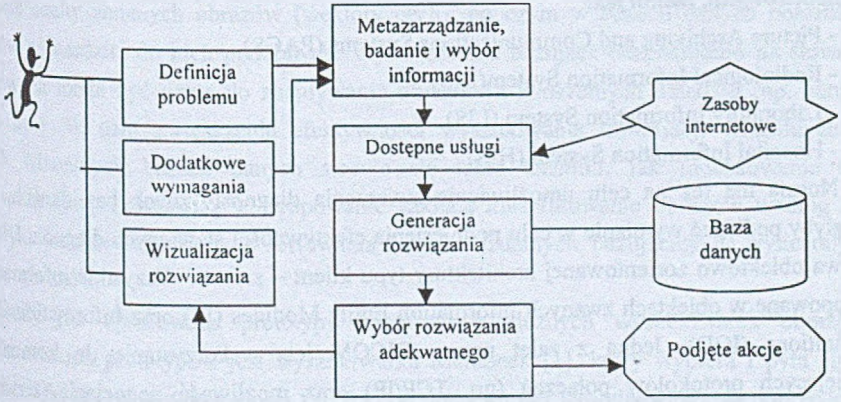
W przeciwieństwie do inteligentnych wyszukiwarek inteligentni agenci wykonują swoje zadania w tle, podczas gdy użytkownik wykonuje inne zadania. Powinni mieć hierarchiczną i czasową inteligencję. W szczególności powinni umieć dziedziczyć reguły od innych strukturalnie powiązanych agentów.

3. Serwery oferujące wiedzę

Przykładem takiego serwera jest WebCokace [4] - serwer, który może być wykorzystany jako komponent do zarządzania dostępną pamięcią, czy też do inteligentnego dostępu do indeksowanych dokumentów. Jego implementacja bazuje na środowisku Cokace do rozwijania modeli ekspertowych. Użytkownicy mogą przeglądać wyszukane biblioteki, aplikacje i modele według zadanego typu. Poza indeksowaniem i narzędziami do wyszukiwania przepływu danych WebCokace oferuje zaawansowane funkcje wyszukiwania

oparte na standardowych technikach obiektowych baz danych, takich jak Object Query Language. Domyślną formą prezentacji danych jest tekst.

Najwyższą formą rozwoju systemów wyszukiwania informacji są tzw. systemy rekomendacji [7] (*ang. recommender systems*). Systemy te wspomagają znajdowanie dopuszczalnych rozwiązań dla podanego problemu oraz dokonywanie wyboru najwłaściwszego wariantu rozwiązania. W przypadkach wielu znanych problemów rolę serwera wiedzy mogą realizować systemy ekspertowe korzystające z posiadanej wiedzy (reguły wnioskowania czy różne przypadki rozwiązań). Jeżeli wymagana wiedza nie jest sprecyzowana, to, jak w życiu, bazuje się wówczas na rekomendacji eksperta, dostępnych recenzjach, dotychczasowych opiniach. Wejściem do takiego systemu są rekomendacje różnych ludzi, które następnie są segregowane i kierowane do odpowiedniego odbiorcy. Oczywiście istotny jest dobór zespołu ekspertów oraz kryteria określające ocenę lub pogląd na daną sprawę. W ogólnym przypadku analizowane mogą być opinie wszystkich zainteresowanych osób bądź grupy ekspertów. Podobnie sposoby agregacji rozwiązań mogą być różne, od ściśle sprecyzowanych ocen do typowych procedur głosowania. Istotne są również kryteria rekomendacji, które mogą być pozytywne i negatywne, swobodnie określone poprzez poprzednich użytkowników, bądź z góry narzucone (np. podstawowe atrybuty jakości) poprzez odpowiednie normy i standardy. Tak więc systemy rekomendacji różnią się od inteligentnych przeglądarek tym, że umożliwiają nie tylko odpowiednie wyszukanie informacji, w przypadkach gdy problem jest sformułowany nieformalnie, ale sugeruje również pewne rozwiązanie tego problemu, możliwe do praktycznego wykorzystania. Oznacza to sformułowanie niezbędnych działań do wykonania w analizowanym środowisku. Dotyczyć to może np. wykonania korekty w funkcjonowaniu danej firmy czy zmiany profilu



Rys. 4. Architektura systemu rekomendacji
Fig. 4. Architecture of recommender system

jej produkcji. Na rysunku 4 został przedstawiony ogólny schemat systemu rekomendacji.

4. Wyszukiwanie informacji w bazach medycznych

Wraz z komputeryzacją przedsiębiorstw, szkół i bibliotek nastąpiła też komputeryzacja branży medycznej. Powstały systemy do obsługi szpitali (HIS) czy też laboratoriów. Do tworzonych baz danych pacjentów zaczęto dołączać historie ich chorób, diagnozy, sposoby leczenia, opisy przeprowadzanych badań oraz zdjęcia obszarów chorobowych. Doprowadziło to do stworzenia wielu baz danych zawierających podobne tematycznie informacje, ale różniących się strukturą przechowywanych danych. Obecnie dąży się do stworzenia jednolitego sposobu udostępniania i prezentacji tych danych w sieci.

Obrazy medyczne [2], zamiast na kliszy filmowej, mogą być przechowywane na nośnikach magnetycznych, CD-ROM'ach oraz taśmach w postaci bazy danych medycznych obrazów diagnostycznych. W tego typu bazach danych znajdują się zarówno obrazy pochodzące z badań specjalistycznych, jak również informacje o pacjentach i sposobach ich leczenia. W celu ujednolicenia zapisu takich danych wykorzystuje się standard DICOM.

DICOM (*ang. Digital Imaging and Communications in Medicine*, 1996) jest to norma opracowana i rozwijana wspólnie przez American College of Radiology (ACR) oraz National Electrical Manufacturers Association (NEMA). Prekursorami tej normy były ACR-NEMA 1.0 (1985) i ACR-NEMA 2.0 (1988). DICOM 3.0 jest nowym standardem sposobu zapisywania obrazów i danych medycznych. Ma on na celu umożliwienie wymiany i przesyłania danych (poprzez sieć) pomiędzy różnymi istniejącymi systemami komputerowymi, takimi jak:

- Picture Archiving and Communications Systems (PACS),
- Radiological Information System(RIS),
- Laboratory Information System (LIS),
- Hospital Information System (HIS).

Norma ma też na celu umożliwienie stworzenia diagnostycznych baz danych, które mogłyby podlegać wymianie w celu podniesienia efektywności stawianych diagnoz. DICOM używa obiektowo zorientowanej architektury typu klient – serwer. Wszystkie informacje są zgrupowane w obiektach zwanych Information Entity Modules (IE) oraz Information Object Definitions (IOD). Jedną z zalet normy DICOM jest wykorzystanie do komunikacji istniejących protokołów połączeń (np. TCP/IP) oraz możliwość negocjacji (serwera z klientem), jak wiadomości mają być przekazywane (zakodowane jako: Little Endian, Big Endian lub JPEG).

Dwa połączone systemy ustalają ze sobą, który z nich jaką pełni rolę (który jest klientem - SCU, a który serwerem – SCP). Klient zgłasza prośbę o połączenie, następuje wymiana danych i zakończenie połączenia.

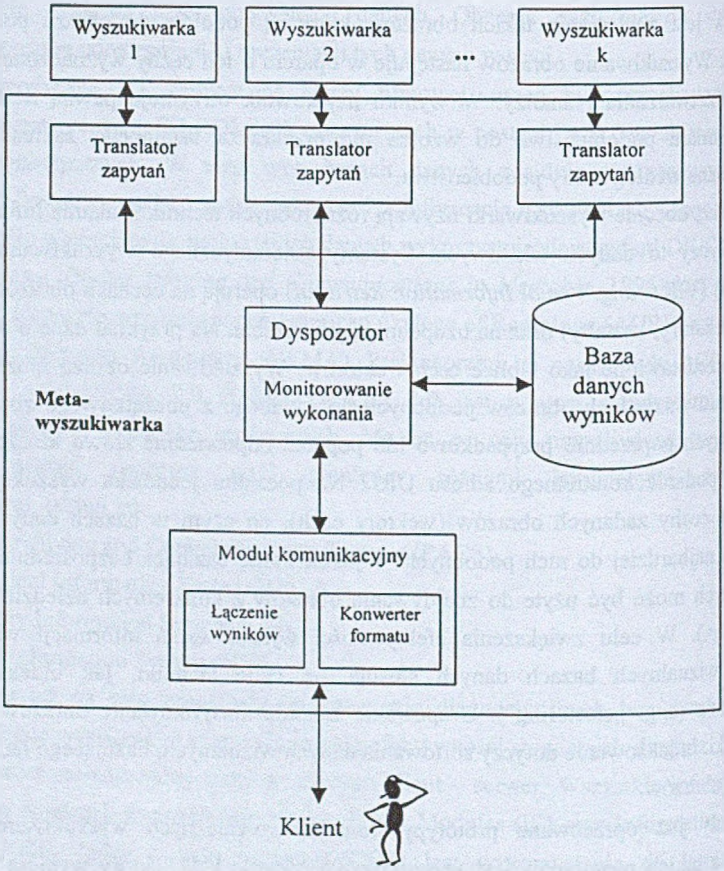
Tworzone zbiory danych mogą być udostępniane poprzez sieć. W związku z tym zachodzi konieczność zabezpieczenia danych klinicznych pacjentów (w celu zachowania tajemnicy lekarskiej) przed nieupoważnionym dostępem.

Istniejące systemy zarządzania bazami danych medycznych umożliwiają wyszukiwanie informacji i obrazów w oparciu o dane tekstowe (czyli opis wyniku przeprowadzonego badania) znajdujące się w rekordach pacjenta. Tworzone są też systemy umożliwiające wyszukiwanie i analizę porównawczą obrazów na podstawie cech obrazu [3]. Celem takich systemów jest znalezienie takich obrazów, które są podobne do obrazu podanego jako wzorzec. Wyszukiwanie obrazów następuje w oparciu o ich cechy wyznaczone przy użyciu metod przetwarzania i analizy. W wyniku użytkownik otrzymuje pewną liczbę obrazów, których miara podobieństwa do wzorca nie przekracza ustalonego zakresu i które są posortowane według miary podobieństwa.

Chociaż obecnie wyszukiwarki używają różnorodnych technik szukania informacji, to na ogół dotyczy to danych tekstowych. Systemy automatycznego wyszukiwania informacji wizualnej (VIR – *ang. Visual Information Retrieval*) operują na cechach obrazów (takich jak kolory, tekstury, kształty) oraz na uzupełniającym tekście. Na przykład dane o obrazie mogą zostać przedstawione jako tablice cech wektorów. Wyszukiwanie obrazu może polegać na znalezieniu wszystkich obrazów podobnych do zadanego z początkowego zbioru obrazów znalezionych poprzednio przypadkowo lub poprzez odpowiednie słowo kluczowe, czy też poprzez podanie konkretnego adresu URL. Na początku jednostka wyszukująca oblicza wizualne cechy zadanych obrazów (wektory cech), po czym w bazach danych poszukuje obrazów najbardziej do nich podobnych. Wyszukiwanie bazujące bezpośrednio na słowach kluczowych może być użyte do znajdywania obrazów z konkretnych dziedzin (np. natura, ludzie itp.). W celu zwiększenia efektywności wyszukiwania informacji wizualnych w dużych wizualnych bazach danych stosuje się takie techniki, jak indeksowanie czy grupowanie (ang. „clustering"). Grupowanie zakłada klasyfikowanie obrazów według ich semantyki. Indeksowanie dotyczy sortowania danych wizualnych, bazującego na wyliczonych cechach obrazów.

Zostały już opracowane prototypy kontekstowo-zależnych wyszukiwarek obrazów. Jednym z takich prototypów jest wyszukiwarka MetaSeek [1], który wybiera i pyta inne wyszukiwarki obrazów w kolejności zależnej od poprawności ich odpowiedzi w poprzednim zapytaniu dotyczącym podobnej sytuacji. MetaSeek zachowuje informacje o poprzednich działaniach przeglądarek wraz z oceną uzyskanych wyników wystawioną przez użytkownika.

MetaSeek jest kontekstowo-zależną wyszukiwarką obrazów w sieci. Dołącza on automatycznie użytkowników do wielu wyszukiwarek w sieci. Schemat architektury „meta-wyszukiwarki" został przedstawiony na rysunku 5. Podstawowe komponenty tego systemu to: translator zapytań, dyspozytor zapytań, moduł do komunikacji z użytkownikiem oraz baza danych z informacjami o poprzednio wykonanych wyszukiwaniach (sukcesach i porażkach). Po otrzymaniu zapytania dyspozytor wybiera najodpowiedniejsze wyszukiwarki poprzez skonsultowanie się z istniejącą bazą danych. Następnie translatory przekształcają zapytanie użytkownika na odpowiednie dla każdej z wybranych docelowych wyszukiwarek. Moduł do komunikacji z użytkownikiem pobiera wyniki wyszukiwania i prezentuje je użytkownikowi. Informację o jakości otrzymanych wyników otrzymuje się poprzez „sprzężenie zwrotne", tj.



Rys. 5. Architektura meta-wyszukiwarki
Fig. 5. Meta-search architecture

pobranie oceny uzyskanych informacji od użytkownika i zaktualizowanie bazy danych. Część z wyszukiwarek akceptuje przykładowe obrazy i dopasowuje je do obrazów w posiadanej

bazie danych według cech obrazu, takich jak kolor czy tekstura. Niektóre z nich pozwalają użytkownikowi ustawiać wagi ważności poszczególnych cech obrazu. Użytkownik ma możliwość ustawienia maksymalnego czasu oczekiwania, dziedziny zainteresowań itp. Po otrzymaniu zapytania dyspozytor wybiera odpowiednie wyszukiwarki w zależności od ustawionych opcji zapytania. Na przykład, jeśli użytkownik zadał pytanie wymagające wyszukiwania po przykładowych obrazach lub po słowach kluczowych, to wybierane są tylko te wyszukiwarki, które zapewniają wykonanie powyższych akcji.

Dla kontekstowo-zależnych zapytań o obrazy wybór odpowiednich wyszukiwarek jest uwarunkowany informacjami z bazy danych o efektach uzyskanych podczas poprzednich zapytań o podobnej tematyce. Kiedy zdarzy się, że użytkownik poszukuje informacji o nowym obrazie, wyszukiwarka porównuje zadany obraz z posiadanymi i wybiera najbardziej podobne, po czym na podstawie informacji o najlepszych wyszukiwarkach dla tych obrazów kieruje zapytanie do tych wyszukiwarek. Nowy obraz jest dodawany wraz z wynikami wyszukiwania do bazy danych przechowującej wyniki zapytań.

Czasami używa się podziału na kategorie wyszukiwań, gdyż wizualne bazy danych zawierają na ogół specyficzne typy obrazów, np. jedna zawiera zdjęcia ludzi, zaś inna zdjęcia z pewnej dziedziny medycyny. Oczywiste jest, że wyszukiwarka z bazą danych zawierającą zdjęcia ludzi szybciej znajdzie zdjęcie dziecka niż zorientowana na organizmy żywe. Na przykład, wyszukiwarka MetaSeek zawiera 10 kategorii semantycznych obrazów: zwierzęta, sztuka, budownictwo, kwiaty, geografia, krajobrazy, natura, obiekty, ludzie i transport.

Baza danych metawyszukiwarki zawiera wektory cech obrazów, wyniki wykonanych zapytań, grupy klas i kategorie semantyczne dla wszystkich obrazów, o które dotąd pytano. Informacje te są potrzebne dla dyspozytora do wybrania jak najodpowiedniejszej wyszukiwarki dla danego obrazu. Baza danych wyszukiwarki MetaSeek ma strukturę hierarchiczną. Obrazy są klasyfikowane względem kategorii semantycznej, następnie w zależności od wektora cech koloru, tekstury czy obu cech na raz są przypisywane do odpowiednich grup. Wektory cech dla koloru i tekstury są liczone przy użyciu histogramu i algorytmów Tamura. Właśnie tego typu wyszukiwarka nadaje się najlepiej również dla zastosowań medycznych.

5. Podsumowanie

Istnieje już wiele cyfrowych baz danych: systemy biblioteczne, medyczne, geograficzne - GIS. Są one udostępniane w sieci Internet. Celowe wydaje się narzucenie jednolitego standardu opisującego sposób dostępu do takich baz przynajmniej związanych z konkretną dziedziną tematyczną. W przypadku baz danych medycznych sprowadza się to do

zapewnienia zgodności informacji z normą DICOM, co umożliwi łatwe udostępnianie danych medycznych czy też wyszukiwanie podobnych przypadków nie tylko we własnej bazie, ale we wszystkich dostępnych w sieci. To z kolei wymaga opracowania problemowo zorientowanych wyszukiwarek, np. wyszukiwarka DICOM dla zastosowań medycznych. Koncepcja metawyszukiwarki może umożliwić nie tylko wspomaganie rozpoznawania chorób, ale także sugerować skuteczne leczenie konkretnej choroby. W rezultacie dalszego rozwoju wyszukiwarki może ona przerodzić się w pewien system rekomendacji, skupiający w sobie wiedzę wielu ośrodków medycznych.

LITERATURA

1. Benitez A. B., Beigi M.: Using Relevance Feedback in Content-Based Image Metasearch, Internet Computing, 1998, July – August, s. 59-69.
2. Piętka E.: Baza danych medycznych obrazów diagnostycznych w zintegrowanym systemie przechowywania i transmisji obrazów, materiały konferencji Infobazy 99, Gdańsk 1999, s. 103-107.
3. Rumiński J., Nowakowski A.: Wyszukiwanie obrazów medycznych z baz danych na podstawie cech obrazu, materiały konferencji Infobazy'99, Gdańsk 1999, s. 114-119
4. Corby O., Dieng R.: The Webcokace Knowledge Server, Internet Computing, 1999, November – December, s. 38-43.
5. Rajaraman A., Norvig P.: Virtual Database Technology: Transfering the Internet into a Database, Internet Computing, 1999, July – August, s.55-58.
6. O'Leary D.: The Internet, Intranets, and the AI Renaissance, Internet Computing, 1997, January, s.71-78.
7. Resnick P., H.R. Varian: Recommender Systems, Comunications of the ACM, Vol. 40, No. 3, 1997.

Recenzent: Dr inż. Katarzyna Stapor

Wpłynęło do Redakcji 27 marca 2000 r.

Abstract

The paper presents two kinds of Internet-based information systems: centralised and distributed ones (see Fig. 1) used for searching and analysing of different kind of information stored in databases or virtual databases (see Fig. 2). The properties of such systems depend on browsers which are basic interfaces with users. Fig. 3 presents the architecture of a typical browser. To analyse millions of WWW pages we need new solutions called intelligent browsers. Artificial intelligence approaches, such as natural language processing, computer learning methods, expert systems and neural networks which are successfully used to improve functionality of intelligent browsers. In consequence intelligent browsers are transformed into so-called recommender systems (see Fig. 4). They improve not only searching methods but also give new possibility of finding solutions of experts or numerous decisions of typical users. In the paper recognition problems of some illnesses basing on medical images are also discussed. To standardise format of searching information the DICOM standard is presented. The suitable architecture of meta-search (see Fig. 5) is proposed. This solution is similar to those used in MetaSeek browser. However, it is shown how such a meta-search can be rebuilt in a recommender system to supporting medical diagnostic decisions and improving its knowledge by storing in databases all representative images for representative classes of illnesses.