

Lech TUZINKIEWICZ  
Politechnika Wrocławska, Instytut Informatyki

## MODELOWANIE DANYCH NA POZIOMIE KONCEPTUALNYM

**Streszczenie.** W pracy przedstawiono proces modelowania danych na poziomie konceptualnym. Omówiono podstawowe etapy procesu modelowania danych oraz wyjaśniono kluczowe problemy występujące w tym procesie. Ponadto, zaproponowano miary do oceny jakości modeli. Przykład ilustrujący realizację podstawowych etapów procesu modelowania danych jest jednocześnie formą dyskusji problemów związanych transformacją modeli na poziomie konceptualnym, wykorzystując refaktoryzację i uściślanie w celu poprawy jakości modeli.

**Słowa kluczowe:** bazy danych, modele danych, jakość modeli

## DATA MODELING AT CONCEPTUAL LEVEL

**Summary.** The paper presents the process of data modeling at conceptual level. There are discussed the basic stages of the data modeling process. Key problems specific for this process are explained. Additionally the measures of quality data model evaluation are defined. Example of development of conceptual data models is shown. The way of improving the quality of the data models by using refactoring and refining is discussed.

**Keywords:** database design, data model, data model quality

### 1. Wprowadzenie

Projektowanie baz danych jest procesem modelowania danych na różnych poziomach abstrakcji. Modele danych reprezentują wycinek rzeczywistości, którego obiekty mają mieć swoją trwałą reprezentację w zasobach systemu informacyjnego (informatycznego). Oczekuje się, by dane przechowywane w bazie danych były w pełni zgodne z faktami wycinka rzeczywistości, a system zarządzania bazą danych był w stanie kontrolować

poprawność przechowywanych danych. Zatem, chcąc zaprojektować bazę danych spełniającą oczekiwania potencjalnych użytkowników zasobów bazodanowych, należy przede wszystkim zrozumieć modelowaną dziedzinę oraz ustalić, na podstawie potrzeb użytkowników, wymagania zarówno funkcjonalne, jak i niefunkcjonalne w stosunku do tworzonego systemu bazodanowego.

O ile tworzenie modeli danych na poziomie logicznym i fizycznym (efekt transformacji modeli konceptualnych) wymaga dobrej wiedzy inżynierskiej i pewnego doświadczenia, o tyle modelowanie na poziomie konceptualnym ma nieco inny charakter. W dalszej części artykułu zostanie przedstawiony proces modelowania danych na poziomie konceptualnym ze wskazaniem podstawowych elementów, mających wpływ na jakość tworzonych modeli. Omawiany przykład, ilustrujący realizację tego procesu, został szerzej omówiony w pracy opisującej metodykę projektowania aplikacji bazodanowych QUAD [3].

Analizując różne propozycje dotyczące metod i narzędzi wykorzystywanych w procesie modelowania danych na różnych poziomach abstrakcji [1, 2, 5, 6] oraz doświadczenia zdobyte przez autora w trakcie realizacji różnych projektów informatycznych skłaniają do szukanie odpowiedzi na następujące pytania:

- Jakimi umiejętnościami należy dysponować, by zaprojektować bazę danych, w szczególności na etapie tworzenia modeli konceptualnych danych?
- Czy w tym przypadku wystarczy wyłącznie wiedza inżynierska?
- Czy umiejętność projektowania baz danych to rodzaj rzemiosła?
- A może to ma jakiś związek ze sztuką, umiejętnością abstrakcyjnego myślenia?

Wnioski wynikające z przedstawionego dalej podejścia do modelowania danych oraz przykład tworzenia modelu danych na poziomie konceptualnym są formą odpowiedzi na postawione wyżej pytania.

## **2. Proces modelowania danych na poziomie konceptualnym**

Jednym z pierwszych etapów projektowania bazy danych, o kluczowym znaczeniu dla jakości bazy danych, jest aktywność związana z opracowaniem konceptualnego modelu danych, który jest reprezentacją bytów rozpatrywanego wycinka rzeczywistości.

Proces tworzenia konceptualnych modeli danych może być realizowany w pięciu etapach, zgodnie z przedstawionym dalej opisem.

## 2.1. Definicja klas

Na podstawie analizy wycinka rzeczywistości należy zidentyfikować klasy reprezentujące byty rozpatrywanej dziedziny. Analizie podlegają procesy biznesowe, które powinny być wspierane za pomocą opracowywanego systemu. Klasy definiuje się korzystając ze słownika pojęć dziedzinowych, przykładów i komentarzy wyjaśniających zasady realizacji procesów. Kluczowe znaczenie dla skuteczności realizacji tego etapu ma zrozumienie dziedziny i umiejętność komunikacji z ekspertami dziedziny. Często, w przypadku analizy rozległej i złożonej dziedziny przedmiotowej, występują problemy w jednoznacznej identyfikacji i interpretacji pojęć będących podstawą definicji klas. Głównie zagrożenie w tym zakresie spowodowane jest używaniem skrótów myślowych w przekazywanej wiedzy eksperckiej, występowaniem synonimów oraz homonimów.

Jednocześnie na tym etapie modelowania konceptualnego należy zdefiniować wewnętrzne własności klas, starając się zachować właściwe nazwy atrybutów. Etap ten jest uściśleniem zidentyfikowanych dotychczas klas. Głównym zagrożeniem, wynikającym ze złych nawyków lub brakiem wiedzy osób zajmujących się analizą, jest używanie nazw atrybutów, które planuje się stosować na poziomie implementacyjnym, oraz agregowanie kilku pojęć, tzn. przypisywanie własności, które charakteryzują obiekty innej klasy pozostające w relacji z definiowaną klasą.

Osobną czynnością, niezwykle ważną w kontekście identyfikacji własności rozpatrywanej dziedziny, jest ustalenie naturalnych identyfikatorów obiektów danej klasy, dalej nazywanych kluczami kandydującymi. Klucze te są podzbiorami własności klasy, których wartości zapewniają rozróżnienie instancji klasy. W wielu w publikacjach, między innymi w [2, 4], proponuje się już w tej aktywności określenie klucza głównego oraz kluczy alternatywnych. Zdaniem autorów jest to zbyt wczesna decyzja, bowiem na wybór klucza głównego, spośród kluczy kandydujących, mają wpływ elementy, które nie są jeszcze określone, np. relacje pomiędzy klasami i obiektami oraz przyjęte zasady realizacji transakcji. Może się okazać, że właściwym rozwiązaniem będzie wybór klucza sztucznego. Stąd też decyzja w tym zakresie będzie podjęta w fazie transformacji modelu konceptualnego do modelu logicznego lub fizycznego. Na pewno naganne jest przypisywanie z automatu wszystkim definiowanym klasom kluczy sztucznych, bowiem w konsekwencji w modelu konceptualnym danych mogą i często pojawiają się klasy, które nie reprezentują bytów wycinka rzeczywistości. Jako zasadę należy przyjąć, iż warunkiem koniecznym uznania definiowanego pojęcia jako klasy jest możliwość rozróżnienia instancji tej klasy (bytów występujących w rozpatrywanej dziedzinie), a zatem konieczność identyfikacji co najmniej jednego naturalnego klucza kandydującego. Natomiast nie wyklucza się uzupełnienia modelu

w późniejszym etapie o klasy niezbędne dla poprawnego funkcjonowania systemu, których obiekty identyfikowane będą wyłącznie na podstawie wartości klucza sztucznego.

## **2.2. Definicja relacji asocjacji pomiędzy klasami**

Obiekty rozpatrywanego wycinka rzeczywistości pozostają w relacjach wynikających z reguł biznesowych. Tworząc model konceptualny, należy uwzględnić możliwość lub konieczność wystąpienia związku między obiektami oraz obowiązujące ograniczenia liczości. Istotnym elementem, na który należy zwrócić uwagę, to własności relacji asocjacji. Częstym błędem popełnianym na tym etapie tworzenia modelu konceptualnego jest unikanie definiowania klas asocjacyjnych reprezentujących własności asocjacji na rzecz zwykłych klas. Po pierwsze, klasy asocjacyjne nie mają kluczy naturalnych, bowiem obiekty tych klas identyfikowane są na podstawie obiektów pozostających w związku, a po drugie czas życia tych obiektów ograniczony jest do czasu istnienia związku. Poza tym rezygnacja z klas asocjacyjnych powoduje zniekształcenie obrazu rzeczywistości i ma wpływ na decyzje dotyczące wyboru reguł transformacji przy odwzorowaniu do kolejnego poziomu abstrakcji. Typowe relacje to asocjacje binarne. Asocjacje wyższego rzędu występują sporadycznie.

## **2.3. Specyfikacja modelu konceptualnego w formie graficznej (np. diagram klas w języku UML)**

Prezentacja modelu w formie graficznej ułatwia zarówno jego interpretację, jak i komunikację z ekspertami rozpatrywanej dziedziny. Powszechnie stosowane techniki prezentacji modeli konceptualnych to diagramy encji i diagramy klas. W monografii do modelowania jest wykorzystywany język UML i zdaniem autorów prezentacja modelu konceptualnego w postaci diagramu klas daje, z racji większej ekspresji tego języka w stosunku do ERD, większą swobodę modelowania i jednocześnie unifikację w zakresie notacji.

## **2.4. Analiza oraz poprawa jakościowa modelu konceptualnego (analiza własności klas i relacji)**

Utworzony model konceptualny danych powinien być poddany dodatkowej analizie w celu identyfikacji relacji generalizacji pomiędzy klasami. Pozwala to określić wspólne cechy pomiędzy klasami oraz specyficzne własności zbiorów generalizacji. W efekcie uzyskuje się „dojrzały model” pojęciowy (ujawnione dodatkowe własności wycinka rzeczywistości) oraz lepszy jakościowo model logiczny, uzyskany w wyniku transformacji modelu konceptualnego.

Etap tworzenia modelu konceptualnego związany jest z wygenerowaniem modelu konceptualnego, który podlega pomiarom własności (miar) związanych z modelem jakości, bowiem jest to podstawą oceny modeli i jednocześnie umożliwia podjęcie decyzji dotyczących wyboru modelu (również modeli na kolejnych poziomach abstrakcji) w kontekście realizacji dalszych działań bądź przejścia do aktywności kolejnych dyscyplin.

Problemy występujące w procesie konstruowania modelu konceptualnego, możliwości refaktoryzacji i uściśleń dokonywanych w celu poprawy jakości modeli zostały omówione na prostym przykładzie w dalszej części rozdziału oraz w rozdziale poświęconym ocenie jakości artefaktów.

### **2.5. Specyfikacja transakcji bazodanowych w kontekście wybranego modelu konceptualnego danych oraz przypadków użycia projektowanego systemu**

Dysponując modelem konceptualnym danych, należy określić wymagane operacje bazodanowe (zapytania i transakcje zmieniające stan bazy danych) na podstawie specyfikacji planowanych usług w systemie (przypadki użycia systemu). Zadanie to ma charakter opcjonalny, aczkolwiek zaleca się zdefiniowanie operacji właśnie w tym kroku z uwagi na jednoczesną możliwość oceny kompletności modelu danych w kontekście realizowalności usług systemu.

## **3. Ocena jakościowa modeli**

W procesie projektowania baz danych powstaje wiele modeli na różnych poziomach abstrakcji. Dla danego poziomu abstrakcji i określonego zbioru reguł biznesowych oraz ograniczeń dziedzinowych można zaproponować wiele modeli danych, z których należy wybrać ten który uznaje się za akceptowalny zarówno ze względów funkcjonalnych, jak i нефункциональных. Ocena funkcjonalna jest dokonywana na podstawie weryfikacji modeli danych w kontekście planowanych usług sytemu (przypadki użycia systemu), natomiast aspekt нефункциональный wiąże się z jakością modeli danych. Jakość ocenia się na podstawie modelu jakości [3, 7, 8], na który składają się charakterystyki określające perspektywę oceny. Metodyka projektowania baz danych sterowana jakością została przedstawiona w [3].

Dla rozpatrywanego dalej przykładu rozważano następujące charakterystyki jakości:

- wydajność,
- pielęgnowalność.

Przygotowując propozycję miar dla tych charakterystyk, wzięto pod uwagę cechy modeli konceptualnych danych, mające wpływ na łatwość interpretacji modeli przez udziałowców

procesu oraz wpływ występujących elementów i konstrukcji modelu konceptualnego na transformację do modelu relacyjnego.

Miary jakości dla konceptualnego modelu danych zostały zdefiniowane z przeznaczeniem do oceny porównawczej modeli. Oczywiście, ma to sens wyłącznie wtedy, gdy modele są reprezentacją (abstrakcją) tego samego wycinka rzeczywistości.

W definicjach podanych dalej miar będą wykorzystywane następujące pojęcia i oznaczenia, związane z diagramem klas (tym diagramem jest opisany model konceptualny):

- $C = \{C_i \mid i \in I\}$ , gdzie  $I$  – podzbiór liczb naturalnych  $1, 2, \dots$ ,  $C_i$  – klasa z modelu konceptualnego,
- $A_i$  – wewnętrzne (nieodziedziczone) atrybuty klasy  $C_i$ ,
- $AI_i$  – zbiór atrybutów dziedziczonych przez klasę  $C_i$ .

Dla charakterystyki wydajności zdefiniowano następujące miary:

- MAD (ang. *Maximum of Association Degree*) – najdłuższa ścieżka pomiędzy klasami, mierzona liczbą asocjacji. Im dłuższa ścieżka dostępu do własności klas, tym bardziej złożone zapytania i dłuższy czas realizacji. Miara ta ma istotne znaczenie w kontekście określonego zbioru operacji bazodanowych.
- MNP (ang. *Maximum Number of Properties*) – maksymalna liczba atrybutów, zdefiniowanych w klasach. Porównanie maksymalnej liczby atrybutów klas umożliwia wstępną ocenę złożoności klas i sygnalizuje ewentualny problem związany z potrzebą normalizacji modelu (im większa wartość, tym gorzej). Miara ta może być stosowana tylko w przypadku, gdy oceniane modele są na tym samym poziomie semantycznym, tzn. obiekty świata rzeczywistego charakteryzowane są na tym samym poziomie szczegółowości.

Dla charakterystyki pielęgnowalności zdefiniowano następujące miary:

- NoC (ang. *Number of Classes*) – liczba klas występujących w modelu. Im jest większa, tym bardziej złożony jest model, trudniej go zrozumieć i zinterpretować oraz zweryfikować względem reguł obowiązujących w wycinku rzeczywistości.
- NoA (ang. *Number of Associations*) – liczba asocjacji występujących w modelu. Wraz ze wzrostem liczby asocjacji model stwarza więcej problemów interpretacyjnych oraz wymaga większego doświadczenia i wiedzy w fazie transformacji do modelu logicznego lub fizycznego.
- NSA (ang. *Number of Sharing Attributes*) – liczba współdzielonych atrybutów między klasami w relacji generalizacji definiowana jako  $\sum |AI_i|$ , dla  $i \in I$ ; Wyższe wartości tej miary świadczą o dogłębnym zrozumieniu modelowanej dziedziny i pozwalają zastosować właściwe reguły transformacji rozpatrywanego modelu do modelu relacyjnego.

- NRA (ang. *Number of Repeated Attributes*) – liczba powtarzających się atrybutów w różnych klasach modelu zdefiniowana jako  $\sum |A_i \cap A_j|$ , dla  $1 \leq i < j \leq n$ , gdzie  $n$  – liczba klas modelu. Liczba powtarzających się atrybutów w różnych klasach modelu danych świadczy o braku lub potrzebie dalszej analizy danych w kontekście występujących, a nie uwzględnionych generalizacji lub asocjacji z klasami o powtarzających się własnościach (powielanie własności wycinka rzeczywistości). Im większa wartość tej miary, tym większe prawdopodobieństwo, że model wymaga refaktoryzacji lub uściślenia w celu ustalenia, czy można wyodrębnić kolejne klasy pozostające w relacji generalizacji lub asocjacji z klasami, które współdzielą zbiory własności. W rezultacie, może to skutkować redundancją i brakiem integralności danych.

W celu przeprowadzenia porównania modeli, dla przyjętych charakterystyk należy jeszcze określić funkcje oceny. Funkcje oceny będą szeregować modele, biorąc pod uwagę miejsca modelu w rankingu modeli wynikające z obliczonych metryk dla poszczególnych charakterystyk. W przykładzie zaproponowano dwie funkcje oceny:

- ranking modeli oparty na określeniu, ile razy dany model uzyskał pierwsze miejsce w ocenie miary należącej do danej charakterystyki;
- olimpijski ranking modeli, klasyfikujący modele na podstawie sumy uzyskanych punktów wynikających z pozycji modeli w ocenie miary należącej do danej charakterystyki.

Funkcja oceny konceptualnego modelu danych z perspektywy obu charakterystyk (całego przyjętego modelu jakości) jest również zdefiniowana jako funkcja rankingowa.

#### **4. Przykład modelowania danych na poziomie konceptualnym**

W przykładzie będzie analizowany i rozwijany model dotyczący relacji rodzicielskich (ojciec, matka, potomek), zgodnie z opisanymi etapami tworzenia konceptualnych modeli danych.

Przyjęte reguły biznesowe, na podstawie których został zdefiniowany model danych, są następujące:

1. Mężczyzna może pozostawać w związku partnerskim z wieloma kobietami.
2. Kobieta może pozostawać w związku partnerskim z wieloma mężczyznami.
3. Mężczyzna może mieć wielu potomków.
4. Potomek ma jednego ojca.
5. Kobieta może mieć wielu potomków.
6. Potomek ma jedną matkę.
7. Związki partnerskie mogą mieć charakter formalny (małżeństwo) lub nieformalny.

### Etap 1. Definicja klas

Na podstawie opisu wycinka rzeczywistości, słownika dziedzinowego oraz modelu biznesowego należy zidentyfikować i podać definicję klas wraz z ich własnościami. W dalszej części zostały podane przykładowe definicje klas reprezentujących trwałe byty rozpatrywanej dziedziny.

W specyfikacji klas i relacji pomiędzy nimi będzie stosowana następująca notacja:

KLA/<xx>: <nazwa>(<atrybuty>) – identyfikator klasy wraz z jej nazwą i atrybutami  
 REL/<xx> <nazwa>(<rola\_koniec<sub>1</sub>>[<liczność>] : <rola\_koniec<sub>2</sub>>[<liczność>], <atrybuty>) – identyfikator relacji asocjacji lub generalicji (<nazwa>::= isA) wraz z określeniem końców, liczności oraz (ewentualnie dla klas asocjacji) – atrybutów  
 <xx> – numery klas i relacji postaci 01, 02, ...  
 <liczność> ::= \* | 1 – definicja liczności końca asocjacji (znaczenie jak w UML)  
 KK: zbiór<sub>1</sub>, zbiór<sub>2</sub>, ... – definicja kluczy kandydujących dla relacji

Definicja i specyfikacja klas, zgodnie z przyjętą notacją, dla rozpatrywanego wycinka rzeczywistości:

KLA/01: Mężczyzna – dorosły, dojrzały człowiek płci męskiej; potencjalny ojciec.

KLA/01: Mężczyzna (imię, nazwisko, data urodzenia, PESEL, płeć).

KK: {PESEL}.

KLA/02: Kobieta – dorosły, dojrzały człowiek płci żeńskiej; potencjalna matka.

KLA/02: Kobieta (imię, nazwisko, data urodzenia, PESEL, płeć).

KK: {PESEL}.

KLA/03: Potomek – każdy bezpośredni następca w prostej linii w stosunku do swojego przodka; syn lub córka w stosunku do rodziców.

KLA/03: Potomek (imię, nazwisko, data urodzenia, PESEL, płeć).

KK: {PESEL}.

Uwaga: W przypadku gdy istnieje wiele naturalnych identyfikatorów obiektów (kluczy kandydujących), to należy je wszystkie umieścić w zbiorze kluczy kandydujących, są to bowiem własności modelowanej dziedziny. Pominięcie jakiegokolwiek identyfikatora jest rów-



noznaczne z brakiem zgodności modelu, na tym poziomie abstrakcji, z modelowaną dziedziną.

### Etap 2. Definicja relacji asocjacji pomiędzy klasami

Na podstawie reguł biznesowych i zbioru zdefiniowanych klas należy określić i wyspecyfikować dopuszczalne relacje asocjacji pomiędzy obiektami klas.

Specyfikacja własności asocjacji:

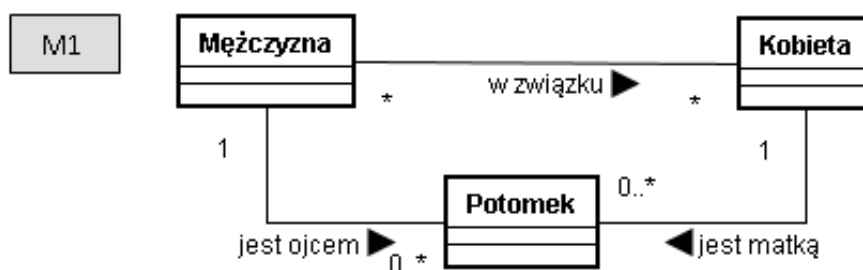
REL/01: w związku (Mężczyzna [\*] : Kobieta [\*])

REL/02 jest ojcem (Mężczyzna [1] : Potomek [\*])

REL/03 jest matką (Kobieta [1] : Potomek [\*])

### Etap 3. Specyfikacja modelu konceptualnego w formie graficznej

Model konceptualny spełniający powyższe reguły biznesowe przedstawiono na rys. 1.



Rys. 1. Model konceptualny *M1* reprezentujący relacje rodzicielskie  
Fig. 1. Conceptual model *M1* which represents parental relationships

W celu porównania modeli danych pod względem jakościowym, dla każdego z proponowanych modeli danych będzie wykonany pomiar miar zdefiniowanych w poprzednim rozdziale. Dla modelu M1 wyniki pomiarów umieszczono w tabeli 1.

Tabela 1

Wyniki pomiarów miar dla modelu M1

Model	NoC	NoA	NSA	NRA	MAD	MNP
<b>M1</b>	<b>3</b>	<b>3</b>	<b>0</b>	<b>15</b>	<b>3</b>	<b>5</b>

### Etap 4. Analiza oraz poprawa jakościowa modelu konceptualnego (analiza własności klas i relacji)

Analiza modelu M1 (zdefiniowanych klas i relacji między nimi) nie budzi większych zastrzeżeń, ale zastosowanie refaktoryzacji w tym procesie pokaże, jak bardzo można zmienić model rozpatrywanego wycinka rzeczywistości przez jego analizę i zastosowanie reguł refaktoryzacji (zachowując jego semantykę) [1], oraz w wyniku kolejnych uściśleń, poprawiając jego jakość.

W przykładzie wykorzystano następujące reguły refaktoryzacji modelu konceptualnego danych: **RRf/01** Uogólnienie – jeżeli dwie klasy współdzielą ten sam zbiór atrybutów, to wspólna część może być reprezentowana za pomocą osobnej klasy, z którą klasy bazowe (źródłowe) pozostają w relacji generalizacji/specjalizacji.

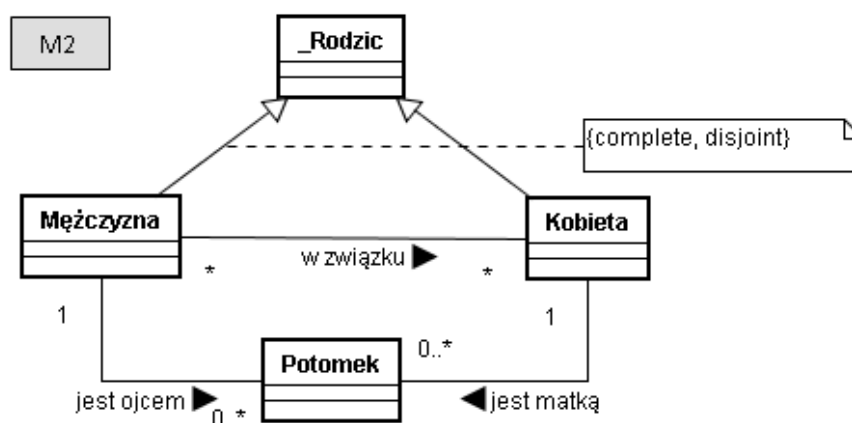
- **RRf/02** Eliminacja klas pośrednich w drzewie generalizacji – jeżeli w drzewie generalizacji występuje klasa pośrednia (tj. taka która, jest jednocześnie potomkiem i rodzicem), której wszystkie własności reprezentuje klasa będąca jej generalizacją, to można tę klasę usunąć, zachowując relację generalizacji pomiędzy jej potomkami i rodzicem.
- **RRf/03** Eliminacja relacji generalizacji dla klas mających wyłącznie dziedziczone własności wewnętrzne – jeżeli zbiór generalizacji o własnościach {Complete, Disjoint} [5] dotyczy wyłącznie klas, których zbiór wewnętrznych własności jest identyczny i dziedziczony ze wspólnej klasy, to te klasy i ich obiekty mogą być reprezentowane w modelu konceptualnym klasą, która jest ich generalizacją.

Przedstawione reguły refaktoryzacji zostały zdefiniowane w celu pokazania możliwości zastosowania refaktoryzacji w procesie projektowania baz danych w fazie modelowania konceptualnego i nie stanowią one kompletnego zbioru możliwych reguł.

Analiza wyjściowego modelu danych, zaprezentowanego na rys. 1, doprowadziła do następujących wniosków:

- Klasy Kobieta i Mężczyzna współdzielą te same atrybuty, które mogą być reprezentowane za pomocą własności klasy Rodzic (obiekty klasy są traktowani jako potencjalni rodzice). Przyjmuje się, że Kobieta i Mężczyzna będą dziedziczyć cechy Rodzica (zastosowanie reguły RRf/01).
- Na podstawie faktu, iż mężczyźni i kobiety reprezentowani są za pomocą osobnych klas, wyprowadza się własność zbioru generalizacji Rodzic {Complete, Disjoint}, tzn. Rodzic musi być kobietą lub mężczyzną.

Efektom procesu refaktoryzacji jest model M2 przedstawiony na rys. 2.



Rys. 2. Model konceptualny M2 utworzony w wyniku refaktoryzacji modelu M1

Fig. 2. Conceptual model M2 - the result of refactoring of model M1

Wyniki pomiarów miar dla modelu M2 umieszczono w tabeli 2.

Tabela 2

Wyniki pomiarów miar dla modelu M2

Model	NoC	NoA	NSA	NRA	MAD	MNP
<b>M2</b>	<b>4</b>	<b>3</b>	<b>10</b>	<b>10</b>	<b>3</b>	<b>5</b>

Specyfikacja własności klas z rys. 2.:

KLA/01: Rodzic (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/02: Mężczyzna (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/03: Kobieta (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/04: Potomek (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

Uwaga: Własności napisane kursywą oznaczają własności dziedziczone.

Specyfikacja własności asocjacji:

REL/01: w związku (Mężczyzna [\*] : Kobieta [\*])

REL/02 jest ojcem (Mężczyzna [1] : Potomek [\*])

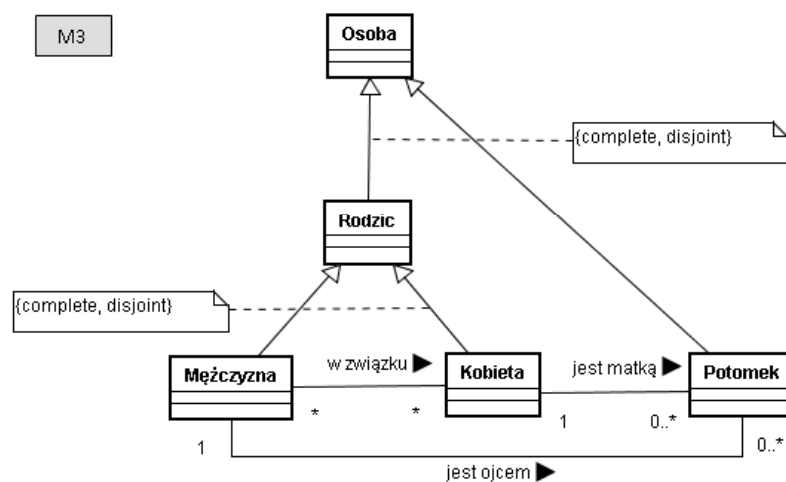
REL/03 jest matką (Kobieta [1] : Potomek [\*])

Zbiór generalizacji: Osoby {Complete, Disjoint}

REL/04 isA (Rodzic: Kobieta)

REL/05 isA (Rodzic: Mężczyzna)

Stosując konsekwentnie regułę refaktoryzacji RRf/01 dla modelu M2, tym razem do klas Rodzic i Potomek, otrzymuje się model M3, który został przedstawiony na rys. 3.



Rys. 3. Model konceptualny M3 utworzony w wyniku zdefiniowania relacji generalizacji dla klas Rodzic i Potomek w modelu M2

Fig. 3. Conceptual model M3 - gained by generalization of classes Parent and Descendant in the model M2

Wyniki pomiarów miar dla modelu M3 umieszczono w tabeli 3.

Tabela 3

Wyniki pomiarów miar dla modelu M3

<b>Model</b>	<i>NoC</i>	<i>NoA</i>	<i>NSA</i>	<i>NRA</i>	<i>MAD</i>	<i>MNP</i>
<b>M3</b>	<b>5</b>	<b>3</b>	<b>20</b>	<b>0</b>	<b>3</b>	<b>5</b>

Specyfikacja własności klas modelu M3 z rys. 3.:

KLA/01: Osoba (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/02: Rodzic (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/03: Mężczyzna (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/04: Kobieta (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/05: Potomek (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

Specyfikacja własności asocjacji:

REL/01: w związku (Mężczyzna [\*] : Kobieta [\*])

REL/02 jest ojcem (Mężczyzna [1] : Potomek [\*])

REL/03 jest matką (Kobieta [1] : Potomek [\*])

Zbiór generalizacji: Osoby {Complete, Disjoint}

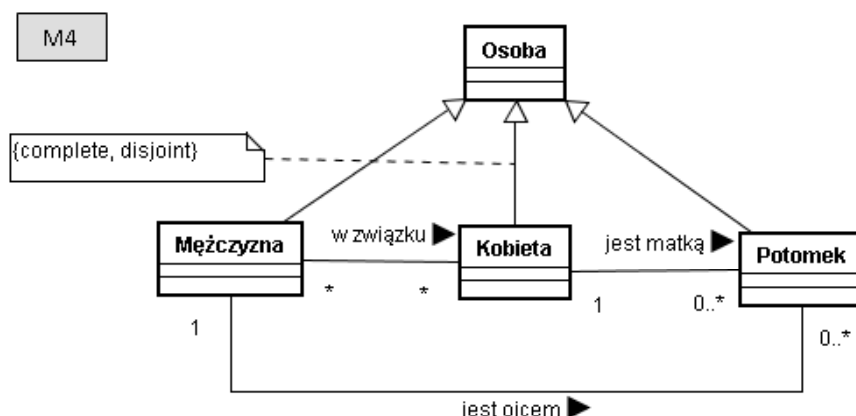
REL/04 isA (Osoba : Rodzic)

REL/05 isA (Rodzic : Kobieta)

REL/06 isA (Rodzic : Mężczyzna)

REL/07 isA (Osoba : Potomek)

Kolejną regułą refaktoryzacji, która ma zastosowanie w przypadku rozpatrywanych modeli, jest reguła RRf/02. W wyniku zastosowania reguły RRf/02 w modelu M3 została pominięta klasa Rodzic (jej własności przejęła klasa Osoba). Efekty wykonania refaktoryzacji zostały przedstawione w postaci modelu M4 na rys. 4.



Rys. 4. Model konceptualny *M4* uzyskany poprzez zastosowanie reguły *RRf/02* dla modelu *M3*  
 Fig. 4. Conceptual model *M4* - gained by using *RRf/02* rule for model *M3*

Wyniki pomiarów miar dla modelu *M4* umieszczono w tabeli 4.

Tabela 4  
 Wyniki pomiarów miar dla modelu *M4*

Model	NoC	NoA	NSA	NRA	MAD	MNP
<b>M4</b>	<b>4</b>	<b>3</b>	<b>15</b>	<b>0</b>	<b>3</b>	<b>5</b>

Specyfikacja własności klas modelu *M4* z rys. 4.:

KLA/01: Osoba (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/02: Mężczyzna (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/03: Kobieta (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/04: Potomek (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

Specyfikacja własności asocjacji:

REL/01: w związku (Mężczyzna [\*] : Kobieta [\*])

REL/02 jest ojcem (Mężczyzna [1] : Potomek [\*])

REL/03 jest matką (Kobieta [1] : Potomek [\*])

Zbiór generalizacji: Osoby {Complete, Disjoint}

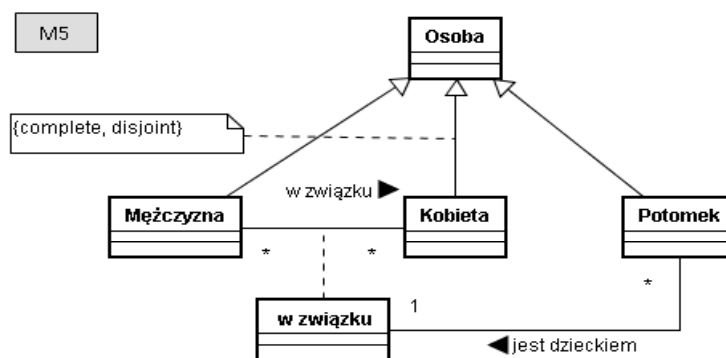
REL/04 isA (Osoba: Kobieta)

REL/05 isA (Osoba: Mężczyzna)

REL/06 isA (Osoba : Potomek)

Dalsza analiza konceptualnego modelu danych *M4* umożliwia przyjęcie założenia, iż relacje asocjacji „jest ojcem” i „jest matką”, które są obligatoryjne dla obiektów klasy Potomek, są własnością asocjacji „w związku” występującej pomiędzy klasami Mężczyzna i Kobieta. Powyższe stwierdzenie jest już traktowane jako uściślenie dotyczące wiedzy na

temat modelowanej dziedziny (dodatkowa wiedza) i w związku z tym wymaga akceptacji ze strony ekspertów dziedzinowych. Powyższa analiza może wydawać się zbędna, bowiem rozpatrywany przykład jest oczywisty ze względu na posiadaną wiedzę ogólną. Natomiast w przypadku modelowania specyficznej dziedziny, spostrzeżenia i wnioski prezentowane w postaci konceptualnego modelu danych nie muszą już być tak oczywiste. Model M5, uzyskany przez uściślenie posiadanych faktów został przedstawiony na rys. 5.



Rys. 5. Model konceptualny *M5* uzyskany przez modyfikację modelu *M4* (wynik zdefiniowania relacji generalizacji dla klas *Mężczyzna*, *Kobieta* i *Potomek*)

Fig. 5. Conceptual model *M5* - gained from model *M4* (the result of generalization of classes *Man*, *Woman* and *Descendant*)

Wyniki pomiarów miar dla modelu M5 umieszczono w tabeli 5.

Tabela 5

Wyniki pomiarów miar dla modelu M5

Model	NoC	NoA	NSA	NRA	MAD	MNP
<b>M5</b>	<b>5</b>	<b>2</b>	<b>15</b>	<b>0</b>	<b>2</b>	<b>5</b>

Specyfikacja własności klas z rys. 5.:

KLA/01: Osoba (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/02: Mężczyzna (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/03: Kobieta (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

KLA/04: Potomek (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

Specyfikacja własności asocjacji:

REL/01: w związku (Mężczyzna [\*] : Kobieta [\*], od, do, rodzaj)

REL/02 jest dzieckiem (Potomek [\*] : w związku [1])

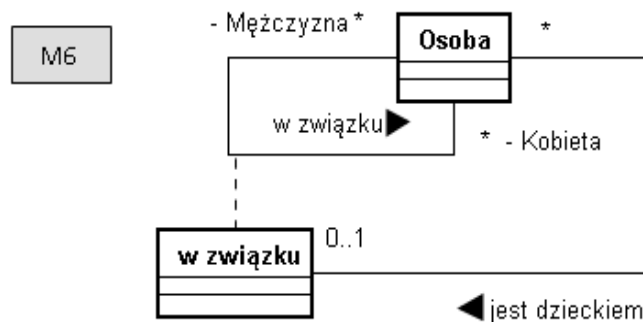
Zbiór generalizacji: Osoby {Complete, Disjoint}

REL/03 isA (Osoba : Kobieta)

REL/04 isA (Osoba : Mężczyzna)

REL/05 isA (Osoba : Potomek)

Analiza modelu M5 pozwala zastosować regułę refaktoryzacji RRf/03. Efekt zastosowania reguły RRf/03 dla modelu M5 został przedstawiony na rys. 6 w postaci modelu M6.



Rys. 6. Model konceptualny *M6* utworzony z modelu *M5* w wyniku zastosowania reguły refaktoryzacji *RRf/03*

Fig. 6. Conceptual model *M6* - gained by using *RRf/03* rule for model *M5*

Wyniki pomiarów miar dla modelu M6 umieszczono w tabeli 6.

Tabela 6

Wyniki pomiarów miar dla modelu M6

Model	NoC	NoA	NSA	NRA	MAD	MNP
<b>M6</b>	<b>2</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>2</b>	<b>5</b>

Specyfikacja własności klas modelu M6 z rys. 6.:

KLA/01: Osoba (imię, nazwisko, data urodzenia, PESEL, płeć)

KK: {PESEL}

Specyfikacja własności asocjacji:

REL/01: w związku (Osoba.Mężczyzna [\*] : Osoba.Kobieta [\*], od, do, rodzaj)

REL/02 jest dzieckiem (Osoba.Dziecko [\*] : w związku [1])

Po utworzenie zbioru alternatywnych modeli konceptualnych można poddać je ocenie, aby wybrać najlepszy do opracowania w kolejnych etapach.

Wyniki pomiarów miar dla rozważanych modeli konceptualnych danych (tabele 1-6) wykorzystano do obliczenia funkcji oceny. Umieszczone w kolumnach tabeli 7 wartości reprezentują pozycje modeli M1-M6 w rankingu ze względu na miarę zdefiniowaną w wierszu.

Dane z tabeli 7 są podstawą obliczenia funkcji oceny dla charakterystyki Pielęgowalność (patrz tabela 8) oraz Wydajność (patrz tabela 9). W obu przypadkach oceniano miejsce modelu w rankingu.

Bez względu na stosowaną funkcję oceny (suma wystąpień modelu na pierwszej pozycji, ranking olimpijski) modele M5 i M6 uzyskały najwyższą ocenę. Z uwagi na to, że różnica

w ocenie jakościowej modeli M5 i M6 jest nieznaczna, należałoby oba modele uwzględnić przy transformacji do modelu logicznego i/lub fizycznego.

Tabela 7

Ranking modeli konceptualnych danych z podziałem na charakterystyki

Model Miara	M1	M2	M3	M4	M5	M6
Pielęgowalność						
<i>Noc</i>	2	3	4	3	4	1
<i>NoA</i>	2	2	2	2	1	1
<i>NSA</i>	3	2	1	2	2	3
<i>NRA</i>	3	3	1	1	1	1
<b>Σ pozycji</b>	<b>10</b>	<b>10</b>	<b>8</b>	<b>8</b>	<b>8</b>	<b>6</b>
Wydajność						
<i>MAD</i>	2	2	2	2	1	1
<i>MNP</i>	1	1	1	1	1	1
<b>Σ pozycji</b>	<b>3</b>	<b>3</b>	<b>3</b>	<b>3</b>	<b>2</b>	<b>2</b>

Tabela 8

Ranking modeli dla charakterystyki *Pielęgowalność*

Model	M1	M2	M3	M4	M5	M6
Σ wystąpień modelu na pierwszej pozycji	0	0	2	1	2	3
<b>Pozycja w rankingu</b>	<b>V</b>	<b>V</b>	<b>II</b>	<b>IV</b>	<b>II</b>	<b>I</b>
Ranking olimpijski	10	10	8	8	8	6
<b>Pozycja w rankingu</b>	<b>V</b>	<b>IV</b>	<b>II</b>	<b>II</b>	<b>II</b>	<b>I</b>

Tabela 9

Ranking modeli dla charakterystyki *Wydajność*

Model	M1	M2	M3	M4	M5	M6
Σ wystąpień modelu na pierwszej pozycji	0	0	0	0	1	1
<b>Pozycja w rankingu</b>	<b>II</b>	<b>II</b>	<b>II</b>	<b>II</b>	<b>I</b>	<b>I</b>
Ranking olimpijski	3	3	3	3	2	2
<b>Pozycja w rankingu</b>	<b>II</b>	<b>II</b>	<b>II</b>	<b>II</b>	<b>I</b>	<b>I</b>

### **Etap 5. Specyfikacja transakcji bazodanowych w kontekście wybranego modelu konceptualnego danych oraz przypadków użycia projektowanego systemu**

Etap ten ma charakter opcjonalny, ale w przypadku, gdy stosowana metodyka projektowania baz danych zakłada weryfikację konceptualnego modelu danych w kontekście realizowalności transakcji, wówczas specyfikacja transakcji jest niezbędna. Zdefiniowana w tabeli 10 operacja bazodanowa OPE/001 LiczbaNarozdin, która ma umożliwić odczyt z bazy danych liczbę dzieci urodzonych w kolejnych latach w określonym okresie, jest jedynie przykładem takiej specyfikacji.

Tabela 10

Przykładowa specyfikacja operacji bazodanowej



Ope/001 LiczbaNarozdin

Opis: Należy wyznaczyć liczbę dzieci ze związkow formalnych i nieformalnych, urodzonych w poszczególnych latach, w określonym przez użytkowników okresie <od – do>

Klasy: Osoba(data urodzenia)

Związki: w związku(rodzaj), jest dzieckiem

Kryteria selekcji: (Osoba.dataUrodzenia >= :od) AND (Osoba.dataUrodzenia <= :do)

## 5. Podsumowanie

W rozdziale skupiono się na zagadnieniu modelowania danych na poziomie konceptualnym, jako że ten etap ma kluczowe znaczenie dla użyteczności i jakości projektowanej bazy danych. W celu oceny jakościowej modeli danych zaproponowano miary jakości, pozwalające klasyfikować modele danych pod względem jakościowym. Przedstawiony przykład modelowania danych na poziomie konceptualnym jest ilustracją problemów występujących w tym procesie. Do analizy został wybrany bardzo ograniczony wycinek rzeczywistości (Dane demograficzne ludności) (można go uznać za powszechnie znany i niewymagający szczegółowego objaśnienia). Pomimo że jest to perspektywa wąskiego problemu, analiza modelu danych utworzonego na podstawie reguł biznesowych pozwoliła, przez stosowanie reguł transformacji, wygenerować wiele konceptualnych modeli danych. Każdy z nich umożliwia realizację wymagań funkcjonalnych aplikacji bazodanowej (możliwość gromadzenia i udostępniania niezbędnych danych do realizacji wymaganych usług). Celem tego przykładu było zwrócenie uwagi na możliwość transformacji modeli danych na poziomie konceptualnym, wykorzystując refaktoryzację i uściślenie.

Reasumując, można stwierdzić, że:

- Modelowanie na poziomie konceptualnym wymaga dobrego zrozumienia modelowanej dziedziny.
- Modelowana dziedzina może być reprezentowana za pomocą wielu modeli danych.

Wybór modelu danych ze zbioru równoważnych modeli pod względem funkcjonalnym powinien być dokonany na podstawie kryteriów jakościowych (ocena w kontekście wymagań niefunkcjonalnych).

Do modelowanie danych wymagane są:

- jest wiedza inżynierska,
- umiejętność abstrakcyjnego myślenia,
- umiejętność identyfikacji bytów świata rzeczywistego, oraz
- precyzja w definiowaniu pojęć wycinka rzeczywistości.

## BIBLIOGRAFIA

1. Ambler S.: Refactoring Database: Evolutionary Database Design, Addison Wesley Professional, 2006.
2. Connolly T., Begg C.: Database Systems. A practical Approach to Design, Implementation, and Management, , 4<sup>th</sup> Edition, Pearson Education, Inc., 2005.
3. Dubielewicz I., Hnatkowska B., Huzar Z., Tuzinkiewicz L.: Metodyka projektowania aplikacji bazodanowych, Wydawnictwo Politechniki Wrocławskiej, 2009 (w druku).
4. Elmasri R., Navathe S.B.: Fundamentals of database systems, 5<sup>th</sup> Edition , Pearson Education, Inc., 2007.
5. Fowler M.: Refactoring. Improving the Design of Existing Code, Addison-Wesley, 2004.
6. Halpin T., Morgan T.: Information Modeling and Relational Databases, 2<sup>nd</sup> Edition, Morgan Kaufmann, 2008.
7. ISO/IEC 9126-1:2001(E), Software engineering – Product quality – Part 1: Quality model.
8. ISO/IEC 25000:2005, Software engineering – Software Quality and Requirements Evaluation (SQuaRE) – Guide to SQuaRE.

Recenzent: Dr inż. Piotr Bajerski

Wpłynęło do Redakcji 3 marca 2009 r.

## Abstract

Database design may be considered as a data modeling process at different levels of abstraction. Initial conceptual model represents a domain problem. Next data model, at logical level, results from transformation of conceptual model. Similarly, data model at implementation level results from transformation of logical model. The conceptual model has

different nature from the other models. It is expected that the model should be consistent with described domain, and serves for understanding the domain; its construction depends on domain-oriented knowledge. The other models are expected to have high quality; their construction depends strongly on engineering-oriented knowledge. Functional and non-functional requirements being inseparable element of database design form the base for definition of models quality.

In the article different aspects of data modeling are discussed. Key problems specific for this process are explained. Additionally the measures of quality data model evaluation are defined. An example of development of conceptual data models is shown. An approach to improving the quality of the data models by using refactoring and refining (?) is discussed.

On the background of the presented approach the paper brings some conclusions relating to the following questions:

- • What skills are necessary to model data at the conceptual level?
- • Is it enough to have an engineering skills only in database design?
- • To what extent database design may be considered as engineering and as an art?

All the questions have implied from the experience gained by the author in realization of various IT projects.

## **Adres**

Lech TUZINKIEWICZ: Politechnika Wrocławska, Instytut Informatyki, ul. Wybrzeże Wyspiańskiego 27, 50-327 Wrocław, Polska, lech.tuzinkiewicz@pwr.wroc.pl.