

**Prace Naukowe
Wydziałowego Zakładu Informatyki
Politechniki Wrocławskiej**



3

BAZY DANYCH

Redakcja naukowa

Zygmunt Mazur



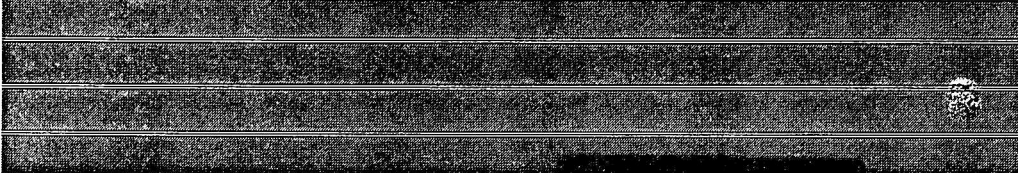
**Oficyna Wydawnicza
Politechniki Wrocławskiej**





Prace Naukowe
Wydziałowego Zakładu Informatyki
Politechniki Wrocławskiej

3 BAZY DANYCH



Redakcja naukowa

Zygmunt Mazur



Oficyna Wydawnicza Politechniki Wrocławskiej
Wrocław 2002



WYDZIAŁOWY ZAKŁAD INFORMATYKI

WYDZIAŁ INFORMATYKI I ZARZĄDZANIA
POLITECHNIKI WROCŁAWSKIEJ

Wybrzeże Wyspiańskiego 27, 50-370 Wrocław

Recenzent

Jerzy SURMA

Opracowanie redakcyjne materiałów

Hanna MAZUR

Projekt okładki

Zygmunt MAZUR

© Copyright by Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2002

OFICyna WYDAWNICZA POLITECHNIKI WROCŁAWSKIEJ
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław



3 1 2 5 0 6

/6

ISBN 83-7085-678-0

Drukarnia Oficyny Wydawniczej Politechniki Wrocławskiej, Zam. nr 922/2002.

**PROJEKTOWANIE
ROZMYTYCH MODELI
RELACYJNYCH BAZ DANYCH**

PROJEKTOWANIE ROZMYTYCH MODELI RELACYJNYCH BAZ DANYCH

Prace przedstawiają metodologie projektowania rozmytych modeli relacyjnych baz danych. Główna uwaga została poświęcona projektowaniu koncektualnemu i logicznemu. Zaprezentowany został rozmyty model związków encji, sposoby reprezentacji różnego rodzaju rozmytych danych w modelu relacyjnym przedstawiające rozszerzenia modelu klasycznego o elementy rozmyte. Dużo uwagi zostało również poświęcone projektowaniu schematów rozmytych relacyjnych baz danych. Podano zasady oraz wskazówki, jak powinny być tworzone schematy, aby zminimalizować problem redundancji i anomalii przechowywanych danych. Dodatkowo zostały przedstawione próby rozszerzenia algebry relacji i języka SQL, aby móc operować rozmytym modelem relacyjnych baz danych.

FUZZY RELATIONAL DATABASE MODELING

Works presents methodologies of fuzzy relational database modeling. Most attention has been paid to conceptual and logical modeling. Fuzzy entity-relationship model and representations of various forms of fuzzy data in relational model addressing extensions of classical model with fuzzy elements have been presented. A great deal of the work concerns schemas of fuzzy relational database design. Principles and suggestions how to create schema in order to minimize redundancy and avoid anomalies in stored data have been included. In addition, the approaches of extending relational algebra and SQL allowing to operate on fuzzy relational database have been mentioned.

SPIS TREŚCI

<i>Wstęp</i>	7
Biedrzycka B., Mazur Z., <i>Podstawy projektowania relacyjnych baz danych</i>	11
Biedrzycka B., Mazur Z., <i>Logika rozmyta</i>	27
Biedrzycka B., Mazur Z., <i>Rozmyty model związków encji</i>	43
Biedrzycka B., Mazur Z., <i>Reprezentacja danych w rozmytych modelach relacyjnych baz danych</i>	59
Biedrzycka B., Mazur Z., <i>Operowanie danymi rozmytymi</i>	77
Biedrzycka B., Mazur Z., <i>Projektowanie schematów rozmytych relacyjnych baz danych</i>	91
<i>Literatura</i>	121

CONTENTS

<i>Introduction</i>	7
Biedrzycka B., Mazur Z., <i>Principles of Relational Databases Design</i>	11
Biedrzycka B., Mazur Z., <i>Fuzzy Logic</i>	27
Biedrzycka B., Mazur Z., <i>Fuzzy Entity-Relationship Model</i>	43
Biedrzycka B., Mazur Z., <i>Representation of Data in Fuzzy Database Relational Model</i>	59
Biedrzycka B., Mazur Z., <i>Operating of Fuzzy Data</i>	77
Biedrzycka B., Mazur Z., <i>Design of Fuzzy Relational Database</i>	91
<i>Bibliography</i>	121

*Cel zamierzony
i cel osiągnięty,
to nigdy nie jest to samo*
• Sofokles

WSTĘP

Tematyka trzeciego Zeszytu Naukowego „Bazy Danych” wiąże się z projektowaniem rozmytych modeli relacyjnych baz danych.

Klasyczne modele danych nie umożliwiają reprezentacji i operowania niepewną, nieprecyzyjną informacją. Od 1980 roku logika rozmyta, wprowadzona przez Zadeha w 1965 roku, jest używana do rozszerzenia modeli danych tak, aby mogły one modelować różnego rodzaju niepewne i nieprecyzyjne informacje, którymi posługują się ludzie w życiu codziennym. Rozmytość własności leży w braku dobrze zdefiniowanych granic zbioru obiektów, do których odnosi się własność. Główna uwaga naukowców została poświęcona modelowi relacyjnemu Codda, w wyniku czego powstało kilka rozmytych relacyjnych modeli danych.

Celem prac jest przedstawienie przede wszystkim metodologii projektowania konceptualnego i logicznego dla rozmytych modeli relacyjnych baz danych. Projektowanie konceptualne ukazuje sposób postępowania z danymi rozmytymi na etapie tworzenia modelu konceptualnego. Uwaga została tu skoncentrowana na rozmytym modelu związków encji. W ramach projektowania logicznego zostały zaprezentowane sposoby reprezentacji różnego rodzaju rozmytych danych, które przedstawiają rozszerzenia klasycznego modelu reprezentacji danych o elementy rozmyte. Dużo uwagi zostało również poświęcone teorii projektowania schematów rozmytych baz danych, która dostarcza zasady i wskazówki, jak powinny być one tworzone, aby zminimalizować problem redundancji danych i anomalii. Aby móc operować rozmytym modelem relacyjnych baz danych, zostały zaprezentowane próby rozszerzenia algebry relacji i języka SQL.

W artykule *Podstawy projektowania relacyjnych baz danych* podano podstawowe pojęcia dotyczące projektowania modelu relacyjnego i opisano model związków encji – syntaktykę i semantykę elementów składowych, zilustrowano jego użycie przy tworzeniu modelu konceptualnego rozważanej rzeczywistości. Przedstawiono również podstawowe elementy relacyjnego modelu danych.

Artykuł *Logika rozmyta* opisuje podstawowe idee logiki rozmytej, wprowadza pewne elementy istotne dla rozmytego rozszerzenia modeli danych, m.in. definiuje zbiór rozmyty i rozkład możliwości, przedstawia jednostkę reprezentacji wiedzy jaką jest zmienna lingwistyczna, wprowadza relacje rozmyte.

Artykuł *Rozmyty model związków encji* podejmuje tematy związane z rozmytym modelowaniem konceptualnym. Opisuje, jak logika rozmyta została użyta w modelu związków encji. Przedstawia rozmyte rozszerzenie pojęć klasycznego modelu związków encji takich jak: encje, związki, atrybuty na trzech poziomach – poziomach typów i ich instancji. Dostarcza szerszego spojrzenia na związki pomiędzy encjami w kontekście uczestnictwa i liczebności.

Artykuł *Reprezentacja danych w rozmytych modelach relacyjnych baz danych* ukazuje sposób rozszerzenia klasycznego modelu reprezentacji danych o elementy rozmyte. Przedstawia różne podejścia do reprezentacji rozmytych danych, a także porusza dla nich problem bliskości i redundancji danych, wskazując istniejące rozwiązania.

W artykule *Operowanie danymi rozmytymi* uwaga została skoncentrowana na operowaniu danymi w rozmytych modelach relacyjnych baz danych. Została przedstawiona rozszerzona algebra relacji dla każdego z podejść do reprezentacji rozmytych danych oraz próby rozszerzenia języka SQL, których wynikiem jest język SQLf.

Artykuł *Projektowanie schematów rozmytych relacyjnych baz danych* podejmuje w ramach rozszerzonego podejścia opartego na możliwości (ang. *extended possibility-based framework*) tematykę projektowania schematów rozmytych relacyjnych baz danych, która jest związana z teorią normalizacji dla rozmytych modeli relacyjnych baz danych. Opiera się ona na pojęciach rozmytych zależności funkcyjnych i rozkładzie schematu relacji. Artykuł ten przedstawia rozmyte zależności funkcyjne, które opisują ograniczenia semantyczne pomiędzy atrybutami, odzwierciedlają związek pomiędzy nieprecyzyjnymi wartościami atrybutów, wprowadza pojęcie θ -klucza oraz rozszerzone reguły integralności konkretności i odwołań, prezentuje rozmyte rozszerzenie aksjomatów Amstronga, tematykę domknięcia zbioru rozmytych zależności funkcyjnych i równoważności zbiorów rozmytych zależności funkcyjnych, które mają istotne znaczenia dla teorii projektowania rozmytych relacyjnych baz danych. Zostały zaprezentowane również dwa rodzaje rozkładów schematu: odwracalny i zachowujący zależności. Praca przedstawia problem anomalii, pojawiających się przy operacjach wstawiania, usuwania i modyfikacji. Definiuje także θ -rozmyte postacie normalne (F1NF, F2NF, F3NF, FBCNF), nakładające ograniczenia na obecność częściowych i przechodnich rozmytych zależności funkcyjnych w schemacie w taki sposób, że problem redundancji danych i anomalii może zostać usunięty. W pracy zostały również umieszczone algorytmy projektowania schematu rozmytej bazy danych, w wyniku których uzyskuje się schematy w odpowiednich rozmytych postaciach normalnych, uwzględniając własność odwracalności lub zachowania zależności.

Na końcu Zeszytu zamieszczono spis literatury do wszystkich artykułów.

Zaprezentowane w niniejszym zeszycie artykuły są wynikiem prac prowadzonych w Wydziałowym Zakładzie Informatyki na Wydziale Informatyki i Zarządzania Politechniki Wrocławskiej.

Autorzy składają podziękowanie Recenzentowi za uwagi przedstawione w recenzjach, które zostały uwzględnione w trakcie przygotowywania ostatecznych wersji artykułów.

Zygmunt Mazur

*baza danych, model relacyjny,
schemat relacji*

Barbara BIEDRZYCKA
Zygmunt MAZUR*

PODSTAWY PROJEKTOWANIA RELACYJNYCH BAZ DANYCH

W niniejszym artykule przedstawiono podstawowe pojęcia dotyczące projektowania modelu relacyjnego i opisano model związków encji – syntaktykę i semantykę elementów składowych oraz zilustrowano jego użycie przy tworzeniu modelu konceptualnego rozważanej rzeczywistości. Przedstawiono również podstawowe elementy relacyjnego modelu danych.

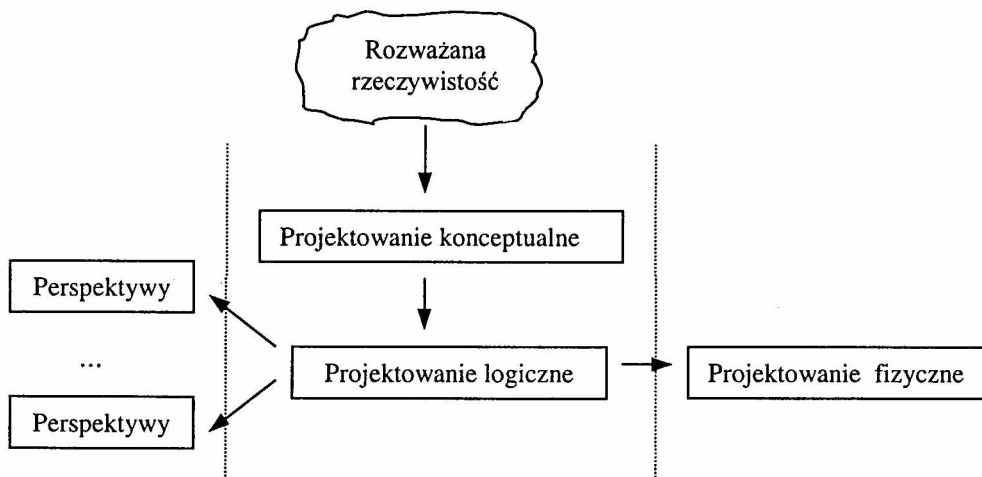
1. WPROWADZENIE

Projektowanie modeli baz danych jest procesem mającym na celu utworzenie modelu bazy danych, który reprezentowałby odpowiedni wycinek rzeczywistości. Na jego przebieg składają się takie etapy jak: analiza rozważanej rzeczywistości, projektowanie konceptualne, projektowanie logiczne, definiowanie perspektyw i projektowanie fizyczne (rysunek 1).

Celem analizy rozważanej rzeczywistości jest rozpoznanie i opisanie rzeczywistości, która ma zostać zamodelowana, tzn. zebranie informacji umożliwiających zbudowanie modelu bazy danych, określenie granic dziedziny, którą należy uwzględnić w systemie. Projektowanie konceptualne ma na celu zbudowanie modelu konceptualnego (ang. *Conceptual Data Model*), który przedstawia, jakie obiekty występują w rozważanej rzeczywistości, opisuje ich cechy i definiuje występujące między nimi związki. Popularnym modelem konceptualnym jest model związków encji. Celem projektowania logicznego jest zbudowanie modelu logicznego. Na tym etapie dokonuje się wyboru modelu danych, na przykład modelu relacyjnego, określa się dokładnie typy danych. Dla modelu relacyjnego definiuje się

* Wydziałowy Zakład Informatyki, Wydział Informatyki i Zarządzania Politechniki Wrocławskiej, 50-370 Wrocław, Skwer Idaszewskiego 1, mazur@ci.pwr.wroc.pl

schemat bazy danych, przy czym istotną rolę odgrywa zaprojektowanie schematów relacji w taki sposób, aby nie pojawiły się problemy typu redundancja i anomalie danych.



Rys. 1. Etapy projektowania baz danych

Projektowanie schematów relacyjnych baz danych jest związane z teorią normalizacji. Definicja perspektyw polega na wyszczególnieniu grup użytkowników, ich uprawnień decydujących o możliwości dostępu do odpowiednich danych i wykonywania odpowiednich operacji. Przy projektowaniu fizycznym, w wyniku którego powstaje fizyczny model danych (ang. *Physical Data Model*) określa się sposób organizacji pamięci, definiuje się indeksy.

2. MODEL ZWIĄZKÓW ENCJI

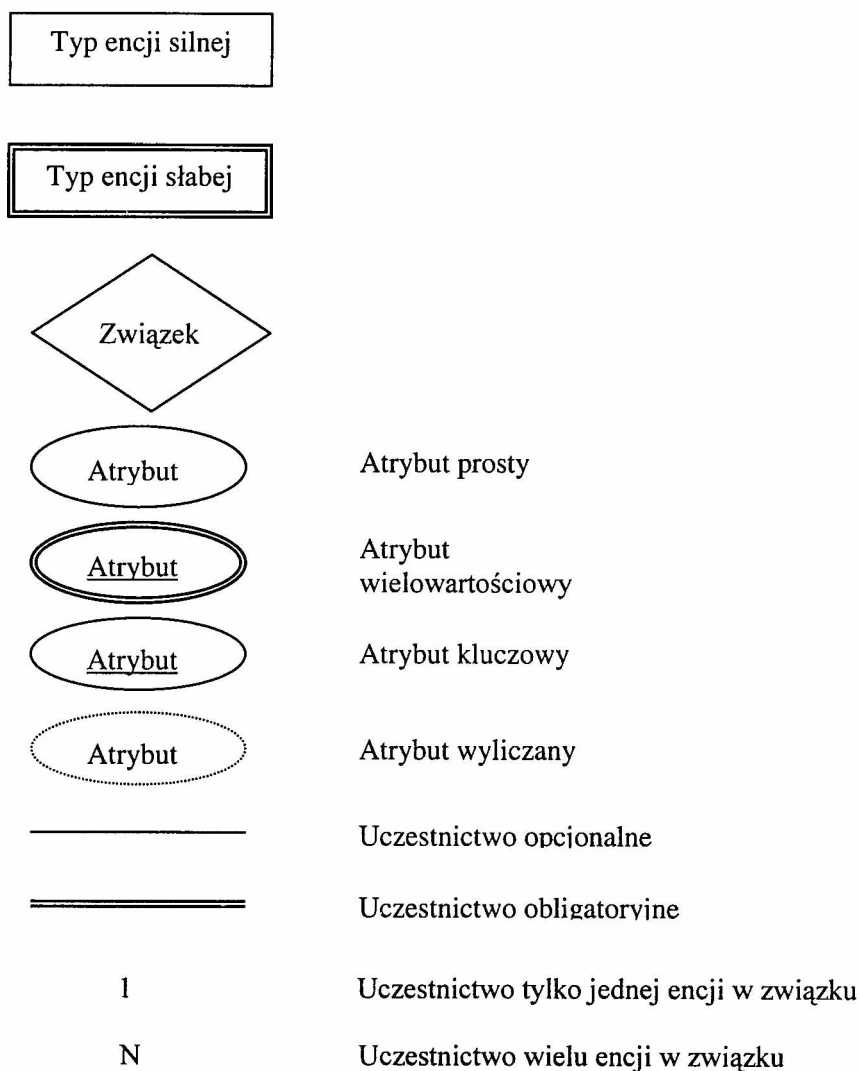
Popularnie stosowanym modelem koncepcyjnym jest model związków encji (ang. *Entity Relationship Model*). Jest on szczególnie stosowany w sytuacjach, kiedy wiadomo, że modelem logicznym będzie model relacyjny, znane są powszechnie reguły transformacji.

Relacyjny model danych jest modelem danych, który reprezentuje bazę danych w postaci zbioru relacji.

Model związków encji został wprowadzony przez Chena w 1976 roku [CHEN 76]. Przez ostatnie dekady był on stale rozwijany. Wprowadzono nowe pojęcia, głównie związane z dziedziczeniem, które wzbogaciły jego możliwości przedstawiania świata rzeczywistego.

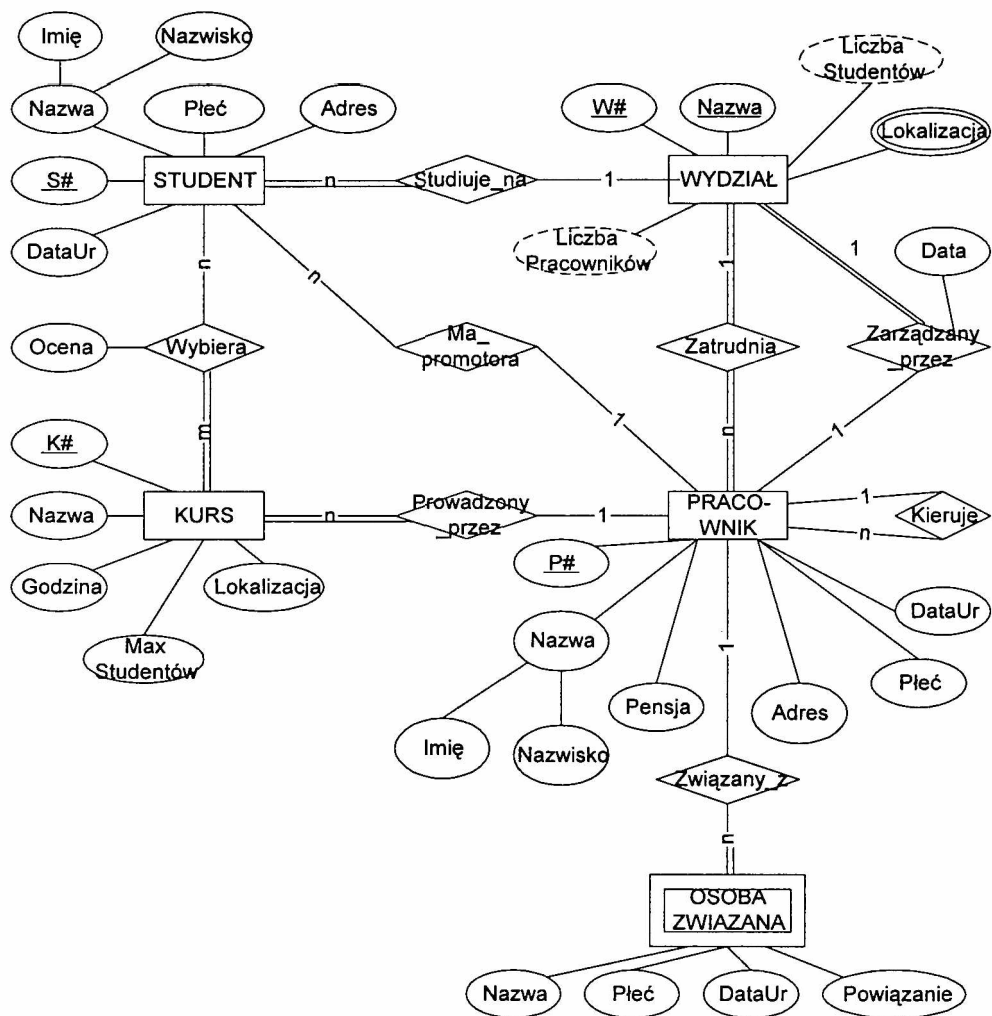
2.1. NOTACJA DIAGRAMÓW ZWIĄZKÓW ENCJI

Model danych można reprezentować przy użyciu konstrukcji diagramu związków encji. Nie istnieje jednak żadna ich standardowa notacja. W pracy będzie używana notacja zaproponowana przez Chena [CHEN 98], przedstawiona na rysunku 2.



Rys. 2. Notacja diagramów związków encji

Przykład diagramu związków encji przedstawia rysunek 3 [CHEN 98].



Rys. 3. Przykładowy diagram związków encji, reprezentujący uniwersytet
(Na podstawie [Chen 98])

Diagram przedstawiony na rysunku 3 reprezentuje rzeczywistość uniwersytetu obejmującą informację o studentach, kursach, wydziałach oraz pracownikach naukowych wydziałów, opartą na następujących założeniach:

1. Uniwersytet jest zorganizowany w wydziały, mające unikalną nazwę oraz numer, kierowane są przez jednego pracownika. Wydział może mieć kilka lokalizacji.

2. Pracownik jest zatrudniony przez jeden wydział i charakteryzowany jest takimi cechami jak: nazwa (imię, nazwisko), numer identyfikacyjny, adres, płeć, pensja, data urodzenia; może kierować grupą pracowników.
3. Pracownik może prowadzić kilka kursów mających swoją nazwę, unikalny numer, godzinę, maksymalne ograniczenie liczby studentów, lokalizację (dany kurs jest prowadzony tylko w jednym miejscu).
4. Student jest charakteryzowany przez nazwę (imię, nazwisko), numer, płeć, datę urodzenia, adres. Studiuje on na jednym wydziale i może uczęszczać na kilka kursów. Z każdego wybranego przez niego przedmiotu posiada ocenę. Student, przygotowując pracę dyplomową, jest związany z promotorem.

Pożądana jest również informacja o osobach powiązanych z pracownikiem, z którymi można byłoby się skontaktować w nagłych wypadkach. Osoba ta jest charakteryzowana nazwą, płcią, datą urodzenia oraz rodzajem powiązania z pracownikiem.

2.2. PODSTAWOWE POJĘCIA

Wycinek świata rzeczywistego w modelu związków encji jest reprezentowany przez jego składowe, jakimi są: encja, związek i atrybut.

Encja jest wielkością („rzeczą”) występującą w świecie rzeczywistym, która istnieje niezależnie, jest jednoznacznie identyfikowana. Przykładem encji jest dany student czy kurs. Encje charakteryzujące się tymi samymi własnościami stanowią typ encji. Można powiedzieć, że typ encji jest kategorią, a encja jego instancją. Na rysunku 3 typami encji są między innymi STUDENT, KURS.

Czasami encje nie mogą istnieć niezależnie lub samodzielnie nie są całkowicie identyfikowane. Takie encje nazywa się encjami słabymi. Zbiór encji słabych, które są do siebie podobne, stanowi typ encji słabej, którego przykładem jest OSOBA ZWIĄZANA połączona z typem encji PRACOWNIK (rysunek 3).

Związek jest pewnym powiązaniem między encjami. Typ związku wiążący typy encji definiuje zbiór związków (instancji związku). Związek zawiera dokładnie jedną encję każdego uczestniczącego typu encji, reprezentując istniejący w rozważanej rzeczywistości fakt, że encje przynależne do związku są w pewien sposób wzajemnie ze sobą powiązane. Na przykład typ związku Wybiera między typami encji STUDENT i KURS wyraża semantykę świata rzeczywistego, że studenci wybierają kursy.

Typ związku może być stopnia n , co oznacza, że wiąże n typów encji i jest wówczas nazywany n -arnym typem związku. Związek między n przynależnymi encjami jest nazywany związkiem n -arnym. W większości zastosowań znaczną część semantyki świata rzeczywistego reprezentuje binarny typ związku. Istnieje szczególnie

jego rodzaj – typ związku między tymi samymi typami encji. W takich związkach dwie encje jednego typu mogą odgrywać różne role. Na rysunku 3 typ związku *Kieruje* reprezentuje rolę grupy liderów i rolę członka grupy pełnione przez PRACOWNIKA.

Istnieją dwa pojęcia, które charakteryzują każdy związek. Noszą one nazwę uczestnictwa i liczebności.

Uczestnictwo określa udział encji w związku. Jeżeli każda encja typu E uczestniczy w odpowiednim związku typu R, wtedy E **obligatoryjnie** (całkowicie) uczestniczy w R. Jeżeli istnieje co najmniej jedna wartość typu E, która nie przynależy do żadnego związku typu R, to E **opcjonalnie** (częściowo) uczestniczy w R. Rozważając przykład uniwersytetu (rysunek 3), każdy kurs musi być wybrany przez studentów i prowadzony przez pracownika, a stąd wynika, że Kurs obligatoryjnie uczestniczy w typach związków Wybiera, Prowadzony_przez. Ponadto każda encja słaba musi być związana z encją ją identyfikującą, tak więc każdy typ encji słabej ma obligatoryjne uczestnictwo w swoim „identyfikującym” typie związku. Stąd OSOBA ZWIĄZANA obligatoryjnie uczestniczy w Związany_z. Na rysunku 3 występują również opcjonalne uczestnictwa, na przykład student może nie zapisać się na żaden kurs, a pracować nad pracą dyplomową, dlatego też STUDENT ma opcjonalne uczestnictwo w Wybiera. Obligatoryjne lub opcjonalne uczestnictwo zależy od kontekstu i powinno być właściwie wyznaczone, żeby odzwierciedlać semantykę świata rzeczywistego.

Liczebność (stopień związku) określa liczbę encji uczestniczących w związkach danego typu związku z daną encją. Typ związku R pomiędzy typami encji E_1 i E_2 może mieć liczebność:

- 1:1** (jeden do jeden) – z każdą encją, z któregośkolwiek uczestniczącego typu, może być związana co najwyżej jedna encja drugiego uczestniczącego typu.
- 1:n** (jeden do wiele) – encja e typu E_1 może być związana z pewną liczbą encji typu E_2 , a encja typu E_2 może być związana z co najwyżej jedną encją typu E_1 ,
- n:m** (wiele do wiele) – nie występują tu żadne ograniczenia dotyczące liczby encji mogących uczestniczyć w związkach danego typu związku z daną encją.

Liczebność 1:n dla typu związku $R(E_1, E_2)$ jest przedstawiona na diagramie poprzez wprowadzenie 1 po stronie E_1 i n po stronie E_2 , co oznacza, że dla jednej instancji encji e_1 typu E_1 może istnieć kilka instancji typu E_2 związanych z e_1 , a dla instancji encji e_2 typu E_2 może istnieć dokładnie jedna instancja typu E_1 uczestnicząca w związku z e_2 . Liczebność często wyraża się w zapisie związku, na przykład rozważany związek R może zostać zapisany w postaci $R(E_1(1,1), E_2(0,n))$. Dla każdego uczestniczącego typu encji jest zapisywana minimalna i maksymalna liczba encji mogących wystąpić w związkach danego typu związku z daną encją. Liczebność

zależy od kontekstu. Kurs musi być prowadzony przez jednego pracownika, co prowadzi do liczebności tego związku 1:n. Jeżeli kilku pracowników mogłoby prowadzić ten sam kurs, wtedy poprawną liczebnością między KURSEM i PRACOWNIKIEM byłaby n:m.

Atrybuty są własnościami encji lub związków. Na przykład encja student jest opisywana przez nazwę, S# (numer studenta), płeć, datę urodzin, adres. Atrybuty (nazwy atrybutów) są związane z typami encji, a wartości atrybutów z instancjami typów encji (encjami).

Typ encji zwykle ma jeden lub więcej atrybutów identyfikujących encje, tzn. ich wartości są różne dla każdej encji. Takie atrybuty są nazywane atrybutami **kluczowymi**. Na przykład na rysunku 3 dla typu encji STUDENT atrybutem kluczowym jest S#.

Mówiąc o atrybutach, wyróżnia się różne ich rodzaje: jednowartościowe i wielowartościowe, proste i złożone, przechowywane i wyprowadzane.

Jeżeli wartość atrybutu dla każdej encji ma tylko jedną wartość, wówczas atrybut jest nazywany jednowartościowym, jeżeli zaś może mieć kilka wartości, to jest to atrybut wielowartościowy. Przykładem atrybutu jednowartościowego jest Płeć będący atrybutem typu encji STUDENT, ponieważ dla każdego studenta istnieje tylko jedna wartość opisująca tę cechę, atrybutem wielowartościowym jest Lokalizacja, będąca atrybutem WYDZIAŁ, ponieważ wydział może mieć kilka lokalizacji (rysunku 3).

Atrybuty, które są niepodzielne, biorąc pod uwagę znaczenie, nazywają się atrybutami prostymi. Jeżeli atrybut może być podzielony na prostsze komponenty, to jest nazywany atrybutem złożonym. Różnica między atrybutami prostymi a złożonymi często jest względna i zależy od kontekstu. Atrybut złożony jest użyteczny jako konstrukcja modelowa, kiedy czasami trzeba odwołać się do niego jako do całości, a w innych przypadkach do jego komponentów. Atrybutem złożonym jest na przykład Nazwa, będąca atrybutem STUDENTA, podzielona na Imię i Nazwisko.

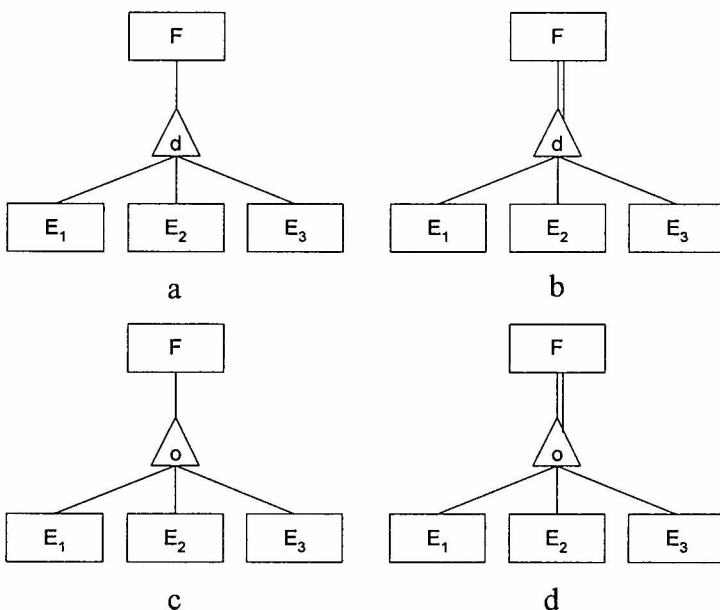
W pewnych przypadkach wartość atrybutu może wywodzić się z wartości innych atrybutów powiązanych encji lub związków. Taki atrybut jest nazywany atrybutem wyprowadzanym. Każdy atrybut, którego wartości stanowią podstawę dla otrzymania wartości atrybutu wyprowadzanego, nazywa się atrybutem przechowywanym. Na rysunku 3 przykładem atrybutu wyprowadzanego jest LiczbaPracowników, będąca atrybutem typu encji WYDZIAŁ, jego wartość jest uzyskiwana przez zliczenie liczby pracowników należących do danego wydziału.

Obok atrybutów związanych z jednym typem encji są również atrybuty związane jednocześnie z kilkoma typami encji – atrybuty typów związków. Przykładem może być atrybut Ocena typu związku Wybiera.

2.3. DZIEDZICZENIE

Model związków encji był przez lata rozwijany, aby można było lepiej opisać przy jego pomocy złożoność świata rzeczywistego. Wprowadzono do niego pojęcia podklasy (ang. *subclass*) i nadklasy (ang. *superclass*), specjalizacji (ang. *specialization*) i generalizacji (ang. *generalization*), podklasy dzielonej (ang. *shared subclass*) i kategorii (ang. *category*), dziedziczenia (ang. *inheritance*) atrybutów.

Podklasa jest typem encji reprezentującym podzbiór encji innego typu encji. Na przykład encje typu PRACOWNIK obejmują zarówno encje będące sekretarkami, technikami, jak i inżynierami. Te trzy podgrupy mogą być reprezentowane jako trzy podklasy PRACOWNIKA zwane SEKRETARKA, TECHNIK i INŻYNIER [ELMASRI 94]. Typ encji, do którego należą encje podklas jest nadklasą, stąd PRACOWNIK jest nadklasą SEKRETARKI, TECHNIKA i INŻYNIERA. Zgodnie z definicją, każda encja podklasy jest encją nadklasy. Notacje podklas E_1 , E_2 i E_3 i ich nadklasy F na diagramie związków encji przedstawia rysunek 4.



Rys. 4. Podklasa i nadklasa

W przypadku a i b podklasy E_1 , E_2 i E_3 są rozłączne (ang. *disjoint*), co oznacza, że encja nadklasy F może przynależeć maksymalnie do jednej z trzech podklas E_1 , E_2 i E_3 . Różnica między przypadkiem a i b polega na tym, że w przypadku a encja nadklasy F może nie należeć do żadnej ze swoich podklas E_1 , E_2 i E_3 , podczas gdy w przypadku b musi ona należeć do jednej z nich. Przypadki c i d na rysunku 4

reprezentują nakrywanie się (ang. *overlapping*) podklas E_1 , E_2 i E_3 . Encja nadklasy F może należeć do więcej niż jednej podklasy. Różnica między przypadkami c i d jest analogiczna do różnicy między a i b.

Proces definiujący podklasy z danego typu encji nazywa się **specjalizacją**, a proces do niego odwrotny, tzn. definiujący nadklasę z typów encji, nazywa się **generalizacją** (terminy specjalizacja i generalizacja są używane w dwóch podobnych znaczeniach:

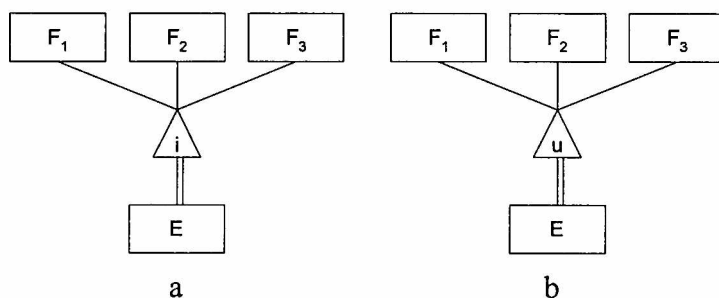
(1) do oznaczenia relacji pomiędzy klasą i jej klasą podrzędną/nadrzędną;

(2) do oznaczenia procesu tworzenia nowej podklasy/nadklasy [SUBIETA 99]).

Każdy z przypadków zaprezentowanych na rysunku 4 może być widziany jako specjalizacja $\{E_1, E_2 \text{ i } E_3\}$ nadklasy F lub jako generalizacja F podklas $E_1, E_2 \text{ i } E_3$.

Rysunki 4a i 4c reprezentują tak zwaną specjalizację lub generalizację opcjonalną (częściową), a rysunki 4b i 4d specjalizację lub generalizację obligatoryjną (całkowitą). Z danej nadklasy można zdefiniować kilka specjalizacji, biorąc pod uwagę różne własności nadklasy. Na przykład opierając się na rodzaju wykonywanego zawodu, PRACOWNIK może być specjalizowany na podklasy SEKRETARKA, TECHNIK i INŻYNIER, a uwzględniając liczbę godzin pracy, może być specjalizowany na podklasy PRACOWNIK PEŁNOETATOWY i PRACOWNIK PÓŁETATOWY.

Do modelu związków encji zostały również wprowadzone **podklasy dzielone** i **kategorie**, mające więcej niż jedną nadklasę. Każda encja podklasy dzielonej należy do wszystkich swoich nadklas, podczas gdy encja kategorii należy do co najmniej jednej ze swoich nadklas. Przykładem podklasy dzielonej może być ASYSTENT będący podklasą OSOBY ZATRUDNIONEJ i STUDENTA, ponieważ asystent jest studentem prowadzącym zajęcia. Przykładem kategorii może być UCZESTNIK PROJEKTU będący podklasą STUDENTA i PRACOWNIKA. Jak pisze Chen [CHEN 98], podklasa dzielona jest podzbiorem iloczynu (ang. *intersections*) swoich nadklas, a kategoria jest podzbiorem sumy (ang. *union*) nadklas. Rysunek 5 przedstawia notacje dzielonej podklasy i kategorii.



Rys. 5. Dzielona podklasa i kategoria

W modelu związków encji mówi się również o **dziedziczeniu atrybutów**. Podklasy mają dwa rodzaje atrybutów: atrybuty, które są charakterystyczne tylko dla nich i atrybuty odziedziczone z ich nadklas. Podklasa SEKRETARKA posiada zatem wszystkie atrybuty nadklasy PRACOWNIK i dodatkowo swoje własne, na przykład PisanieNaMaszynie.

Podklasy dzielone dziedziczą wszystkie atrybuty nadklas z nimi związanych. Taki rodzaj dziedziczenia atrybutów jest nazywany **dziedziczeniem wielokrotnym** (ang. *multiple inheritance*). Inny rodzaj dziedziczenia atrybutów jest związany z kategoriami. Kategoria dziedziczy atrybuty swoich nadklas w taki sposób, że encja kategorii ma wartość dla każdego atrybutu nadklasy, do której encja należy. Inaczej mówiąc, kategoria dziedziczy selektywnie atrybuty jej nadklas, co daje podstawę dla **dziedziczenia selektywnego** (ang. *selective inheritance*). Na przykład podklasa dzielona ASYSTENT dziedziczy wszystkie atrybuty STUDENTA i OSOBY ZATRUDNIONEJ (dziedziczenie wielokrotne), podczas gdy kategoria UCZESTNIK PROJEKTU dziedziczy selektywnie atrybuty STUDENTA lub PRACOWNIKA, zależnie od indywidualnej przynależności encji albo do STUDENTA, albo do PRACOWNIKA. Jeżeli na przykład Kowalski Jan uczestniczy w danym projekcie będąc studentem, wówczas będzie posiadał wartości dla wszystkich atrybutów STUDENTA ale nie będzie miał żadnej wartości atrybutów PRACOWNIKA.

3. RELACYJNY MODEL DANYCH

Relacyjny model danych, wprowadzony przez Codd'a w 1970 roku [CODD 70], reprezentuje bazę danych w postaci zbioru relacji. Model relacyjny powstał jako alternatywne rozwiązanie w stosunku do modelu sieciowego i hierarchicznego. Po okresie dużej fascynacji modelem relacyjnym w latach siedemdziesiątych, co wynikało głównie z jego dobrych podstaw teoretycznych i elegancji matematycznej, w latach osiemdziesiątych nastąpiła eksplozja systemów zarządzania bazami danych opartych na tym modelu. Prace teoretyczne zainspirowane relacyjnym modelem danych, jak również realizowane systemy na nim oparte, wniosły bardzo wiele do lepszego zrozumienia procesów formalnego reprezentowania wiedzy o świecie rzeczywistym.

3.1. RELACJE

Model relacyjny oparty jest na zaczerpniętym z teorii mnogości pojęciu relacji, będącej podzbiorem iloczynu kartezjańskiego kolekcji dziedzin.

Dziedzina (ang. *domain*) jest zbiorem wartości, na przykład zbiorem liczb całkowitych, zbiorem łańcuchów znaków, zbiorem $\{0, 1\}$.

Iloczyn kartezjański dziedzin $D_1, D_2, \dots, D_n$, zapisywany w postaci

$$D_1 \times D_2 \times \dots \times D_n,$$

jest zbiorem wszystkich n -krotek (ang. *tuples*) $(d_1, d_2, \dots, d_n)$ takich, że d_i należy do D_i (dla $i = 1, \dots, n$).

Relacja R jest dowolnym podzbiorem iloczynu kartezjańskiego jednej lub więcej dziedzin $R \subseteq D_1 \times D_2 \times \dots \times D_n$.

W odniesieniu do baz danych rozważane są tylko relacje skończone. Elementy relacji są nazywane krotkami. O każdej relacji będącej podzbiorem $D_1 \times D_2 \times \dots \times D_n$ mówi się, że ma stopień n (relacja n -arna). Relacja jest identyfikowana przez nazwę relacji.

Relację często reprezentuje się jako tabelę, w której każdy wiersz jest krotką, a każda kolumna odpowiada jednej składowej. Kolumnom nadaje się nazwy zwane atrybutami. Nazwa relacji wraz ze zbiorem nazw atrybutów relacji nazywa się **schematem relacji** (ang. *relation schema*).

Schemat relacji R o atrybutach $A_1, A_2, \dots, A_n$, zapisujemy w postaci

$$R(A_1, A_2, \dots, A_n) \text{ lub } R(U), \text{ gdzie } U = \{A_1, A_2, \dots, A_n\}.$$

Krotkę t można zapisać jako $t = (t(A_1), t(A_2), \dots, t(A_n))$, gdzie $t(A_i) \in D_i$ dla $i=1, \dots, n$.

Tabela 1 przedstawia relację, której atrybutami są: $K\#$, Nazwa, Płeć, Wiek, Wzrost, Zdrowie, stąd stopień jej równa się sześć. Wartości krotek dla atrybutów $K\#$, Nazwa, Płeć, Wiek, Wzrost, Zdrowie należą odpowiednio do dziedzin $D_{K\#}$, D_{Nazwa} , $D_{Płeć}$, D_{Wiek} , D_{Wzrost} , $D_{Zdrowie}$, przykładem krotki jest (138, Nowak, M, 59, 180, dobre). Jeżeli relacja ta zostanie nazwana Kontrahenci, to jej schemat można zapisać w postaci **Kontrahenci**($K\#$, Nazwa, Płeć, Wiek, Wzrost, Zdrowie), przy czym można go również zapisać jako Kontrahenci($K\#$, Nazwa, Płeć, Wiek, Wzrost, Zdrowie), drugą notację stosuje się zazwyczaj mając konkretny schemat odzwierciedlający pewną rzeczywistość.

Tabela 1. Relacja Kontrahenci

K#	Nazwa	Płeć	Wiek	Wzrost	Zdrowie
138	Nowak	M	59	180	Dobre
140	Anioł	K	17	175	Średnie

Relacje charakteryzują się następującymi własnościami:

1. Nie ma powtarzających się krotek.
2. Krotki występują w dowolnym porządku.
3. Atrybuty występują w dowolnym porządku.
4. Wartości atrybutów są wartościami atomowymi.

Zbiór schematów relacji użytych do przedstawienia informacji jest nazywany schematem bazy danych, a bieżące wartości odpowiednich relacji noszą nazwę bazy danych (stanem bazy danych).

Dla danej relacji definiuje się **zbiór identyfikujący relację** (ang. *identifying set of relation*) jako zbiór atrybutów danej relacji, których kombinacje jednoznacznie identyfikują każdą krotkę relacji [CELL 88]. Z kolei **klucz** (ang. *key*) jest to taki zbiór identyfikujący relacji, którego żaden podzbiór nie jest zbiorem identyfikującym, czyli jest to atrybut (lub kilka atrybutów), którego wartość jednoznacznie identyfikuje każdą krotkę w relacji. Dla jednej relacji może istnieć kilka kluczy, które jednoznacznie identyfikują krotki relacji. Nazywa się je **kluczami kandydującymi** (ang. *candidate key*). Klucz kandydujący nie może przyjmować wartości pustej (null). Spośród kluczy kandydujących wybiera się jeden tzw. **klucz główny** (ang. *primary key*) oznaczany jako PK. Przykładem klucza jest atrybut K# dla schematu relacji Kontrahenci przedstawionej w tabeli 1. Klucz może być **prosty** (ang. *simple key*) lub **złożony** (ang. *complex key*). Klucz jest prosty, jeśli zbiór identyfikujący relacji jest jednoelementowy, czyli jeśli kluczem jest jeden atrybut. Klucz jest złożony, jeśli zbiór identyfikujący relacji nie jest zbiorem jednoelementowym, czyli jeśli klucz składa się z kilku relacji. W przypadku gdy klucze kandydujące są bardzo złożone lub brak jest klucza głównego, wprowadza się **sztuczny klucz główny** (ang. *artificial primary key*). Słowo „sztuczny” oznacza, że klucz ten nie pojawił się w sposób naturalny i trzeba go dołączyć do atrybutów relacji, np. Numer_utworu. Zbiór atrybutów, który zawiera klucz nazywany jest **nadkluczem** (ang. *superkey*) i jest oznaczany przez SK.

3.2. REGUŁY INTEGRALNOŚCI

Wartości występujące w bazie danych mają sens, jeżeli reprezentują określoną rzeczywistość. W związku z tym definicja bazy danych wymaga rozszerzenia o reguły (reguły integralności), których zadaniem jest informowanie systemu zarządzania bazą danych o ograniczeniach świata rzeczywistego. Wyróżnia się reguły specyficzne dla konkretnej bazy danych, na przykład ograniczenia dziedzinowe oraz reguły ogólne, stosowane dla wszystkich baz danych – reguły klucza głównego i obcego. Pojęcie integralność związane jest z określeniem dokładności i poprawności w bazie danych [Date 00].

Ograniczenia dziedzinowe specyfikują naturę, typy, zakres wartości atrybutów. Każda wartość atrybutu musi być wartością atomową z odpowiadającej jej dziedziny, na przykład atrybut Rok studiów musi być liczbą całkowitą z przedziału [1, 5].

Z pojęciem klucza jest związana jedna z dwóch reguł integralności zwana regułą integralności konkretności. Mówi ona, że żaden komponent klucza głównego relacji bazowej nie może przyjmować wartości nieokreślonej (null). Wynika to z tego, iż relacja bazowa i jej krotki odpowiadają konkretnym obiektom w świecie rzeczywistym, z definicji każdy obiekt musi być rozróżnialny, tzn. możliwy do zidentyfikowania, a identyfikację zapewnia klucz główny. Jeżeli null oznacza wartość „niedotyczy”, tzn. nie można jej określić, to nie może być elementem klucza głównego. Jeśli null oznacza wartość nieznaną, to taki obiekt również nie jest znany.

Do modelu relacyjnego zostało wprowadzone również pojęcie **klucza obcego** (ang. *foreign key*), oznaczanego przez FK. Intuicyjnie kluczem obcym jest zbiór atrybutów, w szczególności jednego, pewnego schematu relacji **R'**, których wartości są wymagane dla wskazania krotki w relacji **R**, będącej bieżącą wartością schematu relacji **R** o kluczu głównym, którego wartości są zgodne z wartościami atrybutów klucza obcego. Inaczej mówiąc, atrybut FK, który może być złożony, jest kluczem obcym wtedy i tylko wtedy, jeśli posiada następujące, czasowo niezależne, własności:

1. Każda wartość FK jest równa null lub w pełni różna od null, w pełni oznacza, że w przypadku, gdy FK jest wieloatrybutowy, to każdy atrybut składowy przyjmuje wartość null lub jest różny od null.
2. Istnieje relacja bazowa **R** będąca bieżącą wartością schematu relacji **R** o kluczu głównym PK, że wartość FK różna od null jest identyczna z wartością PK krotki relacji **R**.

Na przykład schemat bazy danych może składać się ze schematów relacji:

Kontrahenci(K#, Nazwa, Płeć, Wiek, Wzrost, Zdrowie),

Asortyment(A#, Nazwa, Rozmiar, Kolor, Model)

Zamówienia(K#, A#, Cena, Jakość).

Relacje będące ich wartościami reprezentują informacje odpowiednio o kontrahentach firmy, asortymencie, którym się zajmuje i zamówieniach składanych przez kontrahentów na danych asortyment (dany produkt może być zamówiony przez więcej niż jednego kontrahenta i kontrahent może zamówić więcej niż jeden produkt). Atrybut K#, podobnie jak A#, jest kluczem obcym schematu relacji Zamówienia, ponieważ K# jest kluczem schematu relacji Kontrahenci i A# jest kluczem schematu relacji Asortyment. Podsumowując dotychczasowe rozważania można powiedzieć, że klucz obcy jest to atrybut lub kilka atrybutów relacji, które są kluczem głównym w innej relacji.

Z pojęciem klucza obcego jest związana reguła integralności zwana regułą integralności odwołań. Mówi ona, że baza danych nie może zawierać kluczy obcych, których wartości nie mają odpowiedników w kluczach głównych.

3.3. OPERACJE NA RELACJACH

Operowanie danymi, m. in. wyszukiwanie danych, w relacyjnym modelu danych jest wykonywane przy użyciu algebry relacji. Podstawowymi operatorami są:

- projekcja (PROJECT) zwana również rzutem i oznaczana symbolem Π ,
- selekcja (SELECT) oznaczana symbolem σ ,
- łączenie naturalne (JOIN) oznaczane symbolem $*$.

Dzięki tym operatorom można wykonać większość operacji na danych wymaganych od systemu relacyjnego. Dodatkowe operatory: suma ($\cup$), różnica ($-$), iloczyn ($\cap$) i iloczyn kartezjański (produkt kartezjański) ($\times$) są modelowane na

tradycyjnych operatorach teorii zbiorów. Każdy operator bierze jedną lub więcej relacji jako argument i daje w wyniku relację. Definicje operatorów można znaleźć w pracach [DATE 00, ULLMAN 88, PANKOWSKI 92, BEYNON-DAVIES 98, ULLMAN 99].

Najbardziej popularnym mechanizmem definiowania poleceń zapytań i modyfikacji danych w relacyjnych systemach baz danych jest język SQL. Podstawowy jego rdzeń jest implementacją algebry relacji, ale zawarte są w nim również elementy daleko odbiegające od tego, co można znaleźć w algebrze relacji, na przykład możliwość agregowania danych, a także możliwość modyfikowania bazy danych. Omówienie zapytań języka SQL znajduje się w pracach [GRUBER 96, ULLMAN 99].

3.4. PROJEKTOWANIE SCHEMATÓW RELACYJNYCH BAZ DANYCH

Podstawową ideą projektowania schematów relacyjnych baz danych jest zaprojektować model bazy danych w postaci zbioru relacji (schematów) spełniających ograniczenia świata rzeczywistego, utrzymujących semantyczną o nim wiedzę oraz minimalizujących problem anomalii i redundancji danych (m.in. wykorzystując proces normalizacji). Podczas projektowania schematów relacyjnych baz danych głównym pojęciem jest zależność danych, co oznacza więzy nałożone na relacje, mogące być wartością schematu. Podstawowe znaczenie mają zależności funkcyjne – więzy, które zależą od tego, czy określone składowe dwóch krotek są zgodne czy nie.

W pracy [CODD 70] autor zdefiniował reguły projektowania relacyjnych baz danych, wyrażone jako pierwsza, druga i trzecia postać normalna (ang. *normal forms*), oznaczane odpowiednio 1NF, 2NF i 3NF. W połowie lat siedemdziesiątych spostrzeżono pewne braki w trzeciej postaci normalnej i dało to podstawę dla zdefiniowania postaci normalnej Boyce'a-Codda (BCNF) [CODD 74]. Proces kolejnego przekształcania projektu bazy danych przez te postacie normalne jest znany jako normalizacja. Można powiedzieć, że normalizacja jest procesem, którego celem jest wyeliminowanie anomalii i prowadzi do pewnych postaci usuwając odpowiednie, niepożądane zależności funkcyjne. Normalizacja niższych postaci normalnych do wyższych, które narzucają mocniejsze ograniczenia na schemat relacji, jest wykonywana poprzez rozkład schematów relacyjnych (przez projekcje). Należy dodać, że została również sformułowana przez Fagina czwarta i piąta postać normalna (4NF i 5NF) [FAGIN 77, FAGIN 79].

Teoria projektowania schematów relacyjnych baz danych została dobrze przedstawiona w pracy [ULLMAN 88]. Znajdują się w niej między innymi definicje zależności funkcyjnych, rozkładów schematów relacji, postaci normalnych, algorytmy projektowania schematu bazy danych.

4. PODSUMOWANIE

Poprawnie zaprojektowana baza danych musi spełniać wszystkie oczekiwania przyszłego użytkownika jednocześnie musi zapewniać bezbłędne, wydajne i przyjazne w obsłudze działanie. W artykule omówiono w sposób systematyczny i uporządkowany najważniejsze pojęcia z teorii relacyjnych baz danych oraz przedstawiono kolejne etapy projektowania relacyjnych baz danych. Pierwszym i zasadniczym elementem projektowania baz danych jest precyzyjne określenie w jakim celu baza danych jest projektowana. Jasno określony cel umożliwia przystąpienie do fazy analizy projektowania bazy danych. Celem analizy rozważanej rzeczywistości jest rozpoznanie i opisanie rzeczywistości, która ma zostać zamodelowana, tzn. zebranie informacji umożliwiających zbudowanie poprawnego modelu bazy danych. Często jednak fakty dotyczące otaczającego świata są niepełne, nieznane, niedokładne. Model relacyjny nie pozwala na przechowywanie takich danych poza jedną dozwoloną wartością null, która też nie zawsze jest dopuszczana. W modelu klasycznym (relacyjnym) informacja nieprecyzyjna musi być przekształcona do postaci danej dokładnej co często powoduje znaczne zniekształcenie rzeczywistości i utratę informacji.

PRINCIPLES OF RELATIONAL DATABASES DESIGN

The work presents methodologies of relational database modeling. Most attention has been paid to a conceptual and logical modeling. This paper presents basic terms of relational database, entity relationship model and representations data in relational model, relation schema.

*logika rozmyta,
relacje rozmyte,
zbiory rozmyte*

Barbara BIEDRZYCKA
Zygmunt MAZUR*

LOGIKA ROZMYTA

W pracy przedstawiono podstawowe pojęcia logiki rozmytej, omówiono pojęcia niepewności i nieprecyzyjności informacji. Przedstawiono własności zbiorów rozmytych i działania na zbiorach rozmytych. Omówiono zasadę rozszerzania. Wprowadzono pewne elementy istotne dla rozmytego rozszerzenia modeli danych, m.in. zdefiniowano zbiór rozmyty i rozkład możliwości, przedstawiono jednostkę reprezentacji wiedzy jaką jest zmienna lingwistyczna, wprowadzono relacje rozmyte. Znajomość tych pojęć jest niezbędna do projektowania rozmytych relacyjnych modeli baz danych.

1. WPROWADZENIE

Logika rozmyta, zamiennie nazywana teorią zbiorów rozmytych, datuje swój początek na rok 1965. Wówczas to Lofti Zadeh opublikował innowacyjną pracę „Zbiory rozmyte” w *Information and Control* [ZADEH 65]. Rzeczywistość prawie nigdy nie można dokładnie opisać. Logika rozmyta pomaga reprezentować i operować rozmytością, która występuje w świecie rzeczywistym. W życiu codziennym człowiek bardzo często posługuje się wyrażeniami rozmytymi, na przykład trochę, kilka, gruby, chudy, stary, młody, mały, duży, ciepło, zimno, mniej więcej, możliwe, chyba.

Modelowanie danych w szczególności może wymagać reprezentowania i operowania rozmytością, gdyż jest ona właściwa dla atrybutów, obiektów i związków. Klasyczne modele nie umożliwiają operowania takimi zdaniami jak np. „Wyszukaj towary o cenie około 100 zł” lub „Wyszukaj dane osób starych”. Dana przechowywana w bazie musi być precyzyjnie zdefiniowana, jedyną możliwością odzwierciedlenia niepewności jest wartość nieokreślona – null, która może oznaczać zarówno wartość nieznaną (np. nie znamy daty urodzenia starożytnego filozofa), nieokreśloną (nazwa projektu, w którym uczestniczy pracownik) lub pustą (wartość atrybutu `Nazwisko_panięskie` dla mężczyzny). Logika rozmyta wprowadza teoretyczne podstawy umożliwiające reprezentowanie i operowanie rozmytymi danymi w sposób naturalny i efektywny.

* Wydziałowy Zakład Informatyki, Wydział Informatyki i Zarządzania Politechniki Wrocławskiej, 50-370 Wrocław, Skwer Idaszewskiego 1, mazur@ci.pwr.wroc.pl

2. NIEPEWNOŚĆ I NIEPRECYZYJNOŚĆ

Niepewność i nieprecyzyjność powodują niekompletność informacji, częściową wiedzę o rozważanym elemencie. W kontekście zdań taka niekompletność informacji zwykle odnosi się do jego dwóch aspektów: prawdziwości i treści. Mówi się, że zdanie jest **niepewne**, jeżeli nie można wyraźnie określić, czy jest prawdziwe, czy fałszywe, natomiast zdanie jest **nieprecyzyjne**, jeżeli jego treść nie może być dokładnie zdefiniowana [KERRE 93]. Na przykład, jeżeli nie ma pewności, czy zdanie „Wrocław ma 700 000 mieszkańców” jest prawdziwe, wówczas zdanie to jest niepewne. Jeżeli nie można dokładnie określić liczby mieszkańców Wrocławia, ale można określić, że to jest „duże” miasto, to otrzymuje się niedokładną wartość, co jest przypadkiem nieprecyzyjności. Oba aspekty nakładają się na siebie nawzajem. Na przykład można spotkać się z niepewnym zdaniem z precyzyjnie zdefiniowanymi predykatami (niepewne, precyzyjne) czy też z niepewnym zdaniem z nieprecyzyjnymi predykatami (niepewne, nieprecyzyjne), np. „nie wiadomo, czy Wrocław jest dużym miastem”.

Istnieją różne źródła niepewności i nieprecyzyjności i jedno źródło może stanowić podstawę drugiego. Dobrze poznanym źródłem jest losowość, z którą można sobie radzić posługując się rachunkiem prawdopodobieństwa. Innym, istotnym dla tematu pracy, jest rozmytość, z którą można postępować używając teorii zbiorów rozmytych.

Rozmytość jest związana z problemem określenia granicy (progu). Jak zostało to już wcześniej podkreślone, rozmytość własności leży w braku dobrze zdefiniowanych granic zbioru obiektów, do których odnosi się własność. Na przykład, wiedząc, że do miasta A jest 800 km mówimy, że jest daleko, czy jeżeli do miasta B jest 799 km, to też powiemy, że jest daleko? A 790 km? Nasuwa się pytanie, czy w ogóle można wyznaczyć wyraźną granicę odległości, o której powiemy jeszcze „daleko”? Podobnych przykładów można przytoczyć wiele. Wiążą się one z używaniem określeń potocznych takich jak nie gruby, wysoki, bogaty, stary. Jak wyznaczyć wiek, który należy do kategorii „stary”? Jeżeli można powiedzieć, że osoba powyżej 80 lat jest „stara”, to jak określić osobę 79-letnią? W klasycznych modelach danych linia graniczna musi być sztucznie określona, aby można było odpowiedzieć na powyższe pytania.

Zadeh zaproponował rozwiązanie powyższych problemów poprzez wprowadzenie stopniowego przejścia od elementu nie należącego do zbioru elementów o pewnej własności (nie spełniającego rozważanej własności) do elementu w pełni należącego do tego zbioru (w pełni charakteryzującego się daną własnością). Stopniowe przejście pozwala na częściowy stopień przynależności do zbioru, tzn. częściowe spełnienie rozważanej własności.

3. ZBIORY ROZMYTE I ROZKŁAD MOŻLIWOŚCI

Pojęcie zbiór ma dwa znaczenia [ŁACHWA 01]:

- (1) zbiór dystrybutywny (ostry) – wielość rzeczy, o których można myśleć jak o jedności (według definicji Cantora);
- (2) zbiór kolektywny – obiekt przestrzenny np. linia, odcinek czasu.

Ponadto można wyróżnić zbiory rozmyte (ang. *fuzzy sets*), zbiory przybliżone (ang. *rough sets*), zbiory probabilistyczne (ang. *probabilistic sets*) i inne. Z pojęciem zbioru wiąże się pojęcie należenia elementu do zbioru. Zbiór rozmyty jest uogólnieniem zbioru w podstawowym znaczeniu i obejmuje elementy z pewnej przestrzeni, przy czym dany element może należeć do zbioru, nie należeć do zbioru albo należeć w pewnym stopniu do zbioru (dla zbiorów klasycznych czyli ostrych są tylko dwie możliwości – dany element należy do zbioru lub nie należy).

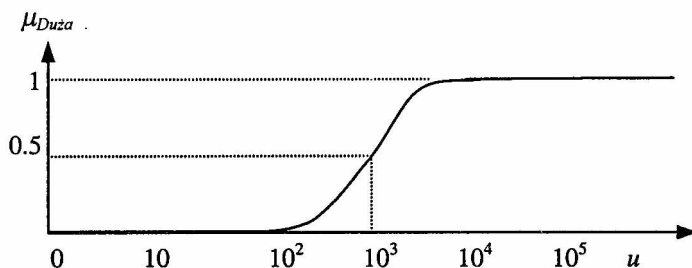
Niech U będzie obszarem odniesienia (uniwersum), tzw. obszarem rozważań. Rozmyty zbiór F będący podzbiorem U jest charakteryzowany przez funkcję przynależności $\mu_F: U \rightarrow [0,1]$, która wiąże każdy element u należący do U z liczbą $\mu_F(u)$ reprezentującą stopień przynależności u do F . Zapis $\mu_F(u) = 0$ oznacza, że u nie należy do F , $\mu_F(u) = 1$ oznacza, że u w pełni należy do F , a μ_F przyjmujące wartość z przedziału $(0, 1)$ ($0 < \mu_F < 1$) oznacza częściową przynależność. Zbiór rozmyty F można zapisać wzorem $F = \{ \mu_F / u \mid u \in U \}$ lub

$$F = \int_{u \in U} \mu_F(u) / u$$

Rozważając przykład dużej liczby, można wprowadzić dla niej zbiór rozmyty $Duża$ definiowany przez funkcję przynależności $\mu_{Duża}: \mathbb{N}^* \rightarrow [0,1]$, $\forall n \in \mathbb{N}^*$

$$\mu_{Duża}(n) = \frac{1}{1 + \left(\frac{1000}{n}\right)^2}$$

Graficznie funkcję $\mu_{Duża}$ przedstawia rysunek 1 [CHEN 98].



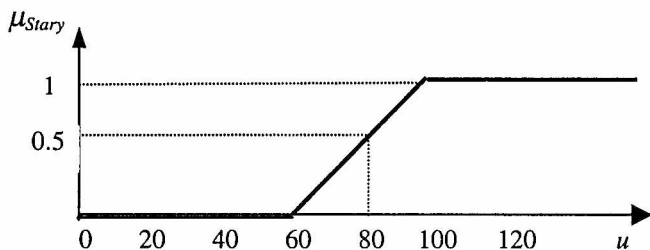
Rys. 1. Funkcja przynależności zbioru rozmytego $Duża$

Każda liczba naturalna ma stopień przynależności do zbioru rozmytego Duża. Na przykład 1000 należy do tego zbioru (dużych liczb) ze stopniem 0.5, 10^4 ze stopniem $100/101$, 10^5 ze stopniem $10000/10001$.

Zdefiniujmy zbiór rozmyty *Stary*, reprezentujący zbiór określający wiek ludzi starych, przez funkcję przynależności $\mu_{Stary}: U \rightarrow [0,1]$, gdzie U jest zbiorem liczb rzeczywistych określających wiek, należących do przedziału $[0,120]$, $\forall u \in U$

$$\mu_{Stary}(u) = \begin{cases} 0 & \text{dla } 0 \leq u \leq 60 \\ \frac{u-60}{20} & \text{dla } 60 < u < 80 \\ 1 & \text{dla } 80 \leq u \leq 120 \end{cases}$$

Rysunek 2 przedstawia graficznie funkcję przynależności zbioru rozmytego *Stary*.



Rys. 2. Funkcja przynależności zbioru rozmytego *Stary*

Zbiór rozmyty *Stary* klasyfikuje osoby od 80 roku życia (włącznie) jako osoby stare, osoby do roku 60 (włącznie) jako nie stare i osoby pomiędzy 60 a 80 rokiem życia jako stare z odpowiednim częściowym stopniem przynależności. Na przykład osoba 62-letnia ma stopień przynależności do zbioru *Stary* równy 0.1, a dla osoby w wieku 70 lat stopień przynależności równa się 0.5.

Opierając się na zbiorach rozmytych, Zadeh wprowadził teorię rozkładu możliwości [ZADEH 78]. Niech U będzie obszarem rozważań, niech X będzie zmienną o wartościach ze zbioru U i F będzie zbiorem rozmytym na U z funkcją przynależności μ_F , wówczas możliwość, że $X = u$, oznaczana przez $\pi_X(u)$, jest równa $\mu_F(u)$, tj. $\pi_X(u) = \mu_F(u)$, gdzie przyjmuje się, że X jest ograniczona w sposób rozmyty przez zbiór F . Rozkład możliwości zmiennej X na zbiorze U (względem F) wyraża się w wzorem:

$$\Pi_X = \{\pi_X(u) / u \mid u \in U \wedge \pi_X(u) = \mu_F(u) \in [0,1]\}$$

Rozkład możliwości Π_X jest znormalizowany wtedy i tylko wtedy, gdy $\exists u' \in U$ takie, że $\pi_X(u') = 1$.

Należy zwrócić uwagę na interpretację $\mu_F(u)$ i $\pi_X(u)$. Na przykład, jeżeli U jest zbiorem średnich stopni studentów, $U = \{5, 4.5, 4, 3.5, 3, 2\}$, *Dobra* jest zbiorem

rozmytym na U reprezentującym dobre oceny, zdefiniowanym w postaci zbioru $\{1/5, 1/4.5, 0.85/4, 0.4/3.5, 0.1/3, 0.0/2\}$ i X jest zmienną o wartościach ze zbioru U , to $\mu_{Dobra}(4) = 0.85$ reprezentuje stopień zgodności oceny B z pojęciem *Dobra*, podczas gdy $\pi_x(4) = Poss(X = 4) = 0.85$ jest interpretowane następująco: możliwość, że stopień równa się 4 wynosi 0.85 pod warunkiem, że wynik studenta jest *Dobry*.

Pojęcie rozkładu możliwości jest bliskie pojęciu rozkładu prawdopodobieństwa. W artykule [ZADEH 78] autor zamieścił interesujący przegląd różnic między teorią zbiorów rozmytych a teorią prawdopodobieństwa. Główna różnica polega na tym, że nie jest potrzebny taki rozkład możliwości, aby spełniony był warunek, że suma możliwości wynosi 1, jak w przypadku sumy prawdopodobieństw.

Mając zdefiniowany rozkład możliwości π_x ($\pi_x(u) = \mu_F(u)$), gdzie X jest zmienną o wartościach ze zbioru U i mając określony zbiór rozmyty A na U z funkcją przynależności μ_A , można je porównać obliczając miarę możliwości wyrażoną wzorem:

$$Poss(A) = \sup_{x \in U} \min[\pi(x), \mu_A(x)]$$

3.1. WŁASNOŚCI ZBIORÓW ROZMYTYCH

Podstawowymi własnościami zbiorów rozmytych są: nośnik (ang. *support*), jądro (ang. *kernel*) zwane zamiennie nukleus (ang. *nucleus*), zbiór progowy (α -cut), punkt rozgraniczenia (ang. *crossover point*), wysokość (ang. *height*), podstawa (ang. *plinth*), które można zdefiniować następująco [DRIANKOOV 96, CHEN 98]: niech U będzie obszarem rozważań, A będzie zbiorem rozmytym zdefiniowanym na U z funkcją przynależności μ_A :

- 1) **Nośnikiem** zbioru rozmytego A , oznaczanym przez $\text{supp}(A)$, nazywa się zbiór ostry zawierający wszystkie elementy A mające niezerowy stopień przynależności:

$$\text{supp}(A) = \{u \mid u \in U \wedge \mu_A(u) > 0\}$$

- 2) **Jądrem** zbioru rozmytego A , oznaczanym przez $\text{ker}(A)$, nazywa się zbiór ostry zawierający wszystkie wartości ze stopniem przynależności 1:

$$\text{ker}(A) = \{u \mid u \in U \wedge \mu_A(u) = 1\}$$

- 3) **Punkt rozgraniczenia** zbioru rozmytego A , jest to taki element x ze zbioru A , dla którego

$$\mu_A(x) = 0.5$$

Zbiór rozmyty może nie mieć punktu rozgraniczenia lub może mieć jeden lub wiele takich punktów.

- 4) **Zbiorem progowym** dla rozmytego zbioru rozmytego A , oznaczanym przez A_α , nazywa się zbiór zawierający wszystkie elementy ze stopniem przynależności większym lub równym α , gdzie $0 \leq \alpha \leq 1$:

$$A_\alpha = \{u \mid u \in U \wedge \mu_A(u) \geq \alpha\}$$

- 5) **Wysokością** zbioru rozmytego A , oznaczaną przez $\text{height}(A)$, jest największy stopień przynależności μ_A :

$$\text{height}(A) = \sup_{u \in U} \mu_A(u)$$

- 6) **Podstawa** zbioru rozmytego A , oznaczana przez $\text{plinth}(A)$, jest zdefiniowana zależnością:

$$\text{plinth}(A) = \inf_{u \in U} \mu_A(u)$$

- 7) Zbiór rozmyty A jest **pusty**, jeśli żaden element uniwersum nie należy do tego zbioru. Nośnik pustego zbioru rozmytego jest pustym zbiorem ostrym.
- 8) **Normalny** zbiór rozmyty jest to taki zbiór rozmyty A , którego wysokość ma wartość 1. Każdy niepusty zbiór rozmyty, który nie jest normalny, można znormalizować, dzieląc wartość funkcji przynależności przez wysokość tego zbioru.

3.2. DZIAŁANIA NA ZBIORACH ROZMYTYCH

Pojęcia takie jak równość i inkluzja (zawieranie) dwóch zbiorów rozmytych są bezpośrednio wyprowadzane z klasycznej teorii zbiorów.

Niech U będzie obszarem rozważań. Dwa zbiory rozmyte A i B są równe ($A = B$) wtedy i tylko wtedy, gdy

$$\forall u \in U \quad \mu_A(u) = \mu_B(u)$$

Zbiór rozmyty A jest podzbiorem zbioru rozmytego B ($A \subseteq B$) wtedy i tylko wtedy, gdy

$$\forall u \in U \quad \mu_A(u) \leq \mu_B(u)$$

W klasycznej teorii zbiorów suma, iloczyn i dopełnienie zbiorów są prostymi działaniami, które są jednoznacznie zdefiniowane. Jednoznaczność wynika z faktu, że operatory logiczne AND, OR i NOT mają dobrze zdefiniowaną semantykę, opartą na logice twierdzeń. W teorii zbiorów rozmytych ich interpretacja nie jest tak prosta, gdyż używa się stopniowych funkcji przynależności. Proste rozszerzenie tych klasycznych działań zaprezentował Zadeh [ZADEH 65], a są one zdefiniowane następująco: niech A i B będą zbiorami rozmytymi na tym samym obszarze rozważań U z funkcjami przynależności μ_A i μ_B :

- 1) $A \cup B$ jest zbiorem rozmytym na U z funkcją przynależności $\mu_{A \cup B}: U \rightarrow [0,1]$

$$\forall u \in U \quad \mu_{A \cup B}(u) = \max(\mu_A(u), \mu_B(u))$$

- 2) $A \cap B$ jest zbiorem rozmytym na U z funkcją przynależności $\mu_{A \cap B}: U \rightarrow [0,1]$,

$$\forall u \in U \quad \mu_{A \cap B}(u) = \min(\mu_A(u), \mu_B(u))$$

- 3) Dopełnieniem zbioru rozmytego, oznaczanym przez A' , jest zbiór rozmyty na U z funkcją przynależności $\mu_{A'}: U \rightarrow [0,1]$,

$$\forall u \in U \quad \mu_{A'}(u) = 1 - \mu_A(u)$$

Operacje $\cup$ i $\cap$ na dwóch zbiorach rozmytych mogą być rozszerzone na skończoną liczbę zbiorów rozmytych.

Iloczyn kartezjański może być również rozszerzony na zbiory rozmyte. Niech A i B będą zbiorami rozmytymi zdefiniowanymi odpowiednio na U i V z funkcjami przynależności μ_A i μ_B . $A \times B$ jest zbiorem rozmytym na $U \times V$ z funkcją przynależności $\mu_{A \times B}: U \times V \rightarrow [0,1]$:

$$\forall (u, v) \in U \times V \quad \mu_{A \times B}(u, v) = \min(\mu_A(u), \mu_B(v))$$

Definicja iloczynu kartezjańskiego dwóch zbiorów rozmytych może być rozszerzona dla skończonej liczby zbiorów rozmytych.

4. ZASADA ROZSZERZANIA

Zasada rozszerzania jest jednym z najważniejszych pojęć w teorii zbiorów rozmytych, podaje ona ogólną metodę łączenia nierozmytych i rozmytych idei różnego rodzaju, na przykład łączenia zbiorów rozmytych i relacji, jak również operacji funkcji matematycznych na zbiorach rozmytych. Zbiory rozmyte można również interpretować jako liczby rozmyte [DRIANKOOV 96, BOLC 91]. W tym wypadku można stosować zasadę rozszerzania do dodawania lub mnożenia liczb rozmytych.

Niech $A_1, \dots, A_n$ będą zbiorami rozmytymi zdefiniowanymi odpowiednio na $U_1, \dots, U_n$ i niech f będzie nierozmytą funkcją $f: U_1 \times \dots \times U_n \rightarrow V$. Rozszerzenie f działające na $A_1, \dots, A_n$, tj. $f(A_1, \dots, A_n)$, jest zbiorem rozmytym na V z funkcją przynależności:

$$\mu_{f(A_1, \dots, A_n)}(v) = \sup_{\substack{u_1, u_2, \dots, u_n \\ f(u_1, u_2, \dots, u_n) = v}} \min(\mu_{A_1}(u_1), \dots, \mu_{A_n}(u_n))$$

gdy istnieje $f^{-1}(v)$. W przeciwnym przypadku $\mu_{f(A_1, \dots, A_n)}(v) = 0$.

Inną metodą otrzymania tego wyniku jest:

$$f(A_1, \dots, A_n) = \int_{U_1 \times \dots \times U_n} \min(\mu_{A_1}(u_1), \dots, \mu_{A_n}(u_n)) / f(u_1, \dots, u_n)$$

5. ROZMYTE OPERATORY IMPLIKACJI

Rozmyty operator implikacji jest rozszerzeniem operatora implikacji ($\rightarrow$) tradycyjnej logiki dwuwartościowej, odwzorowującego $\{0,1\}^2$ na $\{0,1\}$ (Tabela 1).

Rozmyte operatory implikacji I są odwzorowaniem $[0, 1] \times [0, 1]$ w $[0,1]$ i istnieją różne ich definicje. Na przykład, uwzględniając fakt występujący w logice

dwuwartościowej, że $a \rightarrow b$ jest równoważne $\neg a \vee b$ i używając „max” dla „ $\vee$ ” i „1-” dla „ $\neg$ ”, uzyskuje się $I(a,b) = KD(a,b) = \max(1-a,b)$.

Tabela 1. Tablica wartościowania dla implikacji ($\rightarrow$)

a	b	$a \rightarrow b$
0	0	1
0	1	1
1	0	0
1	1	1

Tabela 2. Rozmyte operatory implikacji

Oznaczenie	Definicja
$S^\#$	$S^\#(a,b) = \begin{cases} 1 & a < 1 \vee b = 1 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$
S	$S(a,b) = \begin{cases} 1 & a \leq b \\ 0 & a > b \end{cases}$
S^*	$S^*(a,b) = \begin{cases} 1 & a \leq b \\ b & a > b \end{cases}$
G	$G(a,b) = \begin{cases} 1 & a \leq b \\ b/a & a > b \end{cases}$
G^*	$G^*(a,b) = \begin{cases} 1 & a = 0 \vee b = 1 \\ \min(1, b/a, (1-a)/(1-b)) & \text{w przeciwnym przypadku} \end{cases}$
L	$L(a,b) = \min(1, 1-a+b)$
KDL	$KDL(a,b) = 1 - a + ab$
KD	$KD(a,b) = \max(1-a, b)$
EA	$EA(a,b) = \max(1-a, \min(a,b))$
W	$W(a,b) = \max(\min(1-a,b), \min(a, 1-a), \min(b, 1-b))$
M	$M(a,b) = \min(a,b)$
SS	$SS(a,b) = \min(S(a,b), S(1-a, 1-b))$
S^*S^*	$S^*S^*(a,b) = \min(S^*(a,b), S^*(1-a, 1-b))$
SS^*	$SS^*(a,b) = \min(S(a,b), S^*(1-a, 1-b))$
S^*S	$S^*S(a,b) = \min(S^*(a,b), S(1-a, 1-b))$

Tabela 2 [CHEN 98] przedstawia piętnaście rozmytych operatorów implikacji nazywanych odpowiednio:

EA	– operator Zadeha (ang. <i>Early Zadeh</i>) [ZADEH 75a]
L	– operator Łukasiewicza
M	– operator Mamdani’ego [MAMDANI 77]
KD	– operator Kleene–Dienes’a
KDL	– operator Kleene–Dienes–Łukasiewicza [BANDLER 80]
W	– operator Willmotta [WILLMOTT 80]
S [#]	– operator standardowy ostry (ang. <i>Standard Sharp</i>) [MUZIMOTO82]
S	– operator standardowy ścisły (ang. <i>Standard Strict</i>)
S*	– operator standardowy gwiazdzisty (ang. <i>Standard Star</i>) lub operator Gödel–Browera
SS*	– operator standardowy ścisło-gwiazdzisty (ang. <i>Standard Strict-Star</i>)
S*S*	– operator standardowy gwiazdzisto-gwiazdzisty (ang. <i>Standard Star-Star</i>)
S*S	– operator standardowy gwiazdzisto-ścisły (ang. <i>Standard Star-Strict</i>)
SS	– operator standardowy ścisło-ścisły (ang. <i>Standard Strict-Strict</i>)
G	– operator Gaines’a
G*	– zmodyfikowany operator Gaines’a (ang. <i>Modified Gaines</i>)

Różne rozmyte operatory implikacji mają różne własności. Na przykład L i KD są, w przeciwieństwie do S*, operatorami ciągłymi. Omówienie operatorów można znaleźć w [KERRE 93].

6. ZMIENNE LINGWISTYCZNE

Podstawowym pojęciem logiki rozmytej jest zmienna lingwistyczna (ang. *linguistic variable*). Zmienne lingwistyczne pozwalają na łatwą i naturalną specyfikację wartości dla nieprecyzyjnych pojęć. Zadeh [ZADEH 75b] stwierdza: „Przez zmienną lingwistyczną rozumiemy zmienną, której wartościami są słowa lub zdania w języku naturalnym lub formalnym. Dla przykładu Wiek jest zmienną lingwistyczną, jeśli jej wartości są wyrażone słowami, a nie liczbami, to znaczy młody, niemłody, bardzo młody, całkiem młody, stary, nie bardzo stary i nie bardzo młody itd. zamiast 20, 21, 22, 23, ...”.

Zmienna lingwistyczna X jest definiowana przez piątkę:

$$\langle X, T(X), U, G, M \rangle$$

gdzie:

- X – symboliczna nazwa zmiennej lingwistycznej,
na przykład Temperatura, Prędkość, Wiek.
- T(X) – zbiór wartości lingwistycznych, które może przyjąć X.
- U – obszarem rozważań – rzeczywista dziedzina, nad którą zmienna lingwistyczna X przybiera swoje jakościowe wartości (precyzyjne).

U może być dyskretne lub ciągłe.

Dla zmiennej lingwistycznej Temperatura może być to np. przedział $[-10^{\circ}\text{C}, 35^{\circ}\text{C}]$, dla Prędkości – $[0 \text{ km/h}, 200 \text{ km/h}]$.

G – zbiór reguł syntaktycznych generujących elementy należące do zbioru $T(X)$.

M – jest zbiór reguł semantycznych nadających znaczenie wartościom zmiennej X , odwzorowujących $T(X)$ na odpowiadające zbiory rozmyte zbioru U :

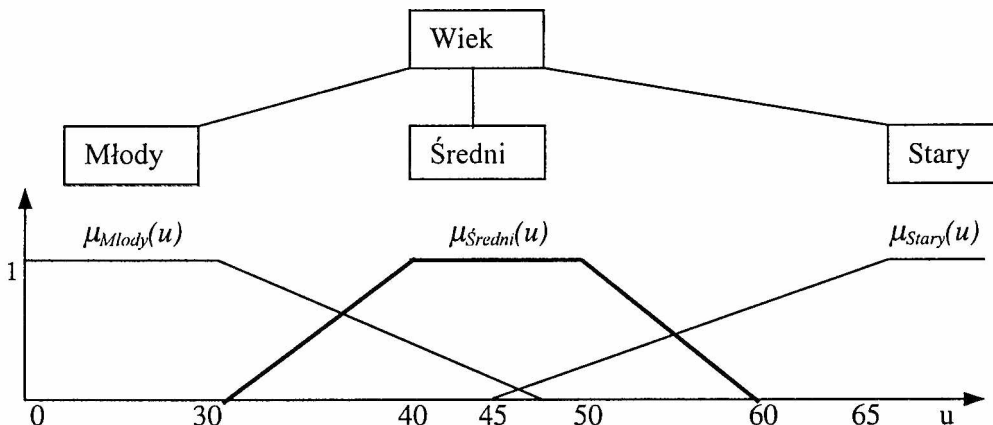
$$M(x) = \sum_U \mu_x(u)/u \text{ dla przypadku } U - \text{dyskretny}$$

$$M(x) = \int \mu_x(u)/u \text{ dla przypadku } U - \text{ciągły}$$

M jest funkcją, która bierze symbol swego argumentu, na przykład zimny, i zwraca znaczenie lub symbol-zimna w warunkach zbioru rozmytego.

Wartość lingwistyczna inaczej wartość rozmyta oznacza symbol dla szczególnej własności X . W wypadku zmiennej lingwistycznej Temperatura zbiór $T(\text{Temperatura})$ może mieć postać {zimny, chłodny, komfortowy, ciepły, gorący}. $T(X)$ jest nazywany również zbiorem termów X lub zbiorem odniesień X . Element takiego zbioru oznacza się często przez x .

Rozważając zmienną lingwistyczną $X = \text{Wiek}$, $T(X)$ może zostać wygenerowany (przez G) w postaci zbioru {młody, średni, stary}, każda wartość ze zbioru $T(X)$ ma szczególne znaczenie wyrażone (przez M) jako zbiór rozmyty [CHEN 98]. Graficznie, zmienna lingwistyczna Wiek może zostać przedstawiona tak jak na rysunku 3.



Rys. 3. Zmienna lingwistyczna Wiek [CHEN 98]

W języku naturalnym główne słowo lub wartość (term), na przykład młody, wysoki, dobry, mogą być modyfikowane przez modyfikatory (ang. *hedges*)

lingwistyczne takie jak: bardzo, mniej więcej, trochę, całkiem, itd. Modyfikatory te, w przeciwieństwie do podstawowych termów, nie są modelowane przez zbiory rozmyte, ale jako operatory działające na zbiorach rozmytych reprezentujących wartości, których dotyczą.

Niech $F(U)$ będzie zbiorem wszystkich zbiorów rozmytych na U , H_λ będzie operatorem modyfikacji, gdzie $\lambda \in [0, \infty)$, wówczas H_λ jest odwzorowaniem $F(U)$ w $F(U)$ takim, że $\forall A \in F(U)$

$$H_\lambda(A) = A^\lambda \in F(U) \vee \forall a \in U \quad \mu_{H_\lambda(A)}(a) = [\mu_A(a)]^\lambda \in [0, 1]$$

Gdy $\lambda > 1$, H_λ zmniejsza stopień przynależności elementów do modyfikowanego zbioru rozmytego. Dla $\lambda < 1$, H_λ zwiększa ów stopień przynależności elementów do modyfikowanego zbioru rozmytego.

Dla $\lambda = 2$, H_2 jest nazywany operatorem koncentracji (CON), tj. $\forall a \in U$

$$\mu_{H_2(A)}(a) = [\mu_A(a)]^2$$

Operator CON jest często używany do reprezentacji modyfikatora lingwistycznego **Bardzo**, tzn. $CON(A) = H_2(A)$ wyraża „bardzo A”, a $H_4(A)$ wyraża „bardzo bardzo A”.

Dla $\lambda = 1/2$, $H_{1/2}$ jest nazywany operatorem rozszerzania (rozciągania) (DIL), tj. $\forall a \in U$

$$\mu_{H_{1/2}(A)}(a) = [\mu_A(a)]^{1/2}$$

DIL lub $H_{1/2}$ jest często używany do reprezentacji modyfikatora lingwistycznego **Mniej-więcej**, tzn. $DIL(A) = H_{1/2}(A)$ wyraża „mniej więcej A”.

Innym interesującym operatorem jest operator intensyfikacji przeciwieństwa (INT). INT jest często używany do reprezentacji modyfikatora lingwistycznego **Trochę** (Całkiem). Najpopularniejszą jego postacią jest:

$$INT_z(A^{1.25} \wedge Co(CON(A)))$$

Inne modyfikatory lingwistyczne są zaprezentowane m.in. w pracach [LAKOFF 73, ZADEH 72, ZADEH 75a, SCHMUCKER 84, WANG 83, KERRE 89].

7. MIARY BLISKOŚCI MIĘDZY ZBIORAMI ROZMYTYMI

W bardzo wielu zastosowaniach wymaga się porównania dwóch zbiorów rozmytych, dlatego też zostało wprowadzone pojęcie bliskości ($\approx$) zbiorów rozmytych. Można powiedzieć, że wyraża ono stopień ich „równości”. Niech $F(U)$ będzie zbiorem wszystkich zbiorów rozmytych na U , wówczas $\forall A, B \in F(U)$ miara bliskości pomiędzy A i B , oznaczana przez $\approx$, jest odwzorowaniem $F(U) \times F(U)$ w $[0, 1]$, tj. $\forall A, B \in F(U) \approx(A, B) \in [0, 1]$. Miary bliskości są definiowane w różny sposób. Miary przedstawione w bieżącym rozdziale zostały opracowane na podstawie [CHEN 98].

Miara bliskości może zostać zdefiniowana opierając się na inkluzji zbiorów ($\subseteq$). Dla zbiorów normalnych dwa zbiory A i B są równe, tj. $A = B$, wtedy i tylko wtedy, gdy $A \subseteq B$ i $B \subseteq A$, gdzie $A \subseteq B$ jest definiowane w następujący sposób: jeżeli $x \in A$, to $x \in B$. Miara bliskości oparta na inkluzji $\approx_I(A, B)$ jest rozszerzeniem powyższej definicji. $A \subseteq B$ uzyskuje się poprzez wprowadzenie do zdania „jeżeli $x \in A$, to $x \in B$ ” rozmytego operatora implikacji (I). Otrzymuje się wówczas:

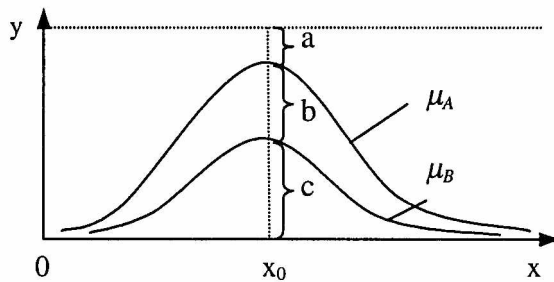
$$\text{Degree}(A \subseteq B) = \inf_x I(\mu_A(x), \mu_B(x))$$

Stosując jako rozmyty operator implikacji L (gdzie $L(a,b) = \min(1, 1-a+b)$), miara $\approx_I(A, B)$ może zostać zdefiniowana następująco:

$$\begin{aligned} \approx_I(A, B) &= \min(\text{Degree}(A \subseteq B), \text{Degree}(B \subseteq A)) \\ &= \min(\inf_x (L(\mu_A(x), \mu_B(x))), \inf_x (L(\mu_B(x), \mu_A(x)))) \\ &= \inf_x \min(L(\mu_A(x), \mu_B(x)), L(\mu_B(x), \mu_A(x))) \\ &= \inf_x \min(\min(1, 1 - \mu_A(x) + \mu_B(x)), \min(1, 1 - \mu_B(x) + \mu_A(x))) \\ &= \inf_x (1 - |\mu_A(x) - \mu_B(x)|) \end{aligned}$$

Rysunek 4 przedstawia funkcję przynależności μ_A i μ_B dla zbiorów rozmytych A i B. Bliskość pomiędzy A i B mierzona w punkcie x_0 wynosi:

$$\approx_I(A, B) = 1 - |\mu_A(x_0) - \mu_B(x_0)| = 1 - b = a + c$$

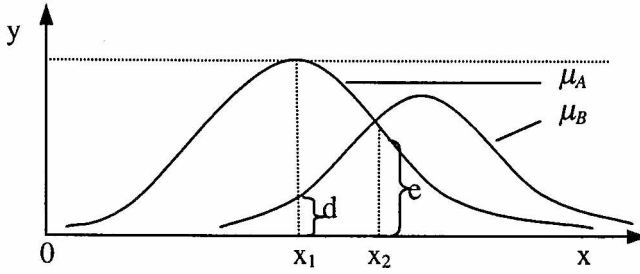


Rys. 4. Miara bliskości oparta na inkluzji [CHEN 98]

Inna miara bliskości, nazywana wysokością $A \cap B$, oznaczana przez $\approx_H$, opiera się na części wspólnej zbiorów rozmytych. Definiuje się ją zależnością:

$$\approx_H(A, B) = \sup_x \min(\mu_A, \mu_B)$$

Na rysunku 4 $\approx_H(A, B) = c$. Rozważając przykład zaprezentowany na rysunku 5, stopień bliskości pomiędzy A i B zgodnie z $\approx_I$ równa się d (w punkcie x_1) a stopień bliskości pomiędzy A i B zgodnie z $\approx_H$ równa się e (w punkcie x_2).



Rys. 5. Reprezentacja miar bliskości $\approx_I(A,B) = d$, $\approx_H(A,B) = e$ (CHEN98]

Istnieją również miary bliskości oparte na pojęciu odległości [WANG 83, HE 83, HE 94]. Rodzina miar odległości może być przedstawiona w postaci miary odległości Minkowskiego ($d_M(A, B)$):

$$d_M(A, B) = \left[\frac{1}{n} \sum_x |\mu_A(x) - \mu_B(x)|^s \right]^{1/s}$$

gdzie s jest parametrem. Dla $s = 1, 2, \infty$, uzyskuje się trzy powszechnie używane miary odległości postaci:

$$d_H(A, B) = \frac{1}{n} \sum_x |\mu_A(x) - \mu_B(x)|$$

$$d_E(A, B) = \frac{1}{\sqrt{n}} \sqrt{\sum_x |\mu_A(x) - \mu_B(x)|^2}$$

$$d_C(A, B) = \max_x |\mu_A(x) - \mu_B(x)|$$

Niech $d(A, B)$ będzie miarą odległości, wówczas miara bliskości oparta na odległości, oznaczana przez $\approx_d$, jest odwzorowaniem

$F(U) \times F(U) \rightarrow [0, 1]$ takim, że $\forall A, B \in F(U)$

$$\approx_D(A, B) = 1 - d(A, B)$$

jeżeli $s = \infty$, to $\approx_I(A, B) = \approx_D(A, B)$.

Należy zauważyć, że wszystkie z przedstawionych miar bliskości spełniają własność

$$\approx(A, B) \in [0, 1]$$

oraz

$$\approx(A, B) = \approx(A, B).$$

Własność

$$\approx(A, A) = 1$$

jest utrzymana przez miarę $\approx_H$, jeżeli A i B są znormalizowane, a dla pozostałych miar jest ona zawsze spełniona.

Przegląd miary bliskości można znaleźć w [HE 94].

8. ROZMYTE RELACJE

Niech $X_1, X_2, \dots, X_n$ będą zbiorami normalnymi (ostrymi). Relacja rozmyta R na zbiorach $X_1, X_2, \dots, X_n$ jest zbiorem rozmytym na $X_1 \times X_2 \times \dots \times X_n$ z funkcją przynależności $\mu_R: X_1 \times X_2 \times \dots \times X_n \rightarrow [0, 1]$. μ_R może być interpretowana jako moc powiązania pomiędzy $x_1, x_2, \dots, x_n$ w R lub jako stopień określający przynależność $(x_1, x_2, \dots, x_n)$ do relacji R . Przykładem relacji rozmytej może być relacja Zamówienia przedstawiona w tabeli 3.

Tabela 3. Rozmyta relacja Zamówienia

K#	A#	Cena	Jakość	μ_R
138	04	90	3	1
138	17	150	1	1
140	04	100	1	1
141	04	95	2	0.9
141	17	150	1	0.75

Dla projektowania rozmytych modeli relacyjnych baz danych szczególne znaczenie mają rozmyte relacje binarne spełniające pewne własności. Są one definiowane zwykle dla dziedziny, aby można było ocenić stopień powiązania pomiędzy jej elementami. Takimi relacjami są m.in. relacja podobieństwa (ang. *similarity*) i bliskości (ang. *closness*), będące uogólnieniem relacji identyczności.

Niech R będzie rozmytą relacją binarną na X , tj. na X, X , z funkcją przynależności $\mu_R: X \times X \rightarrow [0, 1]$. R jest relacją podobieństwa wtedy i tylko wtedy, gdy jest:

1) zwrotna:

$$\forall x \in X \quad \mu_R(x, x) = 1$$

2) symetryczna:

$$\forall (x_1, x_2) \in X \times X \quad \mu_R(x_1, x_2) = \mu_R(x_2, x_1)$$

3) przechodnia sup-min:

$$\forall (x_1, x_3) \in X \times X \quad \mu_R(x_1, x_3) \geq \sup_{x_2 \in X} \min(\mu_R(x_1, x_2), \mu_R(x_2, x_3))$$

R jest relacją bliskości wtedy i tylko wtedy, gdy jest:

1) zwrotna:

$$\forall x \in X \quad \mu_R(x, x) = 1$$

2) symetryczna:

$$\forall (x_1, x_2) \in X \times X \quad \mu_R(x_1, x_2) = \mu_R(x_2, x_1)$$

Relacja podobieństwa jest szczególnym przypadkiem relacji bliskości.

9. PODSUMOWANIE

Najczęściej informacje dotyczące otaczającego świata są nieprecyzyjne, niepełne lub niepewne. Klasyczne modele danych (w tym model relacyjny) nie pozwalają na przechowywanie i operowanie takiego rodzaju danymi. Jedyną możliwością odzwierciedlenia niepewności jest wartość nieokreślona – null. Logika rozmyta jest narzędziem opisu danych nieprecyzyjnych, odpowiada naturalnemu językowi opisu i rozumowania [ŁACHWA 2001]. Zbiory rozmyte (relacje rozmyte, liczby rozmyte) stosuje się do określania i reprezentowania zmodyfikowanych pojęć logicznych takich jak: wartość logiczna, negacja zdania, spójniki zdaniowe, predykaty, modyfikatory predykatów, reguły wnioskowania.

FUZZY LOGIC

The work presents basic notations of fuzzy logic, uncertainty and impreciseness information. It presents properties of fuzzy sets and operations on fuzzy sets. The extensions rule is discussed. Some elements that are important to fuzzy extension of data models is introduced, for example fuzzy set, are defined and possibility distribution, unit of knowledge representation (linguistic variable) and fuzzy relation is presented. The knowledge of these notations is required to design fuzzy relational database models.

Barbara BIEDRZYCKA
Zygmunt MAZUR*

ROZMYTY MODEL ZWIĄZKÓW ENCJI

W pracy przedstawiono tematykę związaną z rozmytym modelowaniem konceptualnym. Opisano w jaki sposób logika rozmyta może zostać użyta w modelu związków encji. Przedstawiono rozmyte rozszerzenie pojęć klasycznego modelu związków encji takich jak: encje, związki, atrybuty na trzech poziomach – poziomach typów i ich instancji. Dostarczono szerszego spojrzenia na związki pomiędzy encjami w kontekście uczestnictwa i liczebności.

1. WPROWADZENIE

Najczęściej informacje dotyczące otaczającego świata są nieprecyzyjne, niepełne lub niepewne. Wielokrotnie istnieje potrzeba reprezentowania rozmytych danych, dlatego też dużo uwagi poświęcono ich modelowaniu. Obok prac nad rozmytymi rozszerzeniami modeli danych, które pozwalają na nieprecyzyjne wartości atrybutów i pewne rozmyte ograniczenia, wyłoniły się również próby rozmytego rozszerzenia modeli konceptualnych, wśród których można wyróżnić model związków encji. W 1985 Zvieli i Chen [ZVIELI 85] zastosowali teorię zbiorów rozmytych do podstawowych pojęć owego modelu. Wprowadzili rozmyte zbiory encji, związków i atrybutów jak również rozmytość w występowaniu encji i związków. Inne próby można również znaleźć w pracach [RUSPINI 86, VANDENBERGHE 91]. Przedstawiony w pracy rozmyty model związków encji został opracowany na podstawie [KERRE 95, CHEN 98].

2. POZIOMY POJĘCIOWE

Model związków encji $\mathcal{M}$ może być wyrażony w postaci $\mathcal{M} = (\mathcal{E}, \mathcal{R}, \mathcal{A})$, gdzie $\mathcal{E}$ jest zbiorem typów encji, $\mathcal{R}$ jest zbiorem typów związków między typami encji, a $\mathcal{A}$

* Wydziałowy Zakład Informatyki, Wydział Informatyki i Zarządzania Politechniki Wrocławskiej,
50-370 Wrocław, Skwer Idaszewskiego 1, mazur@ci.pwr.wroc.pl

jest zbiorem atrybutów związanych z typami encji i typami związków. Może być on opisany na trzech poziomach.

Pierwszym poziomem jest poziom modelowy, oznaczany przez $L1(\mathcal{M})$, który odnosi się do typów podstawowych konstrukcji modelowych. Zawiera on wszystkie typy encji, typy związków i atrybuty, które tworzą dany model $\mathcal{M}$. Można zdefiniować go w postaci:

$$L1(\mathcal{M}) = (\mathcal{E}, \mathcal{R}, \mathcal{A}_{\mathcal{E}}, \mathcal{A}_{\mathcal{R}})$$

gdzie: $\mathcal{E} = \{E \mid E \text{ jest typem encji}\}$, $\mathcal{R} = \{R \mid R \text{ jest typem związku}\}$, $\mathcal{A}_{\mathcal{E}} = \{\mathcal{A}_E \mid \mathcal{A}_E \text{ jest zbiorem atrybutów typu encji } E\}$ i $\mathcal{A}_{\mathcal{R}} = \{\mathcal{A}_R \mid \mathcal{A}_R \text{ jest zbiorem atrybutów typu związku } R\}$. Poziom $L1(\mathcal{M})$ jest reprezentowany na diagramie związków encji. Na przykład diagram związków encji pokazany w [BIEMAZUR 02a] na rysunku 3 przedstawia $L1(\mathcal{M}_{\text{Uniwersytet}}) = (\mathcal{E}, \mathcal{R}, \mathcal{A}_{\mathcal{E}}, \mathcal{A}_{\mathcal{R}})$, gdzie

$$\mathcal{E} = \{\text{STUDENT, KURS, WYDZIAŁ, PRACOWNIK, OSOBA ZWIĄZANA}\}$$

$$\mathcal{R} = \{\text{Wybiera, Prowadzony_przez, Zatrudnia, Studiuje_na, Ma_promotora, Zarządzany_przez, Kieruje, Związany_z}\}$$

$$\mathcal{A}_{\mathcal{E}} = \{\mathcal{A}_{\text{STUDENT}}, \mathcal{A}_{\text{KURS}}, \mathcal{A}_{\text{WYDZIAŁ}}, \mathcal{A}_{\text{PRACOWNIK}}, \mathcal{A}_{\text{OSOBA ZWIĄZANA}}\} \\ = \{\{S\#, \text{Nazwa, Płeć, DataUr, Adres}\}, \dots, \{\text{Nazwa, Płeć, DataUr, Powiązanie}\}\}$$

$$\mathcal{A}_{\mathcal{R}} = \{\mathcal{A}_{\text{Wybiera}}, \mathcal{A}_{\text{Zarządzany_przez}}\} = \{\{\text{Ocena}\}, \{\text{Data}\}\}$$

Drugi poziom odnosi się do wartości typów encji, typów związków i atrybutów. Dla każdej encji e danego typu encji lub związku r danego typu związku istnieją wartości atrybutów związanych z encją e lub związkiem r , dlatego też poziom wartości może być dalej podzielony na dwa podpoziomy, oznaczane przez $L2(\mathcal{V}(\mathcal{E}, \mathcal{R}))$ i $L3(\mathcal{V}(\mathcal{A}))$, reprezentujące odpowiednio wartości typów encji i typów związku oraz wartości ich atrybutów.

$L2(\mathcal{V}(\mathcal{E}, \mathcal{R}))$ można zdefiniować w postaci:

$$L2(\mathcal{V}(\mathcal{E}, \mathcal{R})) = (\mathcal{V}(\mathcal{E}), \mathcal{V}(\mathcal{R}))$$

gdzie

$\mathcal{V}(\mathcal{E}) = \{\mathcal{V}(E) \mid \mathcal{V}(E) \text{ jest zbiorem wszystkich wartości typu encji } E, E \in \mathcal{E}\}$
i $\mathcal{V}(\mathcal{R}) = \{\mathcal{V}(R) \mid \mathcal{V}(R) \text{ jest zbiorem wszystkich wartości typu związku } R, R \in \mathcal{R}\}$.
Dla modelu reprezentującego rzeczywistość uniwersytetu ([BIEMAZUR 02a], rysunek 3) $L2(\mathcal{V}(\mathcal{E}, \mathcal{R}))$ może mieć postać:

$$\mathcal{V}(\mathcal{E}) = \{\mathcal{V}(\text{STUDENT}), \mathcal{V}(\text{KURS}), \mathcal{V}(\text{WYDZIAŁ}), \mathcal{V}(\text{PRACOWNIK}), \\ \mathcal{V}(\text{OSOBA_ZWIĄZANA})\} \\ = \{\{T. Nowak, J. Anioł, J. Kowalski, \dots\}, \\ \{\text{Bazy danych, Systemy ekspertowe, Algebra liniowa, \dots}\}, \dots, \\ \{E. Nowak, A. Bąk, H. Kowalski, \dots\}\}$$

$$\begin{aligned}\mathcal{V}(\mathcal{R}) = & \{ \mathcal{V}(\text{Wybiera}), \mathcal{V}(\text{Prowadzony_przez}), \mathcal{V}(\text{Zatrudnia}), \mathcal{V}(\text{Studiuje_na}), \\ & \mathcal{V}(\text{Ma_promotora}), \mathcal{V}(\text{Zarządzany_przez}), \mathcal{V}(\text{Kieruje}), \mathcal{V}(\text{Związany_z}) \} \\ = & \{ \{ (\text{T. Nowak}, \text{Bazy danych}), (\text{T. Nowak}, \text{Systemy ekspertowe}), \\ & (\text{J. Anioł}, \text{Algebra liniowa}), (\text{J. Kowalski}, \text{Bazy danych}), \dots \}, \\ & \{ (\text{B. Bałazy}, \text{Bazy danych}), (\text{P. Grzyb}, \text{Systemy ekspertowe}), \dots \}, \dots \}\end{aligned}$$

Dla przedstawionych $\mathcal{V}(\mathcal{E})$ i $\mathcal{V}(\mathcal{R})$ każda wartość typu encji jest identyfikowana przez swoją nazwę, każda wartość typu związku – przez nazwy przynależnych encji.

$L3(\mathcal{V}(\mathcal{A}))$ można zdefiniować w postaci:

$$L3(\mathcal{V}(\mathcal{A})) = (\mathcal{V}(\mathcal{A}_{\mathcal{E}}), \mathcal{V}(\mathcal{A}_{\mathcal{R}})),$$

gdzie:

$$\begin{aligned}\mathcal{V}(\mathcal{A}_{\mathcal{E}}) = & \{ \mathcal{V}(\mathcal{A}_{\mathcal{E}}) = \mathcal{V}(A_1, A_2, \dots, A_k) \mid E \in \mathcal{E}, A_i \text{ jest atrybutem typu encji } E, \\ & i = 1, \dots, k \text{ oraz } \mathcal{V}(A_1, A_2, \dots, A_k) \subseteq \text{Dom}(A_1) \times \text{Dom}(A_2) \times \dots \times \text{Dom}(A_k) \}, \\ \mathcal{V}(\mathcal{A}_{\mathcal{R}}) = & \{ \mathcal{V}(\mathcal{A}_{\mathcal{R}}) = \mathcal{V}(B_1, B_2, \dots, B_n) \mid R \in \mathcal{R}, B_j \text{ jest atrybutem typu związku } R, \\ & j = 1, \dots, n \text{ oraz } \mathcal{V}(B_1, B_2, \dots, B_n) \subseteq \text{Dom}(B_1) \times \text{Dom}(B_2) \times \dots \times \text{Dom}(B_n) \}.\end{aligned}$$

Powyższe $\mathcal{V}(\mathcal{A}_{\mathcal{E}})$ i $\mathcal{V}(\mathcal{A}_{\mathcal{R}})$ są zdefiniowane zgodnie z założeniem, że wszystkie atrybuty A_i i B_j są jednowartościowe, tzn. że dla danej encji $e \forall e(A_i) \in \text{Dom}(A_i)$, gdzie $e(A_i)$ jest wartością atrybutu A_i , a dla danego związku $r \forall r(B_j) \in \text{Dom}(B_j)$, gdzie $r(B_j)$ jest jego wartością atrybutu B_j . Jeżeli dopuszcza się atrybuty wielowartościowe, wówczas dla $\mathcal{V}(\mathcal{A}_{\mathcal{E}})$ zachodzi zależność:

$$\mathcal{V}(A_1, A_2, \dots, A_k) \subseteq 2^{\text{Dom}(A_1)} \times 2^{\text{Dom}(A_2)} \times \dots \times 2^{\text{Dom}(A_k)}$$

a dla $\mathcal{V}(\mathcal{A}_{\mathcal{R}})$:

$$\mathcal{V}(B_1, B_2, \dots, B_n) \subseteq 2^{\text{Dom}(B_1)} \times 2^{\text{Dom}(B_2)} \times \dots \times 2^{\text{Dom}(B_n)}$$

gdzie $2^{\text{Dom}(A_i)}$ i $2^{\text{Dom}(B_j)}$ oznaczają odpowiednio moc zbioru $\text{Dom}(A_i)$ i $\text{Dom}(B_j)$. Dla danej encji e i związku $r \forall e(A_i) \in \text{Dom}(A_i)$, $\forall r(B_j) \in \text{Dom}(B_j)$. Rozważając atrybuty WYDZIAŁU można założyć, że wydział $w = \text{Elektronika}$ ($W\# = 18$) ma dwie lokalizacje „Baśniowa 2” i „Zielona 69”, wartości atrybutów dla ($W\#$, Nazwa, Lokalizacja) będą wówczas wynosiły:

$$(V_w(W\#), V_w(\text{Nazwa}), V_w(\text{Lokalizacja})) = (\{18\}, \{\text{Elektronika}\}, \{\text{„Cicha 2”}, \text{„Zielona 69”}\})$$

Należy zauważyć, że na poziomie $L1(\mathcal{M})$ mówi się o typie encji E oraz o typie związku R , a na poziomie $L2(\mathcal{V}(\mathcal{E}, \mathcal{R}))$ mówi się o zbiorze wszystkich wartości E , tj. $\mathcal{V}(\mathcal{E})$, oraz o zbiorze wszystkich wartości związków, tj. $\mathcal{V}(\mathcal{R})$. Takie przedstawianie ma pomóc w rozróżnieniu notacji typu i wartości na różnych poziomach abstrakcji. W rzeczywistości jednak typ encji jest zbiorem wszystkich wartości encji, a typ związku jest zbiorem wszystkich wartości związków.

3. ROZMYTE ENCJE, ZWIĄZKI I ATRYBUTY

Rozmytość w modelu związków encji można rozważać na wszystkich trzech poziomach.

Na poziomie $L1(\mathcal{M})$ zbiory $\mathcal{E}$, $\mathcal{R}$, $\mathcal{A}_{\mathcal{E}}$, $\mathcal{A}_{\mathcal{R}}$ mogą być zbiorami rozmytymi, odzwierciedlającymi możliwość częściowej przynależności odpowiadających im typów do modelu. Stąd otrzymuje się:

$$\mathcal{E} = \{\mu_{\mathcal{E}}(E)/E \mid E \in \mathcal{E} \text{ i } \mu_{\mathcal{E}}(E) \in [0, 1]\}$$

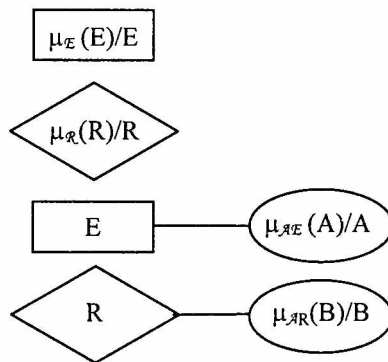
$$\mathcal{R} = \{\mu_{\mathcal{R}}(R)/R \mid R \text{ jest typem związku między typami encji z } \mathcal{R} \text{ i } \mu_{\mathcal{R}}(R) \in [0, 1]\}$$

$$\mathcal{A}_{\mathcal{E}} = \{\mu_{\mathcal{A}_{\mathcal{E}}}(A)/A \mid A \text{ jest atrybutem typu encji } E \text{ i } \mu_{\mathcal{A}_{\mathcal{E}}}(A) \in [0, 1]\}$$

$$\mathcal{A}_{\mathcal{R}} = \{\mu_{\mathcal{A}_{\mathcal{R}}}(B)/B \mid B \text{ jest atrybutem typu związku } R \text{ i } \mu_{\mathcal{A}_{\mathcal{R}}}(B) \in [0, 1]\}$$

gdzie $\mu_{\mathcal{E}}$, $\mu_{\mathcal{R}}$, $\mu_{\mathcal{A}_{\mathcal{E}}}$ i $\mu_{\mathcal{A}_{\mathcal{R}}}$ są funkcjami przynależności odpowiednio dla $\mathcal{E}$, $\mathcal{R}$, $\mathcal{A}_{\mathcal{E}}$, $\mathcal{A}_{\mathcal{R}}$. Przykładowo w rozmytym modelu związków encji, $\mathcal{E}$ może mieć postać: $\{1/\text{Pracownik}, 1/\text{Wydział}, 1/\text{Projekt}, 0.9/\text{Klient}, \dots\}$.

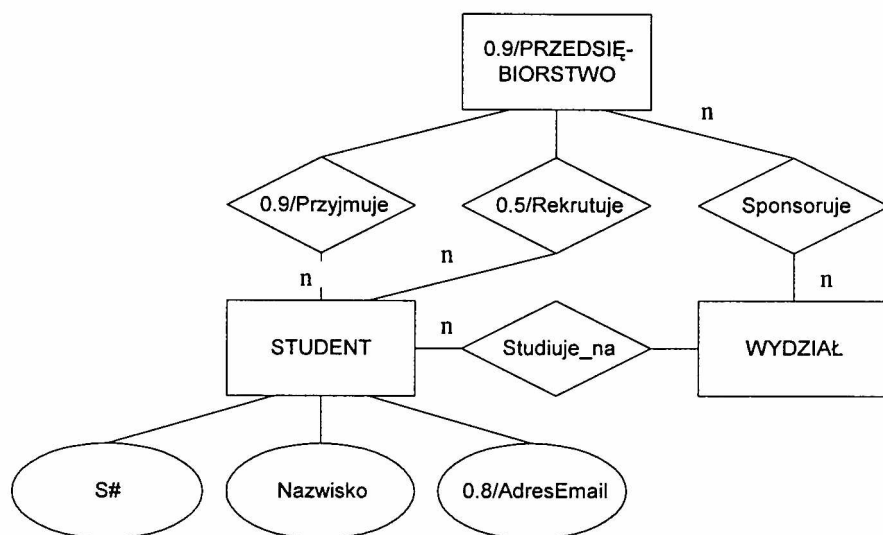
Rozmytość na poziomie $L1(\mathcal{M})$ jest reprezentowana na rozmytym diagramie związków encji, którego notacje przedstawia rysunek 1. Wraz z nazwami typów encji, typów związków i atrybutów umieszczone są stopnie ich przynależności do modelu. Jeżeli stopień przynależności jest równy 1, to jest on pomijany. Na przykład dla typu encji E o stopniu przynależności równym 1, tj. $\mu_{\mathcal{E}}(E) = 1$, zamiast „1/ E ” wprowadza się „ E ”.



Rys. 1. Notacja rozmytego diagramu związków encji

Dla $0 < \mu_{\mathcal{E}}(E) < 1$ mówi się, że E jest związany z modelem $\mathcal{M}$ w stopniu $\mu_{\mathcal{E}}(E)$, co oznacza, że nie ma całkowitej pewności co do roli, jaką odgrywa E w kontekście modelu i/lub istnieje częściowy stopień zgodności między informacją, która ma być reprezentowana a rzeczywistością, którą model ma opisywać. Na przykład [CHEN 98] przedsiębiorstwa mogą dostarczać funduszy dla wydziałów na projekty badawcze lub

przyjmować studentów na praktyki, stąd typ encji PRZEDSIĘBIORSTWO może zostać włączony do modelu. Jeżeli jednak nie jest całkowicie pewna taka rola PRZEDSIĘBIORSTWA lub taka informacja jest częściowo powiązana z rozważaną rzeczywistością, nie jest całkowicie istotna, wówczas wprowadzenie PRZEDSIĘBIORSTWA może być częściowe ($0.9/\text{PRZEDSIĘBIORSTWO}$), co pokazuje rysunek 2.



Rys. 2. Typ encji PRZEDSIĘBIORSTWO z częściowym stopniem 0.9 [CHEN 98]

Typy encji w modelu są razem połączone przez typy związków. Jeżeli częściowe włączenie typu encji E jest związane z niepewnością roli, jaką ma spełniać, to zwykle istnieje typ związku R taki, że $\mu_R(R) = \mu_E(E)$. Na przykład, dla $0.9/\text{PRZEDSIĘBIORSTWO}$ byłby to typ związku $0.9/\text{Przyjmuje}$.

Jeżeli częściowe włączenie typu encji E jest związane z niepewnością jego znaczenia dla rozważanej rzeczywistości, wówczas może istnieć typ związku R, dla którego $\mu_R(R) \geq \mu_E(E)$. Na przykład z całą pewnością wiadomo, że wydziały są związane z przedsiębiorstwami sponsorującymi projekty badawcze, tj. $1/\text{Sponsoruje}$, ale taka informacja o sponsorowaniu może nie być w ogóle interesująca lub częściowo interesująca dla modelu. Stąd możliwe jest istnienie zależności:

$$1 = \mu_R(\text{Sponsoruje}) > \mu_E(\text{PRZEDSIĘBIORSTWO}) = 0.9.$$

Dodatkowo możliwe jest istnienie typu związku R, którego stopień przynależności jest mniejszy od stopni uczestniczących w nim typów encji. Przypadek ten może odzwierciedlać zarówno niepewność mocy typu związku, tzn. powiązania między typami encji, jak i niepewność jego znaczenia dla rozważanej rzeczywistości. Pierwszy przypadek ma miejsce głównie, gdy stopnie przynależności dla typów encji

przedstawiają niepewność ich znaczenia, a drugi przypadek, gdy przedstawiają one niepewność ich ról. Przykładem może być 0.5/Rekrutuje, który łączy 1/STUDENT i 0.9/PRZEDSIĘBIORSTWO (rysunek 2).

Wprowadzenie atrybutu z częściowym stopniem przynależności do $\mathcal{A}_E$ często jest związane z niepewnością, czy własność, którą opisuje atrybut, jest interesująca i potrzebna do reprezentowania. Na rysunku 2 AdresEmail jest atrybutem STUDENTA w stopniu 0.8, co oznacza, że nie ma pewności, czy istnieje potrzeba reprezentowania i operowania taką informacją, jaką jest adres email studenta.

Rozmytość wprowadzana na poziomie $L1(\mathcal{M})$ odzwierciedla częściową wiedzę projektantów modelu i ekspertów dziedziny o granicach modelu. Umożliwia połączenie różnych spojrzeń na rozważaną rzeczywistość.

Często używa się progów $\alpha_E, \alpha_R, \alpha_{AE}, \alpha_{AR}$, aby utrzymać tylko te konstrukcje modelowe w $\mathcal{E}, \mathcal{R}, \mathcal{A}_E, \mathcal{A}_R$, których stopień uczestnictwa jest odpowiednio równy co najmniej $\alpha_E, \alpha_R, \alpha_{AE}, \alpha_{AR}$, gdzie $\alpha_E, \alpha_R, \alpha_{AE}, \alpha_{AR} \in (0, 1]$. Umożliwia to reprezentowanie tylko tych konstrukcji modelowych, które mają odpowiednie znaczenie.

Rozważając rozmytość na poziomie $L2(\mathcal{V}(\mathcal{E}, \mathcal{R}))$, dla każdego typu encji E i typu związku R zbiory ich wartości mogą być zbiorami rozmytymi:

$$\mathcal{V}(E) = \{\mu_E(e)/e \mid e \text{ jest wartością typu encji } E\}$$

$$\mathcal{V}(R) = \{\mu_R(r)/r \mid r \text{ jest wartością typu związku } R\}$$

gdzie $\mu_E(e)$ i $\mu_R(r)$ są stopniami przynależności odpowiednio e i r do E i R .

Dla każdego elementu e należącego do $\text{Dom}(E)$ zbiór rozmyty $\mathcal{V}(E)$ reprezentuje stopień przynależności dla e jako wartości E . Częściowy stopień e w $\mathcal{V}(E)$ często wynika z niepewności, czy charakterystyki (np. wartości atrybutów) encji e są zgodne z własnościami E lub z częściowej o niej wiedzy albo z częściowej wiedzy o informacjach powiązanych z wartością e . Na przykład, jeżeli nie wiadomo, czy przedsiębiorstwo HIJ uczestniczy w jakimś związku (Rekrutuje, Przyjmuje lub Sponsoruje) z innymi encjami w modelu, można przyjąć 0.7/HIJ jako wartość w $\mathcal{V}(\text{PRZEDSIĘBIORSTWO})$. Podobnie dla n -arnego związku R $\mathcal{V}(R)$ jest zbiorem rozmytym, który reprezentuje stopień mocy każdego związku między n uczestniczącymi encjami.

Rozważając model uniwersytetu, $\mathcal{V}(\text{Prowadzony_przez})$ może zawierać (Bazy danych, J. Nowak) ze stopniem 0.8, tj. (Bazy danych, J. Nowak, 0.8), co równoważnie zapisuje się w postaci 0.8/(Bazy danych, J. Nowak, 0.8). Częściowy stopień związku może być interpretowany jako możliwość, że J. Nowak prowadzi Bazy danych albo stopień niepewności, czy informacja „J. Nowak prowadzi Bazy danych” jest prawdziwa.

Trzeci poziom $L3(\mathcal{V}(\mathcal{A}))$ dotyczy wartości atrybutów. Dla każdego atrybutu A jego wartości, tj. $V(A)$, mogą być zbiorami rozmytymi, co odzwierciedla nieprecyzyjność lub częściową o nich wiedzę:

$$V_e(A_i) = \{\mu_{A_i}(a)/a \mid a \subseteq \text{Dom}(A_i)\}, i = 1, 2, \dots, k$$

$$V_r(B_j) = \{\mu_{B_j}(b)/b \mid b \subseteq \text{Dom}(B_j)\}, j = 1, 2, \dots, n$$

Na przykład, jeżeli $A = \text{„Wiek”}$, to wartościami A mogą być: $V_1(A) = \text{„Młody”}$, $V_2(A) = \{1/21, 0.9/22\}$, tzn. 21 albo 22.

Jeżeli A_i jest jednowartościowym atrybutem, wtedy $V_e(A_i)$ jest zbiorem rozmytym, którego wszystkie elementy należą do $\text{Dom}(A_i)$ i wzajemnie się wykluczają albo $V_e(A_i)$ jest wyrażona przez wykluczający rozkład możliwości ([BIEMAZUR 02f]), na przykład $V_e(\text{Wiek}) = \text{„Młody”}$ albo $V_e(\text{Wiek}) = \{1/21, 0.9/22\}$.

Jeżeli A_i jest atrybutem wielowartościowym, wtedy $V_e(A_i)$ może być wyrażona zarówno przez zbiór rozmyty na $\text{Dom}(A_i)$, jak i przez wykluczający rozkład możliwości na $2^{\text{Dom}(A_i)}$. Przykładem jest atrybut JęzykObcy, dla danej encji e

$$V_e(A_i) = \{1/\text{Angielski}, 0.8/\text{Niemiecki}, 0.3/\text{Francuski}\}$$

lub

$$V_e(A_i) = \{1/\{\text{Angielski}\}, 0.8/\{\text{Niemiecki}\}, 0.3/\{\text{Francuski}\}, \\ 0.8/\{\text{Angielski}, \text{Niemiecki}\}, 0.3/\{\text{Angielski}, \text{Francuski}\}, \\ 0.3/\{\text{Niemiecki}, \text{Francuski}\}, 0.3/\{\text{Angielski}, \text{Niemiecki}, \text{Francuski}\}\}.$$

4. ROZMYTE REGUŁY ZWIĄZKÓW

Pojęcie rozmytości zostało również wprowadzone do reguły uczestnictwa i liczebności. Dla przejrzystości zostaną rozważone związki binarne.

Niech R będzie typem związku łączącym typy encji E i F , tj. $R = \{(v, w, \alpha) \mid v \in E, w \in F \text{ i } \alpha \in [0, 1], (\alpha = \mu_R(u, w))\}$. Typ encji E obligatoryjnie uczestniczy w typie związku R , jeżeli $\forall e \in E$ spełniona jest zależność:

$$e \in \{v \mid (v, w, \alpha) \in R \wedge \alpha > 0\}$$

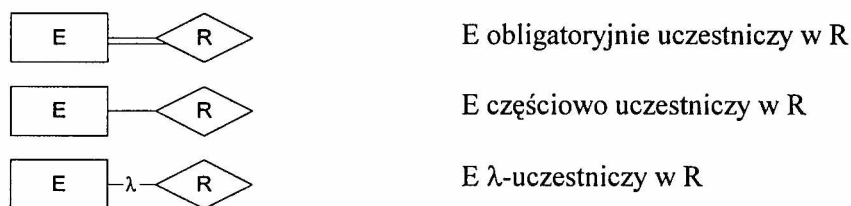
Inaczej mówiąc, E obligatoryjnie uczestniczy w R , jeżeli każda encja e należąca do E uczestniczy w przynajmniej jednym związku należącym do R z dodatnim stopniem. E opcjonalnie uczestniczy w R , jeżeli $\exists e_0 \in E$ takie, że:

$$e_0 \notin \{v \mid (v, w, \alpha) \in R \wedge \alpha > 0\}$$

W rozmytym modelu związków encji istnieje dodatkowo pojęcie λ -uczestnictwa. Typ encji E λ -uczestniczy w R , jeżeli dla każdej encji e należącej do E istnieje f w F taka, że $\alpha = \mu_R(e, f) > \lambda$, gdzie λ jest wartością z przedziału $[0, 1]$. Zapisując wzorem, $\lambda \in [0, 1], \forall e \in E$

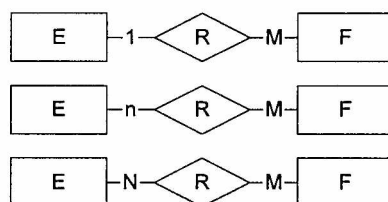
$$e \in \{v \mid (v, w, \alpha) \in R \wedge \alpha \geq \lambda\}$$

Notacje obligatoryjnego, opcjonalnego i λ - uczestnictwa na rozmytym diagramie związków encji przedstawia rysunek 3.



Rys. 3. Notacja diagramu związków encji dla rozmytej reguły uczestnictwa

Do reguły liczebności, która dotyczy liczby encji uczestniczących w związkach danego typu związku z daną encją, zostały wprowadzone elementy rozmytości, aby móc reprezentować więcej informacji o owej liczbie encji, na przykład aby odzwierciedlić wartość „około 20”. W rozmytym modelu związków encji poza liczebnościami 1:1, 1:n, n:m wprowadza się liczebności: 1:M, n:M, N:M, gdzie N, M reprezentują zbiory rozmyte. Notacje rozmytych liczebności na diagramie związków encji przedstawia rysunek 4.



Rys. 4. Notacja rozmytej liczebności

Na przykład, jeżeli na uniwersytecie wolno studentowi wybrać około 10 lub mniej kursów w roku akademickim, wówczas dla typu związku Wybiera(STUDENT, KURS) istnieje liczebność n:M, gdzie M = „około 10” (rysunek 5).



Rys. 5. Przykład liczebności n:M

Reguła liczebności N:M dla typu związku R między typem encji E i typem encji F oznacza, że „liczba” encji f należących do F związanych z e należącą do E musi być w pewien sposób zgodna z M i „liczba” encji e należących do E związanych z f w F musi być w pewien sposób zgodna z N, przy czym dla projektowania rozmytych baz

danych E i F są zawsze zbiorami skończonymi. Rozważając przypadek precyzyjnych, dokładnych danych, tzn. $\mu_R(e_i, f_j) \in \{0, 1\}$, oraz gdy M i N są precyzyjnymi liczbami (integer), liczebność N:M wymaga spełnienia zawsze następujących zależności:

$$\sum_{j=1}^t \mu_R(e_i, f_j) \leq M \quad i = 1, 2, \dots, s$$

$$\sum_{i=1}^s \mu_R(e_i, f_j) \leq N \quad j = 1, 2, \dots, t$$

Rozważając przypadek rozmyty, w którym $\mu_R(e_i, f_j) \in [0, 1]$, M i N są zbiorami rozmytymi reprezentującymi rozmytą liczbą wystąpień, istnieją różne sposoby postępowania z owym rozmytym ograniczeniem liczebności. Przykładem może być rozwiązanie oparte na pojęciu Σ count i środka ciężkości.

Dla każdej encji e_i należącej do zbioru E, związane z nią encje ze zbioru F i ich stopnie $\mu_R(e_i, f_j)$ mogą być traktowane jako zbiór rozmyty:

$$e_i R = \{\mu_{e_i R}(f_j) / f_j \mid \mu_{e_i R}(f_j) = \mu_R(e_i, f_j), f_j \in F\}$$

Wówczas Σ count zbioru $e_i R$, oznaczany przez $\Sigma \text{count}(e_i R)$, oblicza się jako sumę:

$$\sum_{j=1}^t \mu_{e_i R}(f_j)$$

Reprezentuje ona „liczbę” efektywnych encji f w F.

Rozmyty zbiór M może być charakteryzowany przez środek ciężkości. Niech $\mu_M(x)$ będzie stopniem przynależności dla x, $x \in U$, gdzie U jest zbiorem liczb rzeczywistych, wówczas środek ciężkości dla M jest mierzony następująco:

$$G(M) = \frac{\int_U \mu_M(x) \cdot x dx}{\int_U \mu_M(x) dx}$$

gdzie $\int_U \mu_M(x) dx \neq 0$. Gdy $U = \{x_1, x_2, \dots, x_k\}$:

$$G(M) = \frac{\sum_{p=1}^k \mu_M(x_p) x_p}{\sum_{p=1}^k \mu_M(x_p)}$$

gdzie $\sum_{p=1}^k \mu_M(x_p) \neq 0$

Liczebność $N:M$ dla E i F oznacza, że istnieje typ związku R wiele-do-wiele pomiędzy E i F , dla każdego e_i należącego do zbioru E może istnieć M encji należących do F związanych z e_i , tzn. M jest zbiorem rozmytym na $U = \{x_1, x_2, \dots, x_k\}$ i spełniona jest zależność:

$$\sum count(e_i R) \leq G(M)$$

lub

$$\sum_{j=1}^t \mu_{eiR}(f_j) = \sum_{j=1}^t \mu_R(e_i, f_j) \leq \frac{\sum_{p=1}^k \mu_M(x_p) x_p}{\sum_{p=1}^k \mu_M(x_p)}$$

oraz dla każdego f_j należącego do zbioru F może istnieć N encji należących do E związanych z f_j , tzn. N jest zbiorem rozmytym na $U = \{y_1, y_2, \dots, y_h\}$ i spełniona jest zależność:

$$\sum count(f_j R) \leq G(N)$$

lub

$$\sum_{i=1}^s \mu_{fjR}(e_i) = \sum_{i=1}^s \mu_R(e_i, f_j) \leq \frac{\sum_{p=1}^h \mu_N(y_p) y_p}{\sum_{p=1}^h \mu_N(y_p)}$$

5. DZIEDZICZENIE W ROZMYTYM MODELU ZWIĄZKÓW ENCJI

Wprowadzenie do modelu związków encji pojęć takich jak: podklasa i nadklasa, specjalizacja i generalizacja, podklasa dzielona i kategoria, dziedziczenie atrybutów wzbogaciło jego możliwości przedstawiania świata rzeczywistego. Konstrukcje te zostały rozszerzone używając logiki rozmytej, aby uwzględnić niepewność i nieprecyzyjność występującą na poziomie conceptualnym.

Rozmytość nadklas i podklas można rozważać na wszystkich poziomach.

Na poziomie $L2(\mathcal{V}(\mathcal{E}, \mathcal{A}))$ nadklasa F i jej podklasy $E_1, E_2, \dots, E_k$ są zbiorami rozmytymi z funkcjami przynależności odpowiednio $\mu_F, \mu_{E_1}, \mu_{E_2}, \dots, \mu_{E_k}$. Spełniona jest wówczas zależność:

$$\max(\mu_{E_1}(e), \mu_{E_2}(e), \dots, \mu_{E_k}(e)) \leq \mu_F(e)$$

Oznacza to, że stopień wyrażający przynależność encji do podklasy nie może być większy od stopnia jej przynależności do jej nadklasy. Na przykład jeżeli istnieje nadklasa $F = \text{PRACOWNIK}$ i jedną z jej podklas jest INŻYNIER , to jeżeli Nowak jest inżynierem w stopniu 0.8, wtedy jest on pracownikiem w stopniu co najmniej 0.8. Tak więc $0.8 = \mu_{\text{Inżynier}}(\text{Nowak}) \leq \mu_{\text{Pracownik}}(\text{Nowak})$.

Fakt, że nadklasa F ma rozłączne podklasy $E_1, E_2, \dots, E_k$, oznacza, że nie istnieje żadna encja e, E_{i_0} i E_{j_0} takie, że

$$\min(\mu_{E_{i_0}}(e), \mu_{E_{j_0}}(e)) > 0 \quad (i_0, j_0 \in \{1, 2, \dots, k\})$$

Inaczej mówiąc, nie istnieje żadna encja, która należałaby do dwóch lub więcej podklas z niezerowymi stopniami przynależności. Jeżeli nadklasa składa się z nakładających się podklas $E_1, E_2, \dots, E_k$, wówczas istnieje encja e, E_{i_0} i E_{j_0} takie, że

$$\min(\mu_{E_{i_0}}(e), \mu_{E_{j_0}}(e)) > 0 \quad (i_0, j_0 \in \{1, 2, \dots, k\})$$

Rozważając poziom $L1(\mathcal{M})$, stopień przynależności podklasy do modelu, tj. $\mu_E(E_i)$, nie może być większy od stopnia przynależności jej nadklasy. Można zapisać to wzorem:

$$\max(\mu_E(E_1), \mu_E(E_2), \dots, \mu_E(E_k)) \leq \mu_E(F)$$

Z pojęciami nadklas i podklas związane są procesy wobec siebie wzajemnie odwrotne nazywane specjalizacją i generalizacją. Całkowita specjalizacja (generalizacja) może zostać opisana w rozmytym modelu związków encji zależnością:

$$\forall e, \exists E_0 \in \{E_1, E_2, \dots, E_k\}, \mu_F(e) \geq \mu_{E_0}(e) > 0$$

Oznacza to, że jeżeli encja e należy do F z niezerowym stopniem, to musi również należeć z niezerowym stopniem do jednej z jej podklas, E_0 . Częściowa specjalizacja (generalizacja) przedstawia, że $\exists e_0$ takie, że

$$\mu_F(e_0) > \max(\mu_{E_1}(e_0), \mu_{E_2}(e_0), \dots, \mu_{E_k}(e_0)) = 0$$

co oznacza, że encja e_0 należąca do F z $\mu_F(e_0) > 0$ nie należy do żadnej z jej podklas.

Mając zdefiniowaną rozmytą specjalizację (generalizację), a więc dysponując zasadami mówiącymi, kiedy dana encja przynależy do danej podklasy (nadklasy), można określić dokładnie zależności między funkcjami przynależności dla nadklas i podklas występujące na poziomie $L2(\mathcal{V}(\mathcal{E}, \mathcal{A}))$ oraz zależności między stopniami przynależności nadklas i podklas do $\mathcal{E}$ wprowadzane na poziomie $L1(\mathcal{M})$. Na przykład w pracy [CHEN 98] autor przedstawia specjalizację zdefiniowaną na atrybucie (ang. *attribute-defined specialization*) i wprowadza dla niej wzory dla funkcji przynależności podklas, gdy nadklasa jest zbiorem rozmytym, a wartość

atrybutu, na którym opiera się specjalizacja, jest nieprecyzyjna lub rozmyta ($L2(\mathcal{V}(\mathcal{E}, \mathcal{A}))$) oraz wzory określające wartość stopnia przynależności utworzonej podklasy do $\mathcal{E}$ na podstawie stopnia przynależności nadklasy i atrybutu do $\mathcal{E}(L1(\mathcal{M}))$.

Do rozmytego modelu związków encji wprowadzono również pojęcia rozmytych podklas dzielonych i rozmytych kategorii, które również rozważa się na poziomach $L1(\mathcal{M})$ i $L2(\mathcal{V}(\mathcal{E}, \mathcal{A}))$.

Rozważając podklasę dzieloną E i jej n nadklas $F_1, F_2, \dots, F_n$, na poziomie $L2(\mathcal{V}(\mathcal{E}, \mathcal{A}))$ dla każdej encji e jej stopień przynależności do dzielonej podklasy E , tj. $\mu_E(e)$, nie może być większy od stopnia przynależności do jej nadklas F_i ($i = 1, 2, \dots, n$), tj. $\mu_{F_i}(e)$. Stąd

$$\mu_E(e) \leq \min(\mu_{F_1}(e), \mu_{F_2}(e), \dots, \mu_{F_n}(e))$$

Na poziomie $L1(\mathcal{M})$ dla każdej podklasy E spełniona jest zależność:

$$\mu_E(E) \leq \min(\mu_E(F_1), \mu_E(F_2), \dots, \mu_E(F_n))$$

gdzie $\mu_E(F_1), \mu_E(F_2), \dots, \mu_E(F_n)$ i $\mu_E(E)$ są stopniami przynależności odpowiednio $F_1, F_2, \dots, F_n$ i E do $\mathcal{E}$.

Rozważając kategorie E i jej n nadklas $F_1, F_2, \dots, F_n$, na poziomie $L2(\mathcal{V}(\mathcal{E}, \mathcal{A}))$ dla każdej encji należącej do kategorii istnieje przynajmniej jedna jej nadklasa $F_0 \in \{F_1, F_2, \dots, F_n\}$ taka, że stopień przynależności encji do podklasy, tj. $\mu_E(E)$, nie może być większy od stopnia przynależności encji do nadklasy F_0 , tj. $\mu_{F_0}(E)$. Stąd

$$\mu_E(e) \leq \max(\mu_{F_1}(e), \mu_{F_2}(e), \dots, \mu_{F_n}(e))$$

Na poziomie $L1(\mathcal{M})$ dla każdej kategorii E zachodzi zależność:

$$\mu_E(E) \leq \mu_E(F_0) \leq \max(\mu_E(F_1), \mu_E(F_2), \dots, \mu_E(F_n))$$

gdzie $\mu_E(F_1), \mu_E(F_2), \dots, \mu_E(F_n)$ i $\mu_E(E)$ są stopniami przynależności odpowiednio $F_1, F_2, \dots, F_n$ i E do $\mathcal{E}$.

Ważnym procesem w modelu związków encji związanym z pojęciami podklasy i nadklasy jest dziedziczenie przez podklasy związków i atrybuty swoich nadklas.

Dziedziczenie związków oznacza, że encja podklasy dziedziczy wszystkie związki, do których przynależy jako encja nadklasy. Na przykład można rozważyć typ związku Wybiera łączącego typy encji STUDENT i KURS i związek ($s1, c7, 0.9$) typu Wybiera reprezentujący fakt, że stopień możliwości wyboru przez studenta $s1$ kursu $c1$ równa się 0.9 . STUDENT może być specjalizowany na dwie podklasy STUDENT AKTUALNY i ABSOLWENT, jeżeli $s1$ należy do podklasy ABSOLWENT (dla $s1$ TypStudenta = „absolwent”) w stopniu 0.8 , to $s1$ dziedziczy związek ($s1, c7, \min(0.9, 0.8)$). Formalnie niech R będzie typem związku, $R: E \times F \rightarrow [0, 1]$, ($e_i, f_j, \mu_R(i, j)$) będzie związkiem typu R z $\mu_R(i, j)$ będącym stopniem przynależności do R wyrażającym zgodność (e_i, f_j) z R i niech e_i będzie specjalizowany do podklasy przez

predykat p , wtedy e_i dziedziczy związek pomiędzy e_i i f_j ze stopniem równym $\mu_R(i, j) \wedge \text{Prawda}(p)$. Inaczej mówiąc, e_i dziedziczy $(e_i, f_j, \mu_R(i, j) \wedge \text{Prawda}(p))$. Definicje dla związków binarnych można rozszerzyć na związki n -arne.

Rozważając dziedziczenie atrybutów, w tradycyjnym modelu związków encji wszystkie atrybuty nadklasy są dziedziczone przez każdą z jej podklas. Zakładając, że F jest nadklasą mającą n podklas: $E_1, E_2, \dots, E_n$, rozmyty zbiór atrybutów powiązanych z wynikową podklasą E_i , tj. $\mathcal{A}_{E_i}$, składa się z rozmytego zbioru atrybutów dziedziczonych z nadklasy F , tj. $\mathcal{A}_F$, i rozmytego zbioru własnych atrybutów $\mathcal{A}_{O_i}$. Stąd

$$\mathcal{A}_{E_i} = \mathcal{A}_F \cup \mathcal{A}_{O_i}$$

Na przykład dla $E_i = \text{Sekretarka}$, $\mathcal{A}_{O_i}$ może mieć postać: $\{(\text{PisanieNaMaszynie}, 1), (\text{JęzykObcy}, 0.9), \dots\}$. W kontekście rozmytym musi być spełniona dodatkowo zależność $\mu_{AF}(A) \leq \mu_{AE_i}(A)$. Mając zdefiniowaną rozmytą specjalizację (generalizację) można określić dokładnie zależności między stopniami przynależności atrybutu nadklasy i atrybutu podklasy do modelu. Przykład rozmytego dziedziczenia przedstawia Chen [CHEN 98] rozważając rozmyte dziedziczenie dla specjalizacji zdefiniowanej na atrybucie, gdy wartością atrybutu, na którym opiera się specjalizacja jest zbiór rozmyty.

Dziedziczenie atrybutów w przypadku podklas dzielonych jest dziedziczeniem wielokrotnym. Dla podklasy dzielonej z wieloma nadklasami zbiór atrybutów każdej nadklasy jest dziedziczony jako atrybuty podklasy. Niech E będzie dzieloną podklasą nadklas $F_1, F_2, \dots, F_n$, wówczas

$$\mathcal{A}_E = \mathcal{A}_{F_1} \cup \mathcal{A}_{F_2} \cup \dots \mathcal{A}_{F_n} \cup \mathcal{A}_{OE}$$

gdzie $\mathcal{A}_E$ jest rozmytym zbiorem atrybutów dzielonej podklasy E , $\mathcal{A}_{F_i}$ jest zbiorem rozmytym atrybutów F_i , $\mathcal{A}_{OE}$ jest zbiorem rozmytym własnych atrybutów F . Inny rodzaj dziedziczenia, nazywany dziedziczeniem selektywnym, jest związany z kategoriami. Kategoria dziedziczy atrybuty swoich nadklas w ten sposób, że każda instancja kategorii selektywnie (wybiórczo) dziedziczy wszystkie atrybuty nadklasy do której należy. Niech E będzie kategorią nadklas $F_1, F_2, \dots, F_n$, wówczas $\forall e \in E$

$$\mathcal{A}_E(e) = (\mathcal{A}_{F_{j_1}}(e) \cup \mathcal{A}_{F_{j_2}}(e) \cup \dots \mathcal{A}_{F_{j_m}}(e) \cup \mathcal{A}_{OE}(e))$$

gdzie $\mathcal{A}_{F_{j_k}}(e)$ jest rozmytym zbiorem atrybutów F_{j_k} takim, że $e \in F_{j_k}$ ($k=1, 2, \dots, m$ i $j_1, j_2, \dots, j_m \in \{1, 2, \dots, n\}$).

Dla rozmytego modelu związków encji można również rozważyć „dziedziczenie” rozmytości przez wyprowadzane atrybuty. Ten rodzaj „dziedziczenia” oznacza, że rozmytość właściwa dla przechowywanych atrybutów jest „dziedziczona”, a raczej przenoszona, przez odpowiadające atrybuty wyprowadzane. Inaczej mówiąc, jeżeli wartości przechowywanych atrybutów są rozmyte, wówczas wartości ich atrybutów wyprowadzanych są również rozmyte. Jeżeli DataUrodzenia jest atrybutem

przechowywanym, to Wiek może być atrybutem wyprowadzanym, rozmyta wartość DatyUrodzenia, na przykład "około 1975", będzie prowadziła do nieprecyzyjnej wartości Wiek, na przykład „około 25”, jeżeli bieżącym rokiem jest 2000. Niech A będzie atrybutem przechowywanym, B będzie atrybutem wyprowadzanym i Γ będzie funkcją z A w B, tj. $V_e(B) = G(V_e(A))$. Jeżeli wartość atrybutu A, $V_e(A)$, jest zbiorem rozmytym

$$V_e(A) = \{\mu_A(a) / a \mid a \in Dom(A) \wedge \mu_A(a) \in [0, 1]\}$$

wówczas wartość atrybutu B, $V_e(B)$ może być otrzymana również jako zbiór rozmyty

$$V_e(B) = \{\mu_B(b) \mid b \in Dom(B) \wedge b = G((a) \wedge \mu_B(b) = \mu_A(a))\}$$

W niektórych przypadkach atrybut wyprowadzany jest otrzymywany na podstawie powiązanych encji. Na przykład wartość atrybutu wyliczanego LiczbaStudentów dla wydziału #W = 18 może być uzyskany przez „policzenie” liczby studentów studiujących na wydziale #W = 18. Jeżeli związek typu Studiuje_na jest rozmyty, wtedy "policzenie" powinno być operacją uwzględniającą stopnie związków. Atrybut LiczbaStudentów można wyliczyć za pomocą operatora Σ_{count} :

$$LiczbaStudentów = \sum_i \mu_{Stadiuje_na}(S\#_i, 18)$$

gdzie $S\#_i, 18$ są identyfikatorami studenta i wydziału #W = 18,

$(S\#_i, 18, \mu_{Związany_z}(S\#_i, 18))$ jest związkiem typu Studiuje_na.

Wartości wyprowadzanych atrybutów zależą zarówno od sposobu, w który są otrzymywane, jak i od rozmytości wprowadzanej do powiązanych encji, związków i atrybutów.

6. PODSUMOWANIE

Aby poprawnie odzwierciedlić rozważaną rzeczywistość w bazie danych, a więc aby poprawnie stworzyć model bazy danych odpowiadający rzeczywistości, należy przejść od opisu rozważanej rzeczywistości do modelu conceptualnego, a następnie dopiero do modelu logicznego. Chcąc reprezentować rozmyte dane występujące w rzeczywistości nie można jednak posłużyć się klasycznymi modelami, na przykład modelem związków encji, ponieważ nie mają one zdolności posługiwania się tego rodzaju danymi. Można tu jednak zastosować rozmyty model związków encji, który pozwala zaprezentować różnego rodzaju rozmyte dane. Na poziomie $L2(\mathcal{U}\mathcal{E}, \mathcal{R})$ można zaprezentować częściową przynależność encji do typu encji, związku do typu związku, poziom $L3(\mathcal{U}\mathcal{A})$ pozwala odzwierciedlić przyjmowanie przez atrybuty wartości rozmytych, w tym rozkładu możliwości. Dodatkowo rozmyty model związków encji pozwala na wprowadzenie i operowanie rozmytością na poziomie

$L1(\mathcal{M})$, co pozwala wyrażać częściową przynależność typu encji, typu związku lub atrybutu do modelu. Wzbogaca to możliwości modelowania rzeczywistości. Rozmytość wprowadzana na poziomie $L1(\mathcal{M})$ odzwierciedla częściową wiedzę projektantów modelu i ekspertów dziedzinowych o granicach modelu. Umożliwia połączenie różnych spojrzeń na rozważaną rzeczywistość. Dodatkowo do rozmytego modelu związków encji wprowadzona została rozmytość reguły uczestnictwa i liczebności. Zdefiniowano pojęcie λ - uczestnictwa oraz liczebności 1:N, n:M, N:M, gdzie N i M reprezentują zbiory rozmyte, przedstawiając również sposób postępowania z ową rozmytą liczebnością. Rozmytość została również uwzględniona w dziedziczeniu.

FUZZY ENTITY-RELATIONSHIP MODEL

Fuzzy entity-relationship model and representations of various forms of fuzzy data in relational model addressing extensions of classical model with fuzzy elements have been presented.

Barbara BIEDRZYCKA
Zygmunt MAZUR*

REPREZENTACJA DANYCH W ROZMYTYCH MODELACH RELACYJNYCH BAZ DANYCH

W pracy pokazano sposób rozszerzenia klasycznego modelu reprezentacji danych o elementy rozmyte. Przedstawiono różne podejścia do reprezentacji rozmytych danych, a także poruszono problem bliskości i redundancji danych rozmytych, wskazując na istniejące rozwiązania.

1. WPROWADZENIE

W klasycznym modelu relacyjnym informacja przechowywana w bazie danych albo o niej samej musi być precyzyjnie zdefiniowana. Jedyną możliwością odzwierciedlenia niepewności jest wartość nieokreślona – null. Każdy fakt świata rzeczywistego reprezentowany przez krotkę $t = (t(A_1), t(A_2), \dots, t(A_i), \dots, t(A_n))$ należącą do R jest faktem zgodnym z semantyką schematu relacji $R(U)$ i całkowicie należącym do R . Nie jest możliwe przedstawienie elementów mających częściowy stopień przynależności do R , mimo iż może być to zgodne z opisywaną rzeczywistością. Na przykład rozważając zamówienia, które są poleceniami nabycia danych towarów, nie można odzwierciedlić częściowej przynależności danego zamówienia do relacji Zamówienia oznaczającej, że kontrahent zamawiający dany asortyment nie jest pewien, czy je zrealizuje. Podobnie nie można odzwierciedlić częściowego powiązania między danymi obiektami (encjami). Na przykład, jeżeli asortyment „Kubek 0.5l” należy do pewnych grup towarów - rodzajów, to nie można odzwierciedlić częściowego stopnia powiązania między tym towarem a daną grupą, na przykład informacji, że jest on kubkiem w stopniu 1.0 i garnkiem w stopniu 0.7.

W klasycznym modelu relacyjnym wartość atrybutu może być tylko pojedynczym elementem dziedziny atrybutu, stąd nie można wprowadzić danych

* Wydziałowy Zakład Informatyki, Wydział Informatyki i Zarządzania Politechniki Wrocławskiej,
50-370 Wrocław, Skwer Idaszewskiego 1, mazur@ci.pwr.wroc.pl

reprezentujących na przykład fakt, że dany kolor jest niebieski albo zielony, tj. {niebieski, zielony}, czy też informacji, że klient wyraża chęć nabycia danego produktu w liczbie 20 (w stopniu 1.0) lub 22 (w stopniu 0.8), tj. {1.0/20, 0.8/22}.

Dodatkowo w klasycznym modelu relacyjnym elementy dziedziny atrybutu nie są wzajemnie ze sobą związane. Na przykład, jeżeli wprowadzony zostałby atrybut Solidność dla relacji Kontrahenci, który określałby, na ile dany kontrahent wywiązuje się ze swoich zobowiązań wobec przedsiębiorstwa, i którego dziedziną byłby zbiór {solidny, bardzo dobry, dobry, średni, niepewny}, to nie można byłoby reprezentować podobieństwa (bliskości) między elementami dziedziny, które przedstawia tabela 1.

Tabela 1. Relacja bliskości $C_{\text{Solidność}}$ na dziedzinie atrybutu Solidność

$C_{\text{Solidność}}$	Solidny	Bardzo dobry	Dobry	Średni	Niepewny
Solidny	1.0	0.8	0.5	0.1	0.0
Bardzo dobry		1.0	0.6	0.3	0.2
Dobry			1.0	0.8	0.7
Średni				1.0	0.85
Niepewny					1.0

W klasycznych bazach danych nie można wprowadzić jako wartości atrybutu rozkładu możliwości, zmiennej lingwistycznej, itp. Informacje rozmyte i nieprecyzyjne są albo wykluczane, albo sztucznie precyzowane. Aby móc reprezentować i operować rozmytymi danymi w modelu relacyjnym, należy wprowadzić do niego rozmytość. Podstawą jest zdefiniowanie reprezentacji danych, które umożliwiłyby wprowadzenie przedstawionych rodzajów informacji.

2. SPOSOBY REPREZENTACJI DANYCH ROZMYTYCH

Od 1980 roku zostało zaproponowanych kilka podejść do reprezentacji rozmytych danych. Podejścia te różnią się między sobą, dając różne rozmyte modele relacyjnych baz danych. Każdy z nich umożliwia reprezentację pewnych rodzajów rozmytych danych, na przykład częściowej przynależności do relacji, możliwość wprowadzania kilku wartości jako wartości atrybutu czy też wprowadzenia relacji podobieństwa lub bliskości między wartościami elementów dziedziny. W bieżącym rozdziale zostaną przedstawione cztery główne podejścia: podejście oparte na relacji rozmytej (ang. *fuzzy-relation-based framework*), podejście oparte na podobieństwie (ang. *similarity-based framework*), podejście oparte na możliwości (ang. *possibility-based framework*) i rozszerzone podejście oparte na możliwości (ang. *extended possibility-based framework*).

2.1. PODEJŚCIE OPARTE NA RELACJI ROZMYTEJ

Podejście oparte na relacji rozmytej [BALDWIN 84] jest modelem najprostszym, akceptującym różne stopnie przynależności krotki do relacji, stopnie te mogą przyjmować wartości z przedziału $[0, 1]$. Rozszerza ono relację, będącą podzbiorem iloczynu kartezjańskiego dziedzin, do relacji rozmytej, która jest rozmytym podzbiorem iloczynu kartezjańskiego dziedzin.

Opierając się na notacji rozmytych relacji, rozmyta relacja R na $D_1 \times D_2 \times \dots \times D_n$ jest charakteryzowana przez funkcję przynależności

$$\mu_R: D_1 \times D_2 \times \dots \times D_n \rightarrow [0, 1],$$

gdzie D_i ($i = 1, 2, \dots, n$) jest dziedziną atrybutu A_i .

Rozmytą relację binarną R na $D_1 \times D_2$ można przedstawić w postaci krotek wzorem:

$$R = \{(u_1, v_1, \mu_R(u_1, v_1)), \dots, (u_m, v_n, \mu_R(u_m, v_n))\}$$

gdzie $u_j \in D_1$, $j = 1, 2, \dots, m$ i $v_k \in D_2$, $k = 1, 2, \dots, n$.

Ten rodzaj reprezentacji zakłada, że przynależność krotki do relacji jest rozmyta, podczas gdy wartości atrybutów nie są rozmyte (lub mogą być zmiennymi lingwistycznymi, ale są traktowane jako atomowe i jednowartościowe).

Rozszerzenie to pozwala na reprezentację częściowej przynależności krotki do relacji lub na przedstawienie częściowego związku między obiektami (encjami). Przykładem pierwszego przypadku może być relacja Zamówienia przedstawiona w tabeli 2, w której stopień przynależności wyraża niepewność, na ile dana krotka przedstawia zamówienie towarów do zakupu, niepewność jego realizacji.

Tabela 2. Relacja Zamówienia

NrZam	NazwaIdKontrahenta	DataZam	DataRealizacji	$\mu_{\text{Zamówienia}}$
1	NOWAK	2002.03.03	2002.03.18	1.0
2	NOWAK	2002.03.03	2002.03.20	0.8
3	ANDREX	2002.03.04	2002.03.24	0.95

Przykładem drugiego przypadku jest relacja Asortyment (tabela 3), w której stopień przynależności reprezentuje moc związku między danym asortymentem i danym rodzajem asortymentu, przedstawia stopień w jakim dany asortyment należy do danej grupy.

Tabela 3. Relacja Jest_typu

KodAsort	NazwaRodzaju	$\mu_{\text{Jest_typu}}$
KU01	KUBEK	1.0
KU01	GARNEK	0.8
SŁ05	SŁOIK	1.0

2.2. PODEJŚCIE OPARTE NA PODOBIENSTWIE

Podjęcie oparte na podobieństwie [BUCKLES 82] rozszerza klasyczny model reprezentacji danych poprzez założenie, że wartości atrybutów mogą być podzbiorami elementów dziedziny. Ponadto dla każdej dziedziny definiuje się relacje podobieństwa wyrażającą podobieństwo między jej elementami.

Relacja R jest podzbiorem iloczynu kartezjańskiego $2^{D_1} \times 2^{D_2} \times \dots \times 2^{D_n}$, gdzie D_i jest skończoną dziedziną i 2^{D_i} jest mocą zbioru D_i . Krotka t relacji R jest postaci:

$$(d_1, d_2, \dots, d_n), \text{ gdzie } d_i \subseteq D_i, \quad d_i \neq \emptyset$$

Dla każdej skalarnej dziedziny definiuje się relacje podobieństwa.

Relacja podobieństwa s_j na D_j jest odwzorowaniem

$$s_j: D_j \times D_j \rightarrow [0, 1]$$

takim, że dla każdego $x, y, z \in D_j$

$$s_j(x, x) = 1 \quad (\text{zwrotna})$$

$$s_j(x, y) = s_j(y, x) \quad (\text{symetryczna})$$

$$s_j(x, z) \geq \max_{y \in D_j} (\min(s_j(x, y), s_j(y, z))) \quad (\text{przechodnia max-min})$$

Relacja podobieństwa s_j z podanym progiem α_j wprowadza relacje równoważności na D_j .

Przykładem relacji w podejściu opartym na podobieństwie jest relacja Kontrahenci przedstawiona w tabeli 4 z relacją podobieństwa wyspecyfikowaną na dziedzinie atrybutu Solidność (tabela 5).

Tabela 4. Relacja Kontrahenci

NazwaIdK	PełnaNazwa	NIP	Adres	Telefon	Solidność
NOWAK	Nowak Tomasz	525-000-01-27	ul. Zielona 2/12 07-412 Ostrołęka	325-86-59	{bardzo dobry}
ANDREX	Hurtownia art. gosp. dom ANRDEX	525-000-76-03	ul. Targowa 2 07-412 Ostrołęka	324-95-08	{bardzo dobry, solidny}
ANIOŁ	Anioł Joanna	898-000-79-07	ul. Złota 52 07-412 Ostrołęka	(029) 766-64-41	{solidny}
KOWAL	Kowalski Jan	897-001-39-12	ul. Kubusia Puchatka 25		{dobry, średni}

Tabela 5. Relacja podobieństwa $S_{Solidność}$ dla dziedziny atrybutu Solidność

$S_{Solidność}$	Solidny	Bardzo dobry	Dobry	Średni	Niepewny
Solidny	1.0	0.9	0.7	0.5	0.1
Bardzo dobry	0.9	1.0	0.8	0.6	0.2
Dobry	0.7	0.8	1.0	0.8	0.4
Średni	0.5	0.6	0.8	1.0	0.6
Niepewny	0.1	0.2	0.4	0.6	1.0

Możliwość użycia podzbiorów dziedzin jako wartości krotek dostarcza podstawowej struktury reprezentacji niedokładnej lub rozmytej informacji. Im większa niedokładność informacji, tym większy podzbiór. Cała dziedzina D_i odpowiada informacji o najmniejszej precyzji, podczas gdy podzbiór zawierający tylko jeden element dziedziny D_i wyraża precyzyjną informację. Należy zauważyć, że zbiór pusty $\emptyset$ nie ma żadnego znaczenia w tym kontekście i nie jest dopuszczalną wartością dla składników krotki. Model ten ponadto uogólnia relację identyczności zachodzącą między elementami dziedziny w klasycznej relacyjnej bazie danych. Relacja identyczności jest odpowiednia dla informacji dokładnej, gdy żaden element dziedziny nie jest podobny do innego elementu dziedziny (w żadnym stopniu), tylko sam do siebie. Relacje podobieństwa nie tylko wprowadzają niedokładną informację w struktury relacyjnej bazy danych, ale również pozwalają na zachowanie kluczowych własności klasycznego relacyjnego modelu baz danych.

Shenoi i Melton [SHENOI 89] rozszerzyli powyższy model poprzez zastąpienie relacji podobieństwa relacją sąsiedztwa (ang. *proximity relations*) zdefiniowaną na skończonej dziedzinie. Relacja sąsiedztwa p_j na D_j jest odwzorowaniem $p_j: D_j \times D_j \rightarrow [0, 1]$ takim, że dla każdego $x, y \in D_j$

$$p_j(x, x) = 1 \quad (\text{zwrotna})$$

$$p_j(x, y) = p_j(y, x) \quad (\text{symetryczna})$$

Własność zwrotności i symetrii są wystarczające do wyrażania stopnia „podobieństwa”, czy też „sąsiedztwa”, między elementami skończonej dziedziny. Można zauważyć, że relacje podobieństwa są szczególnym rodzajem relacji sąsiedztwa. Wprowadzenie relacji sąsiedztwa jest spowodowane tym, że relacja podobieństwa nie pozwala reprezentować pewnych powiązań. Na przykład, rozważając przyjaźń między osobami, jeżeli osoba x jest bliskim przyjacielem osoby y w stopniu a i osoba y jest bliskim przyjacielem osoby z w stopniu b , to nie oznacza to wcale, że osoba x jest bliskim przyjacielem osoby z w stopniu co najmniej $\min(a, b)$. Wszystko zależy od rzeczywistej sytuacji. Innym przykładem jest odległość między

miejscami. Jeżeli lokalizacja x jest bliska lokalizacji y w stopniu a i lokalizacja y jest bliska lokalizacji z w stopniu b , to nie oznacza to wcale, że lokalizacja x jest bliska lokalizacji z w stopniu co najmniej $\min(a, b)$. Zależy to od konkretnych położań. Powiązania te pokazują, że relacja podobieństwa powinna być w wielu przypadkach zamieniona na relację sąsiedztwa, która potrafi odzwierciedlić te zależności.

Usunięcie ograniczenia przechodniości max-min w relacjach sąsiedztwa dostarcza użytkownikom bazy danych więcej swobody w wyrażaniu ich struktur wartości.

2.3. PODEJŚCIE OPARTE NA MOŻLIWOŚCI

Podejście oparte na możliwości [PRADE 84] rozwiązuje problem opisu wartości pustych, częściowej wiedzy oraz niepewnej i nieprecyzyjnej informacji pozwalając aby rozkład możliwości był wartością atrybutu. Relacja w bazie danych jest definiowana jako relacja normalna na iloczynie kartezjańskim zbiorów, których elementami są rozkłady możliwości. Relacja R jest podzbiorem $\Pi(D_1) \times \Pi(D_2) \times \dots \times \Pi(D_n)$, gdzie $\Pi(D_i) = \{\pi_{A_i} \mid \pi_{A_i} \text{ jest rozkładem możliwości } A_i \text{ na } D_i\}$. Krotka t relacji R jest postaci

$$(\pi_{A_1}, \pi_{A_2}, \dots, \pi_{A_n})$$

Ponadto wprowadzono element e , reprezentujący wartość niedotyczy – przypadki, w których atrybut nie stosuje się do kroki. π_{A_i} jest zdefiniowany jako funkcja

$$\pi_{A_i} : D \cup \{e\} \rightarrow [0, 1]$$

Przykładem relacji w podejściu opartym na możliwości jest relacja Pozycje przedstawiona w tabeli 6, reprezentująca pozycje zamawianego asortymentu znajdujące się na zamówieniu. Wartość dla atrybutu Ilość jest rozkładem możliwości pozwalającym na reprezentację nieprecyzyjnej informacji o ilości zamówionego produktu, na przykład rozkład możliwości $\{1.0/20, 0.9/18\}$ wyraża chęć nabycia produktu w liczbie 20 (w stopniu 1.0) lub 18 (w stopniu 0.8).

Tabela 6. Relacja Pozycje

NrZam	Kod	Ilość
1	KU03	$\{1.0/20, 0.9/18\}$
1	KU05	$\{1.0/30\}$
2	KU03	$\{1.0/12, 0.5/10\}$

Dalej można rozszerzyć strukturę reprezentacji do atrybutów wielowartościowych przez zastosowanie rozkładów możliwości na 2^{D_i} . W tym rozszerzeniu krotka jest postaci:

$$(\pi_{DA_1}, \pi_{DA_2}, \dots, \pi_{DA_n})$$

gdzie $\pi_{DA_i} : 2^{D_i} \rightarrow [0, 1]$.

Rozszerzenie to pozwala na przykład na zaprezentowanie dla atrybutu wielowartościowego JęzykObcy, którego dziedziną jest {Angielski, Niemiecki, Francuski, Hiszpański}, wartości: {Angielski}/1.0, {Angielski, Niemiecki}/1.0, {Angielski, Francuski}/1.0, {Angielski, Niemiecki, Hiszpański}/0.3, {Angielski, Niemiecki, Francuski}/1.0, {Angielski, Hiszpański}/0.3. Nie jest to popularne podejście ze względu na to, iż w rozmytych relacyjnych bazach danych nie rozważa się atrybutów wielowartościowych.

2.4. ROZSZERZONE PODEJŚCIE OPARTE NA MOŻLIWOŚCI

Rozszerzone podejście oparte na możliwości [RUNDENSTEINER 89, CHEN 91] rozszerza podejście oparte na możliwości pozwalając nie tylko, aby rozkład możliwości był wartością atrybutu, ale także aby dziedziny mogły być związane z relacjami bliskości. Zarówno podejście oparte na podobieństwie, jak i podejście oparte na możliwości są szczególnymi instancjami rozszerzonego podejścia opartego na możliwości.

Formalnie, relacja R jest podzbiorem $\Pi(D_1) \times \Pi(D_2) \times, \dots, \times \Pi(D_n)$, gdzie $\Pi(D_i) = \{\pi_{A_i} \mid \pi_{A_i} \text{ jest rozkładem możliwości } A_i \text{ na } D_i\}$. Krotka t relacji R jest postaci:

$$(\pi_{A_1}, \pi_{A_2}, \dots, \pi_{A_n})$$

Z każdą dziedziną D_i atrybutu A_i może być związana relacja bliskości reprezentująca powiązanie między elementami dziedziny. Relacja bliskości c_i na dziedzinie D_i jest odwzorowaniem $D_i \times D_i \rightarrow [0,1]$ takim, że dla każdego $x, y \in D_i$ $c_i(x, x) = 1$ (zwrotność) oraz $c_i(x, y) = c_i(y, x)$ (symetryczność). Relacja bliskości c_i z podanym progiem α_j wprowadza relację tolerancji na D_i . Można zauważyć, że relacja bliskości na taką samą notację jak relacja sąsiedztwa.

Wprowadzenie relacji bliskości jest spowodowane tym, że relacja podobieństwa jest w pewnych przypadkach za bardzo restrykcyjna, co zostało przedstawione przy omawianiu relacji sąsiedztwa. Relacja bliskości pozwala użytkownikom bardziej przyjaźnie i intuicyjnie wyrażać wzajemne powiązania pomiędzy elementami dziedziny.

Przykładem relacji w rozszerzonym podejściu opartym na możliwości jest relacja Asortyment przedstawiona w tabeli 7, której wartość dla atrybutu Rozmiar jest rozkładem możliwości na zbiorze {bardzo mały, mały, średni, duży, bardzo duży}.

Tabela 7. Relacja Asortyment

Kod	Nazwa	ProcVat	JMiary	Rozmiar
SL05	„Słoik 0.5l”	21	sztuka	{1.0/średni. 0.8/mały}
SL10	„Słoik 5l”	21	sztuka	{1.0/duży, 0.6/bardzo duży}

Rozkład możliwości pozwala na zaprezentowanie nieprecyzyjnej informacji o rozmiarze asortymentu, przedstawiając jakie wartości mogą być rzeczywistą daną

i w jakim stopniu. Nieprecyzyjność tej danej wynika z faktu, iż określenie rozmiaru jest subiektywne. Dodatkowo między elementami dziedziny istnieje relacja bliskości zdefiniowana w tabeli 8.

Tabela 8. Relacja bliskości C_{Rozmiar} na dziedzinie atrybutu Rozmiar

C_{Rozmiar}	Bardzo mały	Mały	Średni	Duży	Bardzo duży
Bardzo mały	1.0	0.8	0.3	0.1	0.0
Mały	0.8	1.0	0.7	0.2	0.1
Średni	0.3	0.7	1.0	0.7	0.3
Duży	0.1	0.2	0.7	1.0	0.8
Bardzo duży	0.0	0.1	0.3	0.8	1.0

3. BLISKOŚĆ ROZMYTYCH DANYCH

W bazach danych często porównuje się wartości atrybutów. Dokonuje się tego przy selekcji krotek, której warunki są definiowane na atrybutach i żąda się, aby ich wartości równały się danym wartościom. Można powiedzieć, że w przypadku tym dokonuje się sprawdzenia wartości atrybutu dla krotki znajdującej się w bazie z możliwą jego wartością. Porównywania dokonuje się również, gdy mówi się o zależnościach funkcyjnych i należy odpowiedzieć, czy równym wartościom danego zbioru atrybutów odpowiadają równe wartości innego zbioru atrybutów. Sprawdza się wówczas wartości atrybutów dla krotek znajdujących się w bazie. Należy zwrócić uwagę, że zależnością funkcyjną zawsze występującą jest zależność atrybutów od klucza. Porównania dokonuje się również, aby sprawdzić redundancję danych. W klasycznym modelu relacyjnym krotki są redundantne, jeżeli wartości dla atrybutów klucza są równe, stąd problem ten sprowadza się do porównania wartości atrybutów klucza. Dokładnie problem redundancji jest rozwiązywany przez niedopuszczenie wprowadzenia krotki o wartości klucza istniejącej już w bazie. Rozwiązuje to również problem spełnienia zależności funkcyjnej atrybutów od klucza, zależność będzie zawsze spełniona.

W klasycznym modelu danych porównanie wartości jest prostym zagadnieniem, ponieważ wartości są albo równe (identyczne), albo różne. Gdy rozmytość jest wprowadzana do modelu opierając się o zbiory rozmyte i teorię możliwości, dwie wartości atrybutów (przy czym rozważa tu się wszystkie możliwe wartości dla atrybutu) nie mogą być uważane za równe lub różne, ale tylko jako podobne, bliskie sobie w pewnym stopniu. Wynika to z kilku przyczyn. Po pierwsze wartość atrybutu może być podzbiorem dziedziny, na przykład $\text{Wzrost} = \{170, 175, 180\}$, rozkładem możliwości, na przykład $\text{Wzrost} = \{0.3/150, 1/175, 0.6/200\}$. Dodatkowo problemy pojawiają się, gdy zostają zdefiniowane relacje rozmyte na dziedzinach atrybutów,

których przykładem jest relacja podobieństwa i bliskości. Najczęściej problem bliskości danych jest rozważany w kontekście redundancji danych.

W rozdziale bieżącym i następnym zostanie przedstawione rozwiązanie dla bliskości danych i redundancji zaprezentowane przez Bucklesa i Petry'ego [BUCKLES 82] w ramach podejścia opartego na podobieństwie, Shenoi i Meltona [SHENOI 89] dla rozszerzonego podejścia opartego na podobieństwie, w którym relacja podobieństwa została zastąpiona relacją sąsiedztwa, przez Kerre i Van Schootena [CHEN 98] w ramach podejścia do reprezentacji danych opartego na możliwości, Chena, Vandenbulcke'a i Kerre [CHEN 92] dla rozszerzonego podejścia opartego na możliwości, w ramach którego rozwiązanie dla podejścia opartego na możliwości zostało rozszerzone w sposób taki, że relacje bliskości zdefiniowane na dziedzinach są włączane przy obliczaniu równości elementów dziedziny. Istnieją również dla pewnych podejść inne rozwiązania. Na przykład odmienne rozwiązanie problemu bliskości danych i redundancji dla podejścia opartego na możliwości jest przedstawione w pracy [RUNDENSTEINER 89].

1) Bliskość danych w ramach podejścia opartego na podobieństwie

Mając atrybut A zdefiniowany na dziedzinie D , dla której istnieje relacja podobieństwa s , każda jego wartość d należy wówczas do zbioru 2^D , tj. $d \in 2^D$, inaczej $d \subseteq D$. Dwie jego wartości d i d' są podobne do siebie w stopniu α , gdzie $\alpha \in [0,1]$, jeżeli

$$\min_{x,y \in d \cup d'} (s(x,y)) \geq \alpha$$

Relacja podobieństwa s powoduje podział na D , α -cut relacji podobieństwa zdefiniowanej na D jest relacją równoważności na D . Dla każdego α elementy dziedziny D zostają podzielone na klasy, z których każde dwie nie posiadają wspólnej części i wszystkie pokrywają D . Elementy należące do tej samej klasy są uważane za podobne do siebie w stopniu α , podczas gdy elementy należące do różnych klas nie są podobne do siebie w stopniu α . Oznacza to, że wartość atrybutu d jest podobna do wartości atrybutu d' , jeżeli elementy $d \cup d'$ należą do tej samej klasy.

2) Bliskość danych w ramach podejścia opartego na podobieństwie z relacją sąsiedztwa

Aby zdefiniować bliskość danych dla podejścia opartego na podobieństwie z relacją sąsiedztwa, należy najpierw wprowadzić pojęcie α -sąsiedztwa. Pojęcie to zostało wprowadzone, aby zachować pożądany podział dziedziny.

Jeżeli p jest relacją sąsiedztwa na D , wówczas mając α , $0 \leq \alpha \leq 1$, dwa elementy zbioru D , x , z , tj. $x, z \in D$, są nazywane α -sąsiadami, co oznaczane jest przez $xP_\alpha z$, jeżeli $p(x, z) \geq \alpha$ lub istnieje skończony ciąg $y_1, y_2, \dots, y_r \in D$ taki, że

$$p(x, y_1) \geq \alpha, p(y_1, y_2) \geq \alpha, \dots, p(y_{r-1}, y_r) \geq \alpha, p(y_r, z) \geq \alpha$$

P_α zdefiniowana na relacji sąsiedztwa p jest używana do podziału podanej dziedziny D , ponieważ P_α , będąc zwrotną, symetryczną i przechodnią, jest relacją równoważności. α -sąsiedztwo powoduje podział dziedziny D przez narzucenie sąsiedztwa pomiędzy x i z tak długo, jak $p(x, y) \geq \alpha$ i $(y, z) \geq \alpha$, nie zważając na fakt, że $p(x, z)$ może być mniejsze niż $\min(p(x, y), p(y, z))$. Mając zdefiniowane α -sąsiedztwo można określić podobieństwo danych dla rozważanego podejścia.

Mając atrybut A zdefiniowany na dziedzinie D , dla której istnieje relacja sąsiedztwa p , każda jego wartość d należy wówczas do zbioru 2^D , tj.

$$d \in 2^D, \text{ inaczej } d \subseteq D.$$

Dwie jego wartości d i d' są α -podobne, gdzie $\alpha \in [0, 1]$, jeżeli

$$\text{dla każdego } x, y \in d \cup d', x \text{ i } y \text{ są } \alpha\text{-sąsiadami } (xP_\alpha y).$$

Można powiedzieć, że dwie wartości są podobne do siebie (bliskie), jeżeli elementy $d \cup d'$ należą do tej samej klasy równoważności zdefiniowanej przez P_α .

Można zauważyć, że jeżeli zbiór elementów znajdujących się w bazie danych, oznaczany przez D^{db} , jest tylko właściwym podzbiorem zbioru D , co jest prawdą w większości przypadków, to w przeciwieństwie do rozwiązania w ramach podejścia opartego na podobieństwie, podział wprowadzony przez P_α na podzbiorze D^{db} może różnić się od podziału wprowadzonego przez P_α na całej dziedzinie D . Powodem jest fakt, że

dla $x, y, z \in D$, jeżeli $x, z \in D^{db}$, $y \notin D^{db}$ i $p(x, z) < \min(p(x, y), p(y, z))$, to mając poziom $\mu = \min(p(x, y), p(y, z))$, x, y, z należy do tej samej klasy równoważności (w stopniu $\geq \mu$) dla podziału na D , natomiast x, z należy do różnych klas (w stopniu $\geq \mu$) dla podziału na D^{db} . Ogólnie można powiedzieć, że liczba klas równoważności podziału na D^{db} jest większa niż liczba klas równoważności podziału na D . Fakt, że P_α wprowadza dwa różne podziały na D i D^{db} jest związane z definicją P_α , opierającą się na powiązanych elementach (np. y).

3) Bliskość danych w ramach podejścia opartego na możliwości

Mając atrybut A zdefiniowany na dziedzinie D , każda jego wartość jest rozkładem możliwości π zdefiniowanym na D , tzn. $\pi \in \Pi(D)$ (zakłada się, że π jest znormalizowany).

Bliskość, a raczej równość, między dwoma jego wartościami (wartościami będącymi rozkładami możliwości zdefiniowanymi na D) jest definiowana przez rozkład możliwości E na $\{P, F\}$, gdzie P oznacza Prawdą i F oznacza Fałsz.

Wartości $E(\pi, \pi')(P)$ i $E(\pi, \pi')(F)$ są wyznaczone zgodnie z wzorami:

$$E(\pi, \pi')(P) = \sup_{x \in D} \min(\pi(x), \pi'(x))$$

$$E(\pi, \pi')(F) = \sup_{\substack{x \in D \\ x \neq y}} \min(\pi(x), \pi'(x))$$

Można powiedzieć, że $E(\pi, \pi')(P)$ wyraża możliwość, że $\pi = \pi'$ jest prawdziwe, a $E(\pi, \pi')(F)$ jest możliwością, że $\pi = \pi'$ jest fałszywe.

Dla przykładu można rozważyć dwa rozkłady możliwości

$$\pi = \{1/a, 0.2/b\} \text{ i } \pi' = \{1/c, 0.3/d\}$$

zdefiniowane na dziedzinie $D = \{a, b, c, d\}$.

Możliwość, że $\pi = \pi'$ jest prawdziwe, równa się

$$\begin{aligned} E(\pi, \pi') &= \sup_{x \in D} \min(\pi_1(x), \pi_2(x)) = \\ &= \max(\min(1, 0), \min(0.2, 0), \min(0, 1), \min(0, 0.3)) = 0 \end{aligned}$$

a możliwość, że $\pi_1 = \pi_2$ jest fałszywe, równa się

$$\begin{aligned} E(\pi, \pi') &= \sup_{x \neq y} \min(\pi_1(x), \pi_2(y)) = \\ &= \max(\min(1, 1), \min(1, 0.3), \min(0.2, 1), \min(0.2, 0.3)) = 1 \end{aligned}$$

Zatem uważa się, że niemożliwe jest, aby π i π' były równe.

Jeżeli pożądana jest skalarna wartość prawdy dla $\pi = \pi'$, wówczas najczęściej stosuje się wartość $E(\pi, \pi')(P)$.

4) Bliskość danych w ramach rozszerzonego podejścia opartego na możliwości

Rozważając przykład rozkładów możliwości

$$\pi = \{1/a, 0.2/b\}$$

i

$$\pi' = \{1/c, 0.3/d\}$$

zdefiniowanych na dziedzinie $D = \{a, b, c, d\}$, można zauważyć, że jeżeli istniałaby relacja bliskości między elementami stwierdzająca wysoki stopień bliskości (podobieństwa) między elementami dziedziny D , na przykład, jeżeli a byłaby bliska w wysokim stopniu do c i podobnie b do d , wówczas intuicyjnie oczekiwałoby się wysokiego stopnia możliwości, że $\pi = \pi'$ jest prawdziwe.

Rozwiązanie problemu bliskości danych w ramach rozszerzonego podejścia opartego na możliwości jest rozszerzeniem rozwiązania dla podejścia opartego na możliwości, wprowadzającym bliskość pomiędzy elementami.

Mając atrybut A zdefiniowany na dziedzinie D , każda jego wartość jest rozkładem możliwości π zdefiniowanym na D , π należy wówczas do zbioru $\Pi(D)$, tj. $\pi \in \Pi(D)$, dodatkowo zdefiniowana jest na D relacja bliskości C . Bliskość (równość) między dwoma jego wartościami (wartościami będącymi rozkładami możliwości

zdefiniowanymi na D , dla której istnieje relacja bliskości) jest definiowana przez rozkład możliwości E na $\{P, F\}$, gdzie P oznacza Prawdę i F oznacza Fałsz.

Wartości $E(\pi, \pi')(P)$ i $E(\pi, \pi')(F)$ są wyznaczane zgodnie z wzorami:

$$E_c(\pi, \pi')(P) = \sup_{\substack{x, y \in D \\ c(x, y) \geq \alpha}} \min(\pi(x), \pi'(y))$$

$$E_c(\pi, \pi')(F) = \sup_{\substack{x, y \in D \\ c(x, y) < \alpha}} \min(\pi(x), \pi'(y))$$

Różnica między E_c i E , polega na tym, że E niejawnie zakłada relacje identyczności τ na D , można zatem E zdefiniować wzorami:

$$E(\pi, \pi')(P) = \sup_{\substack{x \in D \\ \tau(x, x)=1}} \min(\pi(x), \pi'(x))$$

$$E(\pi, \pi')(F) = \sup_{\substack{x \in D \\ \tau(x, x)=0}} \min(\pi(x), \pi'(x))$$

E_c uogólnia założenie dotyczące τ akceptując relacje bliskości C na D .

Rozważając ponownie $\pi = \{1/a, 0.2/b\}$ i $\pi' = \{1/c, 0.3/d\}$ zdefiniowane na dziedzinie $D = \{a, b, c, d\}$, można wprowadzić relacje bliskości C na D przedstawioną w tabeli 9.

Tabela 9. Relacja bliskości C

C	a	b	c	d
a	1.0	0.2	0.9	0.1
b		1.0	0.2	0.95
c			1.0	0.1
d				1.0

Rozkład możliwości równości $\pi_1 = \pi_2$, stosując E_c z $\alpha_i = 0.9$, równa się

$$E_c(\pi_1, \pi_2)(F) = \sup_{c_i(x, y) < \alpha_i} \min(\pi_1(x), \pi_2(y)) = \max(\min(1, 0.3), \min(0.2, 1)) = 0.3$$

$$E_c(\pi_1, \pi_2)(P) = \sup_{c_i(x, y) \geq \alpha_i} \min(\pi_1(x), \pi_2(y)) = \max(\min(1, 1), \min(0.2, 0.3)) = 1$$

Przy założeniu bliskości dla elementów dziedziny wynik uzyskany stosując E_c jest bardziej intuicyjny niż wynik uzyskany opierając się na E postaci

$$E(\pi, \pi') = \{0/P \text{ i } 1/F\}.$$

Jeżeli pożądana jest skalarna wartość prawdy dla $\pi = \pi'$, wówczas najczęściej stosuje się wartość $E_c(\pi, \pi')(P)$.

4. REDUNDANCJA DANYCH

Redundancja danych jest jednym z podstawowych zagadnień klasycznych systemów baz danych. Po pierwsze redundancja oznacza duplikację krotek. Nie dozwolone jest występowanie duplikatów krotek w danej tabeli. Nie może istnieć para krotek t i t' w R taka, że wszystkie wartości atrybutów t i t' są identyczne, tzn. dla $t = (t(A_1), t(A_2), \dots, t(A_n))$, $t' = (t'(A_1), t'(A_2), \dots, t'(A_n))$ t i t' są redundantne wtedy i tylko wtedy, gdy $t(A_i) = t'(A_i)$, $i = 1, 2, \dots, n$. Alternatywnie, redundancja krotek może być również definiowana opierając się na zależnościach funkcyjnych i związanych z nimi kluczach głównych. Ponieważ klucz główny K relacji R typu $R(U)$ jest podzbiorem U , tj. $K \subseteq U$, takim, że równe wartości krotek dla atrybutów klucza implikują równe wartości krotek dla każdego $A_i \in U$, dwie krotki t i t' są redundantne, jeżeli ich wartości dla atrybutów klucza są identyczne, tj. $t(K) = t'(K)$. Dodatkowo redundancja danych oznacza duplikację wartości w odniesieniu do pewnych atrybutów z U . Ten rodzaj redundancji danych jest związany z pojęciem zależności funkcyjnych, dekompozycją relacji, o czym mowa będzie w rozdziale 6.

Redundancja dla modeli rozmytych nie jest już tak prostym zagadnieniem, ponieważ nie mówi się w nich o identyczności, ale bliskości wartości atrybutów. Redundancja danych dla rozmytych modeli relacyjnych jest rozszerzeniem definicji dla klasycznego modelu uwzględniającym fakt, że między wartościami atrybutów nie zachodzi relacja identyczności, ale bliskości. Opiera się ona na wprowadzonych w rozdziale 3 definicjach bliskości danych.

1) Redundancja w ramach podejścia opartego na podobieństwie [BUCKLES 82]

Mając dziedziny $D_1, D_2, \dots, D_n$ dla których zdefiniowane są relacje podobieństwa $s_1, s_2, \dots, s_n$ i mając relacje $R \subseteq 2^{D_1} \times 2^{D_2} \times \dots \times 2^{D_n}$, dwie krotki $t = (d_1, \dots, d_i, \dots, d_n)$, $t' = (d_1', \dots, d_i', \dots, d_n')$, gdzie $d_i, d_i' \subseteq D_i$, $d_i \neq \emptyset$, $d_i' \neq \emptyset$ są α -redundantne wobec siebie, gdzie $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in [0, 1]^n$, jeżeli dla $i = 1, \dots, n$ d_i i d_i' są do siebie podobne w stopniu α_i , tzn. jeżeli

$$\min_{x, y \in d_i \cup d_i'} (s_i(x, y)) \geq \alpha_i$$

Zbiór relacji podobieństwa $\{s_1, s_2, \dots, s_n\}$ z $\alpha_1, \alpha_2, \dots, \alpha_n$ powoduje podział na relacji R .

Powtarzające się krotki mogą zostać wyeliminowane przez połączenie wartości atrybutów przez operację sumy zbiorów. Jeżeli krotka t' jest redundantna wobec krotki t , wówczas nowa krotka t'' jest uzyskiwana przez połączenie t i t' : $t'' = (d''_1, \dots, d''_i, \dots, d''_n)$, gdzie $d''_i = d_i \cup d_i'$, dla $i = 1, 2, \dots, n$.

2) Redundancja w ramach podejścia opartego na podobieństwie z relacją sąsiedztwa [SHENOI 89]

Mając dziedziny $D_1, D_2, \dots, D_n$ dla których zdefiniowane są relacje sąsiedztwa $p_1, p_2, \dots, p_n$ i mając relacje $R \subseteq 2^{D_1} \times 2^{D_2} \times \dots \times 2^{D_n}$, dwie krotki

$$t = (d_1, \dots, d_i, \dots, d_n), t' = (d_1', \dots, d_i', \dots, d_n'),$$

gdzie $d_i, d_i' \subseteq D_i, d_i \neq \emptyset, d_i' \neq \emptyset$ są α -redundantne wobec siebie oraz $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in [0, 1]^n$, jeżeli dla każdego $i = 1, \dots, n$, dla każdych $x, y \in d_i \cup d_i'$ x i y są α_i -sąsiadami. Inaczej mówiąc, dwie krotki są redundantne, jeżeli dla wszystkich wartości atrybutów, elementy $d_i \cup d_i'$ należą do tej samej klasy równoważności zdefiniowanej przez α_i -sąsiedztwo (przez relacje P_{α_i}).

3) Redundancja w ramach podejścia opartego na możliwości [CHEN 98]

Mając dziedziny $D_1, D_2, \dots, D_n$, relacje $R \subseteq \Pi(D_1) \times \Pi(D_2) \times \dots \times \Pi(D_n)$ i dwie krotki $t = (\pi_{A_1}, \pi_{A_2}, \dots, \pi_{A_n}), t' = (\pi'_{A_1}, \pi'_{A_2}, \dots, \pi'_{A_n})$, gdzie π_{A_i}, π'_{A_i} są rozkładami możliwości na D_i , tj. $\pi_{A_i}, \pi'_{A_i} \in \Pi(D_i)$, równość między t i t' jest definiowana przez rozkład możliwości G na $\{P, F\}$, gdzie P oznacza Prawdę i F oznacza Fałsz. Można zapisać go wzorem:

$$\begin{aligned} G(t, t') &= Poss((\pi_{A_1} = \pi'_{A_1}) \wedge (\pi_{A_2} = \pi'_{A_2}) \wedge \dots \wedge (\pi_{A_n} = \pi'_{A_n})) \\ &= E(\pi_{A_1}, \pi'_{A_1}) \wedge E(\pi_{A_2}, \pi'_{A_2}) \wedge \dots \wedge E(\pi_{A_n}, \pi'_{A_n}) \\ &= \{g^1 / P, g^2 / F\} \end{aligned}$$

gdzie $0 \leq g^1 \leq 1, 0 \leq g^2 \leq 1$, $\wedge$ jest odwzorowaniem $\Pi(\{P, F\}) \times \Pi(\{P, F\})$ w $\Pi(\{P, F\})$, E jest miarą bliskości (równości) dwóch rozkładów możliwości, tzn. jest odwzorowaniem z $\Pi(D_i) \times \Pi(D_i)$ w $\Pi(\{P, F\})$, którego wartości $E(\pi_{A_i}, \pi'_{A_i})(P)$ i $E(\pi_{A_i}, \pi'_{A_i})(F)$ są definiowane następująco:

$$E(\pi_{A_i}, \pi'_{A_i})(P) = \sup_{x \in D_i} \min(\pi_{A_i}(x), \pi'_{A_i}(x)) \quad i = 1, 2, \dots, n$$

$$E(\pi_{A_i}, \pi'_{A_i})(F) = \sup_{\substack{x \in D_i \\ x \neq y}} \min(\pi_{A_i}(x), \pi'_{A_i}(x)) \quad i = 1, 2, \dots, n$$

Stosując zasadę rozszerzania dla $\wedge$, uzyskuje się:

$$\begin{aligned} g^1 &= [E(\pi_{A_1}, \pi'_{A_1}) \wedge E(\pi_{A_2}, \pi'_{A_2}) \wedge \dots \wedge E(\pi_{A_n}, \pi'_{A_n})](P) \\ &= \sup_{e_1 \wedge e_2 \wedge \dots \wedge e_n = P} \min(E(\pi_{A_1}, \pi'_{A_1})(e_1), E(\pi_{A_2}, \pi'_{A_2})(e_2), \dots, E(\pi_{A_n}, \pi'_{A_n})(e_n)) \\ &= \min(E(\pi_{A_1}, \pi'_{A_1})(T), E(\pi_{A_2}, \pi'_{A_2})(T), \dots, E(\pi_{A_n}, \pi'_{A_n})(T)) \end{aligned}$$

$$\begin{aligned} g^2 &= [E(\pi_{A_1}, \pi'_{A_1}) \wedge E(\pi_{A_2}, \pi'_{A_2}) \wedge \dots \wedge E(\pi_{A_n}, \pi'_{A_n})](F) \\ &= \sup_{e_1 \wedge e_2 \wedge \dots \wedge e_n = F} \min(E(\pi_{A_1}, \pi'_{A_1})(e_1), E(\pi_{A_2}, \pi'_{A_2})(e_2), \dots, E(\pi_{A_n}, \pi'_{A_n})(e_n)) \end{aligned}$$

g^1 jest możliwością, że $t = t'$ jest prawdziwe, g^2 jest możliwością, że $t = t'$ jest fałszywe. Jeżeli pożądana jest skalarna wartość prawdy g dla $t = t'$, to może ona zostać uzyskana na podstawie g^1 i g^2 . g może mieć postać

$$g = \frac{G(T) + 1 - G(F)}{2} = \frac{g^1 + 1 - g^2}{2}$$

Mając próg $\lambda \in [0,1]$, t i t' są równe jeżeli $g \geq \lambda$.

4) Redundancja w ramach rozszerzonego podejścia opartego na możliwości [CHEN 92]

Definicja redundancji w ramach rozszerzonego podejścia opartego na możliwości jest rozszerzeniem równości krotek dla podejścia opartego na możliwości, uwzględniając fakt, że na dziedzinie zdefiniowana jest relacja bliskości, zamiast relacji identyczności, a więc gdy dopuszcza się podobieństwo między elementami dziedziny.

Niech E_c będzie miarą bliskości (równości) danych dla rozszerzonego podejścia opartego na możliwości, tzn. będzie odwzorowaniem $\Pi(D_i) \times \Pi(D_i)$ w $\Pi(\{P, F\})$, $\wedge$ będzie rozszerzoną koniunkcją, $c_1, c_2, \dots, c_n$ będą relacjami bliskości zdefiniowanymi na dziedzinach $D_1, D_2, \dots, D_n$, $\alpha_1, \alpha_2, \dots, \alpha_n$ będą progami odpowiadającymi relacjom bliskości i niech t i t' będą dwoma krotkami relacji $R \subseteq \Pi(D_1) \times \Pi(D_2) \times \dots \times \Pi(D_n)$, tj. $t = (\pi_{A1}, \pi_{A2}, \dots, \pi_{An})$, $t' = (\pi'_{A1}, \pi'_{A2}, \dots, \pi'_{An})$, gdzie π_{Ai} , π'_{Ai} są rozkładami możliwości na D_i , tj. $\pi_{Ai}, \pi'_{Ai} \in \Pi(D_i)$. Równość między t i t' jest definiowana przez rozkład możliwości G_c na $\{P, F\}$, gdzie P oznacza Prawdą i F oznacza Fałsz. Można zapisać go wzorem:

$$G_c(t, t') = \{g_c^1 / P, g_c^2 / F\}$$

gdzie:

$$\begin{aligned} g_c^1 &= [E_c(\pi_{A1}, \pi'_{A1}) \wedge E_c(\pi_{A2}, \pi'_{A2}) \wedge \dots \wedge E_c(\pi_{An}, \pi'_{An})](P) \\ &= \sup_{e1 \wedge e2 \wedge \dots \wedge en = P} \min(E_c(\pi_{A1}, \pi'_{A1})(e_1), E_c(\pi_{A2}, \pi'_{A2})(e_2), \dots, E_c(\pi_{An}, \pi'_{An})(e_n)) \\ &= \min(E_c(\pi_{A1}, \pi'_{A1})(T), E_c(\pi_{A2}, \pi'_{A2})(T), \dots, E_c(\pi_{An}, \pi'_{An})(T)) \end{aligned}$$

$$\begin{aligned} g_c^2 &= [E_c(\pi_{A1}, \pi'_{A1}) \wedge E_c(\pi_{A2}, \pi'_{A2}) \wedge \dots \wedge E_c(\pi_{An}, \pi'_{An})](F) \\ &= \sup_{e1 \wedge e2 \wedge \dots \wedge en = F} \min(E_c(\pi_{A1}, \pi'_{A1})(e_1), E_c(\pi_{A2}, \pi'_{A2})(e_2), \dots, E_c(\pi_{An}, \pi'_{An})(e_n)) \end{aligned}$$

i zgodnie z definicją E_c przedstawioną w rozdziale 3 (punkt 4), wartości $E(\pi_{Ai}, \pi'_{Ai})(P)$ i $E(\pi_{Ai}, \pi'_{Ai})(F)$ są definiowane następująco:

$$E_c(\pi_{Ai}, \pi'_{Ai})(P) = \sup_{\substack{x, y \in D_i \\ c_i(x, y) \geq \alpha_i}} \min(\pi_{Ai}(x), \pi'_{Ai}(y)) \quad i = 1, 2, \dots, n$$

$$E_c(\pi_{Ai}, \pi'_{Ai})(F) = \sup_{\substack{x, y \in D_i \\ c_i(x, y) < \alpha_i}} \min(\pi_{Ai}(x), \pi'_{Ai}(x)) \quad i = 1, 2, \dots, n$$

W wielu przypadkach w ocenianiu równości t i t' preferowana jest jedna wartość skalarna prawdziwości tego zdania. Owa wartość skalarna prawdy g_c może być zdefiniowana jako średnia ważona g_c^1 i g_c^2 , tj.

$$g_c = w_1 \cdot g_c^1 + w_2 \cdot (1 - g_c^2)$$

gdzie $w_1 + w_2 = 1$, $w_1, w_2 \geq 0$. Krańcowym, ale racjonalnym wyborem jest $g_c = g_c^1$ (tj. $w_1 = 1$), ponieważ, zgodnie z definicją, g_c^1 jest możliwością, że $t = t'$ jest prawdziwe.

Podsumowując, a zarazem poprzedzając problematykę rozmytych zależności funkcyjnych i związanych z nim rozmytych kluczy, można powiedzieć, że krotki t i t' nie są redundantne (czyli są 0-redundantne), jeżeli wartości dla atrybutów klucza nie są do siebie podobne (są podobne w stopniu 0) lub możliwość, że są równe równa się 0 (odpowiednio dla danego rodzaju reprezentacji rozmytych danych).

5. PODSUMOWANIE

Rozmytość prawie nigdy nie może być dokładnie opisana. Wprowadzenie czterech różnych podejść do reprezentacji rozmytych danych w modelu relacyjnym umożliwia przechowywanie i operowanie różnego rodzaju rozmytymi danymi. Podejście oparte na relacji rozmytej pozwala na częściową przynależność krotki do relacji. W podejściu opartym na podobieństwie wartość atrybutu może być podzbiorem elementów dziedziny, ponadto dla każdej dziedziny definiuje się relacje podobieństwa między jej elementami. Podejście oparte ma możliwości pozwala, aby rozkład możliwości był wartością atrybutu. W rozszerzonym podejściu opartym na możliwości poza tym, że wartością atrybutu jest rozkład możliwości, definiuje się relacje bliskości dla dziedzin atrybutów.

Praca ukazuje, że zastosowanie logiki rozmytej w bazach danych, w procesie ich projektowania pozwala bardziej naturalnie reprezentować otaczającą rzeczywistość. Potrzeba stosowania logiki rozmytej rośnie wraz z kompleksowością baz danych. Z reguły bowiem rozbudowane (duże) bazy danych zawierają informację dotyczące złożonych obiektów i procesów. Tymczasem tylko proste systemy dają się badać

w sposób jednoznaczny za pomocą logiki dwuwartościowej. Rozmyte bazy danych mają ogromne znaczenie w systemach dotyczących podejmowania decyzji i rozwiązywania problemów.

REPRESENTATION OF DATA IN FUZZY DATABASE RELATIONAL MODEL

Representation of various forms of fuzzy data in relational model addressing extension of classical model with elements has been presented.

*dane rozmyte,
język algebry relacji,
język SQLf*

Barbara BIEDRZYCKA
Zygmunt MAZUR*

OPEROWANIE DANymi ROZMYTYMI

W pracy uwagę skoncentrowano na operowaniu danymi w rozmytych modelach relacyjnych baz danych. Została przedstawiona rozszerzona algebra relacji dla każdego z podejść do reprezentacji rozmytych danych oraz próby rozszerzenia języka SQL, których wynikiem jest język SQLf.

1. WPROWADZENIE

Operowanie danymi jest związane z zapytaniami do bazy danych i utrzymaniem w niej informacji. Zapytania są życzeniem uzyskania interesujących danych na podstawie zadanego kryterium. Jeżeli zapytanie jest formułowane z wyrażeniami rozmytymi, na przykład wyrażeniami lingwistycznymi, warunkami rozmytymi, to mówi się, że jest ono zapytaniem rozmytym niezależnie od „natury” danych w bazie danych. Precyzyjne dane są szczególnym przypadkiem danych rozmytych. Realizacja rozmytych zapytań jest ogólnie wykonywana stosując teorię zbiorów rozmytych i teorię możliwości.

Badania nad rozmytymi zapytaniami wyłoniły dwa kierunki prac. Jeden miał na celu rozszerzenie algebry relacji, aby ułatwić operowanie elementami rozmytymi, m.in. wartościami nieprecyzyjnymi, wyrażeniami lingwistycznymi, zbiorami rozmytymi [BUCKLES 82, PARDE 84, YAGER 91, BOSC 91, CHEN 93]. W klasycznej bazie danych algebra relacji jest fundamentem języka zapytań i jest on nazywany zupełnym, jeżeli umożliwia realizację wszystkich operacji, które mogą być wykonywane przez algebrę relacji (np. $\cup$, $-$, $\cap$, $\times$, Π , σ , $*$). Wprowadzenie rozmytych danych ujawniło konieczność utworzenia rozmytego rozszerzenia algebry relacji [CHEN 93]. Celem drugiego kierunku badań było zbudowanie interfejsu lub języka zapytań ułatwiającego zapytania z rozmytymi wyrażeniami [BUCKLES 82, ANVARI 84, ZEMANKOVA 84, KACPRZYK 86, KERRE 86, BOSC 91, YAGER 88, KACPRZYK 94, CHEN 97]. W ramach tych prac został wprowadzony przez Bosc'a i Piverta [BOSC 91] język zapytań SQLf, Kacprzyk i Zadrozny

* Wydziałowy Zakład Informatyki, Wydział Informatyki i Zarządzania Politechniki Wrocławskiej,
50-370 Wrocław, Skwer Idaszewskiego 1, mazur@ci.pwr.wroc.pl

[KACPRZYK 94] utworzyli FQUERY dla Access'a. W językach tych użytkownicy mogą wprowadzać zapytania formułowane w sposób podobny jak w języku SQL, przy czym po klauzuli WHERE można wprowadzać warunki rozmyte – warunki zawierające m.in. atomowe predykaty rozmyte, na przykład młody, dobrze-płatny, modyfikatory lingwistyczne, których przykładem może być wyrażenie bardzo, względnie.

2. OPERACJE NA DANYCH ROZMYTYCH

Podstawowymi operatorami algebry relacji są: suma, różnica, iloczyn, iloczyn kartezyjański, projekcja, selekcja, złączenie naturalne.

2.1. OPERACJE W RAMACH PODEJŚCIA OPARTEGO NA RELACJI ROZMYTEJ

Podstawowe operacje na relacjach rozmytych są definiowane następująco [DRIANKOV 96, CHEN 98]: niech R_1 i R_2 będą dwoma relacjami rozmytymi na X i Y z odpowiednimi funkcjami przynależności μ_{R_1} i μ_{R_2} :

- 1) suma: $R_1 \cup R_2$ jest relacją rozmytą na X, Y z funkcją przynależności $\mu_{R_1 \cup R_2}: X \times Y \rightarrow [0, 1]$ taką, że $\forall (x, y) \in X \times Y$

$$\mu_{R_1 \cup R_2}(x, y) = \max(\mu_{R_1}(x, y), \mu_{R_2}(x, y))$$

- 2) różnica: $R_1 - R_2$ jest relacją rozmytą na X, Y z funkcją przynależności $\mu_{R_1 - R_2}: X \times Y \rightarrow [0, 1]$ taką, że $\forall (x, y) \in X \times Y$

$$\mu_{R_1 - R_2}(x, y) = \min(\mu_{R_1}(x, y), 1 - \mu_{R_2}(x, y))$$

- 3) iloczyn: $R_1 \cap R_2$ jest relacją rozmytą na X, Y z funkcją przynależności $\mu_{R_1 \cap R_2}: X \times Y \rightarrow [0, 1]$ taką, że $\forall (x, y) \in X \times Y$

$$\mu_{R_1 \cap R_2}(x, y) = \min(\mu_{R_1}(x, y), \mu_{R_2}(x, y))$$

- 4) projekcja: Niech R będzie rozmytą relacją na $X_1, X_2, \dots, X_n$ z funkcją przynależności $\mu_R: X_1 \times X_2 \times \dots \times X_n \rightarrow [0, 1]$.

Projekcja R , oznaczana przez $\Pi_{X_{i1} \dots X_{ik}}(R)$, gdzie $i_1, \dots, i_k \in \{1, 2, \dots, n\}$, jest relacją rozmytą na $X_{i1}, \dots, X_{ik}$ z funkcją przynależności

$$\mu_{\Pi_{X_{i1} \dots X_{ik}}(R)}: X_{i1} \times \dots \times X_{ik} \rightarrow [0, 1] \text{ taką, że } \forall (x_{i1}, \dots, x_{ik}) \in X_{i1} \times \dots \times X_{ik}$$

$$\mu_{\Pi_{X_{i1} \dots X_{ik}}(R)}(x_{i1}, \dots, x_{ik}) = \sup_{\substack{u_{ij} = x_{ij} \\ i=1, \dots, k}} \mu_R(u_1, u_1, \dots, u_n)$$

- 5) złączenie naturalne: Niech R_1 będzie relacją rozmytą na X, Y z funkcją przynależności $\mu_{R_1}: X \times Y \rightarrow [0,1]$ i R_2 będzie relacją rozmytą na Y, Z z funkcją przynależności $\mu_{R_2}: Y \times Z \rightarrow [0,1]$. $R_1 * R_2$ jest relacją rozmytą na X, Y, Z z funkcją przynależności $\mu_{R_1 * R_2}: X \times Y \times Z \rightarrow [0,1]$ taką, że $\forall (x, y, z) \in X \times Y \times Z$

$$\mu_{R_1 * R_2}(x, y, z) = \min(\mu_{R_1}(x, y), \mu_{R_2}(y, z))$$

Innymi interesującymi operacjami na relacjach rozmytych są: dopełnienie, rozszerzenie cylindryczne i kompozycja:

- 6) dopełnienie: CoR jest relacją rozmytą na X, Y z funkcją przynależności $\mu_{CoR}: X \times Y \rightarrow [0,1]$ taką, że $\forall (x, y) \in X \times Y$

$$\mu_{CoR_1}(x, y) = 1 - \mu_{R_1}(x, y)$$

- 7) rozszerzenie cylindryczne: Niech R będzie relacją rozmytą na $X_1, X_2, \dots, X_k$ z funkcją przynależności $\mu_R: X_1 \times X_2 \times \dots \times X_k \rightarrow [0,1]$. Rozszerzenie cylindryczne R w $X_1 \times X_2 \times \dots \times X_k \times X_{k+1} \times \dots \times X_n$, oznaczane przez $Ce(R)$, jest relacją rozmytą na $X_1, X_2, \dots, X_k, X_{k+1}, \dots, X_n$ z funkcją przynależności $\mu_{Ce(R)}: X_1 \times X_2 \times \dots \times X_k \times X_{k+1} \times \dots \times X_n \rightarrow [0, 1]$ taką, że

$$\forall (x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_n) \in X_1 \times X_2 \times \dots \times X_k \times X_{k+1} \times \dots \times X_n$$

$$\mu_{Ce(R)}(x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_n) = \mu_R(x_1, x_2, \dots, x_k)$$

- 8) kompozycja sup-min: niech X, Y, Z będą niepustymi zbiorami normalnymi, R_1 będzie relacją rozmytą na X, Y i R_2 będzie relacją rozmytą na Y, Z . Kompozycja sup-min $R_1 \circ R_2$ jest relacją rozmytą na X, Z taką, że $\forall (x, y) \in X \times Y$

$$\mu_{R_1 \circ R_2}(x, z) = \sup_{y \in Y} \min(\mu_{R_1}(x, y), \mu_{R_2}(y, z))$$

Selekcja jest działaniem definiowanym identycznie jak w algebrze relacji dla modelu klasycznego.

2.2. OPERACJE W RAMACH PODEJŚCIA OPARTEGO NA PODOBIENSTWIE

Dla podejścia opartego na podobieństwie rozmyta algebra relacji niewiele różni się od klasycznej.

Główna różnica występuje przy operacji selekcji [SHENOI 90], dla której specyfikuje się próg stopnia podobieństwa β między elementami dziedziny:

$$\sigma_{P(t)}(R) \quad \text{with level } \beta$$

gdzie $P(t)$ jest prostym lub złożonym predykatem odnoszącym się do krotek t w relacji R . Jeśli wartość progowa nie zostanie określona dla jakiegoś atrybutu, to przyjmuje się wówczas wartość 1. Ogólnie można powiedzieć, że w operacji tej sprawdza się, czy wartości interesujących atrybutów są podobne (bliskie) do danej wartości w stopniu β , podobieństwo jest badane zgodnie z miarą bliskości w ramach podejścia opartego na podobieństwie.

Można na przykład rozważyć relacje Kontrahenci przedstawioną w tabeli 1, reprezentującą dane o kontrahentach hurtowni, wśród których można wyróżnić informacje o ich solidności, czyli wywiązywaniu się z zobowiązań wobec przedsiębiorstwa. Dla dziedziny atrybutu Solidność, $D_{\text{Solidność}} = \{\text{solidny, bardzo dobry, dobry, średni, niepewny}\}$, została zdefiniowana relacja podobieństwa (tabela 2), dla dziedzin pozostałych atrybutów przyjmuje się relacje identyczności.

Tabela 1. Relacja Kontrahenci

NazwaIdK	PełnaNazwa	NIP	Adres	Telefon	Solidność
NOWAK	TOMASZ	525-000-01-27	ul. Zielona 2/12 07-412 Ostrołęka	325-86-59	{bardzo dobry}
ANDREX	Hurtownia art.gosp.dom ANRDEX	525-000-76-03	ul. Targowa 2 07-412 Ostrołęka	324-95-08	{bardzo dobry, solidny}
ANIOŁ	Anioł Joanna	898-000-79-07	ul. Złota 52 07-412 Ostrołęka	(029) 766-64-41	{solidny}
KOWAL	Kowalski Jan	897-001-39-12	ul. Kubusia Puchatka 25		{dobry, średni}

Tabela 2. Relacja podobieństwa $S_{\text{Solidność}}$ na dziedzinie atrybutu Solidność

$S_{\text{Solidność}}$	solidny	bardzo dobry	dobry	średni	niepewny
solidny	1.0	0.9	0.7	0.5	0.1
bardzo dobry	0.9	1.0	0.8	0.6	0.2
dobry	0.7	0.8	1.0	0.8	0.4
średni	0.5	0.6	0.8	1.0	0.6
niepewny	0.1	0.2	0.4	0.6	1.0

Zapytanie: Znajdź kontrahentów, którzy są „mniej więcej” bardzo dobrzy, można wyrazić w postaci:

```
SELECT *
FROM Kontrahenci
WHERE Solidność = „bardzo dobry”
WITH LEVEL (Solidność) = 0.7
```

W zapytaniu tym nieprecyzyjne wyrażenie „mniej więcej” modeluje klauzula **WITH LEVEL** Podanie progu podobieństwa równego 0.7 powoduje, że w odpowiedzi zostaną uwzględnione krotki, których wartość solidności jest podobna do wartości „bardzo dobry” w stopniu co najmniej równym 0.7, tzn. spełniające warunek:

$$\min_{x \in I(Solidnosc)} (s_{Solidnosc}(x, "bardzo\ dobry")) \geq 0.7$$

Inaczej mówiąc, w wyniku zostaną uwzględnione krotki, których wartość solidności posiada tylko elementy dziedziny podobne do „bardzo dobry” w stopniu większym lub równym 0.7, czyli wartości {„bardzo dobry”, „solidny”, „dobry”}. Wynik zapytania przedstawia tabela 3.

Tabela 3. Wynik zapytania o bardzo dobrych kontrahentów

NazwaIdK	PełnaNazwa	NIP	Adres	Telefon	Solidność
NOWAK	TOMASZ	525-000-01-27	ul. Zielona 2/12 07-412 Ostrołęka	325-86-59	{bardzo dobry}
ANDREX	Hurtownia art.gosp.dom ANRDEX	525-000-76-03	ul. Targowa 2 07-412 Ostrołęka	324-95-08	{bardzo dobry, solidny}
ANIOL	Anioł Joanna	898-000-79-07	ul. Złota 52 07-412 Ostrołęka	(029) 766-64-41	{solidny}

Różnica między rozmytą algebrą relacji dla podejścia opartego na podobieństwie a algebrą klasyczną występuje również przy operacjach, dla których usuwa się redundantne krotki. Redundancja dla podejścia opartego na podobieństwie, w przeciwieństwie do modelu klasycznego, nie oznacza, że wartości wszystkich atrybutów są równe, ale że wartości atrybutów są podobne w pewnym stopniu. Badanie redundancji odbywa się zatem w odmienny sposób, dla podejścia opartego na podobieństwie wymaga dodatkowo podania progów podobieństwa dla elementów dziedziny każdego atrybutu.

2.3. OPERACJE W RAMACH ROZSZERZONEGO PODEJŚCIA OPARTEGO NA MOŻLIWOŚCI ORAZ PODEJŚCIA OPARTEGO NA MOŻLIWOŚCI

Podstawowe operatory rozszerzonej algebry relacji dla rozszerzonego podejścia opartego na możliwości są definiowane następująco [CHEN 93]:

1) suma:

$$R \cup S = \{(\pi_{A_1}, \pi_{A_2}, \dots, \pi_{A_n}) \mid (\pi_{A_1}, \pi_{A_2}, \dots, \pi_{A_n}) \in R \vee (\pi_{A_1}, \pi_{A_2}, \dots, \pi_{A_n}) \in S\}.$$

2) różnica :

$$R - S = \{t = (\pi_{A_1}, \pi_{A_2}, \dots, \pi_{A_m}) \mid r \in R \wedge (\neg \exists s)(s \in S \wedge F_c(s, r) \geq \lambda)\}.$$

Fakt, że krotka r ze zbioru R należy do $R-S$ oznacza, że nie istnieje w S krotka bliska r w stopniu $\geq \lambda$.

3) iloczyn:

$$\begin{aligned} R \cap S = \{ & (\pi''_{A_1}, \pi''_{A_2}, \dots, \pi''_{A_n}) \mid r = (\pi_{A_1}, \pi_{A_2}, \dots, \pi_{A_n}) \in R \\ & \wedge s = (\pi'_{A_1}, \pi'_{A_2}, \dots, \pi'_{A_n}) \in S \\ & \wedge F_c(r, s) \geq \lambda \\ & \wedge \pi''_{A_i} = \pi_{A_i} \cup \pi'_{A_i}, i = 1, 2, \dots, n\}. \end{aligned}$$

$R \cap S$ zawiera nie tylko wspólne krotki R i S , ale również krotki będące połączeniem bliskich wzajemnie krotek z R i S zgodnie z miarą bliskości $E_c(r, s)$ i podanym progiem λ .

4) iloczyn kartezjański:

$$\begin{aligned} R \times S = \{ & (\pi_{A_{i1}}, \pi_{A_{i2}}, \dots, \pi_{A_{in}}, \pi_{A_{j1}}, \pi_{A_{j2}}, \dots, \pi_{A_{jm}}) \mid (\pi_{A_{i1}}, \pi_{A_{i2}}, \dots, \pi_{A_{in}}) \in R \\ & \wedge (\pi_{A_{j1}}, \pi_{A_{j2}}, \dots, \pi_{A_{jm}}) \in S\}. \end{aligned}$$

5) projekcja:

$$\Pi_{A_{i1} \dots A_{ik}}(R) = \{(\pi_{A_{i1}}, \pi_{A_{i2}}, \dots, \pi_{A_{ik}}) \mid (\pi_{A_{i1}}, \pi_{A_{i2}}, \dots, \pi_{A_{ik}}) \in R, i_1, i_2, \dots, i_k \in \{1, 2, \dots, n\}\}.$$

Dla projekcji na rozmytych relacjach bazowych dozwolone jest łączenie tylko identycznych krotek.

6) selekcja:

Niech $P(t)$ będzie prostym lub złożonym predykatem odnoszącym się do krotek t w R i niech θ będzie kolekcją operatorów porównania stosowanych do obliczania prawdziwości $P(t)$, na przykład $E_c, \cup$, operatory arytmetyczne, wówczas

$$\sigma_{P(t)\theta}(R) = \{t = (\pi_{A_1}, \pi_{A_2}, \dots, \pi_{A_n}) \mid t \in R \wedge P(t) \mid_{\theta} \geq \lambda\}.$$

7) złączenie naturalne:

$$\begin{aligned}
 R * S = \{ & t = (\pi_{A_1}, \dots, \pi_{A_{k-1}}, \pi_{A_k}, \dots, \pi_{A_n}, \pi_{A_{n+1}}, \dots, \pi_{A_m}) \mid r \in R \wedge s \in S \\
 & \wedge F_c(r(D_k \dots D_n), s(D_k \dots D_n)) \geq \lambda \wedge t(D_1 \dots D_{k-1}) = r(D_1 \dots D_{k-1}) \\
 & \wedge t(D_j) = r(D_j) \cup s(D_j) \text{ dla } j = k, \dots, n \\
 & \wedge t(D_{n+1} \dots D_m) = s(D_{n+1} \dots D_m) \}.
 \end{aligned}$$

Zdefiniowane złączenie jest wykonywane nie tylko na elementach wspólnych R i S , ale również na elementach bliskich należących do R i S . Bliskie elementy są łączone przez sumę zbiorów rozmytych $\cup$. Często jest to cecha niepożądana, dlatego też wprowadza się inną definicję złączenia, która obowiązuje dla relacji bazowych, a wykonywana jest tylko na identycznych elementach (krotek), łączenie krotek dotyczy tylko identycznych elementów. Złączenie to jest definiowane następująco: niech $=_{id}$ oznacza identyczność („identyczne”), wówczas:

$$\begin{aligned}
 R * S = \{ & t = (\pi_{A_1}, \dots, \pi_{A_{k-1}}, \pi_{A_k}, \dots, \pi_{A_n}, \pi_{A_{n+1}}, \dots, \pi_{A_m}) \mid r \in R \wedge s \in S \\
 & \wedge r(D_k \dots D_n) =_{id} s(D_k \dots D_n) \\
 & \wedge t(D_1 \dots D_{k-1}) = r(D_1 \dots D_{k-1}) \\
 & \wedge t(D_k \dots D_n) = r(D_k \dots D_n) \\
 & \wedge t(D_{n+1} \dots D_m) = s(D_{n+1} \dots D_m) \}.
 \end{aligned}$$

Podejście oparte na możliwości jest szczególnym przypadkiem rozszerzonego podejścia opartego na możliwości. W podejściu opartym na możliwości niejawnie zakłada się relacje identyczności τ dla każdej dziedziny, tzn. że relacją bliskości jest relacja identyczności. Dlatego operacje zdefiniowane na danych rozmytych w ramach rozszerzonego podejścia opartego na możliwości są operacjami również dla podejścia opartego na możliwości, przy czym należy pamiętać, że relacją bliskości jest tu relacja identyczności, a miarą bliskości jest E , będąca szczególnym przypadkiem miary E_c uwzględniając, że relacją bliskości jest relacja identyczności.

3. SQLf

Nad rozszerzeniem SQL pracowało wielu naukowców. Po raz pierwszy rozmyte zapytania w języku SQL były rozważane przez Tahaniego w 1979 roku. Do zapytań zostały wówczas wprowadzone proste rozmyte predykaty takie, jak na przykład „młody”, około 40”. Kolejne prace zostały poczynione m.in. przez Bosca, Galibourga i Hamona [BOSC 88], którzy rozważali rozszerzenie i implementacyjne aspekty rozmytych zapytań w języku SQL. W 1991 Bosc i Pivert [BOSC 91] przedstawili język SQLf, umożliwiający tworzenie nieprecyzyjnych zapytań. Opis tego języka został również opisany w [BOSC 94] oraz w [BOSC 95].

1) Podstawowy blok języka SQLf:

Podstawowy blok rozszerzonego języka SQL jest postaci:

SELECT [n | t | n, t] <atrybuty> **FROM** <relacje> **WHERE** <warunki>

gdzie n, t są opcjonalnymi parametrami służącymi do kalibracji ilościowej i jakościowej wyniku, n oznacza żadaną liczbę odpowiedzi, t oznacza minimalny stopień spełnienia warunku, <atrybuty> jest listą atrybutów relacji wynikowej, <relacje> oznacza nazwy relacji, z których mają pochodzić dane, <warunki> są kryteriami wyboru krotek mogącymi być warunkami rozmytymi, tzn. mogącymi zawierać atomowe rozmyte predykaty, na przykład młody, dobrze płatny, modyfikatory, na przykład bardzo, względnie, oraz połączenia, wśród których można wyróżnić AND, OR oraz rozmyte operatory złączenia.

Blok **SELECT ... FROM ... WHERE ...** może być interpretowany jako rozmyte ograniczenie iloczynu kartezjańskiego relacji wymienionych w klauzuli **SELECT** i kalibracja wyniku. Jeżeli R oznacza klasyczną relację, a F jest warunkiem rozmytym, to w odpowiedzi na zapytanie:

SELECT A FROM R WHERE F

otrzymywana jest relacja rozmyta R_f , przy czym stopień przynależności krotki t do R_f definiuje się wzorem:

$$\mu_{R_f}(t) = \sup_{x \in R \wedge x.A=t} \mu_{fc}(x)$$

gdzie $x.A$ oznacza projekcję krotki x na atrybut A.

Mając relacje Pracownicy i Wydziały przedstawione odpowiednio w tabelach 4 i 5, stosując SQLf można zadać różne zapytania posiadające rozmyte warunki

Tabela 4. Relacja Pracownicy

P#	Nazwisko	Imię	W#	Wiek	Stanowisko	Wynagrodzenie
1	Nowak	Tomasz	1	25	dyrektor	1400
2	Anioł	Joanna	1	45	księgową	3400
3	Kowalski	Andrzej	1	75	kierownik	2300
4	Bąk	Anna	2	50	sekretarka	4500

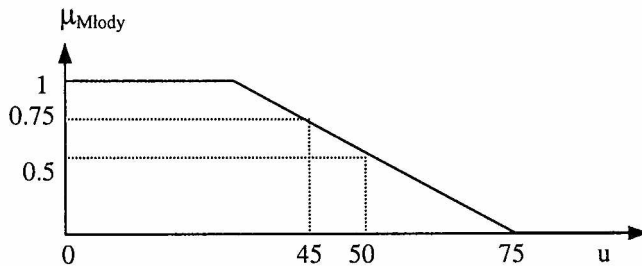
Tabela 5. Relacja Wydziały

W#	Budżet	Wielkość	Miasto
1	2000000	duży	Warszawa
2	300000	średni	Warszawa

Na przykład zapytanie: „znajdź młodych pracowników” można wyrazić w postaci:

SELECT P#, Wiek FROM Pracownicy WHERE Wiek = „Młody”

Funkcje przynależności dla rozmytego predykatu (zbioru rozmytego) „Młody” przedstawia rysunek 1. W wyniku otrzymuje się tabelę 6.



Rys. 1. Funkcja przynależności dla zbioru rozmytego „Młody”

Tabela 6. Wynik zapytania o młodych pracowników

P#	Wiek	$\mu_{\text{Młody}}$
1	25	1
2	45	0.75
4	50	0.5

W podobny sposób można wyrazić zapytanie o nie bardzo młodych pracowników:

SELECT P# FROM Pracownicy WHERE Wiek = „Nie bardzo młody”

gdzie „nie bardzo młody” jest zdefiniowane wzorem:

$$\mu_{\text{Nie bardzo młody}}(x) = 1 - (\mu_{\text{Młody}})^2$$

W wyniku otrzymuje się tabelę 7.

Tabela 7. Wynik zapytania o „nie bardzo młodych” pracowników

P#	Wiek	$\mu_{\text{Młody}}$
3	75	1
4	50	0.75
2	45	0.44

Zapytanie „znajdź młodych pracowników, którzy pracują w wysoko budżetowym wydziale” można wyrazić w postaci:

SELECT P#, Wydziały.W#

FROM Pracownicy, Wydziały

WHERE Wiek = "Młody"

AND Budżet = „Wysoki” **AND** Pracownicy.W# = Wydziały.W#

W zapytaniu tym warunek rozmyty składa się z dwóch rozmytych predykatów i złączenia. Krotka $t = (p, w)$ będzie przynależała do relacji wynikowej R_f w stopniu:

$$\mu_{R_f}(p, w) = \sup_{x \in R \wedge x.P\# = p \wedge y \in S \wedge y.W\# = w} \min(\mu_{\text{młody}}(x), \mu_{\text{wysoki}}(y))$$

Gdyby operator złączenia był rozmytym operatorem złączenia θ , to wówczas krotka $t = (p, w)$ należałaby do relacji wynikowej R_f w stopniu:

$$\mu_{R_f}(p, w) = \sup_{x \in R \wedge x.P\# = p \wedge y \in S \wedge y.W\# = w} \min(\mu_{\text{młody}}(x), \mu_{\text{wysoki}}(y), \mu_{\theta}(x.P\#, y.W\#))$$

Należy zatem zauważyć, że mając złożony warunek rozmyty w celu obliczenia stopnia spełnienia warunku należy zastosować operator min dla wszystkich atomowych predykatów rozmytych.

2) Podzapytania: operatory zagnieżdżające IN, EXISTS

Wyrażenie występujące po **WHERE** może również zawierać podzapytania, czyli zagnieżdżone klauzule **SELECT ... FROM ... WHERE ...**. Bosc i Pivert przedstawili semantykę zapytań zawierających operatory zagnieżdżające takie jak: **IN**, **EXISTS**, **ALL**, **ANY**. Zastosowanie tych operatorów można przedstawić, posługując się relacjami Pracownicy i Wydziały o schematach odpowiednio

Pracownicy(P#, Nazwisko, Imię, W#, Wiek, Stanowisko, Wynagrodzenie)

Wydziały(W#, Budżet, Wielkość, Miasto).

Mając zapytanie „znajdź pracowników pracujących w wydziale, którego budżet jest mniej więcej równy 1000-krotnemu wynagrodzeniu pracownika”, może mieć ono różne postacie:

- a) stosując złączenie:

```
SELECT P#
FROM Pracownicy, Wydziały
WHERE Wydziały.W# = Pracownicy.W#
      AND Budżet ≈ Wynagrodzenie * 1000
```

- b) stosując podzapytanie z operatorem **IN**:

```
SELECT P#
FROM Pracownicy
WHERE W# IN
      (SELECT W#
       FROM Wydziały
       WHERE Budżet ≈ Pracownicy.Wynagrodzenie * 1000)
```

- c) stosując podzapytanie z operatorem **EXISTS**:

```
SELECT P#
FROM Pracownicy
WHERE EXISTS (SELECT *
              FROM Wydziały
              WHERE W# = Pracownicy.W# AND
                    Budżet ≈ Pracownicy.Wynagrodzenie * 1000)
```

gdzie $\approx$ jest rozmytym operatorem reprezentującym mniej więcej.

Rozwiązania a) i b) są równoważne, jeśli przyjmujemy następującą definicję:

$$\mu_{IN}(a, E) = \sup_{b \in \text{supp}(E)} (\min(\mu_=(a, b), \mu_E(b)))$$

Rozwiązania a) i c) są równoważne, jeśli zdefiniujemy:

$$\mu_{EXISTS}(E) = \sup_{x \in \text{supp}(E)} (\mu_E(x))$$

3) Grupowanie

W języku SQL relacja może być podzielona na podrelacje, tzn. zbiory krotek, według wartości jednego lub kilku atrybutów wymienionych w klauzuli **GROUP BY**. Klauzuli **HAVING** używa się wraz z warunkami zorientowanymi na zbiory rozmyte do selekcji podrelacji uzyskanych za pomocą klauzuli **GROUP BY**. W ramach języka

SQLf Bosc i Pivert przedyskutowali różnorodne możliwe postacie warunków – od prostych zawierających funkcje agregujące (MIN, SUM, MAX, COUNT, AVG, itp.) do bardziej złożonych zawierających rozmyte kwantyfikatory.

Mając relacje Pracownicy i Wydziały o schematach odpowiednio

Pracownicy(P#, Nazwisko, Imię, W#, Wiek, Stanowisko, Wynagrodzenie)

Wydziały(W#, Budżet, Wielkość, Miasto)

można zadać zapytanie „znajdź wydziały, w których średnie wynagrodzenie jest duże”. Zapytanie to może mieć różne postacie:

a) **SELECT W#**

FROM Pracownicy

GROUP BY W#

HAVING AVG(Wynagrodzenie) = „Duże”

b) **SELECT W#**

FROM Pracownicy P

WHERE (SELECT AVG(Wynagrodzenie)

FROM Pracownicy

WHERE W# = P.W#) = „Duże”

Można rozważyć również inne, bardziej złożone zapytanie: „znajdź najlepszych 5 wydziałów, w których większość pracowników jest w średnim wieku”. Może mieć ono postać:

SELECT TOP 5 W#

FROM Pracownicy

GROUP BY W#

HAVING MOST ARE „wiek średni”

Podobnie można zadać zapytanie: „znajdź wydziały, w których większość pracowników w średnim wieku jest dobrze opłacanych”:

SELECT W#

FROM Pracownicy

GROUP BY W#

HAVING MOST (Wiek = „Średni”) ARE (Płaca = „Wysoka”)

4. PODSUMOWANIE

W klasycznych nierozmytych modelach baz danych języki zapytań służą do wydobywania informacji takiej, jaka została w bazie zapisana. Baza danych jest miejscem, gdzie się przechowuje dane a język zapytań pełni rolę pośrednika między bazą danych a jej użytkownikiem. Aby zapytanie zostało prawidłowo wykonane wartości atrybutów występujące w warunkach zapytań muszą pochodzić z dziedziny atrybutu. Jeżeli wzrost przechowujemy w bazie danych w postaci liczby rzeczywistej, to nie jest możliwe zadanie zapytania wykorzystującego warunek „wzrost = niski”. Natomiast język zapytań w rozmytych bazach danych musi umieć wykonać zapytania zarówno wykorzystujące warunki typu „wzrost <146” jak i „wzrost = niski”. Dane rozmyte są wykorzystywane w wielu operacjach wykonywanych na bazie danych a w szczególności w operacjach wyszukiwania o warunkach wyspecyfikowanych na owych danych, dodatkowo warunki te są często są również rozmyte. Przykładem takich operacji jest poszukiwanie rodzaju towaru (z określonym minimalnym, stopniem powiązania) spełniającego (w pewnych stopniach) ograniczenia położone na kolor, rozmiar i cenę, przy czym dodatkowo np. dla rozmiaru można wyspecyfikować próg bliskości do zadanych wartości, co daje możliwość bardziej naturalnego i swobodnego zdefiniowania, jakie rozmiary nas interesują. Innymi operacjami np. są poszukiwanie kontrahenta o danej solidności ze zdefiniowanym progiem podobieństwa do określonej wartości, poszukiwanie np. magazynu spełniającego w danym stopniu ograniczenia dotyczące wielkości powierzchni. Istotne znaczenie ma również sporządzanie odpowiednich zestawień, np. dla zaplanowania ilości asortymentu w magazynach, potrzebne jest zestawienie tych zamówień, które mają odpowiedni stopień przynależności do zamówień, aby móc ocenić, dla których trzeba na pewno przygotować zamówiony asortyment. Przedstawiony język SQLf jest próbą rozszerzenia przez Bosc’a i Piverta języka SQL do rozmytych baz danych.

OPERATING OF FUZZY DATA

The approaches of extending relation algebra and SGLf allowing to operate fuzzy relational database have been mentioned. We show now we can formulate fuzzy query in fuzzy database.

Barbara BIEDRZYCKA
Zygmunt MAZUR*

PROJEKTOWANIE SCHEMATÓW ROZMYTYCH RELACYJNYCH BAZ DANYCH

W pracy podjęto w ramach rozszerzonego podejścia opartego na możliwości (ang. *extended possibility-based framework*) tematykę projektowania schematów rozmytych relacyjnych baz danych, która jest związana z teorią normalizacji dla rozmytych modeli relacyjnych baz danych.

1. WPROWADZENIE

W rozmytej relacyjnej bazie danych, podobnie jak w klasycznej, mogą wystąpić anomalie przy wprowadzaniu, modyfikacji i usuwaniu danych, jeżeli schemat bazy (schematy relacji) nie zostanie poprawnie zaprojektowany, tzn. jeżeli pewne atrybuty zostaną umieszczone razem w jednym schemacie relacji.

Projektowanie schematów rozmytych relacyjnych baz danych jest związane z teorią normalizacji dla rozmytych modeli relacyjnych baz danych. Opiera się ona na pojęciach rozmytych zależności funkcyjnych i rozkładzie schematu relacji. Ogólną ideą jest stwierdzenie, czy schemat relacji jest poprawny, a jeżeli nie, to rozłożenie go na kilka schematów nie posiadających niepożądanych własności, zachowujących dane oryginalnego schematu i zachowujących jego semantykę. Uzyskane w rozkładzie schematy są schematami w pewnej rozmytej postaci normalnej (zgodnie z pewnymi kryteriami).

W artykule przedstawiono zasady projektowania schematów rozmytych relacyjnych baz danych w ramach rozszerzonego podejścia opartego na możliwości. Zaprezentowane definicje, przykłady, algorytmy pochodzą z [CHEN 95, CHEN 98], w owych pracach można również znaleźć dowody przytoczonych twierdzeń i dowody poprawności przedstawionych algorytmów.

* Wydziałowy Zakład Informatyki, Wydział Informatyki i Zarządzania Politechniki Wrocławskiej, 50-370 Wrocław, Skwer Idaszewskiego 1, mazur@ci.pwr.wroc.pl

2. ROZMYTE ZALEŻNOŚCI FUNKCYJNE

Podobnie jak w przypadku klasycznych zależności funkcyjnych, rozmyte zależności funkcyjne są rodzajem semantycznej wiedzy, ograniczeniami zależnymi jedynie od bliskości wartości, na przykład „identyczne nazwy implikują identyczny wiek”, „bliskie sobie wynagrodzenia odpowiadają bliskim poziomom wykonania”. Utrzymanie rozmytych zależności funkcyjnych w rozmytej relacyjnej bazie danych może być realizowane przez sprawdzanie ich spełnienia w czasie dokonywania operacji wprowadzania, usuwania czy modyfikacji danych.

2.1. OGÓLNA POSTAĆ ROZMYTYCH ZALEŻNOŚCI FUNKCYJNYCH

Klasyczna zależność funkcyjna jest definiowana następująco:

niech $R(A_1, A_2, \dots, A_n)$ będzie schematem relacji na dziedzinach $D_1, D_2, \dots, D_n$, gdzie $\text{Dom}(A_i) = D_i$, X i Y będą podzbiórmi zbioru atrybutów $U = \{A_1, A_2, \dots, A_n\}$, tj. $X, Y \subseteq U$ i R będzie wartością $R, R \subseteq D_1 \times D_2 \times \dots \times D_n$.

X wyznacza funkcyjnie Y (lub Y zależy funkcyjnie od X), oznaczane przez $X \rightarrow Y$, wtedy i tylko wtedy, gdy $\forall R \in R(U), \forall t, t' \in R$ jeżeli $t(X) = t'(X)$, to $t(Y) = t'(Y)$.

W przypadku danych dokładnych stopień $X \rightarrow Y$ równa się 1. W przypadku rozmytym, gdy dopuszczalny jest jako wartość atrybutu rozkład możliwości, można oczekiwać, że stopień ten będzie należeć do przedziału $[0, 1]$. Pojawiają się tu trzy problemy: co oznaczają równości $t(X) = t'(X)$ i $t(Y) = t'(Y)$, jeżeli $t(X), t'(X), t(Y), t'(Y)$ są nieprecyzyjnymi wartościami takimi, jak na przykład rozkład możliwości, jak wyznaczyć stopień zdania „jeżeli $t(X) = t'(X)$, to $t(Y) = t'(Y)$ ”, jeżeli $t(X) = t'(X)$ i $t(Y) = t'(Y)$ są prawdziwe w stopniu należącym do przedziału $[0, 1]$, jak obliczyć stopień $X \rightarrow Y$, jeżeli różne pary krotek dają różne wartości prawdy dla zdania „jeżeli-to”. Te problemy związane są odpowiednio z problemem bliskości rozmytych danych, rozmytą logiczną implikacją i rozmytym operatorem AND.

Rozmytą zależność funkcyjną definiuje się zatem następująco:

niech $R(A_1, A_2, \dots, A_n)$ będzie schematem relacji na dziedzinach $D_1, D_2, \dots, D_n$, gdzie $\text{Dom}(A_i) = D_i$, X i Y będą podzbiórmi zbioru atrybutów $U = \{A_1, A_2, \dots, A_n\}$, tj. $X, Y \subseteq U$, i R będzie wartością $R, R \subseteq \Pi(D_1) \times \Pi(D_2) \times \dots \times \Pi(D_n)$, gdzie $\Pi(D_i) = \{\pi \mid \pi \text{ jest wykluczającym rozkładem możliwości atrybutu } A_i \text{ na } D_i, i=1, \dots, n\}$. X wyznacza funkcyjnie Y (lub Y zależy funkcyjnie od X) w stopniu θ , oznaczane przez $X \rightarrow_\theta Y$, wtedy i tylko wtedy, gdy $\forall R \in R$

$$\min_{t, t' \in R} I(t(X) =_c t'(X), t(Y) =_c t'(Y)) \geq \theta$$

gdzie $\theta \in [0, 1]$, $=_c: [0, 1] \times [0, 1] \rightarrow [0, 1]$ jest miarą bliskości i

$I: [0, 1] \times [0, 1] \rightarrow [0, 1]$ jest rozmytym operatorem implikacji.

Jest to ogólna postać rozmytej zależności funkcyjnej, wyrażająca semantykę, że bliskim wartościom atrybutów ze zbioru X odpowiadają bliskie wartości atrybutów ze zbioru Y (w stopniu θ). Wybierając szczególną miarę bliskości dla $=_c$ i szczególny rozmyty operator implikacji dla I , otrzymuje się szczególną postać rozmytej zależności funkcyjnej. Przyjmijmy następującą definicję miary bliskości $=_c$:

$$Prawda(t(X) =_c t'(X)) = G_c(t(X), t'(X))(T)$$

$$Prawda(t(Y) =_c t'(Y)) = G_c(t(Y), t'(Y))(T)$$

gdzie

$$\approx(t(A) =_c t'(A)) = E_c(t(A), t'(A))(T) = \sup_{\substack{u, v \in D_A \\ c_A(u, v) \geq \alpha_A}} \min(\pi_A(u), \pi'_A(v)) \quad \forall A \in X$$

$$\approx(t(B) =_c t'(B)) = E_c(t(B), t'(B))(T) = \sup_{\substack{w, z \in D_B \\ c_B(w, z) \geq \alpha_B}} \min(\pi_B(w), \pi'_B(z)) \quad \forall B \in Y$$

Jeżeli rozkład możliwości będący wartością atrybutu nie jest znormalizowany, to $\approx(t(A), t'(A))$ i $\approx(t(B), t'(B))$ może być oszacowywana na podstawie wzorów:

$$\approx(t(A) =_c t'(A)) = \begin{cases} 1 & t(A) \text{ i } t'(A) \text{ są identyczne} \\ \sup_{\substack{u, v \in D_A \\ c_A(u, v) \geq \alpha_A}} \min(\pi_A(u), \pi'_A(v)) & \text{w przeciwnym przypadku} \end{cases}$$

$$\approx(t(B) =_c t'(B)) = \begin{cases} 1 & t(B) \text{ i } t'(B) \text{ są identyczne} \\ \sup_{\substack{w, z \in D_B \\ c_B(w, z) \geq \alpha_B}} \min(\pi_B(w), \pi'_B(z)) & \text{w przeciwnym przypadku} \end{cases}$$

2.2. ROZSZERZONE KLUCZE I REGUŁY INTEGRALNOŚCI

Rozmyte zależności funkcyjne stanowią podstawę dla wprowadzenia θ -kluczy, będących rozszerzeniem klasycznej koncepcji klucza. Definicja θ -kluczy jest związana z pojęciami pełnej i częściowej rozmytej zależności funkcyjnej oraz z logicznymi implikacjami rozmytych zależności.

Niech $U = \{A_1, A_2, \dots, A_n\}$ będzie zbiorem atrybutów, $\mathbf{R}(U)$ będzie schematem relacji na dziedzinach $D_1, D_2, \dots, D_n$, gdzie $\text{Dom}(A_i) = D_i$, X i Y będą podzbiorami U , tj. $X, Y \subseteq U$. Mówi się, że Y jest w pełni zależny funkcyjnie od X w stopniu θ lub że istnieje pełna zależność funkcyjna z X do Y w stopniu θ wtedy i tylko wtedy, gdy $X \rightarrow_{\theta} Y$ i nie istnieje X' ($X' \subset X$, $X' \neq \emptyset$) taki, że $X' \rightarrow_{\theta} Y$. Jeżeli taki X' istnieje, to

mówi się, że Y jest częściowo zależny funkcyjnie od X w stopniu θ (istnieje częściowa zależność z X do Y w stopniu θ).

Warto zauważyć, że dla pełnej zależności $X \rightarrow_{\theta} Y$ dozwolone są zależności $X' \rightarrow_{\alpha} Y$ dla których $\alpha < \theta$, $X' \subset X$. Na przykład, dla $U = \{A, B, C\}$, $F = \{AB \rightarrow_{\theta} C, A \rightarrow_{\alpha} C\}$ i $\alpha < \theta$, C jest w pełni zależy od AB w stopniu θ .

Logiczne implikacje rozmytych zależności są zdefiniowane następująco: niech F będzie zbiorem rozmytych zależności funkcyjnych dla schematu relacji $R(U)$ i $X \rightarrow_{\theta} Y$ będzie rozmytą zależnością funkcyjną, $X, Y \subseteq U$ ($X \rightarrow_{\theta} Y$ może nie należeć do F). $X \rightarrow_{\theta} Y$ jest logiczną implikacją F wtedy i tylko wtedy, gdy $X \rightarrow_{\theta} Y$ jest zachowana przez każdą relację schematu $R(U)$ spełniającą wszystkie rozmyte zależności funkcyjne z F . $X \rightarrow_{\theta} Y$ logicznie implikowana przez F jest oznaczana zależnością $X \rightarrow_{\theta} Y \in F^+$, gdzie F^+ jest zbiorem wszystkich rozmytych zależności funkcyjnych, które są logicznie implikowane przez F .

Mając zdefiniowaną pełną i częściową zależność funkcyjną oraz logiczne implikacje rozmytych zależności można wprowadzić definicję klucza dla rozmytego modelu relacyjnego. Niech $U = \{A_1, A_2, \dots, A_n\}$ będzie zbiorem atrybutów, $R(U)$ będzie schematem relacji na dziedzinach $D_1, D_2, \dots, D_n$, gdzie $\text{Dom}(A_i) = D_i$, K będzie podzbiorem U , tj. $K \subseteq U$. K jest θ -kluczem schematu relacji $R(U)$ wtedy i tylko wtedy, gdy $K \rightarrow_{\theta} U \in F^+$ i $K \rightarrow_{\theta} U$ jest zależnością pełną.

Na przykład niech $R(S\#, \text{Nazwa}, \text{Adres}, \text{Kurs}\#, \text{Ocena})$ i $F = \{S\# \rightarrow_{\alpha} \text{Nazwa}, S\# \rightarrow_{\beta} \text{Adres}, (S\#, \text{Kurs}\#) \rightarrow_{\chi} \text{Ocena}\}$. $(S\#, \text{Kurs}\#)$ jest θ -kluczem schematu relacji R z $\theta = \min(\alpha, \beta, \chi)$, ponieważ $(S\#, \text{Kurs}\#) \rightarrow_{\theta} (S\#, \text{Nazwa}, \text{Adres}, \text{Kurs}\#, \text{Ocena}) \equiv U$. Zależności $S\# \rightarrow_{\theta} U$, $\text{Kurs}\# \rightarrow_{\theta} U$ nie są spełnione, a Nazwa i Adres są częściowo zależne od θ -klucza $(S\#, \text{Kurs}\#)$ w stopniu θ .

Jeżeli $R(U)$ ma więcej niż jeden θ -klucz, wtedy każdy θ -klucz jest θ -kluczem kandydującym. Jeden θ -klucz kandydujący jest wyznaczany na θ -klucz główny. Atrybuty należące do klucza są nazywane atrybutami kluczowymi, a atrybuty nie należące do niego – niekluczowymi. W rozmytej bazie danych zależność $K \rightarrow_{\theta} A_i$ wyraża, że identycznym wartościom dla atrybutów klucza odpowiadają identyczne wartości dla A_i , a bliskim wartościom dla atrybutów klucza odpowiadają bliskie wartości dla A_i . Dla każdego $S \subseteq U$, S jest θ -nadkluczem schematu relacji $R(U)$ wtedy i tylko wtedy, gdy S zawiera θ -klucz.

Z pojęciem klucza jest związana reguła integralności konkretnu. W klasycznych bazach danych reguła ta mówi, że wartości klucza nie mogą przyjmować wartości nieokreślonej (null). W rozmytej bazie danych, w której wartości atrybutów mogą przyjmować wartości nieprecyzyjne, definiuje się dwie reguły integralności konkretnu:

- 1) Słaba θ -reguła integralności konkretnu:
wartości θ -klucza nie mogą być wartościami nieokreślonymi (null).
- 2) Mocna θ -reguła integralności konkretnu:
wartości θ -klucza nie mogą być ani wartościami nieokreślonymi (null), ani wartościami nieprecyzyjnymi.

Narzucenie mocnej θ -reguły integralności konkretności pozwala na występowanie rozmytych danych tylko dla atrybutów niekluczowych.

Obok pojęcia θ -klucza, wprowadza się definicję θ -klucza obcego. Niech $R(U)$ i $R'(U)$ będą dwoma schematami relacji, $K \subseteq U$ i $K' \subseteq U'$. K jest nazywany θ -kluczem obcym schematu relacji $R(U)$ wtedy i tylko wtedy, gdy K nie jest θ -kluczem $R(U)$, ale jest θ -kluczem $R'(U)$.

Na przykład niech

$$R_1(S\#, \text{Nazwa}, \text{Adres}),$$

$$R_2(S\#, \text{Kurs}\#, \text{Ocena}),$$

$$S\# \rightarrow_{\alpha} \text{Nazwa},$$

$$S\# \rightarrow_{\beta} \text{Adres}$$

$$(S\#, \text{Kurs}\#) \rightarrow_{\chi} \text{Ocena},$$

wówczas $S\#$ jest θ -kluczem $R_1(\theta = \min(\alpha, \beta))$ i $S\#$ jest θ -kluczem obcym R_2 .

Z pojęciem klucza obcego związana jest reguła integralności odwołań. W rozmytym modelu θ -reguła integralności odwołań mówi, że każda wartość θ -klucza obcego relacji musi albo „pasować” do wartości odpowiadającego klucza głównego innej relacji, albo przyjmować wartość nieokreśloną (null). Dopasowanie wartości dla θ -reguły integralności odwołań może opierać się albo na identyczności, albo na bliskości. W drugim przypadku, wraz z miarą bliskości powinien być wyspecyfikowany próg $\lambda \in (0, 1]$, aby wprowadzić ograniczenie mówiące, że wartości klucza obcego mają być bliskie wartościom odpowiedniego klucza głównego w stopniu $\geq \lambda$. Jeżeli θ -reguła integralności odwołań opiera się na identyczności wraz z narzuceniem mocnej reguły integralności konkretności dla odpowiadającego klucza głównego, wtedy wartości θ -klucza obcego nie mogą być wartościami nieprecyzyjnymi.

2.3. AKSJOMATY DLA ROZMYTYCH ZALEŻNOŚCI FUNKCYJNYCH

Aby wyznaczyć θ -klucze i zrozumieć logiczne wynikanie zachodzące między rozmytymi zależnościami funkcyjnymi, należy z F obliczyć F^+ lub co najmniej powiedzieć, mając dany zbiór F oraz zależność $X \rightarrow_{\theta} Y$, czy $X \rightarrow_{\theta} Y$ należy do F^+ . Dokonanie tego wymaga znajomości reguł wnioskowania określających, w jaki sposób pewne zależności wynikają z innych. W rzeczywistości można zrobić więcej – zapewnić zupełny zbiór reguł wnioskowania, co oznacza, że jeśli mamy dany zbiór zależności F , reguły umożliwiają wnioskowanie wszystkich zależności należących do F^+ . Co więcej reguły te są poprawne, co oznacza, że stosując je nie można wywnioskować z F żadnej zależności, która nie należałaby do F^+ .

W klasycznym modelu relacyjnym ten zbiór reguł tworzą aksjomaty Armstronga [ULLMAN 88]:

- A_1 : jeżeli $Y \subseteq X$, to $X \rightarrow Y$ (zwrotność)
 A_2 : jeżeli $X \rightarrow Y$, to $XZ \rightarrow YZ$ (rozszerzanie)
 A_3 : jeżeli $X \rightarrow Y$ i $Y \rightarrow Z$, to $X \rightarrow Z$ (przechodność)

Istnieje również kilka innych reguł wnioskowania wynikających z powyższych aksjomatów, wyróżnia się wśród nich:

- D_1 : jeżeli $X \rightarrow Z$ i $X \rightarrow Y$, to $X \rightarrow YZ$ (reguła sumowania)
 D_2 : jeżeli $X \rightarrow Y$ i $WY \rightarrow Z$, to $XW \rightarrow Z$ (reguła pseudoprzechodności)
 D_3 : jeżeli $X \rightarrow Y$ i $Z \subseteq Y$, to $X \rightarrow Z$ (reguła rozkładania)

Zbiór aksjomatów Armstronga jest poprawny i zupełny, a dowód tego można znaleźć w [ULLMAN 88].

Dla rozmytego modelu relacyjnego zbiór aksjomatów Armstronga został rozszerzony do trzech reguł wnioskowania rozmytych zależności funkcyjnych nazywanych rozszerzonymi aksjomatami Armstronga (A'_1 , A'_2 , A'_3), mają one następującą postać:

- A'_1 : jeżeli $Y \subseteq X$, to $X \rightarrow_\theta Y$ dla każdego θ
 A'_2 : jeżeli $X \rightarrow_\theta Y$, to $XZ \rightarrow_\theta YZ$
 A'_3 : jeżeli $X \rightarrow_\alpha Y$ i $Y \rightarrow_\beta Z$, to $X \rightarrow_\chi Z$, przy czym $\chi = \min(\alpha, \beta)$
gdzie $\alpha, \beta, \chi, \theta \in [0, 1]$.

Dodatkowo z A'_1 , A'_2 , A'_3 można otrzymać reguły:

- D'_1 : jeżeli $X \rightarrow_\alpha Z$ i $X \rightarrow_\beta Y$, to $X \rightarrow_\chi YZ$ gdzie $\chi = \min(\alpha, \beta)$
 D'_2 : jeżeli $X \rightarrow_\alpha Y$ i $WY \rightarrow_\beta Z$, to $XW \rightarrow_\chi Z$ gdzie $\chi = \min(\alpha, \beta)$
 D'_3 : jeżeli $X \rightarrow_\alpha Y$ i $Z \subseteq Y$, to $X \rightarrow_\alpha Z$
 D'_4 : jeżeli $X \rightarrow_\alpha Y$, to $X \rightarrow_\beta Y$ dla każdego $\beta \leq \alpha$
gdzie $\alpha, \beta, \chi, \theta \in [0, 1]$.

Ważnym następstwem reguł sumowania i rozkładania jest fakt, że jeśli $A_1, A_2, \dots, A_n$ są atrybutami, to $X \rightarrow_\theta A_1, A_2, \dots, A_n$ zachodzi wtedy i tylko wtedy, gdy dla każdego i , zachodzi $X \rightarrow_\theta A_i$. Warto również wspomnieć, że dla zależności $X \rightarrow_\theta Y$ zakłada się, że $\theta \in (0, 1]$.

Rozszerzone aksjomaty Armstronga A'_1 , A'_2 i A'_3 są poprawne i zupełne, fakt ten dostarcza „równoważności” między pojęciem logicznej implikacji rozmytych zależności funkcyjnych i pojęciem rozmytych zależności funkcyjnych otrzymanych opierając się na aksjomatach. Jeżeli F^A jest zbiorem wszystkich rozmytych zależności funkcyjnych, które mogą być otrzymane z F stosując aksjomaty, tj. $F^A = \{X \rightarrow_\theta Y \mid X \rightarrow_\theta Y \text{ jest wyprowadzona z } F \text{ stosując aksjomaty i } \theta > 0\}$, to poprawność i zupełność aksjomatów Armstronga, oznaczają, że $F^A = F^+$. Wynik ten jest ważną teoretyczną podstawą dla projektowania schematów rozmytych relacyjnych baz danych.

2.4. SZCZEGÓLNE POSTACIE ROZMYTYCH ZALEŻNOŚCI FUNKCYJNYCH

Aby zapewnić spełnienie rozszerzonych aksjomatów Armstronga, zastosowany szczególny rozmyty operator implikacji musi posiadać następujące własności:

C_1 : jeżeli $a \leq b$ wtedy $I(a, b)=1$

C_2 : jeżeli $I(a, b) \geq \theta$ wtedy $I(a', b') \geq \theta$ gdzie $a' = \min(a, c)$, $b' = \min(b, c)$

C_3 : jeżeli $I(a, b) \geq \alpha$ i $I(b, c) \geq \beta$ wtedy $I(a', c) \geq \gamma$ gdzie $\gamma = \min(\alpha, \beta)$

oraz $a, b, c, a', b', c', \alpha, \beta, \gamma, \theta \in [0, 1]$.

W pracy [BIEMAZUR 02b] zostało zaprezentowanych 15 definicji rozmytych operatorów implikacji. Tabela 1 [CHEN 98] przedstawia wyniki badania ich wobec posiadania cech C_1 , C_2 , C_3 , czyli ujawnia, jakiego wyboru można dokonać pomiędzy tymi operatorami dla rozmytej zależności funkcyjnej, aby zostały spełnione reguły wnioskowania.

Tabela 1. Rozmyte operatory implikacji wobec własności C_1 , C_2 , C_3

Oznaczenie	C_1	C_2	C_3
$S^\#$	Tak	Tak	Tak
S	Tak	Tak	Tak
S^*	Tak	Tak	Tak
G	Tak	Tak	Nie
G^*	Tak	Tak	Nie
L	Tak	Tak	Nie
KDL	Nie	Nie	Nie
KD	Nie	Nie	Nie
EA	Nie	Nie	Nie
W	Nie	Nie	Nie
M	Nie	Nie	Tak
SS	Nie	Tak	Tak
S^*S^*	Nie	Tak	Tak
SS^*	Nie	Tak	Tak
S^*S	Nie	Tak	Tak

Ogólna postać rozmytej zależności funkcyjnych wyraża związek pomiędzy atrybutami polegający na tym, że „bliskim” wartościom jednego zbioru atrybutów odpowiadają „bliskie” wartości drugiego zbioru atrybutów. Różne szczególne postacie rozmytych zależności funkcyjnych, uzyskane przez wybór odpowiedniego rozmytego operatora implikacji w miejsce I , reprezentuje ową „bliskość” w różny sposób.

Stosując operator implikacji Gödela S^* ($S^*(a, b) = 1$ dla $a \leq b$; $S^*(a, b) = b$ dla $a > b$), tj.

$$I \equiv S^*$$

$X \rightarrow_{\theta} Y$ oznacza, że bliskim wartościom dla X odpowiadają bliskie wartości dla Y i stopień bliskości wartości dla Y równa się co najmniej stopniowi bliskości wartości dla X lub co najmniej θ .

Stosując operator implikacji Gödela i klasyczny operator implikacji I_c , tj.

$$I \equiv \begin{cases} I_c & \text{t(X) i t'(X) są identyczne} \\ S^* & \text{w przeciwnym przypadku} \end{cases}$$

$X \rightarrow_{\theta} Y$ oznacza, że identycznym wartościom dla X odpowiadają identyczne wartości dla Y , bliskim wartościom dla X odpowiadają bliskie wartości dla Y i stopień bliskości wartości dla Y równa się co najmniej stopniowi bliskości wartości dla X lub co najmniej θ .

Stosując operator implikacji standardowy ścisły S ($S(a, b) = 1$ dla $a \leq b$, $S(a, b) = 0$ dla $a > b$) tj.

$$I \equiv S$$

dla $\theta \geq 0$, $I(t(X) =_c t'(X), t(Y) =_c t'(Y)) \geq \theta$ staje się $S(t(X) =_c t'(X), t(Y) =_c t'(Y)) = 1$.
Stąd

$$X \rightarrow_{\theta} Y \Leftrightarrow X \rightarrow_1 Y \Leftrightarrow t(X) =_c t'(X) \leq t(Y) =_c t'(Y),$$

co oznacza, że bliskim wartościom dla X odpowiadają bliskie wartości dla Y i stopień bliskości wartości dla Y równa się co najmniej stopniowi bliskości wartości dla X .

Stosując operator implikacji standardowy ostry $S^{\#}$ ($S^{\#}(a, b) = 1$ dla $a < 1$ lub $b = 1$, $S^{\#}(a, b) = 0$ w przeciwnym przypadku), tj.

$$I \equiv S^{\#}$$

dla $\theta \geq 0$, $I(t(X) =_c t'(X), t(Y) =_c t'(Y)) \geq \theta$ staje się $S(t(X) =_c t'(X), t(Y) =_c t'(Y)) = 1$.
Stąd

$$X \rightarrow_{\theta} Y \Leftrightarrow X \rightarrow_1 Y \Leftrightarrow t(X) =_c t'(X) < 1 \text{ lub } t(Y) =_c t'(Y) = 1,$$

co oznacza, że bliskim wartościom dla X odpowiadają bliskie wartości dla Y i stopień bliskości wartości dla X jest mniejszy od 1 lub stopień bliskości wartości dla Y równa się 1 (całkowicie bliskie).

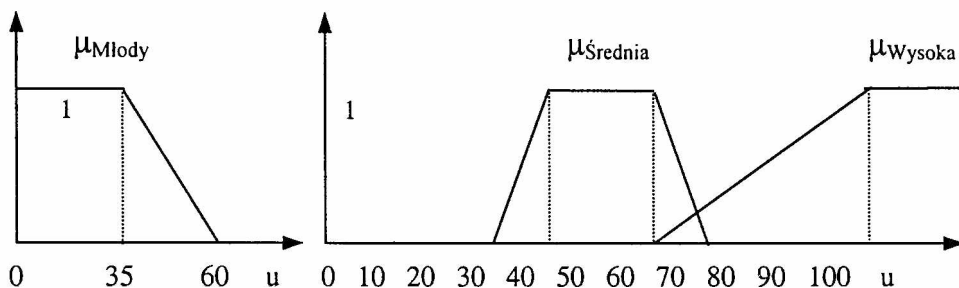
Przedstawione rozmyte operatory implikacji spełniają C_1 , C_2 , C_3 , a więc rozmyte zależności funkcyjne zdefiniowane z nimi spełniają rozszerzone aksjomaty Armstronga.

Dla przykładu można rozważyć przedsiębiorstwo utrzymujące dane swoich klientów. Relacja Kontrahenci może mieć postać zaprezentowaną w tabeli 2.

Tabela 2. Relacja Kontrahencji

	Nazwa	Wiek	Wykształcenie	Konsumpcja(\$)
t_1	Nowak	25	licencjat	Średnia
t_2	Anioł	25	magister	90.000
t_3	Kowalski	{23, 24}	licencjat	36.000
t_4	{Anioł, Bąk}	Młody	magister	Wysoka

Definicje dla wyrażen lingwistycznych: „Młody”, „Średnia” i „Wysoka” przedstawia rysunek 1.



Rys. 1. Funkcje przynależności zbiorów rozmytych "Młody", "Średni" i "Wysoki"

Można założyć, że bliskość wartości atrybutów będących zbiorami rozmytymi jest mierzona jako wysokość iloczynu zbiorów rozmytych:

$$\approx_H(A, B) = \sup_x \min(\mu_A(x), \mu_B(x))$$

a miarą bliskości dla wartości atrybutów będących rozkładami możliwości jest:

$$E(\pi_1, \pi_2)(T) = \sup_x \min(\pi_1(x), \pi_2(x))$$

Rozważając rozmyte zależności funkcyjne reprezentujące semantykę „identycznym wartościom dla X odpowiadają identyczne wartości dla Y, bliskim wartościom dla X odpowiadają bliskie wartości dla Y”, relacja Kontrahencji spełnia zależności:

Nazwa $\rightarrow_{1,0}$ Wiek, Nazwa $\rightarrow_{1,0}$ Wykształcenie, Nazwa $\rightarrow_{0,75}$ Konsumpcja, ponieważ $E(t_2(\text{Nazwa}), t_4(\text{Nazwa}))(T) = E(\{1/\text{Anioł}\}, \{1/\text{Anioł}, 1/\text{Bąk}\})(T) = 1$, $E(t_2(\text{Wiek}), t_4(\text{Wiek}))(T) = E(\{1/25\}, \text{Młody})(T) = 1$, $E(t_2(\text{Wykształcenie}), t_4(\text{Wykształcenie}))(T) = E(\{1/\text{Magister}\}, \{1/\text{Magister}\})(T) = 1$, $E(t_2(\text{Konsumpcja}), t_4(\text{Konsumpcja}))(T) = E(\{1/90.000\}, \text{Wysoka})(T) = 0.75$.

Można również zauważyć, że relacja Kontrahencji nie spełnia zależności: Wiek $\rightarrow_\alpha$ Wykształcenie, Wiek $\rightarrow_\beta$ Konsumpcja, Wykształcenie $\rightarrow_\gamma$ Konsumpcja dla $\alpha, \beta, \gamma \in (0, 1]$.

W niektórych przypadkach nie jest wymagane, aby identycznym wartościom dla X odpowiadały identyczne wartości dla Y, na przykład poziom konsumpcji produktów firmy przez klienta może być związany z poziomem wykształcenia, wówczas semantyka Wykształcenie $\rightarrow_\gamma$ Konsumpcja ($\gamma > 0$) może być reprezentowana używając rozmytej zależności funkcyjnej opartej na operatorze implikacji Gödela. W tym przypadku relacja Kontrahenci spełnia zależność Wykształcenie $\rightarrow_{0.6}$ Konsumpcja, ponieważ

$$E(t_1(\text{Konsumpcja}), t_3(\text{Konsumpcja}))(T) = E(\{\text{Średnia}, \{1/36.000\}\})(T) = 0.6$$

i

$$E(t_2(\text{Konsumpcja}), t_4(\text{Konsumpcja}))(T) = E(\{1/90.000\}, \text{wysoka})(T) = 0.75.$$

Dla następnych rozważań zostanie przyjęta następująca postać rozmytej zależności funkcyjnej:

$$\min_{t, t' \in R} I(t(X) =_c t'(X), t(Y) =_c t'(Y)) \geq \theta$$

gdzie

$$I \equiv \begin{cases} I_c & t(X) \text{ i } t'(X) \text{ są identyczne} \\ S^* & \text{w przeciwnym przypadku} \end{cases}$$

$$\text{Prawda}(t(X) =_c t'(X)) = G_c(t(X), t'(X))(T)$$

$$\text{Prawda}(t(Y) =_c t'(Y)) = G_c(t(Y), t'(Y))(T)$$

z

$$\approx (t(A) =_c t'(A)) = \begin{cases} 1 & t(A) \text{ i } t'(A) \text{ są identyczne} \\ \sup_{\substack{u, v \in D_A \\ c_A(u, v) \geq \alpha_A}} \min(\pi_A(u), \pi'_A(v)) & \text{w przeciwnym przypadku} \end{cases}$$

$$\approx (t(B) =_c t'(B)) = \begin{cases} 1 & t(B) \text{ i } t'(B) \text{ są identyczne} \\ \sup_{\substack{w, z \in D_B \\ c_B(w, z) \geq \alpha_B}} \min(\pi_B(w), \pi'_B(z)) & \text{w przeciwnym przypadku} \end{cases}$$

2.5. DOMKNIĘCIE ZBIORU ROZMYTYCH ZALEŻNOŚCI FUNKCYJNYCH

Czasami konieczne jest określenie, czy dana rozmyta zależność funkcyjna $X \rightarrow_{\theta} Y$ może być wyprowadzona z F stosując aksjomaty (A'_1, A'_2, A'_3). Inaczej mówiąc, czy dana zależność $X \rightarrow_{\theta} Y$ należy do F^A . Prosty sposób rozwiązania jest obliczenie zbioru F^{A+} zdefiniowanego wzorem:

$$F^{A+} = \{X \rightarrow_{\theta} Y \mid X \rightarrow_{\theta} Y \in F^A \wedge \theta = \sup\{\beta \mid X \rightarrow_{\beta} Y \in F^A\}\}$$

Zbiór F^{A+} jest skończony pod warunkiem, że zbiór F jest skończony (U jest skończonym zbiorem zgodnie z definicją $R(U)$). Mając zależność $X \rightarrow_{\theta} Y$, $\theta > 0$ można stwierdzić, czy należy ona do F^A przez sprawdzenie zbioru F^{A+} . $X \rightarrow_{\theta} Y \in F^A$ wtedy i tylko wtedy, gdy $\exists X \rightarrow_{\beta} Y \in F^{A+}$ taka, że $\beta \geq \theta$.

Bezpośrednie obliczanie zbioru F^{A+} przez rozważanie wszystkich atrybutów w U i powtarzające się zastosowywanie wszystkich aksjomatów może być za bardzo nieefektywne. Bardziej efektywnym sposobem stwierdzenia, czy $X \rightarrow_{\theta} Y$ należy do F^{A+} jest zdefiniowanie domknięcia X (względem F), oznaczanego przez X_F^+ , gdzie

$$X_F^+ = \{(A, \theta) \mid A \in U, \theta = \sup\{\beta \mid X \rightarrow_{\beta} A \in F^A\}\}$$

Niech $Y = A_1 A_2 \dots A_k \subseteq U$. $X \rightarrow_{\theta} Y \in F^A$ wtedy i tylko wtedy, gdy $\exists (A_1, \theta_1), (A_2, \theta_2), \dots, (A_k, \theta_k) \in X_F^+$ takie, że $\min(\theta_1, \theta_2, \dots, \theta_k) \geq \theta$.

Do wyznaczania X_F^+ stosuje się algorytm 1.

Algorytm 1. Wejście: $X = A_1 A_2 \dots A_k$. Wyjście: X_F^+

Metoda:

- 1) (Podstawa) $X_F^{(0)} = \{(A_1, 1), (A_2, 1), \dots, (A_k, 1)\}$, $i = 0$.
- 2) $B^{(i)} = \{(A, \theta) \mid (\exists V)(\exists W)(V \rightarrow_{\alpha} W \in F, V \subseteq \text{Dom}(X_F^{(i)}), A \in W, \theta = \min_{(C, \beta) \in X_F^{(i)}} (\alpha, \gamma), \gamma = \min_{C \in V} (\beta))\}$.
- 3) $X_F^{(i+1)} = X_F^{(i)} \cup B^{(i)}$.
- 4) Jeżeli

$$X_F^{(i+1)} = X_F^{(i)}$$

to koniec, w przeciwnym przypadku ($i = i+1$, idź do kroku 2),

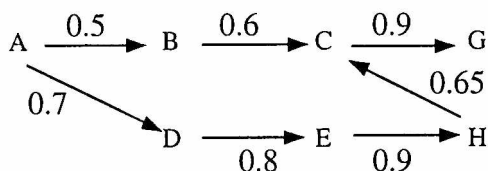
gdzie $\text{Dom}(X_F^{(i)}) = \{A \mid (A, \theta) \in X_F^{(i)}\}$ oraz $\cup$ jest operatorem sumy zbiorów rozmytych.

Należy zauważyć, że $X_F^{(i)}$ jest zbiorem rozmytym, gdzie (A, θ) wyraża, że maksymalny stopień rozmytej zależności funkcyjnej A od X równa się θ . $X_F^{(i)}$ jest konstruowany przez operacje sumy zbiorów rozmytych $\cup$.

Algorytm 1 jest podobny do algorytmu klasycznego [ULLMAN 88]. Różnica polega na tym, że w algorytmie 1 za każdym razem, kiedy obliczany jest B_i , musi

zostać zastosowany operator min dla stopni α i elementów w V (krok 2 linia algorytmu 1). Jest to związane z wprowadzeniem rozmytości.

W celu zaprezentowania działania algorytmu 1 rozważmy przykład: można obliczyć domknięcie A_F^+ dla $U = \{A, B, C, D, E, G, H\}$, $|U| = 7$, $F = \{A \rightarrow_{0.5} B, B \rightarrow_{0.6} C, C \rightarrow_{0.9} G, A \rightarrow_{0.7} D, D \rightarrow_{0.8} E, E \rightarrow_{0.9} H, H \rightarrow_{0.65} C\}$. Diagram zależności względem F przedstawia rysunek 2.



Rys. 2. Diagram zależności

Obliczenie A_F^+ :

$$A_F^{(0)} = \{(A, 1)\}$$

$$B_F^{(0)} = \{(B, 0.5), (D, 0.7)\}$$

$$A_F^{(1)} = A_F^{(0)} \cup B_F^{(0)} = \{(A, 1), (B, 0.5), (D, 0.7)\}$$

$$B_F^{(1)} = \{(B, 0.5), (D, 0.7), (\underline{C, 0.5}), (\underline{E, 0.7})\}$$

$$A_F^{(2)} = A_F^{(1)} \cup B_F^{(1)} = \{(A, 1), (B, 0.5), (D, 0.7), (C, 0.5), (E, 0.7)\}$$

$$B_F^{(2)} = \{(B, 0.5), (D, 0.7), (C, 0.5), (E, 0.7), (\underline{H, 0.7}), (\underline{G, 0.5})\}$$

$$A_F^{(3)} = A_F^{(2)} \cup B_F^{(2)} = \{(A, 1), (B, 0.5), (C, 0.5), (D, 0.7), (E, 0.7), (G, 0.5), (H, 0.7)\}$$

$$B_F^{(3)} = \{(B, 0.5), (C, 0.5), (\underline{C, 0.65}), (D, 0.7), (E, 0.7), (H, 0.7), (G, 0.5)\}$$

$$A_F^{(4)} = A_F^{(3)} \cup B_F^{(3)} = \{(A, 1), (B, 0.5), (C, 0.65), (D, 0.7), (E, 0.7), (G, 0.5), (H, 0.7)\}$$

$$B_F^{(4)} = \{(B, 0.5), (C, 0.5), (C, 0.65), (D, 0.7), (E, 0.7), (H, 0.7), (\underline{G, 0.65})\}$$

$$A_F^{(5)} = A_F^{(4)} \cup B_F^{(4)} = \{(A, 1), (B, 0.5), (C, 0.65), (D, 0.7), (E, 0.7), (G, 0.65), (H, 0.7)\}$$

$$B_F^{(5)} = \{(B, 0.5), (C, 0.5), (C, 0.65), (D, 0.7), (E, 0.7), (H, 0.7), (G, 0.65)\}$$

$$A_F^{(6)} = A_F^{(5)} \cup B_F^{(5)} = \{(A, 1), (B, 0.5), (C, 0.65), (D, 0.7), (E, 0.7), (G, 0.65), (H, 0.7)\}$$

$$= A_F^{(5)}$$

$\Rightarrow$ Koniec.

2.6. RÓWNOWAŻNOŚĆ ZBIORÓW ROZMYTYCH ZALEŻNOŚCI FUNKCYJNYCH

Niech R będzie schematem relacji. Dwa zbiory rozmytych zależności funkcyjnych są równoważne (F pokrywa G i G pokrywa F) dla R wtedy i tylko wtedy, gdy $F^+ = G^+$.

Z pojęciem równoważności zbiorów zależności związane są bezpośrednio dwie idee. Pierwsza dotyczy zagadnienia, czy kolekcja rozmytych zależności funkcyjnych dla schematów uzyskanych w rozkładzie jest równoważna zbiorowi rozmytych

zależności funkcyjnych dla schematu oryginalnego. Temat ten zostanie omówiony w rozdziale 3. Druga zaś, przedstawiona w bieżącym podrozdziale, dotyczy zagadnienia, czy dla zadanego schematu R i zbioru rozmytych zależności funkcyjnych istnieje minimalny zbiór rozmytych zależności funkcyjnych G równoważny z F . Uzyskanie minimalnego zbioru jest użyteczne, gdy rozwija się teorię projektowania schematów relacyjnych baz danych.

Zbiór rozmytych zależności funkcyjnych F jest minimalny, jeżeli:

- 1) każda prawa strona zależności należącej do F jest pojedynczym atrybutem,
- 2) dla żadnej zależności $X \rightarrow_{\alpha} A$ ze zbioru F , zbiór $F - \{X \rightarrow_{\alpha} A\}$ nie jest równoważny ze zbiorem F ,
- 3) dla żadnej zależności $X \rightarrow_{\alpha} A$ ze zbioru F i zbioru Z będącego podzbiorem właściwym F , zbiór $(F - \{X \rightarrow_{\alpha} A\}) \cup \{Z \rightarrow_{\alpha} A\}$ nie jest równoważny z F .

Intuicyjnie warunek 2 zapewnia, że w zbiorze F żadna zależność nie jest nadmiarowa, a warunek 3 gwarantuje, że żaden atrybut w żadnej z lewych stron nie jest nadmiarowy. Z warunku 1 wynika, że po każdej prawej stronie znajduje się tylko jeden atrybut, jest więc oczywiste, że żaden atrybut występujący po prawej stronie nie jest nadmiarowy.

Na przykład niech $F_1 = \{A \rightarrow_{\alpha} B, A \rightarrow_{\beta} B\}$, gdzie $\beta \leq \alpha$, $F_2 = \{A \rightarrow_{\alpha} B, B \rightarrow_{\beta} C, A \rightarrow_{\chi} C\}$, gdzie $\chi < \min(\alpha, \beta)$, $F_3 = \{AB \rightarrow_{\alpha} C, A \rightarrow_{\alpha} B, B \rightarrow_{\alpha} A\}$. F_1 nie jest minimalny zgodnie z punktem 2, ponieważ $F_1 - \{A \rightarrow_{\beta} B\} = \{A \rightarrow_{\alpha} B\}$ jest równoważny z F_1 , tj. $F_1^+ = \{A \rightarrow_{\alpha} B\}^+$. Podobnie F_2 nie jest minimalny zgodnie z punktem 2, ponieważ $F_2 - \{A \rightarrow_{\chi} C\} = \{A \rightarrow_{\alpha} B, B \rightarrow_{\beta} C\}$ jest równoważny z F_2 , tj. $F_2^+ = \{A \rightarrow_{\alpha} B, B \rightarrow_{\beta} C\}^+$. F_3 nie jest minimalny zgodnie z punktem 3, ponieważ $(F_3 - \{AB \rightarrow_{\alpha} C\}) \cup \{A \rightarrow_{\alpha} C\} = \{A \rightarrow_{\alpha} C, A \rightarrow_{\alpha} B, B \rightarrow_{\alpha} A\}$ jest równoważny z F_3 , tj. $F_3^+ = \{A \rightarrow_{\alpha} C, A \rightarrow_{\alpha} B, B \rightarrow_{\alpha} A\}^+$. Ze względu na to, iż F jest zbiorem skończonym i definicja minimalnego zbioru rozmytych zależności funkcyjnych jest konstruktywna, można oczekiwać, że zbiór minimalny zawsze istnieje i może być uzyskany przez usunięcie nadmiarowych zależności.

Każdy zbiór rozmytych zależności funkcyjnych F jest równoważny minimalnemu zbiorowi rozmytych zależności funkcyjnych G (F ma minimalne pokrycie G). Minimalne pokrycie nie musi być unikalne, tzn. zbiór rozmytych zależności funkcyjnych może mieć kilka minimalnych pokryć. Na przykład niech $F = \{AB \rightarrow_{\alpha} C, A \rightarrow_{\alpha} B, B \rightarrow_{\alpha} A\}$. Minimalnymi pokryciami F mogą być zbiory $G = \{A \rightarrow_{\alpha} C, A \rightarrow_{\alpha} B, B \rightarrow_{\alpha} A\}$ i $G' = \{B \rightarrow_{\alpha} C, A \rightarrow_{\alpha} B, B \rightarrow_{\alpha} A\}$, ponieważ $F^+ = G^+$ i $F^+ = G'^+$.

Dla zaprezentowania obliczania minimalnego pokrycia zbioru rozmytych zależności funkcyjnych można rozważyć zbiór $F = \{AB \rightarrow_{0.95} C, C \rightarrow_{0.9} A, BC \rightarrow_{0.1} D, ACD \rightarrow_{0.85} B, D \rightarrow_{0.9} EH, BE \rightarrow_{0.9} C, CH \rightarrow_{0.8} BD, CE \rightarrow_{0.8} AH\}$. Obliczenie minimalnego pokrycia F może mieć następujący przebieg:

Niech $G = \{AB \rightarrow_{0.95} C, C \rightarrow_{0.9} A, BC \rightarrow_{0.1} D, ACD \rightarrow_{0.85} B, D \rightarrow_{0.9} E, D \rightarrow_{0.9} H, BE \rightarrow_{0.9} C, CH \rightarrow_{0.8} B, CH \rightarrow_{0.8} D, CE \rightarrow_{0.8} A, CE \rightarrow_{0.8} H\}$.

Ponieważ $C \rightarrow_{0.9} A$ implikuje $CE \rightarrow_{0.8} A$, stąd $CE \rightarrow_{0.8} A$ jest nadmiarowa. $CH \rightarrow_{0.8} B$ jest również nadmiarowa, ponieważ $CH \rightarrow_{0.8} D$, $C \rightarrow_{0.9} A$ i $ACD \rightarrow_{0.85} B$ implikuje $CH \rightarrow_{0.8} B$. Usuając nadmiarowe zależności zostaje otrzymany następujący zbiór G :

$G = \{AB \rightarrow_{0.95} C, C \rightarrow_{0.9} A, BC \rightarrow_1 D, ACD \rightarrow_{0.85} B, D \rightarrow_{0.9} E, D \rightarrow_{0.9} H, BE \rightarrow_{0.9} C, CH \rightarrow_{0.8} D, CE \rightarrow_{0.8} H\}$.

Następnie, ponieważ $ACD \rightarrow_{0.85} B$ i $C \rightarrow_{0.9} A$ implikuje $CD \rightarrow_{0.85} B$ i $CD \rightarrow_{0.85} B$ implikuje $ACD \rightarrow_{0.85} B$, stąd $ACD \rightarrow_{0.85} B$ w G może zostać zastąpione $CD \rightarrow_{0.85} B$.

Otrzymuje się zatem minimalne pokrycie F postaci:

$G = \{AB \rightarrow_{0.95} C, C \rightarrow_{0.9} A, BC \rightarrow_1 D, CD \rightarrow_{0.85} B, D \rightarrow_{0.9} E, D \rightarrow_{0.9} H, BE \rightarrow_{0.9} C, CH \rightarrow_{0.8} D, CE \rightarrow_{0.8} H\}$.

3. ROZKŁAD SCHEMATÓW RELACJI

Rozkładem schematu relacji R nazywa się układ $\rho = (R_1, R_2, \dots, R_k)$ złożony z podzbiorów R takich, że

$$R = R_1 \cup R_2 \cup \dots \cup R_k$$

Nie jest wymagane, aby R_i były rozłączne.

Rozkład schematów relacji opiera się na operacjach projekcji i złączenia bazujących na relacji identyczności, tzn. uwzględniających, że dane są albo identyczne, albo różne, nie mówi się tu już o ich bliskości. Rozważając rozkłady, mówi się o rozkładach odwracalnych i rozkładach zachowujących zależności.

3.1. ROZKŁADY ODWRACALNE

Własność odwracalności dla rozkładu schematu otrzymanego przez projekcję oznacza, że oryginalna relacja schematu może być zrekonstruowana przez złączenie z relacji schematów uzyskanych w rozkładzie. Bardziej formalnie, niech $R(U)$ będzie schematem relacji, F niech będzie zbiorem rozmytych zależności funkcyjnych dla R , $\rho = (R_1, R_2, \dots, R_k)$ niech będzie rozkładem schematu, gdzie R_i jest schematem $R_i(U_i)$ i $U = U_1 \cup U_2 \cup \dots \cup U_k$. Mówi się, że rozkład ρ jest rozkładem odwracalnym względem F wtedy i tylko wtedy, gdy dla każdej relacji R schematu relacji R spełniającej F , $R = m_\rho(R)$, gdzie $m_\rho(R) = \Pi_{U_1}(R) * \Pi_{U_2}(R) * \dots * \Pi_{U_k}(R)$.

Jeżeli R jest schematem relacji, $\rho = (R_1, R_2, \dots, R_k)$ jest rozkładem schematu R , $R \in \mathbf{R}$ i $R_i = \Pi_{U_i}(R)$, to zachodzą trzy podstawowe zależności:

- $R \subseteq m_\rho(R)$
- Jeżeli $S = m_\rho(R)$, to $\Pi_{U_i}(S) = R_i$
- $m_\rho(m_\rho(R)) = m_\rho(R)$.

Odwracalność rozkładu jest badana względem zbioru rozmytych zależności funkcyjnych. Ponieważ projekcja i złączenie odnoszą się tylko do identycznych (lub

różnych) elementów, rozważane w tym kontekście zależności $X \rightarrow_0 Y$ opierają się również na identyczności ($X \rightarrow_{id} Y$), tzn. że identycznym elementom X odpowiadają identyczne elementy Y . Zależności funkcyjne oparte na identyczności mogą być wyprowadzone z podanego zbioru rozmytych zależności funkcyjnych zgodnie z faktem, że dla $\alpha > 0$, $X \rightarrow_\alpha Y$ implikuje $X \rightarrow_{id} Y$. Badanie odwracalności rozkładu przedstawia algorytm 2.

Algorytm 2. Wejście: Schemat relacji R z $U = \{A_1, A_2, \dots, A_n\}$, zbiór zależności funkcyjnych opartych na identyczności uzyskanych z podanego zbioru zależności funkcyjnych F , dekompozycja $\rho = (R_1, R_2, \dots, R_k)$

Metoda: Konstruujemy tablicę o n kolumnach i k wierszach, j -ta kolumna odpowiada atrybutowi A_j , a i -ty wiersz odpowiada schematowi relacji R_i . Jeśli A_j jest atrybutem R_i , do i -tego wiersz i j -tej kolumny wstawiamy symbol a_j . Jeśli nie, to wstawiamy tam b_{ij} .

Każdą z zależności $X \rightarrow_{id} Y$ z F „rozważa” się wielokrotnie dopóty, dopóki w tablicy nie można dokonać żadnej zmiany. Za każdym razem gdy „rozważa” się zależność $X \rightarrow_{id} Y$, poszukuje się wierszy zgodnych we wszystkich kolumnach odpowiadających atrybutom X . Jeśli zostaną znalezione dwa takie wiersze, zrównuje się w nich symbole odpowiadające atrybutom Y . Gdy zrównuje się dwa symbole, to jeśli jednym z nich jest a_j , drugi też staje się a_j . Jeśli były nimi b_{ij} i b_{lj} , oba stają się równe bądź b_{ij} , bądź b_{lj} . Gdy po zmodyfikowaniu wierszy tablicy w powyższy sposób okaże się, że pewien wiersz przybrał postać $a_1 \dots a_n$, wówczas rozkład jest odwracalny. W przypadku przeciwnym jest to rozkład nieodwracalny.

Algorytm 2 jest podobny do algorytmu sprawdzania odwracalności rozkładu w klasycznych bazach danych, różni się tylko tym, że zamiast zależności funkcyjnych występują zależności funkcyjne oparte na identyczności oraz wartości atrybutów i elementy dziedzin mogą być rozmyte. W tabeli symbole a_i , b_{ij} , itd. reprezentują rozkład możliwości, ale traktowane są jako wartości identyczne albo różne.

W celu zaprezentowania działania algorytmu 2 można zbadać odwracalność rozkładu $\rho = \{R_1, R_2\}$ będącego rozkładem schematu R , gdzie $U = \{A, B, C\}$, $U_1 = \{A, B\}$, $U_2 = \{B, C\}$, odwracalność można zbadać względem dwóch zbiorów rozmytych zależności funkcyjnych $F_1 = \{B \rightarrow_\alpha A\}$, $\alpha > 0$ i $F_2 = \{A \rightarrow_\beta B\}$, $\beta > 0$. Tabelę początkową przedstawia tabela 3.

Tabela 3. Tabela początkowa dla badania odwracalności rozkładu ρ

	A	B	C
R_1	a_1	a_2	b_{13}
R_2	b_{21}	a_2	a_3

Stosując algorytm 2 dla $\mathbf{R}$, ρ względem F_1 , b_{21} staje się a_1 zgodnie z $B \rightarrow_\alpha A$, otrzymuje się wynik pokazany w tabeli 4. Ponieważ uzyskana tabela posiada wiersz postaci $a_1 \dots a_3$, rozkład jest odwracalny.

Tabela 4. Badanie odwracalności rozkładu ρ względem F_1

	A	B	C
$\mathbf{R}_1$	a_1	a_2	b_{13}
$\mathbf{R}_2$	a_1	a_2	a_3

Stosując algorytm 2 dla $\mathbf{R}$, ρ względem F_2 , żadna zmiana nie jest wprowadzana do oryginalnej tabeli (tabela 3). ρ jest dekompozycją nieodwracalną, ponieważ nie posiada żadnego wiersza mającego postać $a_1 \dots a_3$.

3.2. ROZKŁADY ZACHOWUJĄCE ZALEŻNOŚCI FUNKCYJNE

Ważną własnością rozkładu schematu, obok jego odwracalności, jest zachowanie zależności funkcyjnych. Własność ta wymaga zachowania informacji o semantyce danych wyrażonej przez rozmyte zależności funkcyjne. Oznacza ona, że zbiór rozmytych zależności funkcyjnych dla oryginalnego schematu jest przykrywany przez kolekcję zbiorów rozmytych zależności funkcyjnych dla schematów uzyskanych w rozkładzie. Bardziej formalnie, niech $\mathbf{R}$ będzie schematem relacji, $\rho = (\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_k)$ będzie rozkładem schematu $\mathbf{R}$, F będzie zbiorem rozmytych zależności funkcyjnych dla $\mathbf{R}$. Niech $\Pi_{U_i}(F)$ będzie projekcją F na U_i (zbiór atrybutów $\mathbf{R}_i$), tj. $\Pi_{U_i}(F) = \{X \rightarrow_\alpha Y \mid X \rightarrow_\alpha Y \in F^+ \text{ i } XY \subseteq U_i\}$. Mówi się, że rozkład ρ zachowuje F wtedy i tylko wtedy, gdy suma wszystkich rozmytych zależności funkcyjnych w $\Pi_{U_i}(F)$ dla $i = 1, 2, \dots, k$ implikuje logicznie wszystkie rozmyte zależności funkcyjne w F . Inaczej mówiąc, ρ jest rozkładem zachowującym zależności funkcyjne względem F wtedy i tylko wtedy, gdy G pokrywa F , gdzie

$$G = \Pi_{U_1}(F) \cup \Pi_{U_2}(F) \cup \dots \cup \Pi_{U_k}(F).$$

Na przykład mając $U = \{A, B, C\}$, $U_1 = \{A, B\}$, $U_2 = \{B, C\}$ dla schematów odpowiednio $\mathbf{R}$, $\mathbf{R}_1$, $\mathbf{R}_2$, rozkład $\rho = (\mathbf{R}_1, \mathbf{R}_2)$ będący rozkładem schematu $\mathbf{R}$ i zbiór rozmytych zależności funkcyjnych $F_1 = \{A \rightarrow_\alpha B, B \rightarrow_\beta C\}$ dla $\mathbf{R}$, otrzymuje się, że $\{A \rightarrow_\alpha B\} \subseteq \Pi_{U_1}(F_1)$ i $\{B \rightarrow_\beta C\} \subseteq \Pi_{U_2}(F_1)$. Stąd $F_1 = \{A \rightarrow_\alpha B, B \rightarrow_\beta C\} = \{A \rightarrow_\alpha B\} \cup \{B \rightarrow_\beta C\} \subseteq \Pi_{U_1}(F_1) \cup \Pi_{U_2}(F_1) = G_1$, co oznacza, że G_1 pokrywa F_1 , a zatem ρ zachowuje zbiór zależności F_1 . Rozważając zbiór zależności $F_2 = \{B \rightarrow_\phi A, A \rightarrow_\chi C\}$, ρ nie jest rozkładem zachowującym zależności względem F_2 , ponieważ $A \rightarrow_\chi C$ nie wynika logicznie z $G_2 = \Pi_{U_1}(F_2) \cup \Pi_{U_2}(F_2)$.

Aby sprawdzić, czy rozkład zachowuje zbiór zależności funkcyjnych F , można obliczyć F^+ i dokonać projekcji na wszystkie $\mathbf{R}_i$, a następnie wziąć sumę powstałych w ten sposób zbiorów zależności i zbadać, czy zbiór ten pokrywa F . W praktyce obliczenie zbioru F^+ jest zadaniem niebagatelnym, istnieje jednak inna metoda

sprawdzająca zachowanie zależności, mniej złożona czasowo, obliczająca X^+_G dla każdego $X \rightarrow_\phi Y$ należącego do F . Jeżeli $Y = A_1 A_2 \dots A_r$ i $\{(A_i, \phi)\} \subseteq X^+_G$, gdzie $i = 1, \dots, r$, to wiadomo, że $X \rightarrow_\phi Y \in G^+$. Jeżeli dla każdego $X \rightarrow_\phi Y \in F$, $X \rightarrow_\phi Y \in G^+$, to $F \subseteq G^+$, co oznacza, że G pokrywa F . Badanie zachowania zależności przedstawia algorytm 3.

Algorytm 3.

Wejście: schemat relacji R , dekompozycja schematu R : $\rho = (R_1, R_2, \dots, R_k)$ i zbiór zależności funkcyjnych F dla R

Wyjście: rozstrzygnięcie, czy ρ zachowuje F .

Metoda: Niech $U_1, U_2, \dots, U_k$ będą zbiorami atrybutów dla odpowiednio $R_1, R_2, \dots, R_k$.

1) Oblicz X^+_G , jeżeli $X = X_1 X_2 \dots X_m$ jest lewą stroną zależności funkcyjnej należącej do F :

$Z = \{(X_1, 1), (X_2, 1), \dots, (X_m, 1)\}$

while następują zmiany w Z do // wykonanie jednego kroku

for $i = 1$ to k do

begin

$T_i = \{(A_i, \phi) \mid (A_i, \phi) \in (Z_{R_i})^+_F \text{ i } A \in U_i\}$

$Z = Z \cup T_i$

end

gdzie $Z_{R_i} = \{A \mid (A_i, \phi) \in Z \text{ i } A \in U_i\}$, $(Z_{R_i})^+_F$ jest otrzymywany stosując algorytm obliczania domknięć (algorytm 1) względem F tych atrybutów, które są zarówno w Z , jak i w U_i . Zbiorem początkowym dla obliczenia $(Z_{R_i})^+_F$ jest

$(Z_{R_i})^0_F = \{(A, \phi) \mid (A, \phi) \in Z \text{ i } A \in U_i\}$.

2) Jeżeli dla każdego $X \rightarrow_\phi Y$ należącego do F , $X = X_1 X_2 \dots X_m$, $Y = Y_1 Y_2 \dots Y_r$ otrzymujemy $\{(Y_1, \phi), (Y_2, \phi), \dots, (Y_r, \phi)\} \subseteq X^+_G$, to G pokrywa F , tj. ρ zachowuje F , w przeciwnym przypadku G nie pokrywa F , tj. ρ nie zachowuje F .

W celu zaprezentowania działania algorytmu 3 można zbadać zachowanie zależności przez rozkład $\rho = \{AB, BC\}$, gdzie $U = \{A, B, C\}$, $F = \{A \rightarrow_{0.9} B, B \rightarrow_{0.8} C\}$, T_i z algorytmu 3 będzie oznaczane przez $\{(Z_{R_i})^0_F\}^+|_{U_i}$.

Najpierw należy obliczyć A^+_G : Na początku $Z = \{(A, 1)\}$

W kroku pierwszym:

dla AB :

$Z = Z \cup \{(A, 1)\}^+|_{AB}$

$Z = Z \cup \{(A, 1), (B, 0.9), (C, 0.8)\}|_{AB}$

$Z = Z \cup \{(A, 1), (B, 0.9)\} = \{(A, 1), (B, 0.9)\}$

dla BC :

$Z = Z \cup \{(B, 0.9)\}^+|_{BC}$

$Z = Z \cup \{(B, 0.9), (C, 0.8)\}|_{BC}$

$Z = Z \cup \{(B, 0.9), (C, 0.8)\} = \{(A, 1), (B, 0.9), (C, 0.8)\}$

W kroku drugim:

dla AB:

$$Z = Z \cup \{(A, 1), (B, 0.9)\}^+|_{AB}$$

$$Z = Z \cup \{(A, 1), (B, 0.9), (C, 0.8)\}|_{AB}$$

$$Z = Z \cup \{(A, 1), (B, 0.9)\} = Z$$

dla BC:

$$Z = Z \cup \{(B, 0.9), (C, 0.8)\}^+|_{BC}$$

$$Z = Z \cup \{(B, 0.9), (C, 0.8)\}|_{BC}$$

$$Z = Z \cup \{(B, 0.9), (C, 0.8)\} = Z$$

Ponieważ krok drugi nie wprowadza żadnych zmian do Z , zostaje osiągnięty warunek stopu. $A^+_G = \{(A, 1), (B, 0.9), (C, 0.8)\}$, co oznacza, że $A \rightarrow_{0.9} B$ należy do G^+ .

Następnie należy obliczyć B^+_G : Na początku $Z = \{(B, 1)\}$

Dla kroku pierwszego:

dla AB:

$$Z = Z \cup \{(B, 1)\}^+|_{AB}$$

$$Z = Z \cup \{(B, 1), (C, 0.8)\}|_{AB}$$

$$Z = Z \cup \{(B, 1)\} = \{(B, 1)\}$$

dla BC:

$$Z = Z \cup \{(B, 1)\}^+|_{BC}$$

$$Z = Z \cup \{(B, 1), (C, 0.8)\}|_{BC}$$

$$Z = Z \cup \{(B, 1), (C, 0.8)\} = \{(B, 1), (C, 0.8)\}$$

Dla kroku drugiego:

dla AB:

$$Z = Z \cup \{(B, 1)\}^+|_{AB}$$

$$Z = Z \cup \{(B, 1), (C, 0.8)\}|_{AB}$$

$$Z = Z \cup \{(B, 1)\} = Z$$

dla BC:

$$Z = Z \cup \{(B, 1), (C, 0.8)\}^+|_{BC}$$

$$Z = Z \cup \{(B, 1), (C, 0.8)\}|_{BC}$$

$$Z = Z \cup \{(B, 1), (C, 0.8)\} = Z$$

$B^+_G = \{(B, 1), (C, 0.8)\}$, co oznacza, że $B \rightarrow_{0.8} C$ należy do G^+ .

Ponieważ $F \subseteq G^+$, G pokrywa F i ρ zachowuje zależności funkcyjne F .

Dla pewnych szczególnych zbiorów zależności funkcyjnych istnieją warunki, których spełnienie gwarantuje zachowanie zależności przez odpowiedni rozkład. Jeżeli zbiór zależności funkcyjnych F tworzy pętlę (cykl):

$$F = \{X_1 \rightarrow_{\alpha_1} X_2, X_2 \rightarrow_{\alpha_2} X_3, \dots, X_{m-1} \rightarrow_{\alpha_{m-1}} X_m, X_m \rightarrow_{\alpha_m} X_1\}$$

i $\rho = (R_1, R_2, \dots, R_{m-1})$ jest rozkładem schematu R , gdzie $U = \{X_1, X_2, \dots, X_m\}$, U_i schematu R_i jest zbiorem $\{X_i, X_{i+1}\}$, $i = 1, 2, \dots, m-1$, to ρ zachowuje zależności F wtedy i tylko wtedy, gdy $\alpha_m = \min(\alpha_1, \alpha_2, \dots, \alpha_m)$.

Jeżeli zbiór F jest zbiorem składającym się z zależności $X_1 \rightarrow_{\alpha_m} X_m$ i „ścieżki” $X_1 \rightarrow_{\alpha_1} X_2, X_2 \rightarrow_{\alpha_2} X_3, \dots, X_{m-1} \rightarrow_{\alpha_{m-1}} X_m$, czyli

$F = \{X_1 \rightarrow_{\alpha_m} X_m, X_1 \rightarrow_{\alpha_1} X_2, X_2 \rightarrow_{\alpha_2} X_3, \dots, X_{m-1} \rightarrow_{\alpha_{m-1}} X_m\}$,
wówczas ρ zachowuje zależności F wtedy i tylko wtedy, gdy
 $\alpha_m \leq \min(\alpha_1, \alpha_2, \dots, \alpha_{m-1})$.

4. ZAGADNIENIE ANOMALII

Anomalie wynikają z istnienia dwóch rodzajów związków między atrybutami. Jednym z nich jest zależność częściowa, zdefiniowana w podrozdziale 1.2, drugą zaś jest przechodnia zależność funkcyjna. Niech $X, Y, Z \subseteq U$, Y nie zawiera się w X (nie jest podzbiorem właściwym X). Mówi się, że Z przechodnie zależny funkcyjnie od X w stopniu θ wtedy i tylko wtedy, gdy $X \rightarrow_{\alpha} Y$, $Y \rightarrow_{\beta} Z$ i $\min(\alpha, \beta) \geq \theta$.

Jeżeli istnieje zależność częściowa $Y \rightarrow_{\alpha} A$, gdzie X jest kluczem, a Y jest podzbiorem właściwym X , to w każdej krotce użytej do skojarzenia wartości dla X z wartościami innych atrybutów muszą pojawić się te same skojarzenia wartości między Y i A . Sytuację tę można zaobserwować na przykładzie schematu

Zajęcia(IdStudenta, Nazwisko, Wiek, Kurs, Ocena),
reprezentującego informację, że dani studenci wybrali określone kursy i uzyskali z nich odpowiednie oceny. Zależność

$\text{IdStudenta} \rightarrow_{0,8} \text{Wiek}$ ($\text{IdStudenta} \rightarrow_1 \text{NazwaStudenta}$)

jest zależnością częściową i nazwisko studenta (jak również jego wiek) musi być wprowadzane dla każdego kursu wybranego przez studenta. Problem ten może zostać rozwiązany poprzez rozkład owego schematu złożony z dwóch schematów: Studenti(IdStudenta, Nazwisko, Wiek) i Wybiera(IdStudenta, Kurs, Ocena).

Jeżeli istnieje zależność przechodnia $X \rightarrow_{\alpha} Y \rightarrow_{\beta} A$, to z wartością atrybutu ze zbioru X nie można skojarzyć wartości atrybutu ze zbioru Y , chyba że istnieje wartość A skojarzona z wartością atrybutu z Y . Taka sytuacja prowadzi do powstawania anomalii przy wstawianiu i usuwaniu. Gdy skojarzenia X z Y nie można wstawić bez skojarzenia Y z A oraz gdy usunie się wartość A skojarzoną z daną wartością dla Y , można stracić ślad po skojarzeniu X z Y . Na przykład w schemacie relacji

Usytuowanie(NrWydziału, NrBudynku, LPomBudynku),

przedstawiającego informacje, że dane wydziały znajdują się w określonych budynkach mających odpowiednią liczbę pomieszczeń, o zależnościach

$\text{NrWydziału} \rightarrow_1 \text{NrBudynku}$,

$\text{NrBudynku} \rightarrow_{0,9} \text{LPomBudynku}$,

nie daje się zapisać informacji, że dany wydział znajduje się w określonym budynku bez podania liczby pomieszczeń. Rozwiązaniem tego problemu może być wprowadzenie zamiast Usytuowanie rozkładu złożonego z

Wydziały(NrWydziału, NrBudynku)

i

Budynki(NrBudynku, LPomBudynku).

5. NORMALIZACJA

Anomalii można uniknąć, jeżeli schematy relacji będą spełniały pewne ograniczenia na atrybutach – będą schematami w pewnej postaci normalnej. Opierając się na pojęciu rozmytej zależności funkcyjnej, klasyczne postacie normalne mogą być rozszerzone i uogólnione dla rozmytego modelu relacyjnych baz danych. Otrzymuje się wówczas rozmytą pierwszą postać normalną (F1NF, ang. *fuzzy normal form*), nakładającą ograniczenie eliminujące powtarzanie grup wartości atrybutów, θ -rozmytą drugą postać normalną (θ -F2NF), wykluczającą zależności częściowe θ -niekluczowych atrybutów od θ -kluczy, θ -rozmytą trzecią postać normalną (θ -F3NF), wykluczającą zależności przechodnie θ -niekluczowych atrybutów od θ -kluczy i θ -rozmytą postać normalną Boyce'a-Codda (θ -FBCNF), nie pozwalającą na częściową lub przechodnią rozmytą zależność funkcyjną θ -kluczowego atrybutu od θ -klucza nie zawierającego go.

5.1. ROZMYTA PIERWSZA POSTAĆ NORMALNA

Niech D_i będzie dziedziną atrybutu A_i . Mówi się, że schemat relacji R jest w **rozmytej pierwszej postaci normalnej** (F1NF) wtedy i tylko wtedy, gdy dla każdej relacji R należącej do R każda wartość atrybutu, π_{A_i} , jest wykluczającym rozkładem możliwości.

Wykluczający rozkład możliwości jest rozkładem możliwości, w którym elementy należące do D_i wzajemnie się wykluczają. Inaczej mówiąc, rozkład możliwości $\pi = \{\alpha_1/x_1, \alpha_2/x_2, \dots, \alpha_m/x_m\}$ jest wykluczającym rozkładem możliwości, jeśli wyraża, że x_1 jest wartością A_i w stopniu α_1 albo x_2 jest wartością A_i w stopniu α_2 , ..., albo x_m jest wartością A_i w stopniu α_m .

Czasami można się spotkać z rozkładem $\pi = \{\alpha_1/x_1, \alpha_2/x_2, \dots, \alpha_m/x_m\}$ nie będącym wykluczającym rozkładem możliwości. Rozkład taki wyraża, że x_1 jest wartością A_i w stopniu α_1 i x_2 jest wartością A_i w stopniu α_2 , ..., i x_m jest wartością A_i w stopniu α_m . Atrybut A_i jest wówczas traktowany jak atrybut wielowartościowy. Aby móc reprezentować taką informację w rozmytej bazie danych, rozkład będący jego wartością musi zostać przekształcony na kilka wykluczających rozkładów możliwości, na przykład $\pi_i = \{\alpha_i/x_i\}$.

Rozważając schemat relacji $R_1(\text{Nazwa}, \text{Wiek}, \text{JęzykObcy})$, relacja R_1 będąca wartością R_1 może mieć postać pokazaną w tabeli 5.

Krotka t_1 przedstawia informacje, że Nowak ma 21 lat i zna Angielski (A), krotka t_2 reprezentuje fakt, że Anioł ma około 16 albo 17 albo 18 lat i zna Angielski (A) i Francuski (F) i Niemiecki (G) w stopniach odpowiednio 1.0, 0.3 i 0.6, a krotka t_3 przedstawia, że Kowalski ma 25 albo 26 lat i zna język orientalny (jeden z języków: Japoński, Chiński, Koreański albo innego podobnego do nich). R_1 nie jest w rozmytej

pierwszej postaci normalnej ponieważ $\{1/E, 0.3/F, 0.6/G\}$ reprezentuje fakt, że Anioł zna trzy języki.

Tabela 5. Relacja R_1 nie będąca w F1NF

	Nazwa	Wiek	JęzykObcy
t_1	Nowak	21	E
t_2	Anioł	$\{0.7/16, 1/17, 0.8/18\}$	$\{1/E, 0.3/F, 0.6/G\}$
t_3	Kowalski	$\{25, 26\}$	Język orientalny

Aby został spełniony warunek F1NF, R_1 musi zostać przekształcona w R_2 pokazaną w tabeli 6.

Tabela 6. Relacja R_1 po sprowadzeniu do F1NF

	Nazwisko	Wiek	JęzykObcy
t_1	Nowak	21	E
t_2	Anioł	$\{0.7/16, 1/17, 0.8/18\}$	$\{1/E\}$
t'_2	Anioł	$\{0.7/16, 1/17, 0.8/18\}$	$\{0.3\}$
t''_2	Anioł	$\{0.7/16, 1/17, 0.8/18\}$	$\{0.6/G\}$
t_3	Kowalski	$\{25, 26\}$	Język orientalny

5.2. θ -ROZMYTE POSTACIE NORMALNE

Omówimy teraz θ -rozmyte postacie normalne, które są oparte na rozmytych zależnościach funkcyjnych i θ -kluczach.

Niech F będzie zbiorem rozmytych zależności funkcyjnych dla schematu R i K będzie θ -kluczem schematu R .

Mówi się, że R jest w θ -rozmytej drugiej postaci normalnej (θ -F2NF) wtedy i tylko wtedy, gdy R jest w F1NF i dla każdego θ -niekluczowego atrybutu A , A w pełni funkcyjnie zależy od K w stopniu α ($K \rightarrow_{\alpha} A \in F^+$), gdzie $0 < \alpha \leq \theta$.

θ -F2NF uniemożliwia, aby właściwy podzbiór θ -klucza wyznaczał θ -niekluczowy atrybut A . Ograniczenie to nie pozwala na częściową zależność θ -niekluczowego atrybutu od θ -klucza.

Na przykład, niech $R(\text{Nazwisko}, \text{Wiek}, \text{JęzykObcy})$, $F = \{\text{Nazwisko} \rightarrow_{0.9} \text{Wiek}\}$. Ponieważ $(\text{Nazwisko}, \text{JęzykObcy}) \rightarrow_{0.9} \text{Wiek}$, $(\text{Nazwisko}, \text{JęzykObcy}) \rightarrow_{1.0} \text{Nazwisko}$, $(\text{Nazwisko}, \text{JęzykObcy}) \rightarrow_{1.0} \text{JęzykObcy}$, stąd $(\text{Nazwisko}, \text{JęzykObcy})$ jest 0.9-kluczem. R nie jest w 0.9-F2NF, gdyż Wiek nie jest w pełni zależny od $(\text{Nazwisko}, \text{JęzykObcy})$.

Można rozważyć również inny przykład, niech $R = (A, B, C, D)$ i $F = \{AB \rightarrow_{0.9} C, AC \rightarrow_{0.8} D\}$. Ponieważ $AB \rightarrow_{0.9} C$, stąd $AB \rightarrow_{0.9} AC$. Dodatkowo ponieważ $AC \rightarrow_{0.8} D$, otrzymana zostaje zależność $AB \rightarrow_{0.8} D$. AB jest 0.8-kluczem schematu R . C i D są

0.8-niekluczowymi atrybutami i w pełni zależą od AB w stopniu 0.8. **R** jest w 0.8-F2NF.

Niech **F** będzie zbiorem rozmytych zależności funkcyjnych dla schematu **R** i **K** będzie θ -kluczem schematu **R**.

Mówi się, że **R** jest w θ -rozmytej trzeciej postaci normalnej (θ -F3NF) wtedy i tylko wtedy, gdy **R** jest w θ -F2NF i dla każdej $X \rightarrow_{\phi} A$ należącej do F^+ , gdzie **A** nie należy do **X**, **X** jest θ -nadkluczem albo **A** jest θ -niekluczowym atrybutem.

θ -F3NF uniemożliwia, aby θ -niekluczowy atrybut był zależny przechodnie (przez pewne zbiory atrybutów nie będące θ -nadkluczami) od θ -nadklucza w pewnym stopniu. Ograniczenie to nie pozwala na częściową lub przechodnią zależność θ -niekluczowego atrybutu od θ -klucza.

Na przykład niech **R** = (S#, Nazwisko, Adres, Wiek), **F** = {S# $\rightarrow_{0.9}$ Wiek, S# $\rightarrow_{0.8}$ Adres, S# $\rightarrow_1$ Nazwisko, Nazwisko $\rightarrow_1$ S#}. S# jest 0.8-kluczem schematu **R**. Z zależności Nazwisko $\rightarrow_1$ S# wiadomo, że Nazwisko jest również 0.8-kluczem. Każda z zależności należących do F^+ spełnia ograniczenie nakładane przez 0.8-F3NF, stąd **R** jest w 0.8-F3NF. Można rozważyć również inny przykład, niech **R**(A, B, C, D) i **F** = {AB $\rightarrow_{0.9}$ C, AC $\rightarrow_{0.8}$ D}. Ponieważ AB $\rightarrow_{0.8}$ D jest przechodnia przez AC: AB $\rightarrow_{0.9}$ AC, AC $\rightarrow_{0.8}$ D, gdzie AC nie jest 0.8-nadkluczem, **R** nie jest w 0.8-F3NF.

Niech **F** będzie zbiorem rozmytych zależności funkcyjnych dla schematu **R** i **K** będzie θ -kluczem schematu **R**. Mówi się, że **R** jest w θ -rozmytej postaci normalnej Boyce'a-Codda (θ -FBCNF) wtedy i tylko wtedy, gdy **R** jest w θ -F3NF i dla każdej $X \rightarrow_{\alpha} A$ należącej do F^+ , gdzie **A** nie należy do **X**, **X** jest θ -nadkluczem.

θ -FBCNF uniemożliwia, aby atrybut (zbiór atrybutów) był zależny przechodnie przez pewne zbiory atrybutów nie będące θ -nadkluczami i określa, że dla każdego θ -klucza, zbiór atrybutów do niego nie należących i zależnych od niego musi być zależny w pełni. Ograniczenie to nie pozwala na częściową lub przechodnią zależność θ -niekluczowego atrybutu od θ -klucza i nie pozwala na częściową lub przechodnią zależność θ -kluczowego atrybutu od θ -klucza nie zawierającego go.

Na przykład, niech

R(S#, Nazwisko, Adres, Wiek),

F = {S# $\rightarrow_{0.95}$ Wiek, S# $\rightarrow_{1.0}$ Adres, S# $\rightarrow_{0.9}$ Nazwisko, Nazwisko $\rightarrow_{0.9}$ S#}.

S#, podobnie jak Nazwisko, jest 0.9-kluczem. **R** jest w 0.9-FBCNF.

Rozważając

R = (S#, Nazwisko, Kurs#, Ocena),

F = {S# $\rightarrow_1$ Nazwisko, Nazwisko $\rightarrow_1$ S#, (S#, Kurs#) $\rightarrow_{0.9}$ Ocena},

(S#, Kurs#), podobnie jak (Nazwisko, Kurs#), jest 0.9-kluczem schematu relacji **R**. Ocena jest jedynym 0.9-niekluczowym atrybutem, który w pełni zależy od 0.9-klucza schematu **R**, zatem **R** jest w 0.9-F3NF. **R** nie jest jednak w 0.9-FBCNF, ponieważ

S# $\rightarrow_1$ Nazwisko, Nazwisko $\rightarrow_1$ S#

są zależnościami, których lewe strony nie są 0.9-nadkluczami.

θ -rozmyte postacie normalne nie pozwalają na występowanie określonych rozmytych zależności funkcyjnych, dlatego też anomalie wynikające z istnienia częściowych i przechodnich rozmytych zależności mogą być usunięte przez zaprojektowanie schematów w odpowiedniej rozmytej postaci normalnej. Należy też zwrócić uwagę, że im wyższa postać, tym ograniczenia przez nią narzucane są mocniejsze.

6. ALGORYTMY PROJEKTOWANIA SCHEMATU BAZY DANYCH

Rozmyte postacie normalne można uzyskać rozkładając schemat na kilka schematów nie posiadających częściowych i przechodnich rozmytych zależności funkcyjnych. Poprawny projekt wymaga dodatkowo zachowania informacji związanych z schematem oryginalnym, co oznacza zachowanie jego danych i zależności. Uzyskany rozkład powinien być odwracalny i zachowujący rozmyte zależności funkcyjne.

W rozmytych bazach danych istnieje jeszcze jeden aspekt zachowania informacji, jest on związany z zachowaniem zależności związanych ze schematami uzyskanymi w rozkładzie (R_i). Jeżeli U_i jest zbiorem atrybutów schematu relacji R_i , to zbiór rozmytych zależności funkcyjnych związanych z R_i jest zbiorem $\Pi_{U_i}(F) = \{V \rightarrow_{\alpha} W \mid V \rightarrow_{\alpha} W \in F^+, VW \subseteq U_i\}$. Zazwyczaj można zachować tylko podzbiór $\Pi_{U_i}(F)$, ponieważ zachowanie $\Pi_{U_i}(F)$ jest spełnione wtedy i tylko wtedy, gdy F jest zbiorem bez zależności częściowych. Ten aspekt zachowania informacji zostanie omówiony w podrozdziale 6.1.

Rozdział bieżący przedstawia algorytmy projektowania rozkładu schematu zachowującego zależności do F3NF (algorytm 4), zachowującego zależności i odwracalnego do F3NF (algorytm 5) i odwracalnego do FBCNF rozmytej postaci normalnej Boyce'a-Codda (algorytm 6).

6.1. ROZKŁAD ZACHOWUJĄCY ZALEŻNOŚCI DO ROZMYTEJ TRZECIEJ POSTACI NORMALNEJ

Dla rozmytych baz danych algorytm rozkładu schematu relacji zachowującego zależności do F3NF przedstawia algorytm 4.

Algorytm 4. Rozkład zachowujący zależności do F3NF

Wejście: schemat relacji R i zbiór rozmytych zależności funkcyjnych F .

Wyjście: rozkład ρ schematu relacji R .

Metoda: Jeżeli w R istnieją jakiegokolwiek atrybuty nie związane żadnymi zależnościami z F ani po lewej, ani po prawej stronie, to taki atrybut może w zasadzie

sam utworzyć relację i da się go wyeliminować z **R**. Jeżeli jedna z zależności z **F** wiąże wszystkie atrybuty **R**, to uzyskuje się sam schemat **R** ($\rho = \{\mathbf{R}\}$). W przeciwnym przypadku otrzymany w wyniku rozkład ρ składa się ze schematów **XA** dla każdej zależności $X \rightarrow_{\phi} A$ z **F** (tj. $R_i \in \rho$, gdzie $U_i = XA$).

Rozkład $\rho = \{\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_k\}$, uzyskany w wyniku zastosowania algorytmu 4, jest rozkładem zachowującym zależności względem **F** i każdy $\mathbf{R}_i$ (dla $i = 1, 2, \dots, k$) jest w postaci ϕ -F3NF względem $\Pi_{U_i}(\mathbf{F})|_{\phi}$, gdzie $YB=U_i$ wynika z zależności $Y \rightarrow_{\phi} B$ należącej do **F** i $\Pi_{U_i}(\mathbf{F})|_{\phi} = \{V \rightarrow_{\alpha} W \mid V \rightarrow_{\alpha} W \in F^+, VW \subseteq U_i, \alpha \geq \phi\}$.

Na przykład niech

$$\mathbf{R}(A, B, C, D), F = \{CD \rightarrow_{\theta} A, CD \rightarrow_{\theta} B, D \rightarrow_{\alpha} B\} \text{ i } \alpha < \theta.$$

R nie jest w θ -F2NF względem **F**, ponieważ istnieje częściowa zależność $D \rightarrow_{\alpha} B$. Stosując algorytm 4 uzyskuje się $\rho = \{CAD, CDB, DB\}$. ρ zachowuje zależności, **CDA** jest w θ -F3NF względem $\Pi_{CDA}(\mathbf{F})|_{\theta}$, **CDB** jest w θ -F3NF względem $\Pi_{CDB}(\mathbf{F})|_{\theta}$, **BD** jest w α -F3NF względem $\Pi_{DB}(\mathbf{F})|_{\alpha}$. $D \rightarrow_{\alpha} B$ nie należy do $\Pi_{CDB}(\mathbf{F})|_{\theta}$ zgodnie z nierównością $\alpha < \theta$. Można również rozważyć inny przykład, niech $\mathbf{R}(A, B, C, D, E)$, $F = \{CD \rightarrow_{\theta} A, CD \rightarrow_{\theta} B, AD \rightarrow_{\phi} E, CD \rightarrow_{\theta} E, A \rightarrow_{\alpha} B, A \rightarrow_{\beta} E\}$, $\phi > \min(\alpha, \beta)$. **CD** jest θ -kluczem **R**. **R** jest w θ -F2NF względem **F**, ponieważ żaden niekluczowy atrybut nie jest częściowo zależny od **CD**, ale **R** nie jest w θ -F3NF względem **F**, ponieważ istnieją $AD \rightarrow_{\phi} E$, $A \rightarrow_{\alpha} B$, $A \rightarrow_{\beta} E$. Stosując algorytm 4 uzyskuje się $\rho = \{CDA, CDB, ADE, CDE, AB, BE\}$, gdzie **CDB**, **CDA**, **CDE** są w θ -F3NF względem odpowiednio $\Pi_{CDB}(\mathbf{F})|_{\theta}$, $\Pi_{CDA}(\mathbf{F})|_{\theta}$, $\Pi_{CDE}(\mathbf{F})|_{\theta}$, **ADE** jest w θ -F3NF względem $\Pi_{ADE}(\mathbf{F})|_{\phi}$, **AB** jest w α -F3NF względem $\Pi_{AB}(\mathbf{F})|_{\alpha}$ i **BE** jest w α -F3NF względem $\Pi_{BE}(\mathbf{F})|_{\beta}$. $A \rightarrow_{\min(\alpha, \beta)} E$ jest w $\Pi_{ADE}(\mathbf{F})$, ale nie należy do $\Pi_{ADE}(\mathbf{F})|_{\phi}$ dla $\phi > \min(\alpha, \beta)$.

Należy zauważyć, że dla rozkładu $\rho = \{\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_k\}$ zachowującego zależności, uzyskanego stosując algorytm 4 $\mathbf{R}_i$ dla $i = 1, 2, \dots, k$ jest w ϕ -F3NF względem $\Pi_{U_i}(\mathbf{F})|_{\phi}$, gdzie $YB=U_i$ wynika z zależności $Y \rightarrow_{\phi} B$ należącej do **F**, nie zaś względem $\Pi_{U_i}(\mathbf{F})$.

Rozważając ponownie $\mathbf{R}(A, B, C, D)$, $F = \{CD \rightarrow_{\theta} A, CD \rightarrow_{\theta} B, D \rightarrow_{\alpha} B\}$ i $\alpha < \theta$, stosując algorytm 4 uzyskuje się $\rho = \{CAD, CDB, DB\}$. Schemat **CDB** jest w θ -F3NF względem $\Pi_{CDB}(\mathbf{F})|_{\theta}$, ale nie jest w θ -F3NF względem $\Pi_{CDB}(\mathbf{F})$, ponieważ $D \rightarrow_{\alpha} B$ należy do $\Pi_{CDB}(\mathbf{F})$, co powoduje częściową zależność **B** od **CD** (**CDB** nie jest tu nawet w θ -F2NF względem $\Pi_{CDB}(\mathbf{F})$).

Aby $\mathbf{R}_i$ dla $i = 1, 2, \dots, k$ był w ϕ -F3NF względem $\Pi_{U_i}(\mathbf{F})$, to zbiór **F** musi być takim zbiorem rozmytych zależności funkcyjnych, że dla każdej zależności $Y \rightarrow_{\phi} B$ należącej do **F**, nie może istnieć $X \rightarrow_{\alpha} B$ należąca do F^+ taka, że $X \subset Y$ i $\alpha < \theta$. Algorytm sprawdzania, czy dany zbiór **F** nie posiada takich częściowych rozmytych zależności funkcyjnych znajduje się w [CHEN 98].

6.2. ROZKŁAD ZACHOWUJĄCY ZALEŻNOŚCI I ODWRACALNY DO ROZMYTEJ TRZECIEJ POSTACI NORMALNEJ

Algorytm odwracalnego i zachowującego zależności rozkładu schematu do F3NF opiera się na algorytmie 4 i fakcie, że, jeżeli F jest minimalnym pokryciem zbioru rozmytych zależności funkcyjnych dla schematu relacji R i K jest θ -kluczem schematu R , to schemat R' z $U'=K$ jest w 1-F1NF względem $\Pi_K(F)$. Uzyskanie rozkładu zachowującego zależności i odwracalnego do F3NF przedstawia algorytm 5.

Algorytm 5. Odwracalny i zachowujący zależności rozkład do F3NF

Wejście: schemat relacji R i zbiór rozmytych zależności funkcyjnych F , θ -klucz K schematu R i rozkład ρ uzyskany stosując algorytm 4.

Wyjście: rozkład σ schematu relacji R .

Metoda:

$$\sigma = \rho \cup \{R'\}, \text{ gdzie } U'=K.$$

Rozkład $\sigma = \{R_1, R_2, \dots, R_k\}$, uzyskany w wyniku zastosowania algorytmu 5, spełnia własności:

- 1) $\sigma = \rho \cup \{R'\}$, gdzie $U'=K$, jest rozkładem zachowującym zależności należące do F i każdy R_i dla $i = 1, 2, \dots, k$ jest w ϕ -F3NF względem $\Pi_{U_i}(F)|_{\phi}$, gdzie $YB=U_i$ wynika z zależności $Y \rightarrow_{\phi} B$ należącej do F i R' jest 1-F3NF względem $\Pi_K(F)$.
- 2) jeżeli F jest zbiorem bez częściowych rozmytych zależności funkcyjnych, to $\sigma = \rho \cup \{R'\}$, gdzie $U'=K$, jest rozkładem zachowującym zależności należące do F i każdy R_i dla $i = 1, 2, \dots, k$ jest w ϕ -F3NF względem $\Pi_{U_i}(F)$, gdzie $YB=U_i$ wynika z zależności $Y \rightarrow_{\phi} B$ należącej do F i R' jest 1-F3NF względem $\Pi_K(F)$.
- 3) σ jest rozkładem odwracalnym względem F .

Na przykład niech

$$R(A, B, C, D, E, G, H),$$

$$F = \{ABC \rightarrow_{0.8} D, ABC \rightarrow_{0.9} E, ABC \rightarrow_{0.9} G, DE \rightarrow_{0.8} E, D \rightarrow_{0.7} H, H \rightarrow_{0.6} G\}.$$

ABC jest 0.7-kluczem schematu R , F jest minimalnym pokryciem.

$$\sigma = \rho \cup \{ABC\} = \{ABCD, ABCE, ABCG, DEG, DH, HG, ABC\}$$

zachowuje zależności, wszystkie schematy należące do σ są w F3NF w stopniu odpowiednio 0.8, 0.9, 0.9, 0.8, 0.7, 0.6 i 1.0 względem odpowiednich zbiorów rozmytych zależności funkcyjnych:

$$\Pi_{ABCD}(F)|_{0.8}, \Pi_{ABCE}(F)|_{0.9}, \Pi_{ABCG}(F)|_{0.9}, \Pi_{DEG}(F)|_{0.8}, \Pi_{DH}(F)|_{0.7}, \Pi_{HG}(F)|_{0.6} \text{ i } \Pi_{ABC}(F).$$

σ jest odwracalny, a własność tą można sprawdzić stosując algorytm 2. Tabelę początkową dla badania odwracalności przedstawia tabela 7.

Tabela 7. Tabela początkowa dla badania odwracalności rozkładu σ

R	A	B	C	D	E	G	H
ABCD:	a ₁	a ₂	a ₃	a ₄	b ₁₅	b ₁₆	B ₁₇
ABCE:	a ₁	a ₂	a ₃	b ₂₄	a ₅	b ₆	B ₂₇
ABCG:	a ₁	a ₂	a ₃	b ₃₄	b ₃₅	a ₆	B ₃₇
DEG:	b ₄₁	b ₄₂	b ₄₃	a ₄	a ₅	a ₆	B ₄₇
DH:	b ₅₁	b ₅₂	b ₅₃	a ₄	b ₅₅	b ₅₆	A ₇
HG:	b ₆₁	b ₆₂	b ₆₃	b ₆₄	b ₆₅	a ₆	A ₇
ABC:	a ₁	a ₂	a ₅	b ₇₄	b ₇₅	b ₇₆	B ₇₇

$ABC \rightarrow_{id} D$ powoduje, że b_{24} , b_{34} , b_{74} stają się a_4 , $ABC \rightarrow_{0.9} E$ powoduje, że b_{15} , b_{35} , b_{75} stają się a_5 , $ABC \rightarrow_{0.9} G$ powoduje, że b_{16} , b_{26} , b_{76} stają się a_6 i $D \rightarrow_{0.7} H$ powoduje, że b_{17} , b_{27} , b_{37} , b_{47} , b_{77} stają się a_7 . Zostaje uzyskana tabela 8, w której istnieje wiersz (ABC) z $a_1, \dots, a_7$, stąd σ jest rozkładem odwracalnym.

Tabela 8. Badanie odwracalności rozkładu σ

R	A	B	C	D	E	G	H
ABCD:	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	A ₇
ABCE:	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	A ₇
ABCG:	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	A ₇
DEG:	b ₄₁	b ₄₂	b ₄₃	a ₄	a ₅	a ₆	A ₇
DH:	b ₅₁	b ₅₂	b ₅₃	a ₄	b ₅₅	b ₅₆	A ₇
HG:	b ₆₁	b ₆₂	b ₆₃	b ₆₄	b ₆₅	a ₆	A ₇
ABC:	a ₁	a ₂	a ₅	a ₄	a ₅	a ₆	A ₇

6.3. ROZKŁAD ODWRACALNY DO ROZMYTEJ POSTACI NORMALNEJ BOYCE'A-CODDA

Dla rozmytych baz danych algorytm odwracalnego rozkładu schematu relacji do FBCNF przedstawia algorytm 6.

Algorytm 6. Odwracalny rozkład do FBCNF

Wejście: schemat relacji **R** i zbiór rozmytych zależności funkcyjnych **F**,

$\theta = \min\{\beta \mid X \rightarrow_{\beta} Y \text{ należy do } F\}$.

Wyjście: rozkład ρ schematu relacji **R**.

Metoda:

$\rho = \{\mathbf{R}\}$

while $\exists$ schemat $S \in \rho$ nie będący w θ -FBCNF

begin

dla rozmytej zależności $X \rightarrow_{\alpha} A$ należącej do $\{V \rightarrow_{\beta} W \mid V \rightarrow_{\beta} W \in F^+, VW \subseteq S\}$, która narusza θ -FBCNF dla S :

$$S_1 = XA$$

$$S_2 = S - A$$

$$\rho = (\rho - \{S\}) \cup (\{S_1\} \cup \{S_2\})$$

end.

$X \rightarrow_{\alpha} A$ narusza θ -FBCNF oznacza, że X nie jest θ -nadkluczem schematu S (A nie należy do X i XA należy do S).

Rozkład $\rho = \{R_1, R_2, \dots, R_k\}$ uzyskany stosując algorytm 6 jest rozkładem odwracalnym względem F i każdy R_i dla $i = 1, 2, \dots, k$ jest w θ -FBCNF względem $\Pi_{U_i}(F)$.

Na przykład niech

$$R(A, B, C, D, E, G, H),$$

$$F = \{ABC \rightarrow_{0.8} D, ABC \rightarrow_{0.9} E, ABC \rightarrow_{0.9} G, DE \rightarrow_{0.8} E, D \rightarrow_{0.7} H, H \rightarrow_{0.6} G\}.$$

ABC jest 0.7-kluczem schematu R , $\theta = \min(0.8, 0.9, 0.9, 0.8, 0.7, 0.6)$. ABC jest również 0.6-kluczem schematu R . Ponieważ DE dla $DE \rightarrow_{0.8} E$, D dla $D \rightarrow_{0.7} H$ i H dla $H \rightarrow_{0.6} G$ nie są 0.6-nakluczami, R nie jest w 0.6-FBCNF. Wybór zależności naruszającej θ -FBCNF do rozważenia jest dowolny. Zaczynając przykładowo od zależności $DE \rightarrow_{0.8} E$ uzyskuje się $\rho = \{ABCDEH, DEG\}$. W $ABCDEH$, gdzie ABC jest 0.6-kluczem, $D \rightarrow_{0.7} H$ narusza 0.6-FBCNF. $ABCDEH$ jest zatem rozkładany na $ABCDE$ i DH . W DEG , gdzie DE jest 0.6-kluczem, $D \rightarrow_{0.6} G$ należąca do $\Pi_{DEG}(F)$ narusza 0.6-FBCNF, stąd z DEG zostaje rozłożony na DE i DG . Ostatecznie otrzymany zostaje rozkład

$$\rho = \{ABCDE, DH, DE, DG\},$$

który prowadzi do schematów w 0.6-FBCNF. σ jest odwracalny, a własność tą można sprawdzić stosując algorytm 2. Tabelę początkową dla badania odwracalności przedstawia tabela 9.

Tabela 9. Tabela początkowa dla badania odwracalności rozkładu σ

R	A	B	C	D	E	G	H
ABCDE:	a ₁	a ₂	a ₃	a ₄	a ₅	b ₁₆	b ₁₇
DH:	b ₂₁	b ₂₂	b ₂₃	a ₄	b ₂₅	b ₂₆	a ₇
DE:	b ₃₁	b ₃₂	b ₃₃	a ₄	a ₅	b ₃₆	b ₃₇
DG:	b ₄₁	b ₄₂	b ₄₅	a ₄	b ₄₅	a ₆	b ₄₇

$D \rightarrow_{0.6} H$ powoduje, że b_{17} , b_{37} , b_{47} stają się a_7 , $D \rightarrow_{0.6} G$ powoduje, że b_{16} , b_{26} , b_{36} stają się a_6 . Zostaje uzyskana tabela 10, w której istnieje wiersz ($ABCDE$) z $a_1, \dots, a_7$, stąd σ jest rozkładem odwracalnym.

Tabela 10. Badanie odwracalności rozkładu σ

R	A	B	C	D	E	G	H
ABCDE:	a_1	a_2	a_3	a_4	a_5	a_6	a_7
DH:	b_{21}	b_{22}	b_{23}	a_4	b_{25}	a_6	a_7
DE:	b_{31}	b_{32}	b_{33}	a_4	a_5	a_6	a_7
DG:	b_{41}	b_{42}	b_{45}	a_4	b_{45}	a_6	a_7

7. PODSUMOWANIE

Projektowany model rozmytej relacyjnej bazy danych musi być poprawny, co oznacza, że schematy muszą być tak zaprojektowane, aby nie występowały problemy redundancji czy też anomalii przy wstawianiu, modyfikacji i usuwaniu danych. Ponieważ wartościami atrybutów mogą być wartości nieprecyzyjne, teoria projektowania dla modelu klasycznego została tu rozszerzona. Teoria projektowania schematów rozmytych baz danych została dobrze sformułowana dla rozszerzonego podejścia opartego na możliwości. Wprowadzono rozmyte zależności funkcyjne, pojęcie θ -klucza oraz rozszerzone reguły integralności konkretnego i odwołań, rozmyte rozszerzenie aksjomatów Armstronga, zdefiniowano domknięcie zbioru rozmytych zależności funkcyjnych i równoważność zbiorów rozmytych zależności funkcyjnych. Rozważone zostały również dwa rodzaje rozkładu schematu: odwracalny i zachowujący zależności. Przede wszystkim jednak zdefiniowano θ -rozmyte postacie normalne (F1NF, F2NF, F3NF, FBCNF), nakładające ograniczenia na obecność w schemacie częściowych i przechodnich rozmytych zależności funkcyjnych, aby problem redundancji danych i anomalii nie pojawił się. Zostały opracowane algorytmy projektowania schematu rozmytej bazy danych, w wyniku których uzyskuje się schematy w odpowiednich rozmytych postaciach normalnych, uwzględniając własność odwracalności lub zachowania zależności. Należy podkreślić, że teoria projektowania schematów rozmytych baz danych dla rozszerzonego podejścia opartego na możliwości stosuje się do podejścia opartego na możliwości, który jest szczególnym jego przypadkiem. Dla pozostałych podejść rozważono tylko niektóre aspekty teorii projektowania. Dla podejścia opartego na relacji rozmytej zdefiniowane zostały zależności funkcyjne i rozkład odwracalny [RAJU 87]. Dla podejścia opartego na podobieństwie opracowano postać rozmytej zależności funkcyjnej i wyspecyfikowano rozszerzone aksjomaty Armstronga [CHEN 95, YAZICI 99].

Uzyskane w projekcie schematy relacji powinny być w rozmytej trzeciej postaci normalnej. Praktyka projektowa pozwala uniknąć konieczności korzystania z teorii projektowania schematów rozmytych relacyjnych baz danych. Utworzenie poprawnego modelu koncepcyjnego, a następnie jego transformacja do odpowiedniego rozmytego modelu relacyjnego pozwala uzyskać schemat bazy danych, w której nie będzie występowała redundancja oraz anomalie danych. Normalizacja jest zazwyczaj używana do sprawdzenia poprawności lub do weryfikacji

krytycznych obszarów modelu danych. Jeżeli każda reguła funkcjonowania i ograniczenia wynikają z zależności funkcyjnych lub ograniczeń dziedzinowych, wówczas schematy relacji mają „właściwą” postać normalną.

Dalsze prace nad projektowaniem rozmytych modeli relacyjnych baz danych powinny być zwrócone na rozszerzenia innych modeli konceptualnych, które mogłyby reprezentować rozmyte dane. Można byłoby również rozbudować rozmyty model związków encji, ponieważ, jak można zauważyć, nie pozwala on odzwierciedlić pewnych nieprecyzyjnych i niepewnych danych. Dla modelowania logicznego wartość rozważenia mogą okazać się podejścia łączone.

DESIGN OF FUZZY RELATIONAL DATABASE

In this paper the definition of normal forms for the fuzzy relational database was been investigated. A great deal of the work concerns schemas of fuzzy relational database design. Principles and suggestion how to create schemas in order to minimize redundancy and avoid anomalies in store data have been included.

LITERATURA

- [ANVARI 84] Anvari M., Rose G.F., *Fuzzy relational databases*. Proc. of 1st Intl. Conf. on FIP, Hawaii, 1984.
- [BADUREK 99] Badurek J., *Logika rozmyta w bazach danych*. Informatyka nr 1/99, s. 14–17.
- [BALDWIN 84] Baldwin J.F., Zahou S.Q., *A fuzzy relational inference language*. Fuzzy Sets and Systems 1984, vol. 14, s. 155–174.
- [BANDLER 80] Bandler W., Kohout L., *Fuzzy power sets and fuzzy implication operators*. Fuzzy Sets and Systems 1980, vol 4, s. 13–30.
- [BEYNON_DAVIES 98] Beynon-Davies P., *Systemy baz danych*. Warszawa: WNT 1998.
- [BIEMAZUR 02a] Biedrzycka B., Mazur Z., *Podstawy projektowania relacyjnych baz danych*, Prace Naukowe WZI PWr., Zeszyt 3, Wrocław, 2002
- [BIEMAZUR 02b] Biedrzycka B., Mazur Z., *Logika rozmyta*, Prace Naukowe WZI PWr., Zeszyt 3, Wrocław, 2002
- [BIEMAZUR 02c] Biedrzycka B., Mazur Z., *Rozmyty model związków encji*, Prace Naukowe WZI PWr., Zeszyt 3, Wrocław, 2002
- [BIEMAZUR 02d] Biedrzycka B., Mazur Z., *Reprezentacja danych w rozmytych modelach relacyjnych baz danych*, Prace Naukowe WZI PWr., Zeszyt 3, Wrocław, 2002
- [BIEMAZUR 02e] Biedrzycka B., Mazur Z., *Operowanie danymi rozmytymi*, Prace Naukowe WZI PWr., Zeszyt 3, Wrocław, 2002
- [BIEMAZUR 02f] Biedrzycka B., Mazur Z., *Projektowanie schematów rozmytych relacyjnych baz danych*, Prace Naukowe WZI PWr., Zeszyt 3, Wrocław, 2002
- [BOLC 91] Bolc L., Borodziejewicz W., Wójcik M. *Podstawy przetwarzania informacji niepewnej i niepełnej*. Warszawa, PWN 1991.
- [BOSC 88] Bosc P., Galibourg M., Namon G., *Fuzzy Querying with SQL: Extensions and Implementation Aspects*. Fuzzy Sets and Systems 1988, vol 28, s. 333–349.
- [BOSC 91] Bosc P., Privert O., *About equivalents in SQLf: relational language supporting imprecise querying*. Proc. of Intl. Fuzzy Engineering Symposium, Japan, 1991, s. 309–320.
- [BOSC 94] Bosc P., Pivert O., *Imprecise Data Management and Flexible Querying in Databases*. W: Fuzzy Sets and Neural Networks, New York 1994, s. 368–395.
- [BOSC 95] Bosc P., Privet O., *SQLf: a Relational Database Language for Fuzzy Querying*. IEEE Transactions on Fuzzy Systems 1995, vol. 3, nr 1, s. 1–17.
- [BUCKLES 82] Buckles B.P., Petry F.E., *A fuzzy representation of data for relational databases*. Fuzzy Sets and Systems 1982, vol. 7, s. 213–226.

- [CELL 88] Cellary W., Królikowski Zbyszko, *Wprowadzenie do projektowania baz danych. dBase III.*, WNT, Warszawa, 1988
- [CHEN 76] Chen P.P. *The entity-relationship model: towards a unified view of data.* ACM Transactions on Database Systems 1976, vol. 1, nr 1, s. 9–36.
- [CHEN 91] Chen G.Q., Vandenbulcke J., Kerre E.E., *A step towards the theory of fuzzy relational database design.* Proceedings of IFSA'91 World Congress 1991, s. 44–47.
- [CHEN 92] Chen G.Q. , Vandenbulcke J. i Kerre E.E., *A general treatment of data redundancy in a fuzzy relational data model.* Journal of The American Society for Information Science 1992, s. 304–311.
- [CHEN 93] Chen G.Q., Kerre E.E., Vandenbulcke J., *On the lossless-join decomposition in fuzzy relational data model.* Proceedings of International Symposium on Uncertainty Modelling and Analysis (ISUMA'93), IEEE Press, Maryland (USA), 1993, s. 40–446.
- [CHEN 95] Chen G., *Fuzzy Functional Dependency and Series of Design Issues of Fuzzy Relational Databases.* W: Bosc P., Kacprzyk J. (eds), *Fuzziness in Database Management Systems*, Physica-Verlag 1995, s. 165–185.
- [CHEN 97] Chen G.Q., Kerre E.E., *Designing a general-purpose system for fuzzy data representation and queries.* Proc. of IFSA'97 Prague 1997, s. 255–260.
- [CHEN 98] Chen G., *Fuzzy logic in data modeling: Semantics, Constraints, and Database Design.* Kluwer Academic Publishers 1998.
- [CODD 70] Codd E.F., *A relational model for large shared data blanks.* Communications of The ACM 1970, vol. 13, nr 1, s. 377–387.
- [CODD 74] Codd E.F., *Recent investigations into relational database systems.* Proc. IFIP Congress, 1974.
- [DATE 00] Date C.J., *Wprowadzenie do systemów baz danych.* Warszawa, WNT 2000.
- [DE CALUWE 95] De Caluwe R., Vandenberghe R., Van Gyseghem N., Van Schooten A., *Integrating Fuzziness in Database Models.* W: Bosc P., Kacprzyk J.(eds), *Fuzziness in Database Management Systems*, Physica-Verlag 1995, s. 114–138.
- [DRIANKOOV 96] Driankov D., Hellendoorn H., Reinfrank M., *Wprowadzenie do sterowania rozmytego.* Warszawa, WNT 1996.
- [ELMASRI 94] Elmasri R., Navathe S.B. *Fundamentals of Database Systems.* Benjamin/Cummings Publishing Company, Inc. 1994.
- [FAGIN 77] Fagin R., *Multivalued dependencies and a new normal form for relational databases.* ACM Transactions on Database Systems 1977, vol 2, nr 1.
- [FAGIN 79] Fagin R. *Normal forms and relational database operators.* ACM SIGMOD Int. Symposium on the Management of Data, 1979, s. 153–160.
- [GRUBER 96] Gruber M., *SQL.* Gliwice, HELION 1996.
- [HE 83] He Z., *Fuzzy Mathematics and Applications.* China, Tianjin Science & Technology Publishing Co. 1983.

- [HE 94] He X., *Fuzzy Database Systems*. China, Tshinghua University Publishing House 1994.
- [KACPRZYK 86] Kacprzyk J., Ziółkowski A., *Database queries with fuzzy linguistic quantifiers*. IEEE Trans. on Sys. Man and Cybern. 1986, vol. 16, s. 474–479.
- [KACPRZYK 94] Kacprzyk J., Zadrozny S., *Fuzzy quering for Microsoft Acces*. Proc. of 3rd IEEE Conf. on Fuzzy Systems. Ornaldo 1994; vol. 1, s. 167–171.
- [KERRE 86] Kerre E.E., Zenner R.B.R.C., De Culwe R.M.M., *The use of fuzzy set theory in information retrieval and databases: a survaey*. Journal of the American Society for Information Science 1986, vol. 37, nr 5, s. 341–345.
- [KERRE 89] Kerre E.E., *Outlines of the expert system for ECG-diagnosis using fuzzy sets*. Artificial Intelligence in Medicine 1989, vol 1, s. 139–144.
- [KERRE 93] Kerre E.E. (ed.), *Introduction to Basic Principles of Fuzzy Set theory and Some of its Applications*. 2nd edition. Gent, Belgium: Communication & Cognition 1993.
- [KERRE 95] Kerre E.E., Chen G., *An Overview of Data Models*. In Studies in Fuzziness: Fuzziness in Database Management Systems. Bosc P. i Kacprzyk J., eds. Physica-Verlag 1995.
- [LAKOFF 73] Lakoff G., *Hedges: a study in meaning criteria and the logic of fuzzy concepts*. Journal of Philosophical Logic 1973, vol. 2, s. 458–508.
- [ŁACHWA 2001] Łachwa A., *Rozmyty świat zbiorów, liczb, relacji, faktów, reguł i decyzji*. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2001
- [MAMDANI 77] Mamdani E., *Application of fuzzy logic to approximate reasoning using linguistic systems*. IEEE Trans. On Computer 1977, vol. 26, s. 1182–1191.
- [MUZIMOTO 82] Muzimoto M., Zimmermann H., *Comparison of fuzzy reasoning*. Fuzzy Sets and Systems 1982, vol. 8, s. 253–283.
- [PANKOWSKI 92] Pankowski T. *Podstawy baz danych*. Warszawa, PWN 1992.
- [PRADE 94] Prade H., Testemale C., *Generalizing Database Relational Algebra for the Treatment of Incomplete or Uncertain Information and Vague Queries*. Information Sciences 1984 vol 34, s. 115–143.
- [RAJU 87] Raju K., Majumdar A., *The study of joins in fuzzy relational databases*. Fuzzy Sets and Systems 1987, vol. 21, s. 19–34.
- [RUNDENSTEINER 89] Rundensteiner E.A., Hawkes L.W., Bandler W., *On nearness measures in fuzzy relaional data models*. Int J. of Approximate Reasoning 1989, vol. 3, s. 267–298.
- [RUSPINI 86] Ruspini E. *Imprecision and uncertainty in the entity-relationship model*. In Fuzzy Logic in Knowledge Engineering. Prade H. i Negoita V., eds. Verlag TUV Rheinland 1986.
- [SCHMUCKER 84] Schmucker K.J., *Fuzzy Sets, Natural Language Computations, and Risk Analysis*. Maryland: Computer Science Press 1984.
- [SHENOI 89] Sheno S., Melton A., *Proximity relations in the fuzzy relational databases*. Fuzzy Set and Systems 1989, vol. 31, s. 285–296.

- [ŠNENOI 90] Shenoi S. i Melton A., Fan L., *An equivalence classes model od fuzzy relational databases*. Fuzzy Set and Systems 1989, vol. 31, s. 285–296.
- [SUBIETA 99] Subieta K., *Słownik terminów z zakresu obiektowości*. Akademicka Oficyna Wydawnicza PLJ, Warszawa, 1999
- [ULLMAN 88] Ullman J.D., *Systemy baz danych*. Warszawa, WNT 1988.
- [ULLMAN 99] Ullman J.D., Widom J. *Podstawowy wykład z systemów baz danych*. Warszawa, WNT 1999.
- [VANDENBERGHE 91] Vandenberghe R.M., *An extended in the entity-relationship model for fuzzy databases based on fuzzy truth values*. Proceedings of IFSA'91, Bruselss, 1991, s. 280–283.
- [WANG 83] Wang P. *Fuzzy Set Theory and Applications*. China, Shanghai Science & Technology Publishing Co. 1983.
- [WILLMOTT 80] Willmott R., *Two fuzzier implication operators in the theory of fuzzy power sets*. Fuzzy Sets and Systems 1980, vol. 4, s. 31–36.
- [YAGER 88] Yager R.R., *On ordered weighted average aggregation operators in multicriteria decisionmaking*. IEEE Trans. on Sys., Man., Cybern. 1988, vol. 18, nr 1, s. 183–190.
- [YAGER 91] Yager R.R., *Fuzzy quotient operators for fuzzy relation databases*. Proc. of Intl. Fuzzy Engineering Symposium. Japan 1991, vol. 289–296.
- [YAGER 95] Yager R.R., Dimitar P.F., *Podstawy modelowania i sterowania rozmytego*. Warszawa, WNT 1995.
- [YAZICI 99] Yazici A., George R. *Fuzzy Database Modeling*. New York, Physica-Verlag 1999.
- [ZADEH 65] Zadeh L.A., *Fuzzy Sets*. Information and Control 1965, vol. 8, s. 338–358.
- [ZADEH 72] Zadeh L.A., *A fuzzy-sets-theoretic interpretation of linguistic hedges*. Journal of Cybernetics 1972, vol. 3, s. 4–34.
- [ZADEH 75a] Zadeh L.A., *Calculus of fuzzy restrictions*. In Fuzzy Sets and Their Applications to Cognitive and Decision Processes, L. A. Zadeh, et al., ed. New York: Academic Press 1975.
- [ZADEH 75b] Zadeh L.A. *The Concept of a Linguistic Variable and its Application to Approximate Reasoning*. Information Sciences 1975, vol. 8, s. 199–249, vol. 8 s. 301–357, vol. 9, s. 43–80.
- [ZADEH 78] Zadeh L.A., *Fuzzy sets as a basis for a theory of possibility*. Fuzzy Sets and Systems 1978; vol. 1, nr 1, s. 3–28.
- [ZEMANKOVA 84] Zemankova M., Kandel A., *Fuzzy Relational Database – a key to expert system*. Verlag TUV Rheinland 1984.
- [ZVIELI 85] Zvieli A., Chen P.P., *Entity-relationship modeling and fuzzy databases*. Proceedings of 2nd Conference on Data Engineering, LA 1985.





BIBLIOTEKA GŁÓWNA

312506/6

Wydawnictwa Politechniki Wrocławskiej
są do nabycia w następujących księgarniach:
„Politechnika”

Wybrzeże Wyspiańskiego 27, 50-370 Wrocław
budynek A-1 PWr., tel. (0-71) 320-25-34;
„Tech”

plac Grunwaldzki 13, 50-377 Wrocław
budynek D-1 PWr., tel. (0-71) 320-32-52

Prowadzimy sprzedaż wysyłkową

ISBN 83-7085-678-0