

УДК 517.9

Івохін Є.В.¹, д.ф.-м.н, доцент,
Кучерявий М.Г.², студент

Про один приклад реалізації нечітких баз даних

^{1,2} Київський національний університет
імені Тараса Шевченка, м.Київ, пр-т Глушкова,
4д, e-mail: ivohin@univ.kiev.ua

Ivohin E. V.¹, D.Sci.,
Kucheryavyu M.G.², student

About one example of fuzzy databases realization

^{1,2} Taras Shevchenko National University of Kyiv,
Kyiv, Glushkova st., 4d,
e-mail: ivohin@univ.kiev.ua

Для обробки неточно заданої та нечітко формалізованої інформації запропоновано підхід до формалізації нечітких баз даних. Рівень нечіткості атрибутів таблиць визначається на основі чітких даних у вигляді спеціального представлення, яке створюється за допомогою операції розширення. Наведено приклад створення нечіткої бази даних.

Ключові слова: реляційні бази даних, нечіткі множини, мова запитів.

The relational databases and basic operations with them are considered. For treatment unexactly set and the unexpressly formalized information offered approach to the fuzzy databases formalization. It is introduced the concept of fuzzy database based on the composite fuzzy sets and fuzzification crisp data. The definition of the fuzzy database is given. The tool for the creation and updating of the fuzzy relational database is a query language SQL. The level of tables attributes fuzzyness accounts on the basis of clear information as the special presentation which is created by the expansion operation. The evaluation of belonging measure is performed by using the trapezoidal membership function. An example of creating a fuzzy database, resources for the manipulation (updating) tables and scheme of user interface are considered. For the example implementation was used MS Access application. This approach demonstrates a constructive tool for creating fuzzy data base on the operative formation of membership parameters for the data of corresponding fuzzy sets.

Key words: relational databases, fuzzy datasets, query language.

Статтю представив д.т.н., проф. Волошин О.Ф.

Сучасну інформаційну систем важко уявити без використання баз або сховищ даних. Під цими поняттями, як правило, розуміють «набір структурованих даних, що підтримуються системою управління баз даних (СУБД), яка забезпечує різним додаткам потрібні дані у заданому вигляді» [1].

Однак, дане визначення є недостатньо конструктивним і слабо формалізованим. З іншого боку зауважимо, що кожна програмна система, яка призначається для накопичення та обробки (актуалізації) даних, використовує інформаційну модель предметної області. В загальному випадку характерними категоріями, що формують таку модель, є об'єкти та відношення між ними [2].

Найбільш популярною серед моделей баз даних (БД) є реляційна модель предметної області. У цій моделі вважається, що всі дані БД інформаційної системи представлені у вигляді таблиць [3]. Рядок кожної таблиці являє собою кортеж неструктурованих одиниць даних або атрибутів. Набір кортежів, що складає таблицю,

задає відношення: таким чином, модель даних визначається множиною таблиць-відношень (R -таблиць) або «реляцій».

Значення атрибутів кортежів і таблиць-відношень визначаються заданими разом з таблицями областями визначення (доменами). В рамках реляційної теорії визначений список операцій, які можна здійснювати над R -таблицями, отримуючи результат у вигляді R -таблиці. Це означає, що при виконанні операцій на реляційною базою даних отримуємо нову реляційну БД. До таких операцій над базою даних відносять:

1) базові операції: обмеження (виключення з таблиць деяких кортежів); проекція (виключення з таблиці деяких стовбців); декартовий добуток двох таблиць (за правилом декартового добутку двох множин рядків R -таблиць); об'єднання (об'єднання множин рядків двох таблиць); різниця (різниця множин рядків); присвоювання (іменованій таблиці присвоюється значення виразу над R -таблицями).

2) похідні операції: з'єднання таблиць за деяким правилом; перетин таблиць; розширення таблиць новими стовбцями; підсумовування (зменшення рядків в таблиці за рахунок агрегування вихідних даних).

Операції над R -таблицями дозволяють отримувати таблиці-представлення для формування динамічних наборів кортежів з додатково заданими правилами відбору.

Таким чином, можна констатувати, що будь-яка прикладна проблемна область описується в термінах таблиць достатньо просто, що дозволяє суттєво спростити розробку інформаційних систем. Крім цього, реляційна концепція баз даних може бути легко формалізованою, базуючись на математичній моделі відношень. Припустимо, що задано n множин D_1, D_2, \dots, D_n . Тоді реляція R є відношення над цими множинами, якщо R є множина впорядкованих наборів виду $\langle d_1, d_2, \dots, d_n \rangle$, де d_1 - елемент з D_1 , d_2 - елемент з D_2 , ..., d_n - елемент з D_n . При цьому набори вигляду $\langle d_1, d_2, \dots, d_n \rangle$ визначають кортежі, а множини D_1, D_2, \dots, D_n - домени.

Універсальним засобом актуалізації сховищ даних є мова запитів SQL [4], яка дозволяє зручно обробляти інформацію в різних реалізаціях моделей БД. Зрозуміло, що коли з бази даних необхідно вибрати дані, що мають неточний або нечітко визначений характер, стандартними засобами SQL це виконати неможливо.

Традиційно мова SQL використовується для опису запитів до БД в спеціальних термінах. Останнім часом багато уваги приділяється різним способам формалізації нечітких запитів. Звичайні сховища чітких даних з підтримкою чітких запитів називають чіткими базами даних (Crisp DataBases), а в реляційних базах даних з обробкою слабо формалізованих запитів все частіше вживаються терміни «нечіткі бази даних» (Fuzzy DataBases), «нечіткі запити» (Fuzzy Queries for DataBases). В БД з підтримкою механізмів нечітких запитів використовується підхід, що базується на використанні нечітких множин Заде [5] і нечітких відношень з підтримкою формул нечіткої логіки. Підхід, який було запропоновано у 1984 році, надалі отримав розвиток в роботах Ягера Р. [6], Дюбуа Д. та Прада Г. [7].

Означення 1. [5] Нечіткою множиною \tilde{A} в універсальній множині X називається сукупність пар виду

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x))\}, \quad (1)$$

де $x \in X$, а $\mu_{\tilde{A}} : X \rightarrow [0,1]$ - функція належності нечіткої множини \tilde{A} .

Означення 2. Нечітким відношенням \bar{F} на універсальному просторі $X \times Y$ називається нечітка множина декартового добутку $X \times Y$, що характеризується функцією належності $\mu_{\bar{F}} : X \times Y \rightarrow [0,1]$.

Виходячи з традиційних понять, представлень та операцій з нечіткими множинами, в існуючих варіантах нечітких баз даних реалізується підтримка виконання нечітких запитів до чітких даних сховища. Продукційні правила нечіткої бази знань формулюються на основі реалізації відомих формул нечіткої логіки, а база даних про проблемну область подається у вигляді типового чіткого сховища інформації.

Покажемо, що поняття нечітких баз даних може бути розширено та узагальнено. В багатьох сучасних інформаційних системах, окрім нечітких запитів, приходиться мати справу з неточною, нечіткою інформацією. Це можуть бути дані про результати експериментальних досліджень на основі вимірювань; такими є дані для систем підтримки прийняття рішень, в яких зберігається і ведеться інформація, наприклад, про рівні компетентності експертів, важливість критеріїв тощо. Нарешті, це можуть бути дані з чітких таблиць БД, що оформлюються у вигляді представлення, яке доповнюється полями величин відповідності значень кожного поля усіх кортежів деяким динамічним критеріям.

Іншими словами, для довільної чіткої бази даних може бути побудоване нечітке представлення, в якому інформація про характеристики (поля) об'єктів проблемної області подаються парою стовбців: у першому – конкретні значення показника з таблиці чіткої бази даних (вік, показання вимірювального приладу, номер зайнятого місця, значення критерія якості і т.і.), у другому – значення відповідності даних першого стовця деякому правилу (нечіткому поняттю). Ця процедура має назву фазифікації чітко визначених даних [8].

Зрозуміло, що результат процесу фазифікації чіткого відношення $R = \{\langle d_1, d_2, \dots, d_n \rangle\}$, можна виразити у вигляді R -таблиці. Це нове нечітке відношення

$$\bar{R} = \{\langle d_1, d_2, \dots, d_i | \mu_{\tilde{A}_i}(d_i), \dots, d_n \rangle\}, \quad (2)$$

що формується як множина кортежів з $n+1$ елемента, в яких величини $\mu_{\tilde{A}_i}(d_i)$ задають ступінь належності значень елементів d_i деякій нечіткій множині $\tilde{A}_i = \{(d, \mu_{\tilde{A}_i}(d)), d \in D_i\}$.

Таким чином, список базових та похідних

операцій над R -таблицями може бути розширений операцією фазифікації як різновидом операції розширення.

Потрібно зауважити, що відношення \bar{R} , сформоване за i -тим атрибутом, $i = \overline{1, n}$, буде представлятися у вигляді таблиці з p додатковими стовбцями, якщо на i -ий атрибут накладено p нечітких понять. Таким чином, відношення \bar{R} за i -им атрибутом можна представити у вигляді таблиці з $n+p$ стовбцями

$$\bar{R} = \{ \langle d_1, d_2, \dots, d_i | \mu_{\bar{A}_i^1}(d_i) | \mu_{\bar{A}_i^2}(d_i) | \dots | \mu_{\bar{A}_i^p}(d_i), \dots, d_n \rangle \}.$$

Остаточно, якщо частина доменів (крім i -ого) не потребує фазифікації, то відношення \bar{R} формально можна записати у вигляді

$$\bar{R} = \{ \langle d_1 | 1, d_2 | 1, \dots, d_i | \mu_{\bar{A}_i^1}(d_i) | \mu_{\bar{A}_i^2}(d_i) | \dots | \mu_{\bar{A}_i^p}(d_i), \dots, d_n | 1 \rangle \}.$$

Означення 3 [9]. Нечітку множину \bar{R} , яка визначає нечітке табличне відношення

$$\bar{R} = \{ \langle d_1 | \mu_{\bar{A}_1}(d_1) | \dots | \mu_{\bar{A}_{p_1}}(d_1), d_2 | \mu_{\bar{A}_2}(d_2) | \dots | \mu_{\bar{A}_{p_2}}(d_2), \dots, d_i | \mu_{\bar{A}_i}(d_i) | \dots | \mu_{\bar{A}_p}(d_i), \dots, d_n | \mu_{\bar{A}_n}(d_n) | \dots | \mu_{\bar{A}_{p_n}}(d_n) \rangle \}, \quad (3)$$

назвемо складеною нечіткою множиною в універсальному просторі $D_1 \cup \dots \cup D_i \dots \cup D_n$.

Означення 4. Базу даних інформаційної системи будемо називати нечіткою (НБД), якщо для представлення даних предметної області використовується хоча б одне нечітке відношення \bar{R} (3) у вигляді R -таблиці або R -представлення.

Розглянемо приклад НБД інформаційної системи. Розглянемо базу даних про нафтові родовища, інформація про які містить слабо-формалізовані поняття «відносна потужність» та «перспективність розробки». Ці поняття дають можливість оцінити інвестору привабливість проекту розробки кожного родовища з точки зору очікуваного прибуток від інвестування. Процедура фазифікації базується на результатах обробки експертних даних про характеристики запасів родовищ. Експерти на основі наданої їм інформації проводять оцінку об'ємів нафти, що знаходиться у надрах родовищ. Відносна потужність вимірюється у відсотках відносно потужності найбільших діючих родовищ України

(наприклад, Лесяківського [10]) на початок їх розробки. Цей підхід дає можливість оцінити величину покладів нафти у відсотковому еквіваленті, а також спрогнозувати, як довго дане нафтове родовище зможе функціонувати у режимі виробничого добутку. Перспективність розробки – інший якісний параметр, що враховує декілька позицій, які впливають на позитивність оцінки вкладання інвестицій у розвиток добутку нафти на заданій території. Серед них, наприклад, зручність розташування родовища відносно основних шляхів транспортування нафти, сприятливість природних умов для будівництва (рельєф, кліматичні умови, глибина залягання покладів, загроза несприятливих природних явищ), інвестиційна привабливість регіону в економічному сенсі та інші. Для числового оцінювання згаданих параметрів використовується шкала від 0 до 1.

Не обмежуючи загальності визначення НБД, будемо вважати, що довільна інформаційна система базується на нечітких відношеннях \bar{R} , які подаються у вигляді сукупності кортежів

$$\bar{R} = \{ \langle d_1 | \mu_{\bar{A}_1}(d_1) | \dots | \mu_{\bar{A}_n}(d_2), \dots, d_i | \mu_{\bar{A}_i}(d_i) | \dots | \mu_{\bar{A}_n}(d_i), \dots, d_m | \mu_{\bar{A}_1}(d_m) | \dots | \mu_{\bar{A}_n}(d_m) \rangle \}, \quad (4)$$

де $d_i \in D_i \subset R_+^1$, $i = \overline{1, m}$, m - кількість родовищ, належність кожної величини $\mu_{\bar{A}_j}(d_i)$, $j = \overline{1, n}$,

$i = \overline{1, m}$, до відповідної нечіткої множини обчислюється за допомогою трапецієподібних функцій, які мають вигляд

$$\mu_{\bar{A}_j}(d_i) = \begin{cases} 1 - (b_j - d_i)/(b_j - a_j), & a_j \leq d_i < b_j, \\ 1, & b_j \leq d_i \leq c_j, \\ 1 - (d_i - c_j)/(e_j - c_j), & c_j < d_i \leq e_j, \\ 0, & \text{інакше,} \end{cases} \quad (5)$$

з заданими параметрами a_j, b_j, c_j, e_j , $j = \overline{1, n}$, що визначаються експертами та користувачами.

В якості середовища розробки використано СУБД Access, одну з середньопотужних СУБД, представлених сьогодні на ринку. Система Access – це набір потужних інструментів кінцевого користувача для керування базами даних. До її складу входять конструктори таблиць, форм, запитів і звітів. Цю систему досить зручно розглядати і як середовище розробки додатків. Використовуючи макроси або модулі для автоматизації розв'язання завдань,

можна створювати орієнтовані на користувача додатки, що є такими ж потужними, як і додатки, написані безпосередньо мовами програмування.

При цьому, слід зауважити про відносно складні дії з реструктуризації таблиць даних для отримання фазифікованих даних. Однак, незважаючи на окремі незручності розробки, створено засоби маніпуляції (актуалізації) таблиць та інтерфейс користувача, що забезпечив наповнення, модифікацію бази даних та пошук необхідних даних з урахуванням визначених параметрів функцій належності. Усі розроблені засоби оформлено у вигляді сукупності макросів та запитів до таблиць БД з вхідними параметрами.

Наведений приклад нечіткої БД демонструє конструктивність використання підходу для отримання нечітко визначених даних на основі оперативного формування показників належності даних відповідним нечітким множинам. Запропонований підхід забезпечує можливість створення та обробки слабоформалізованої інформації, дозволяє розробити спосіб отримання кортежів нечіткого відношення нечіткої бази даних на основі формування SQL-запитів до нечітких даних і процедур їх реалізації. Для їх виконання використовуються різні засоби, що дозволяють записати необхідний запит у вигляді спеціальної послідовності операторів, яка потім транслюється у відповідну послідовність SQL-запитів [9].

Список використаних джерел

1. *Pascal, F.* Understanding Relational Data-bases with Examples in SQL-92 / F. Pascal. – John Wiley & Sons, 1993.
2. *Won K.* On Object-Oriented Database Technology / K.Won // Open systems. – 1994. – №4 (8). – P. 14.
3. *Codd, E. F.* An Evaluation Scheme for Database Management Systems that are Claimed to be Relational / E.F.Codd // Proc. IEEE CS International Conference on Data Eng., LA, USA, February, 1986.
4. *Dudly, K.* SQL compliance / K.Dudly // World of Oracle. – 1996. – №1(39). – P. 7-16.
5. *Орловский С.А.* Проблемы принятия решения при нечеткой исходной информации/ С.А.Орловский. – М.: Наука, 1981. – 206с.
6. *Yager, R.R.* Fuzzy sets, probabilities and decision / R.R.Yager // J.of Cybern. – 1980. – №10. – P.1-18.
7. *Дюбуа Д.* Теория возможностей. Приложения к представлению знаний в информатике/ Д. Дюбуа, А.Прад. – М.: Радио и связь, 1990. – 287с.
8. *Осовский С.* Нейронные сети для обработки информации / С.Осовский – М.:Финансы и статистика, 2002. – 344с.
9. *Івохін Є.В.* Про підхід до реалізації нечітких баз даних / Є.В. Івохін, К.О.Косинський // Вісник Київського університету. Серія: фіз.-мат. науки. – 2008. – Вип.2. – С. 83 – 87.
10. *Ладийчук Д.О., Пічура В.І.* Геоінформаційні бази даних. – Херсон: Вид-во ХДУ, 2007. – 103 с.

References

1. PASCAL, F. (1993) *Understanding Relatio-nal Data-bases with Examples in SQL-92.* John Wiley & Sons.
2. WON K. (1994) *On Object-Oriented Database Technology*// Open systems. №4 (8). p. 14.
3. CODD, E. F. (1986) *An Evaluation Scheme for Database Management Systems that are Claimed to be Relational* //Proc. IEEE CS International Conference on Data Eng., LA, February, 1986.
4. DUDLY, K. (1996) *SQL compliance*// World of Oracle. № 1 (39). p. 7-16.
5. ORLOVSKIY, S. (1981) *Problemy prinyatiya reshenyya pri nechetkoy ishodnoy informaciyi.* Moskva: Nauka.
6. YAGER, R. (1980) *Fuzzy sets, probabilities and decision* // J.of Cybern. №10. p.1-18.
7. DUBOIS, D. and PRADE, H. (1988) *Theorie des possibilites. Applications a la representation des connaissances en informatique.* Paris: MASSON.
8. OSOVSKIY, S. (2002) *Neyronnyye seti dlya obrabotki informaciyi.* Moskva: Finansy i statistika.
9. IVOHIN, E. and KOSINSKIY, K. (2008) About realization of fuzzy databases // Bulletin of Taras Shevchenko National University of Kyiv. Series: Physics&Mathematics. – 2. – pp. 83 – 87.
10. LADYCHUK, D. and PICHURA, V. (2007) *Geoinformaciyni bazy danykh.* Kherson: Kherson State Univ.

Надійшла до редколегії 28.02.2014