

УДК 517.9

Івохін Є.В.¹, д.ф.-м.н, доцент,
Вадньов Д.О.², пошукач

Ivohin E.V.¹, Dr.Sci., associate professor,
Vadnyev D.O.², researcher

Про реалізацію кількісних запитів в нечітких базах даних

About realization quantitative requests in fuzzy databases

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

^{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 формування, накопичення та актуалізації нечітких баз даних. Ступінь нечіткості атрибутів таблиць розраховується на основі чітких даних у вигляді спеціального представлення, яке створюється за допомогою операції розширення. Пропонується підхід для реалізації кількісних запитів до нечітких баз даних.

Ключові слова: реляційні бази даних, нечіткі множини, нечіткі трикутні числа, мова запитів, динамічні представлення, відношення переваги.

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 dataset, fuzzy triangle numbers, query language, dynamic view representations, relations of preference.

Статтю представив доктор технічних наук, с.н.с. Кудін В.І.

Сучасну інформаційну систем важко представити 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. У загальному випадку характерними категоріями, що формують таку модель, є об'єкти та відношення між ними. Такий підхід дозволяє визначити базу даних (БД) як об'єктно-орієнтований опис моделі, що базується на поняттях об'єктів та відношень, з повним або частковим фізичним розташуванням змістовних даних предметної

області на комп'ютерних носіях.

Використання моделі даних при роботі з БД дає можливість відображати особливості предметної області. За час існування розробок інформаційних систем запропоновано багато різних моделей з власними перевагами та недоліками.

Найбільш популярною є реляційна модель предметної області. В реляційній моделі вважається, що всі дані БД інформаційної системи представлені у вигляді таблиць [2]. Рядок кожної таблиці представляє 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). Множини значень різних стовбців в одній і тій

самій таблиці або в різних таблицях можуть задаватися одним доменом, а можуть – і різними. При чому, значення атрибутів в таблиці-відношенні мають лише один конкретний вигляд функціональної залежності один від іншого. Іншими словами, всі значення в довільному кортежі повинні окремо залежати лише від значення стовпця або групи стовбців – одних для всього відношення. Такий стовбець або група стовбців називається ключовими, а значення атрибутів в них – ключами.

В реляційній теорії визначений список операцій, які можна здійснювати над R -таблицями, отримуючи результат у вигляді R -таблиці. До таких операцій над базою даних відносять:

1) базові операції: обмеження; проекція; декартовий добуток двох таблиць; об'єднання; різниця; присвоювання (іменованій таблиці присвоюється значення виразу над R -таблицями).

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

Крім постійних таблиць, що визначені в моделі інформаційної системи, операції над R -таблицями дозволяють отримувати таблиці-представлення, що дають можливість динамічно формувати набори кортежів з додатково заданими правилами їх формування.

Таким чином, констатуємо, що будь-яка прикладна проблемна область описується в термінах таблиць достатньо просто, що дозволяє суттєво спростити розробку інформаційних систем. Реляційна концепція баз даних може бути легко формалізована, базуючись на математичній моделі відношень. Якщо задано n множин D_1, D_2, \dots, D_n , тоді реляція R є множиною впорядкованих наборів виду $\langle d_1, d_2, \dots, d_n \rangle$, де d_1 - елемент з D_1 , ..., 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.

Ідея реляційних баз даних стала популярною і серед розробників СУБД. Виходячи з того, що більша частина даних, які обробляються в інформаційних системах, носять кількісний характер, зростає зацікавленість в створенні способів та механізмів отримання конкретних даних. Універсальним засобом доступу до сховищ даних є мова запитів SQL, яка дозволяє зручно маніпулювати інформацією в різних реалізаціях моделей баз даних.

Мова SQL використовується для опису запитів до БД в спеціальних термінах. Припустимо, що потрібно отримати відомості про догово-

ри з усіма фірмами заданого регіону, в яких сума кошторисів складає приблизно 900000 гривень. Перелік можна отримати за допомогою наступного запиту мовою SQL:

*Select AgreementID from Agreements
Where Agreement.Sum >= 900000 and
Agreement.FirmRegion = 1.*

Запити до БД формуються у вигляді чітких вимог до шуканих даних. Чіткі вимоги дозволяють отримувати конкретні результати пошуку в БД. Однак в запитах, що формулює користувач, досить часто присутні неточності та невизначеності. Якщо повернутися до наведеного прикладу, то договори з сумою 899000 гривень не попадають в результат виконання запиту, хоча їх характеристики є дуже близькими до вимог пошуку. Для забезпечення включення таких договорів в результат пошуку у запиті потрібно змінювати величини сум укладених договорів.

Зрозуміло, що в загальному випадку, коли з бази даних необхідно вибрати лінгвістично задану інформацію («молодий», «приблизно» «висока ефективність» і т.і.), реалізація такого запиту можлива лише за умов формалізації понять, які відносяться до категорії неточних, невизначених даних. Такі запити неможливо виконати стандартними засобами SQL.

Останнім часом багато уваги приділяється реалізації нечітких запитів. При цьому, сховища чітких даних з підтримкою чітких запитів були виділені групу так званих Crisp DataBases. В той же час, в реляційних базах даних з підтримкою нечітких запитів широко вживаються терміни «нечіткі бази даних» (Fuzzy DataBases). В БД з підтримкою механізмів нечітких запитів використовується підхід, що базується на ідеях нечітких множин Заде [3] і нечітких відносин з підтримкою формул нечіткої логіки.

Означення 1. Нечіткою множиною \tilde{A} в універсальному просторі X називається сукупність пар вигляду $\{(x, \mu_{\tilde{A}}(x))\}$, де $x \in X$, а $\mu_{\tilde{A}}(x) : X \rightarrow [0,1]$ - функція належності нечіткої множини \tilde{A} .

Звичайна множина A з X може бути представлена у вигляді нечіткої множини, яка визначається характеристичною функцією:

$$\chi_A(x) = \begin{cases} 0 & x \notin A \\ 1 & x \in A \end{cases} \quad (1)$$

Означення 2. [4] Нечітким трикутним числом \tilde{b} називають впорядковану трійку чисел (a, b, c) , $a \leq b \leq c$, для якої функція належності $\mu_{\tilde{b}}(x)$ має вигляд:

$$\begin{aligned}\mu_{\tilde{a}}(x) &= (x - a)/(b - a), x \in [a, b], \\ \mu_{\tilde{b}}(x) &= (c - x)/(c - b), x \in [b, c], \\ \mu_{\tilde{b}}(x) &= 0, x \notin [a, c].\end{aligned}\quad (2)$$

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

В роботі [5] запропоновано узагальнення поняття нечітких баз даних. В багатьох сучасних інформаційних системах, окрім нечітких запитів, приходить мати справу з неточною, нечіткою інформацією. Це можуть бути дані про результати експериментальних досліджень на основі вимірювань з різними величинами точності (похибками); 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. Наприклад, для запиту про договори можна сформувати таблицю-представлення з полем «сума договору» (A), до якого занести значення міри відповідності суми категорії «приблизно» 900000. В термінах нечітких множин, ці значення є величинами міри належності конкретних числових даних $x, x \in A$ до нечіткої множини $\tilde{Y} = \{(x, \mu_{\tilde{Y}}(x))\}$ з функцією належності $\mu_{\tilde{Y}}(x) \in [0, 1]$ для заданого числового значення суми. Формування нового представлення у цьому випадку можна записати, наприклад, у вигляді послідовності операторів мовою Linq [6], які потім транлюються у відповідну послідовність SQL-запитів.

Іншими словами, для довільної чіткої бази даних може бути побудоване нечітке представлення (нечітка БД, НБД), в якому інформація про характеристики (поля) об'єктів проблемної області подаються парою стовбців: в першому – конкретні значення показника з таблиці чіткої бази даних (вік, показання вимірювального приладу, номер зайнятого місця, значення критерія якості і т.і.), в другому – значення відповідності даних першого стовбця деякому

правилу (нечіткому поняттю). Ця процедура має назву фаззифікації чітко визначених даних.

Зрозуміло, що результат процесу фаззифікації чіткого відношення $R = \{ \langle d_1, d_2, \dots, d_n \rangle \}$, де d_1 - елемент з D_1 , d_2 - елемент з D_2 , ..., d_n - елемент з D_n за i -тим атрибутом, $i = \overline{1, n}$, можна виразити у вигляді \bar{R} -таблиці. Це нове нечітке відношення

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

що формується як множина кортежів з $n+1$ елемента, в яких величини $\mu_{\tilde{A}_i}(d_i)$ задають міру 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 фаззифікації 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 -тий атрибут накладено p нечітких понять. Наприклад, для наведеного вище випадку, крім нечіткої формалізації поняття «приблизно» аналогічно будується представлення нечіткого поняття «суттєво більше» і т.і. Таким чином, відношення \bar{R} за i -тим атрибутом можна представити у вигляді таблиці з $n+p$ стовбцями

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

Частина доменів, що не потребує фаззифікації, тобто є чіткими множинами з функцією належності у вигляді характеристичної функції (1), у відношенні \bar{R} може бути представлена у вигляді

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

Визначення 3. Нечітку множину \bar{R} , яка визначає нечітке табличне відношення

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

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

Визначення 4. [7] Базу даних інформаційної системи будемо називати нечіткою (НБД), якщо для формалізації даних предметної області використовується хоча б одне нечітке відношення \bar{R} (6) у вигляді R -таблиці або R -представлення.

Не обмежуючи загальності, вважатимемо, що у НБД нечітке відношення \bar{R} представлено у вигляді сукупності кортежів $\bar{R} = \{ \langle d_1 | \mu_{\bar{A}_1}(d_1),$

$d_2 | \mu_{\bar{A}_2}(d_2), \dots, d_i | \mu_{\bar{A}_i}(d_i), \dots, d_n | \mu_{\bar{A}_n}(d_n) \rangle \}$,

$d_i \in D_i \subset R_+^1, i = \overline{1, n}$. З практичної точки зору

постає важливе питання про формування запитів та реалізацію механізмів отримання даних з НБД.

Для його розв'язання пропонується підхід на основі використання нечітких трикутних чисел для визначення діапазону значень, що реалізує задане нечітке поняття. У наведеному вище випадку відбір договорів з нечітким атрибутом «приблизно» може бути здійснений за допомогою SQL-оператора

*Select AgreementID from Agreements
Where Agreement.FirmRegion=1 and
Agreement.Sum between Amin and Amax,*

Список використаних джерел

1. *Pascal F.* Understanding relational databases with examples in SQL-92. Wiley & Sons, 1993. 278pp.
2. *Codd E. F.* An evaluation scheme for database management systems that are claimed to be relational // Proc. of second International Conf. on Data Engineering, Los Angeles, February, 1986. – P.720-729.
3. *Zadeh L.A.* Fuzzy sets // Inf. Contr., 1965. – 8. – P.338-353.
4. *Bablu Jana, Tapan Kumar Roy.* Multi-objective fuzzy linear programming and its application in transportation model // Tamsui Oxford Journal of Math. Sc. – 2005. – V.21(2). – P.243-268.
5. *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.
6. *Fabrice Marguerie, Steve Eichert, Jim Wooley.* LINQ in Action. Manning Publ., 2008. 600 p.
7. *Івохін Є.В.* Про нечітке представлення дійсних чисел у формі триплетів// Вісник Київського університету. Серія: фіз.-мат. науки. – 2014. – №3. – С.118-121.

де $Amin$ та $Amax$ – мінімальне та максимальне значення інтервалу представлення суми як нечіткого числа. Методика подання деякої величини в умовах невизначеності з допомогою її нечіткого представлення на основі нечітких трикутних чисел та використання спеціальних послідовностей простих чисел запропоновано і детально викладено у [8].

Проведені чисельні експерименти дозволили говорити про конструктивність і ефективність розглянутого способу. Слід також зазначити, що у випадках, коли діапазон представлення не задовольняє користувача, відбір кортежів можна здійснювати на основі суб'єктивного визначення нечіткого трикутного числа.

Наведений підхід до використання нечітко визначених даних на основі формування нечітких представлень і формалізації нечітких операцій вибору за кількісними атрибутами надає можливість створення та обробки слабоформалізованої інформації, визначає спосіб пошуку кортежів нечіткого відношення нечіткої бази даних, які формують результати реалізації SQL-запитів до нечітких даних.

References

1. *PASCAL F.* (1993) *Understanding Relational Databases with Examples in SQL-92.* John Wiley & Sons.
2. *CODD E. F.* (1986) *An Evaluation Scheme for Database Management Systems that are Claimed to be Relational.* Proc. of second International Conference on Data Engineering, Los Angeles. – P.720-729.
3. *ZADEH L.A.* (1965) *Fuzzy sets.* Inf. Contr. – 8. – P.338-353.
4. *BABLU JANA and TAPAN KUMAR ROY.* (2005) *Multi-objective fuzzy linear programming and its application in transportation model.* Tamsui Oxford Journal of Math. Sc. – V.21 (2). – P.243-268.
5. *IVOHIN E. and KOSINSKIY K.* (2008) *About realization of fuzzy databases* // Bulletin of Taras Shevchenko National University of Kyiv. Series: Phys.&Math. – 2. – P. 83 – 87.
6. *MARGUERIE F., EICHERT S. and WOOLEY J.* (2008). *LINQ in Action.* Manning Publ.
7. *IVOHIN E.* (2014) *About fuzzy real numbers presentation in triplet form* // Bulletin of Taras Shevchenko National University of Kyiv. Series: Phys.&Math. – 3. – P. 118 – 121.