ДОПОВІДІ
НАЦІОНАЛЬНОЇ
АКАДЕМІЇ НАУК
УКРАЇНИ

6 • 2015

ІНФОРМАТИКА ТА КІБЕРНЕТИКА

УДК 004.655

Академік НАН України **В. Н. Редько, Д. Б. Буй, А. В. Пузікова**

Аксіоматика багатозначних залежностей табличних баз даних

Розглядаються аксіоматика багатозначних залежностей в табличних базах даних і аксіоматика функціональних та багатозначних залежностей; встановлюється повнота цих аксіоматик через збіжність відношень синтаксичного та семантичного прямування; наводяться критерії повноти вказаних аксіоматик в термінах потужностей універсального домену та множини атрибутів.

Ключові слова: табличні бази даних, функціональні залежності, багатозначні залежності, повнота аксіоматики, критерій повноти.

Аксіоматика багатозначних залежностей. Всі неозначувані поняття та позначення використовуються в розумінні [1]. Зокрема, скажемо, що на таблиці t схеми R виконується багатозначна залежність (БЗЗ) $X \rightarrow\rightarrow Y$, якщо для двох довільних рядків s_1, s_2 таблиці t, які збігаються на множині атрибутів X, існує рядок $s_3 \in t$, який дорівнює об'єднанню обмежень рядків s_1, s_2 на множини атрибутів $X \bigcup Y$ і $R \setminus (X \bigcup Y)$ відповідно:

$$(X \rightarrow\rightarrow Y)(t) = \text{true} \overset{\text{def}}{\Leftrightarrow} \forall\, s_1, s_2 \in t(s_1 \mid X = s_2 \mid X \Rightarrow$$
$$\Rightarrow \exists\, s_3 \in t(s_3 = s_1 \mid (X \bigcup Y) \bigcup s_2 \mid R \setminus (X \bigcup Y)))\ [1].$$

Структура таблиці t, на якій задана БЗЗ $X \rightarrow\rightarrow Y$, може бути зображена за допомогою такого відношення. Скажемо, що рядки s_1, s_2 таблиці t знаходяться у відношенні $=_X$, якщо вони збігаються на множині атрибутів X:

$$s_1 =_X s_2 \overset{\text{def}}{\Leftrightarrow} s_1 \mid X = s_2 \mid X.$$

Зрозуміло, що відношення $=_X$ є відношенням еквівалентності і тому розбиває множину рядків таблиці t на класи еквівалентності, які мають таке зображення: $[s]_{=_X} = \{s \mid X\} \otimes \otimes\, \pi_Y([s]_{=_X}) \otimes \pi_{R\setminus(X\cup Y)}([s]_{=_X})$, де s — довільний представник класу.

Таблиця $t(R)$ (таблиця t схеми R) називається *моделлю* множини БЗЗ G, якщо кожна БЗЗ $X \to\to Y \in G$ виконується на таблиці $t(R)$:

$$t(R) \text{ модель } G \overset{\text{def}}{\Leftrightarrow} \forall\, (X \to\to Y)(X \to\to Y \in G \Rightarrow (X \to\to Y)(t) = \text{true}).$$

Виконуються такі аксіоми і правила виведення [2].

Аксіома рефлексивності: $\forall\, t(X \to\to Y)(t) = \text{true}$, де $Y \subseteq X$.

Аксіома: $\forall\, t(X \to\to Y)(t) = \text{true}$, де $X \bigcup Y = R$.

Правило повноти: $(X \to\to Y)(t) = \text{true} \Rightarrow (X \to\to R \setminus (X \bigcup Y))(t) = \text{true}$.

Правило поповнення: $(X \to\to Y)(t) = \text{true} \& Z \subseteq W \Rightarrow (X \bigcup W \to\to Y \bigcup Z)(t) = \text{true}$.

Правило транзитивності:

$$(X \to\to Y)(t) = \text{true} \& (Y \to\to Z)(t) = \text{true} \Rightarrow (X \to\to Z \setminus Y)(t) = \text{true}.$$

БЗЗ $X \to\to Y$ семантично прямує з множини БЗЗ G за означенням, якщо на кожній таблиці $t(R)$, яка є моделлю множини БЗЗ G, виконується також БЗЗ $X \to\to Y$: $G| {=} X \to\to Y \overset{\text{def}}{\Leftrightarrow} \forall\, t(R)(t \text{ модель } G \Rightarrow (X \to\to Y)(t) = \text{true})$.

З наведених аксіом і правил виведення випливають такі наслідки:

1) $\varnothing | {=} X \to\to Y$ для $Y \subseteq X$;

2) $\varnothing | {=} X \to\to Y$ для $X \bigcup Y = R$;

3) $G| {=} X \to\to Y \Rightarrow G| {=} X \to\to R \setminus (X \bigcup Y)$;

4) $G| {=} X \to\to Y \& Z \subseteq W \Rightarrow G| {=} X \bigcup W \to\to Y \bigcup Z$;

5) $G| {=} X \to\to Y \& G| {=} Y \to\to Z \Rightarrow G| {=} X \to\to Z \setminus Y$;

6) $G| {=} X \to\to Y \& G| {=} Y \to\to Z \& Z \bigcap Y = \varnothing \Rightarrow G| {=} X \to\to Z$.

БЗЗ $X \to\to Y$ синтаксично прямує з множини БЗЗ G відносно схеми R за означенням ($G|{-}_R X \to\to Y$), якщо існує скінченна послідовність БЗЗ $\varphi_1, \varphi_2, \ldots, \varphi_{m-1}, \varphi_m$, така, що $\varphi_m = X \to\to Y$ і для $\forall\, i = \overline{1, m-1}$ кожна φ_i є або аксіома рефлексивності, або належить G, або отримана за будь-яким правилом виведення для БЗЗ з попередніх у цій послідовності БЗЗ φ_j, φ_k, $j, k < i$. Згідно з традиціями математичної логіки, послідовність $\varphi_1, \varphi_2, \ldots, \varphi_m$ назвемо *доведенням* [3].

Нехай задана множина БЗЗ G. *Замиканням* $[G]_R$ називається множина БЗЗ, які синтаксично прямують з G відносно схеми R:

$$[G]_R \overset{\text{def}}{=} \{X \to\to Y | G|{-}_R X \to\to Y\}.$$

З наведених властивостей випливає, що, за термінологією [4], оператор $G \mapsto [G]$ є оператором замикання:

1) $G \subseteq [G]$ (*зростання*);

2) $[[G]] = [G]$ (*ідемпотентність*);

3) $G \subseteq H \Rightarrow [G] \subseteq [H]$ (*монотонність*).

З описаних вище аксіоми рефлексивності і правил виведення будуються доведення похідних правил виведення для БЗЗ [2, 5].

Аксіоматика багатозначних і функціональних залежностей. Нагадаємо, що на таблиці t виконується функціональна залежність (ФЗ) $X \to Y$, якщо для двох довільних рядків s_1, s_2 таблиці t, які збігаються на множині атрибутів X, має місце їх рівність і на множині атрибутів Y:

$$(X \to Y)(t) = \text{true} \quad \overset{\text{def}}{\Leftrightarrow} \quad \forall\, s_1, s_2 \in t(s_1|X = s_2|X \Rightarrow s_1|Y = s_2|Y)\ [1].$$

Нехай задано множини ФЗ F і БЗЗ G. Таблиця $t(R)$ є *моделлю* множини $F \bigcup G$, якщо кожна залежність $\varphi \in F \bigcup G$ виконується на таблиці t:

$$t(R) \text{ модель } F \bigcup G \quad \overset{\text{def}}{\Leftrightarrow} \quad \forall\, \varphi(\varphi \in F \bigcup G \Rightarrow \varphi(t) = \text{true}).$$

ФЗ або БЗЗ φ *семантично* прямує з множини $F \bigcup G$ відносно схеми R, якщо на кожній таблиці $t(R)$, яка є моделлю множини залежностей $F \bigcup G$, виконується також залежність φ: $F \bigcup G |{=} \varphi \overset{\text{def}}{\Leftrightarrow} \forall\, t(R)(t(R) -$ модель $F \bigcup G \Rightarrow \varphi(t) = \text{true})$.

Для ФЗ та БЗЗ виконуються спільні правила виведення [2].

1) Правило *розширення* ФЗ до БЗЗ: $(X \to Y)(t) = \text{true} \Rightarrow (X \to\to Y)(t) = \text{true}$.

2) $(X \to\to Z)(t) = \text{true} \& (Y \to Z')(t) = \text{true} \& Z' \subseteq Z \& Y \bigcap Z = \varnothing \Rightarrow (X \to Z')(t) = \text{true}$.

Наслідками з наведених правил є такі властивості відношення семантичного прямування:

1) $F |{=} X \to Y \Rightarrow F |{=} X \to\to Y$;

2) $G |{=} X \to\to Z \& F |{=} Y \to Z' \& Z' \subseteq Z \& Y \bigcap Z = \varnothing \Rightarrow F \bigcup G |{=} X \to Z'$.

ФЗ або БЗЗ φ *синтаксично* прямує з об'єднання множин ФЗ і БЗЗ $F \cup G$ відносно схеми R за означенням $(F \bigcup G |{-}_R \varphi)$, якщо існує скінченна послідовність ФЗ або БЗЗ $\varphi_1, \varphi_2, \ldots, \varphi_{m-1}, \varphi_m$, така, що $\varphi_m = \varphi$ і для $\forall\, i = \overline{1, m-1}$ кожна φ_i є або аксіома рефлексивності (для ФЗ або БЗЗ), або належить $F \cup G$, або отримана за будь-яким правилом виведення (повноти (для БЗЗ), поповнення (для ФЗ або БЗЗ), транзитивності (для ФЗ або БЗЗ), спільних правил для ФЗ і БЗЗ) з попередніх у цій послідовності ФЗ або БЗЗ φ_j, φ_k, $j, k < i$. Послідовність $\varphi_1, \varphi_2, \ldots, \varphi_{m-1}, \varphi_m$ називається доведенням φ з об'єднання множин $F \bigcup G$.

Нехай задано деякі множини F і G ФЗ та БЗЗ відповідно. *Замикання* $[F \bigcup G]_R$ — це множина усіх ФЗ і БЗЗ, які синтаксично прямують з $F \bigcup G$ відносно схеми R: $[F \bigcup G]_R \overset{\text{def}}{=} \{\varphi | F \bigcup G |{-} \varphi\}$.

Виконуються властивості:

1) $F \bigcup G \subseteq [F \bigcup G]$ (*зростання*);

2) $[[F \bigcup G]] = [F \bigcup G]$ (*ідемпотентність*);

3) $F' \bigcup G' \subseteq F \bigcup G \Rightarrow [F' \bigcup G'] \subseteq [F \bigcup G]$ (*монотонність*);

4) $[F] \subseteq [F \bigcup G]$, $[G] \subseteq [F \bigcup G]$, $[F] \bigcup [G] \subseteq [F \bigcup G]$.

З властивостей *1–3* випливає, що, за термінологією [4], оператор $F \bigcup G \mapsto [F \bigcup G]_R$ є оператором замикання.

Замиканням $[X]_{F \cup G, R}$ множини атрибутів X (відносно множини залежностей $F \bigcup G$ і схеми R) називається сім'я усіх множин атрибутів, що є правими частинами БЗЗ, які синтаксично прямують з множини $F \bigcup G$:

$$[X]_{F \cup G, R} \overset{\text{def}}{=} \{Y | X \to\to Y \in [F \bigcup G]_R\}.$$

Нехай $[X]_F$ — замикання множини атрибутів X відносно множини ФЗ F [5]. Мають місце властивості:

1) $Y \subseteq [X]_F \Rightarrow Y \in [X]_{F\cup G,R}$;

2) $[X]_{F\cup G,R} = [[X]_F]_{F\cup G,R}$.

Зауважимо, що оператор $X \mapsto [X]_{F\cup G,R}$ не є оператором замикання; для обгрунтування досить вказати, що даний оператор не має властивості ідемпотентності (поняття множини атрибутів і замикання множини атрибутів відносно об'єднання множин ФЗ і БЗЗ $F\bigcup G$ і схеми R мають різну природу, тому вираз $[[X]_{F\cup G,R}]_{F\cup G,R}$ не має сенсу в означених термінах).

Базисом $[X]^{\text{bas}}_{F\cup G,R}$ множини атрибутів X відносно множини залежностей $F\cup G$ і схеми R називається підсім'я замикання $[X]_{F\cup G,R}$, така, що:

1) $\forall W (W \in [X]^{\text{bas}}_{F\cup G,R} \Rightarrow W \neq \varnothing)$ (тобто базис містить тільки непорожні множини атрибутів);

2) $\forall W_i,\ W_j (W_i, W_j \in [X]^{\text{bas}}_{F\cup G,R} \Rightarrow W_i \cap W_j = \varnothing)$ (тобто множини атрибутів базису попарно не перетинаються);

3) $\forall Y \left(Y \in [X]_{F\cup G,R} \Rightarrow \exists i,\ 1 \leqslant i \leqslant n \left(Y = \bigcup\limits_{W_i\in[X]^{\text{bas}}_{F\cup G,R}} W_i\right)\right)$ (тобто кожна множина атрибутів з замикання $[X]_{F\cup G,R}$ є скінченним об'єднанням деяких множин атрибутів з базису).

Виконуються властивості:

1) $\bigcup\limits_{W\in[X]^{\text{bas}}_{F\cup G,R}} W = R$ для $X \subseteq R$ (тобто базис є розбиттям множини атрибутів R);

2) $A \in [X]_F \Rightarrow \{A\} \in [X]^{\text{bas}}_{F\cup G,R}$.

Нехай φ — ФЗ або БЗЗ.

Твердження 1 (коректність аксіоматики БЗЗ і ФЗ). *Якщо залежність φ синтаксично виводиться з множини залежностей $F\bigcup G$, то φ виводиться з $F\bigcup G$ семантично: $F\bigcup G|\!-\varphi \Rightarrow F\bigcup G|\!=\varphi$.*

Твердження 2 (повнота аксіоматики БЗЗ і ФЗ). *Якщо залежність φ семантично виводиться з множини залежностей $F\bigcup G$, то φ виводиться з $F\bigcup G$ синтаксично: $F\bigcup G|\!=\varphi \Rightarrow F\bigcup G|\!-\varphi$.*

Теорема 1 . *Відношення семантичного та синтаксичного прямувань для аксіоматики БЗЗ та ФЗ збігаються: $F\bigcup G|\!=\varphi \Leftrightarrow F\bigcup G|\!-\varphi$ за умови $|D| \geqslant 2$ та $|R| \geqslant 2$.*

Критерій повноти аксіоматики ФЗ і БЗЗ. Аналіз доведення основного результату про збіжність відношень синтаксичного та семантичного прямувань для аксіоматики БЗЗ та ФЗ (теорема 1) показує, що воно проведено в припущенні $|D| \geqslant 2$ та $|R| \geqslant 2$.

Залежність збіжності відношень синтаксичного та семантичного прямувань при різних значеннях потужностей множин R та D вказана відповідно у табл. 1 і 2. Символ "+" (відповідно "$-$") в комірці означає, що при вказаних припущеннях відношення $|\!=$ та $|\!-$ збігаються (не збігаються відповідно).

З наведених таблиць випливають такі основні результати.

Таблиця 1. Всі варіанти потужностей множин R та D для аксіоматики БЗЗ

D	R		
	$\|R\| = 0$	$\|R\| = 1$	$\|R\| \geqslant 2$
$\|D\| = 0$	+	+	$-$
$\|D\| = 1$	+	+	$-$
$\|D\| \geqslant 2$	+	+	+

Таблиця 2. Всі варіанти потужностей множин R та D для аксіоматики БЗЗ та ФЗ

D	R		
	$\|R\| = 0$	$\|R\| = 1$	$\|R\| \geqslant 2$
$\|D\| = 0$	+	—	—
$\|D\| = 1$	+	—	—
$\|D\| \geqslant 2$	+	+	+

Теорема 2. *Відношення семантичного* $\models$ *та синтаксичного* $\vdash$ *прямувань для аксіоматики БЗЗ і ФЗ збігаються тоді і тільки тоді, коли* $|D| \geqslant 2$ *або* $|R| = 0$.

Теорема 3. *Відношення семантичного* $\models$ *та синтаксичного* $\vdash$ *прямувань для аксіоматики БЗЗ збігаються тоді і тільки тоді, коли* $|R| \leqslant 1$ *або* $|R| \geqslant 2$ *та при цьому* $|D| \geqslant 2$.

Подальша робота полягає у встановленні взаємної незалежності складових наведених аксіоматик.

Цитована література

1. *Редько В. Н., Брона Ю. И., Буй Д. Б., Поляков С. А.* Реляційні бази даних: табличні алгебри та SQL-подібні мови. – Київ: ВД "Академперіодика", 2001. – 198 с.
2. *Beeri C., Fagin R., Howard J.* A complete axiomatization for functional and multivalued dependences // Proceedings of the ACM-SIGMOD Conference, August 3–5, 1977. – Toronto, Canada. – P. 47–61.
3. *Линдон Р.* Заметки по логике. – Москва: Мир, 1968. – 128 с.
4. *Скорняков Л. А.* Элементы теории структур. – Москва: Наука, 1982. – 160 с.
5. *Мейер Д.* Теория реляционных баз данных. – Москва: Мир, 1987. – 608 с.

References

1. *Redko V. N., Brona Yu. J., Buy D. B., Polyakov S. A.* Reliatsiini bazy danykh: tablychni alhebry ta SQL-podibni movy, Kiev: Akademperiodyka, 2001 (in Ukrainian).
2. *Beeri C., Fagin, R., Howard J.* In Proceedings of the ACM-SIGMOD Conf., August 3–5, 1977, Toronto, Canada: 47–61.
3. *Lyndon R.* Notes on Logic. D. Van Nostrand, Princeton, 1966.
4. *Skornyakov L. A.* Elementy teorii struktur, Moskva: Nauka, 1982 (in Russian).
5. *Maier D.* The theory of relational databases, Computer Science Press, 1983.

Київський національний університет
ім. Тараса Шевченка

Надійшло до редакції 15.01.2015

Академик НАН Украины **В. Н. Редько, Д. Б. Буй, А. В. Пузикова**

Аксиоматика многозначных зависимостей табличных баз данных

Киевский национальный университет им. Тараса Шевченко

Рассматривается аксиоматика многозначных зависимостей в табличных базах данных и аксиоматика функциональных и многозначных зависимостей; устанавливается, полнота этих аксиоматик в терминах совпадения отношений синтаксического и семантического следований, а также критерий полноты аксиоматик в терминах мощностей универсального домена и множества атрибутов.

Ключевые слова: табличные базы данных, функциональные зависимости, многозначные зависимости, полнота аксиоматик, критерий полноты.

Academician of the NAS of Ukraine **V. N. Redko, D. B. Bui, A. V. Puzikova**

An axiomatics for multivalued dependences in tabular databases

Taras Shevchenko National University of Kiev

Axiomatics for multivalued dependences in table databases and for functional and multivalued dependences is considered. The completeness of these axiomatic systems is established in terms of the coincidence of syntactic and semantic consequence relations. The completeness criteria for these axiomatic systems are formulated in terms of cardinalities of the universal domain and the set of attributes.

Keywords: table databases, functional dependencies, multivalued dependences, completeness of axiomatic system, completeness criteria.