Аксиоматика функциональных зависимостей с неопределенными значениями в базах данных

С.В. Зыкин

Институт Математики им. С.Л. Соболева СО РАН (ОФ)

Пример отношения БД

№ паци-ента	ФИО пациента	Группа пациентов	№ показате-ля	Показатель	№ дня получения показателя	Значение показателя
31	Иванов И.И.	2	5	Креатинин	1	97
56	Петров П.П.	3	5	Креатинин	3	101
78	Сидоров С.С.	2	9	Белок	2	64

$\boldsymbol{A_1}$ – № пациента, $\boldsymbol{A_2}$ – ФИО пациента,

$\boldsymbol{A_3}$ – № показателя, $\boldsymbol{A_4}$ – Показатель,

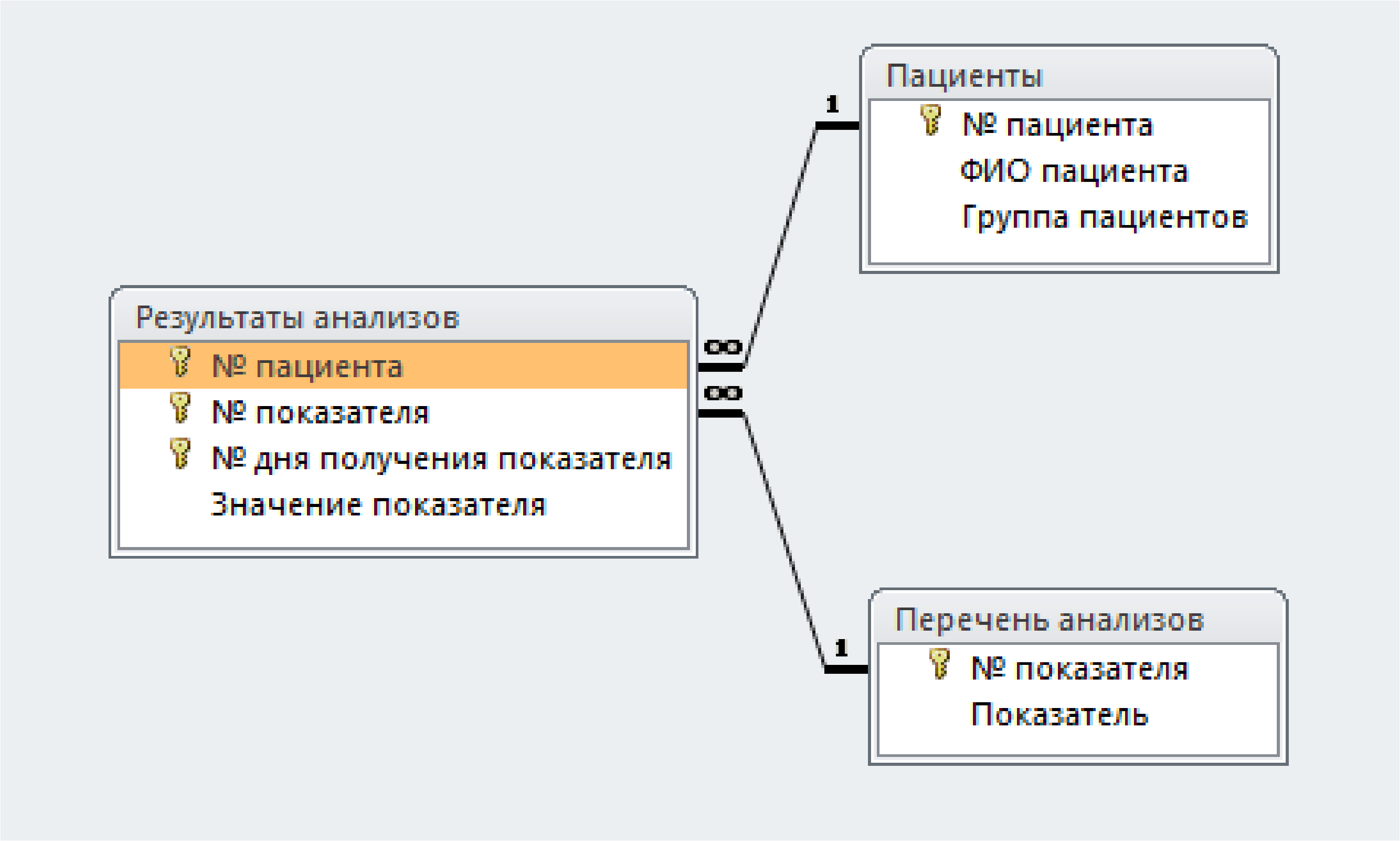
$\boldsymbol{A_5}$ – Значение показателя,

$\boldsymbol{A_6}$ – № дня получения показателя,

$\boldsymbol{A_7}$ – Группа пациентов.

$\boldsymbol{DEP=\{A_1 \rightarrow A_2A_7,\ A_3 \rightarrow A_4,\ A_1A_3A_6 \rightarrow A_5\}}$

Схема БД в примере 1



Основы формальной теории

Отношение R задано на атрибутах $U = \{A_1, A_2, \ldots, A_n\}$.

Определение. Пусть $X \subseteq U$ и $Y \subseteq U$, тогда X функционально определяет Y: $X \to Y$, если для любых кортежей $t, u \in R$, не выполнено $t[X] = u[X]$ и $t[Y] \neq u[Y]$.

Аксиомы Армстронга:

- А1. Рефлексивность. Если $Y \subseteq X \subseteq U$, то $X \to Y$.
- А2. Пополнение. Если $X \to Y$ и $Z \subseteq U$, то $XZ \to YZ$ (где $XZ = X \cup Z$).
- А3. Транзитивность. Если $X \to Y$ и $Y \to Z$, то $X \to Z$.

Пусть F множество зависимостей в R.

Пример 2: отдел кадров

A_1 – № сотрудника,

A_2 – ФИО сотрудника,

A_3 – должность сотрудника,

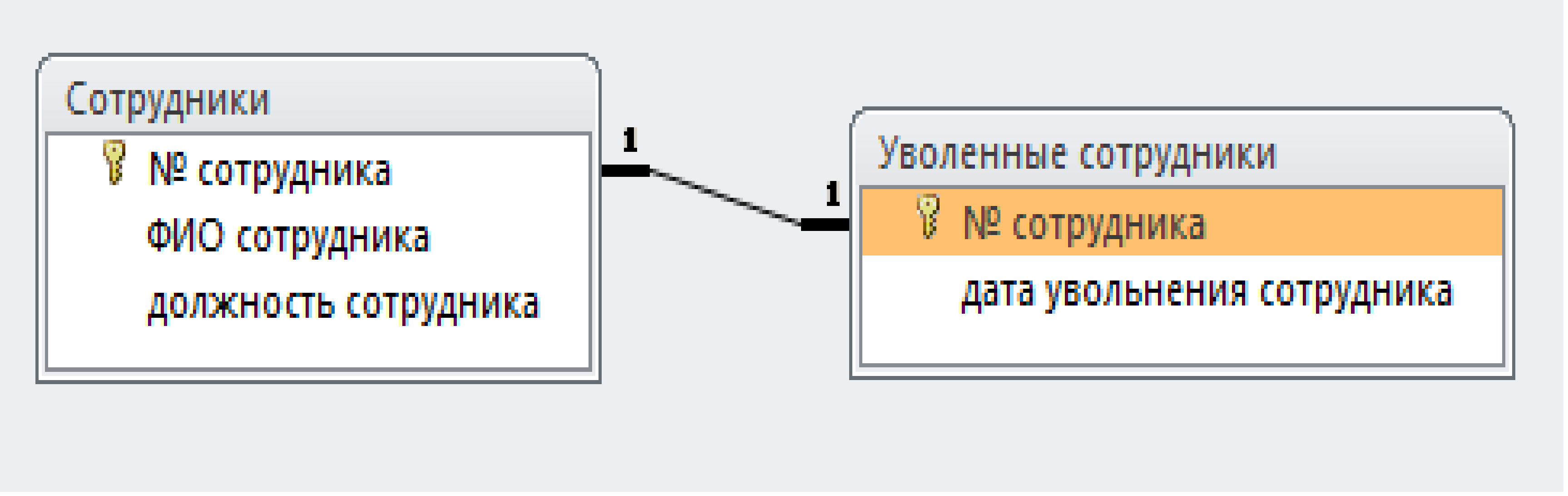
A_4 – дата увольнения сотрудника.

$DEP=\{A_1 \rightarrow A_2,\ A_1 \rightarrow A_3,\ A_1 \rightarrow A_4\}$

$dom(A_1 \rightarrow A_4)=\{$увоненные сотрудники$\}$

Схема БД в примере 2

$\boldsymbol{R}(\underline{A}_1, A_2, A_3, A_4)$ – схема БД в соответствии с классической теорией проектирования



Связанные работы

- Zaniolo C. Database Relations with Null Values // Journal of Computer and System Sciences. 1984. No.~28. P. 142-166.
  R=(*Номер сотрудника*, *ФИО сотрудника, Номер руководителя)*
  R*=(*Номер сотрудника*, *ФИО сотрудника, Должность сотрудника, Место работы)*
- Vassiliou Y. Functional dependencies and incomplete information // VLDB '80 Proceedings of the sixth international conference on Very Large Data Bases. 1980. V.~6. P. 260-269.

Связанные работы

- Atzeni P., Morfuni N. Functional dependencies and constraints on Null values in database relations // Information and Control. 1986. V.~70. No~1. P. 1-31.
- Levene M., Loizou G. Axiomatisation of functional dependencies in incomplete relations // Theoretical Computer Science. 1998. V.~206. No~1-2. P. 283-300.
- Hartmann S., Link S. The implication problem of data dependencies over SQL table definitions: axiomatic, algorithmic and logical characterizations // ACM Transactions on Database Systems. 2012. V.~37. No~2. P. 1-40.

Связанные работы

- A1. <u>Рефлексивность</u>. Если $\boldsymbol{Y \subseteq X \subseteq U}$, то $\boldsymbol{X \to Y}$.
- A2. <u>Декомпозиция</u>. Если $\boldsymbol{X \to Y}$ и $\boldsymbol{Z \subseteq Y}$, то $\boldsymbol{X \to Z}$.
- A3. <u>Объединение</u>. Если $\boldsymbol{X \to Y}$ и $\boldsymbol{X \to Z}$, то $\boldsymbol{X \to YZ}$.

- Cubero J.C., Medina J.M., Pons O., Vila M.A. Non-transitive fuzzy dependencies (I) // Fuzzy Sets and Systems. 1999. V.~106. No~3. P. 401-431.

$$\boldsymbol{0 \le \beta \le 1}$$

Области определения

- <u>Дано</u>: произвольная реализация отношения $\boldsymbol{R}$, множества атрибутов $\boldsymbol{X,Y\subseteq U}$, зависимость $\boldsymbol{X\to Y}$ на $\boldsymbol{R}$.
- <u>Определение</u>. Областью определения $\boldsymbol{dom(X)}$ в $\boldsymbol{R}$ является множество кортежей $\boldsymbol{T_R(X)=\{t_1, t_2, \ldots, t_s\}\subseteq R}$, для которых выполнено условие $\boldsymbol{t[X]}\neq$Null.
- <u>Определение</u>. Областью определения зависимости $\boldsymbol{dom(X\to Y)}$ в текущей реализации $\boldsymbol{R}$ является множество кортежей $\boldsymbol{T_R(X\to Y)=\{t_1, t_2, \ldots, t_m\}\subseteq R}$, для которых выполнено условие $\boldsymbol{t[XY]}\neq$Null.

Свойства областей определения

- П1. Включение. Если $\boldsymbol{Y \subseteq X \subseteq U}$, то $\boldsymbol{dom(X) \subseteq dom(Y)}$.
- П2. Зависимость. Если $\boldsymbol{X \to Y}$, то $\boldsymbol{dom(X \to Y) = dom(X) \cap dom(Y)}$.
- П3. Объединение. $\boldsymbol{dom(XY) = dom(X) \cap dom(Y)}$.
- П4. Логическое следствие. $\boldsymbol{dom(F_1 \vDash F_2) = dom(F_2 | F_1) = dom(F_1) \cap dom(F_2)}$.

Области определения логических операций

- Конъюнкция. $dom(F_1 \wedge F_2) = dom(F_1) \cap dom(F_2)$.
- Дизъюнкция. $dom(F_1 \vee F_2) = dom(F_1) \cup dom(F_2)$.
- Отрицание. $dom(\neg F_i) = R - dom(F_i)$.
- Импликация. $dom(F_1 \rightarrow F_2) = dom(\neg F_1 \vee F_2) = (R - dom(F_1)) \cup dom(F_2) \neq dom(F_1) \cap dom(F_2)$.

Модифицированная система аксиом

R задано на атрибутах $\boldsymbol{U = \{A_1, A_2, ..., A_n\}}$.

Аксиомы:

•<u>А*1.</u> (рефлексвность) Если $\boldsymbol{Y \subseteq X \subseteq U}$, то $\boldsymbol{X \to Y}$ и $\boldsymbol{dom(X \to Y) = dom(X)}$.

•<u>А*2.</u> (пополнение) Если $\boldsymbol{X \to Y}$ и $\boldsymbol{Z \subseteq U}$, то $\boldsymbol{XZ \to YZ}$ и $\boldsymbol{dom(XZ \to YZ) = dom(X) \cap dom(Y) \cap dom(Z)}$.

•<u>А*3.</u> (транзитивность) Если $\boldsymbol{X \to Y}$ и $\boldsymbol{Y \to Z}$, то $\boldsymbol{X \to Z}$ и $\boldsymbol{dom(X \to Z) = dom(X) \cap dom(Y) \cap dom(Z)}$.

Надежность системы аксиом

Теорема. Аксиомы **A*1** – **A*3** надежны (непротиворечивы).

Доказательство. **А*1**. Существуют два кортежа $\boldsymbol{t_1, t_2 \in R}$. Пусть $\boldsymbol{t_1[X]=t_2[X]}$. Поскольку $\boldsymbol{Y \subseteq X}$, то $\boldsymbol{t_1[Y]=t_2[Y]}$. Следовательно, зависимость $\boldsymbol{X \to Y}$ выполнена. По правилу **П2** имеем: $\boldsymbol{dom(X \to Y)=dom(X) \cap dom(Y)}$. Поскольку $\boldsymbol{Y \subseteq X}$, то по правилу П1 $\boldsymbol{dom(X) \subseteq dom(Y)}$. Следовательно, $\boldsymbol{dom(X \to Y)=dom(X)}$. Заметим, что правило **П2** является частным случаем правила **П4** (логическое следствие).

Выводимость из аксиом

- **Лемма.** Если $X \to Y$ логически следует из F, то для любой области определения $dom(V) \subseteq dom(X \to Y)$, зависимость $X \to Y$ также следует из F.
- **Алгоритм** построения замыкания X^+_V
- **Теорема.** Зависимость $X \to Y$ выводима из аксиом A^*1 – A^*3, в области $dom(V)$, если $Y \subseteq X^+_V$.

Полнота системы аксиом

Теорема. Система аксиом A^*1 – A^*3 полна для конечномерных баз данных.

Для доказательства полноты надо показать, что если $\boldsymbol{F}$ – заданное множество функциональных зависимостей и зависимость $\boldsymbol{X}\rightarrow\boldsymbol{Y}$ не может быть выведена из системы аксиом, то должно существовать отношение, в котором справедливы все зависимости из $\boldsymbol{F}$, кроме $\boldsymbol{X}\rightarrow\boldsymbol{Y}$, то есть $\boldsymbol{X}\rightarrow\boldsymbol{Y}$ логически не следует из $\boldsymbol{F}$.

	Атрибуты $\boldsymbol{X}^+$				Другие атрибуты			
$\boldsymbol{t}_1$	1	1	…	1	1	1	…	1
$\boldsymbol{t}_2$	1	1	…	1	0	0	…	0

Исходные данные алгоритма нормализации

- Отношение $\boldsymbol{R}$, определенное на множестве атрибутов:
- $\boldsymbol{U}=\{\boldsymbol{A}_1, \boldsymbol{A}_2,\ldots, \boldsymbol{A}_n\}$,
- $\boldsymbol{F}=\{\boldsymbol{F}_1, \boldsymbol{F}_2,\ldots, \boldsymbol{F}_k\}$ – множество функциональных зависимостей и
- $\boldsymbol{D}=\{\boldsymbol{D}_1, \boldsymbol{D}_2,\ldots, \boldsymbol{D}_k\}$ – соответствующие области определения зависимостей.

Алгоритм нормализации

- Цикл: (Сочетания без повторений из $\boldsymbol{n}$ элементов по $\boldsymbol{m}$, $\boldsymbol{m}$=1,2,…,$\boldsymbol{n}$-1.
- Для атрибутов $\boldsymbol{X}$, соответствующих текущему сочетанию, строим замыкания по областям $\boldsymbol{D_i}$, начиная с минимальных.
- Для каждого замыкания, если оно не совпадает со всем множеством $\boldsymbol{U}$, производим декомпозицию: $\boldsymbol{[R]=[R]-Y}$, $[R^*]=XY$.
- Конец цикла.

Направление исследований

- Декомпозиция, построенная по алгоритму, удовлетворяет нормальной форме Бойса-Кодда (НФБК).

- Построение нормальных форм на основе минимального покрытия множества зависимостей.

Благодарю за внимание!

Вопросы?