MODELING AND ANALYSIS OF INFORMATION SYSTEMS, VOL. 27, NO. 3, 2020
journal homepage: www.mais-journal.ru

DISCRETE MATHEMATICS IN RELATION TO COMPUTER SCIENCE

A Generalization of the Inference Rules for Join Dependencies in Databases

S. V. Zykin[1]

DOI: 10.18255/1818-1015-2020-3-356-365

[1]Sobolev Institute of Mathematics SB RAS, 4 Acad. Koptyug avenue, Novosibirsk, 630090, Russia.

MSC2020: 68P15
Research article
Full text in Russian

Received July 6, 2020
After revision September 7, 2020
Accepted September 9, 2020

In this paper a generalisation of the inference rules of the join dependencies that are used in the design of database schemas that meets the requirements of the fifth normal form is considered. In the previous works devoted to this problem, attempts to construct systems of the axioms of such dependencies based on inference rules are made. However, while the justification for the consistency (soundness) of the obtained axioms does not cause difficulties, the proof of completeness in general has not been satisfactorily resolved. First of all, this is due to the limitations of the inference rules themselves. This work focuses on two original axiom systems presented in the works of Sciore and Malvestuto. For the inclusion dependencies a system of rules that generalises existing systems and has fewer restrictions has been obtained. The paper presents a proof of the derivability of known systems of axioms from the presented inference rules. In addition, evidence of the consistency (soundness) of these rules is provided. The question of the completeness of the formal system based on the presented rules did not find a positive solution. In conclusion, the theoretical and practical significance of the inference rules for the join dependencies is noted.

Keywords: relational databases; join dependencies; inference rules.

INFORMATION ABOUT THE AUTHORS

Sergey V. Zykin correspondence author	orcid.org/0000-0002-0576-2149. E-mail: szykin@mail.ru Leading researcher, Doctor of Science.

Funding: The work was supported by the program of fundamental scientific researches of the SB RAS № I.5.1., project № 0314-2019-0020.

For citation: S. V. Zykin, "A Generalization of the Inference Rules for Join Dependencies in Databases", *Modeling and analysis of information systems*, vol. 27, no. 3, pp. 356-365, 2020.

МОДЕЛИРОВАНИЕ И АНАЛИЗ ИНФОРМАЦИОННЫХ СИСТЕМ, ТОМ 27, № 3, 2020
сайт журнала: www.mais-journal.ru

DISCRETE MATHEMATICS IN RELATION TO COMPUTER SCIENCE

Обобщение правил вывода для зависимостей соединения в базах данных

С. В. Зыкин[1]

DOI: 10.18255/1818-1015-2020-3-356-365

[1]Институт математики им. С. Л. Соболева СО РАН, пр. ак. Коптюга, 4, г. Новосибирск, 630090 Россия.

УДК 004.652.4
Научная статья
Полный текст на русском языке

Получена 6 июля 2020 г.
После доработки 7 сентября 2020 г.
Принята к публикации 9 сентября 2020 г.

В работе рассматривается обобщение правил вывода зависимостей соединения, которые используются при проектировании схемы базы данных, удовлетворяющей требованиям пятой нормальной формы. В предшествующих работах, посвященных данной проблематике, делаются попытки построить системы аксиом таких зависимостей, основанных на правилах вывода. Однако, если обоснование непротиворечивости (надежности) полученных аксиом не вызывает затруднений, то доказательство полноты в общем случае не получило удовлетворительного решения. Прежде всего, это связано с ограниченностью самих правил вывода. В данной работе акцентировано внимание на двух оригинальных системах аксиом, представленных в работах Sciore и Malvestuto. Для зависимостей включения получена система правил, которая обобщает существующие системы и при этом имеет меньше ограничений. В работе представлено доказательство выводимости известных систем аксиом из представленных правил вывода. Кроме того, приводится доказательство непротиворечивости (надежности) этих правил. Вопрос о полноте формальной системы, основанной на представленных правилах, не нашел положительного решения. В заключение отмечена теоретическая и практическая значимость правил вывода для зависимостей соединения.

Ключевые слова: реляционные базы данных; зависимости соединения; правила вывода.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Сергей Владимирович Зыкин автор для корреспонденции	orcid.org/0000-0002-0576-2149. E-mail: szykin@mail.ru вед. науч. сотр., доктор техн. наук.

Финансирование: Работа выполнена при поддержке программы фундаментальных научных исследований СО РАН № I.5.1., проект № 0314-2019-0020.

Для цитирования: S. V. Zykin, “A Generalization of the Inference Rules for Join Dependencies in Databases”, *Modeling and analysis of information systems*, vol. 27, no. 3, pp. 356-365, 2020.

Введение

Основой проектирования логической структуры базы данных (БД) являются зависимости данных в прикладной области. В настоящее время актуальными при проектировании являются функциональные зависимости, многозначные зависимости, зависимости соединения [1—3] и зависимости включения [3—7]. Для рассмотрения материала введем необходимые обозначения. Пусть задана конечная совокупность всех отношений реляционной БД: $(R_1, R_2, ..., R_k)$ – результат декомпозиции при нормализации отношений, $[R_i]$ – совокупность атрибутов (схема отношения R_i), $R_i[V]$ – проекция отношения R_i на множество атрибутов V, $U = \{A_1, A_2, \dots, A_n\}$ – конечное множество всех атрибутов, на которых заданы отношения БД.

Данная работа посвящена исследованию свойств зависимостей соединения, необходимых при построении пятой нормальной формы. Для того, чтобы подчеркнуть практическую значимость указанной проблемы, рассмотрим пример, позаимствованный из аналитических материалов С. Д. Кузнецова. Пусть A_1 – сотрудник, A_2 – проект, A_3 – этап проекта. Если предположить наличие зависимости соединения $\bowtie(A_1A_2, A_2A_3)$, где $\bowtie$ – знак операции естественного соединения [1, 2], то в результате декомпозиции отношения $R[A_1, A_2, A_3]$ будет получено два отношения:

$$R_1 = R[A_1, A_2], R_2 = R[A_2, A_3].$$

Наличие зависимости $\bowtie(A_1A_2, A_2A_3)$ гарантирует выполнение равенства

$$R = R_1 \bowtie R_2$$

для любой реализации R. Соответствующее зависимости бизнес-правило можно сформулировать следующим образом: “Если сотрудник участвует в каком-либо проекте, то он автоматически участвует в выполнении всех этапов этого проекта”. Заметим, что зависимость $\bowtie(A_1A_2, A_2A_3)$ эквивалентна многозначной зависимости $A_2 \twoheadrightarrow A_1|A_3$, следовательно, отношения декомпозиции (R_1, R_2) удовлетворяют четвертой нормальной форме (4НФ).

Если предположить наличие зависимости соединения $\bowtie(A_1A_2, A_2A_3, A_1A_3)$, то в результате декомпозиции отношения $R[A_1, A_2, A_3]$ будет получено три отношения:

$$R_1 = R[A_1, A_2], R_2 = R[A_2, A_3], R_3 = R[A_1, A_3].$$

Наличие зависимости $\bowtie(A_1A_2, A_2A_3, A_1A_3)$ гарантирует выполнение равенства

$$R = R_1 \bowtie R_2 \bowtie R_3$$

для любой реализации R. Соответствующее зависимости бизнес-правило: “Сотрудник участвует в выполнении только некоторых этапов проекта”, о чем записываются соответствующие данные в отношение R_3. В отношениях R_1 и R_2 также должна быть информация об участии сотрудника в проекте и наличии соответствующего этапа в проекте. Отношения декомпозиции (R_1, R_2, R_3) удовлетворяют пятой нормальной форме (5НФ).

Допустим, что нам надо оценить участие сотрудника в проекте. Для реляционной схемы БД, сформированной по зависимостям, дополнения выполняются без разрушения существующей схемы. Для зависимости $\bowtie(A_1A_2, A_2A_3)$ надо добавить соответствующий атрибут в отношение R_1. В этом случае оцениваться будет участие сотрудника во всем проекте. Для зависимости $\bowtie(A_1A_2, A_2A_3, A_1A_3)$ соответствующий атрибут должен быть добавлен в отношение R_3. В этом случае оцениваться будет участие сотрудника в каждом этапе проекта.

Распространенным является утверждение, что зависимости соединения редко встречаются на практике. Рассмотренный пример опровергает это утверждение. Аналогичные зависимости возникают везде, где требуется сопоставить объекты с ограниченным множеством событий в системе.

1. Связанные работы

Риссанен [8] впервые ввел понятие зависимости соединения и определил важное свойство данных: соединение множества отношений $R_1, R_2, \ldots, R_k$ без потери информации для любой реализации отношения R:

$$R = R_1 \bowtie R_2 \bowtie \ldots \bowtie R_k,$$

что соответствует зависимости соединения $\bowtie (V_1, V_2, \ldots, V_k)$, где $V_i = [R_i]$.

В работе [9] предложена система правил для зависимостей соединения. Пусть $\mathbf{S} = \{X_1, X_2, \ldots, X_p\}$, $\mathbf{R} = \{Y_1, Y_2, \ldots, Y_q\}$, $attr(\mathbf{S}) = X_1 \cup X_2 \cup \ldots \cup X_p$ и $attr(\mathbf{R}) = Y_1 \cup Y_2 \cup \ldots \cup Y_q$, $\models$ – логическое следствие, тогда:
A0: $\varnothing \models \bowtie (X)$ для любого множества $X \subseteq U$.
A1: $\bowtie (\mathbf{S}) \models \bowtie (\mathbf{S}, Y)$, если $Y \subseteq attr(\mathbf{S})$.
A2: $\bowtie (\mathbf{S}, Y, Z) \models \bowtie (\mathbf{S}, YZ)$.
A3: $\bowtie (\mathbf{S}, Y), \bowtie (\mathbf{R}) \models \bowtie (\mathbf{S}, \mathbf{R})$, если $attr(\mathbf{R}) = Y$.
A4: $\bowtie (\mathbf{S}, YA) \models \bowtie (\mathbf{S}, Y)$, если $A \nsubseteq attr(\mathbf{S})$.

В [9] показано, что правила A0 – A4 непротиворечивы (надежность – soundness), однако не обладают свойством полноты (completeness). Кроме того, рассмотрено обобщение рассмотренных правил при расширении множества U атрибутами специального вида.

В статье [10] сделана попытка найти надежную и полную формальную систему для зависимостей соединения. Получена ограниченная формальная система, позволяющая выводить зависимости соединения с использованием только обобщенных зависимостей соединения. Также получена неограниченная формальная система, которая полна для расширенных зависимостей соединения, получена надежная и полная неограниченная система в стиле Gentzen для зависимостей соединений. При этом остается открытым вопрос существования надежной и полной ограниченной формальной системы для расширенных зависимостей соединения.

Метод разложения зависимостей соединения в реляционной базе данных с использованием понятия шарнира (hinge) рассматривается в работе [11]. Показано, что метод декомпозиции может быть использован для повышения эффективности проверки целостности. Под проверкой целостности для зависимостей соединения подразумевается проверка выполнения этой зависимости в отношении после обновления. В общем случае такая задача является NP-полной. В работе показано, что этой проверки целостности можно избежать за счет разложения отношения на набор меньших отношений. Рассмотрено понятие n-цикличности для зависимостей соединения. Показано, что для фиксированного значения n проверка целостности может быть выполнена за полиномиальное время.

В работе [12] предложена надежная и полная формальная система для полных (full) ациклических зависимостей соединения, которая получена добавлением трех правил к аксиомам многозначных зависимостей. При определении ацикличности используется модель гиперграфа [2]. Аксиомы в этой системы запишем в следующем виде:
B1. $\bowtie (X_1, X_2) \models \bowtie (X_1 \cup Y, X_2 \cup Y)$.
B2. $\bowtie (X_1, X_2), \bowtie (Y, X_1 \cap X_2) \models \bowtie (X_1, X_2 \cap Y)$.
B3. $\models \bowtie (U)$.
B4. $\bowtie (X_1, X_2, \ldots, X_k) \models \bowtie (Y_1, Y_2)$, где $\{Y_1, Y_2\}$ – компонент (двухреберный гиперграф).
B5. $\bowtie (X, \{X_1, X_2, \ldots, X_k\}), \bowtie (Y_1, Y_2) \models \bowtie (X \cap Y_1, X \cap Y_2, \{X_1, X_2, \ldots, X_k\})$, где $Y_1 \cap Y_2 \subseteq X, X \cap X_i \subseteq Y_1$ или $X \cap X_i \subseteq Y_2$, $i = \overline{1, k}$.

Далее мы вернемся к обсуждению этих аксиом.

Связь между аксиоматизацией зависимостей в реляционных базах данных и цилиндрическими алгебрами рассмотрена в работе [13]. Рассмотрены зависимости соединения и соответствующий класс (представимых) цилиндрических полурешеток. Поскольку представимые цилиндрические полурешетки имеют бесконечную аксиоматизируемую квазиэквивалентную теорию, то не существует конечной аксиоматизации для n-мерных неограниченных зависимостей соединения. В работе получены контрпримеры к утверждению, что цилиндрические зависимости конечно аксиоматизируемы.

В статьях [14, 15] рассматриваются полные иерархические зависимости (FHD), которые соответствуют естественному соединению, по крайней мере, двух отношений БД. В статье [14] определяется полный набор правил вывода для FHD. Первый результат устанавливает конечную аксиоматизацию для FHD. Второй основной результат устанавливает конечную аксиоматизацию для выводимости FHD. В статье [15] показано, как выводы полных иерархических зависимостей можно моделировать с помощью выводов многозначных зависимостей и наоборот. Это позволяет применять оба основных результата к комбинированному классу функциональных и многозначных зависимостей. Кроме того, установлена новая аксиоматизация для класса нетривиальных функциональных зависимостей.

В работе [16] получено обобщение для задачи выводимости атомов соединения и атомов включения, что дает полную аксиоматизацию для проблемы выводимости встроенных зависимостей соединения и зависимостей включения в специальном случае. В более общем смысле статья представляет способ поиска аксиоматизации для различных коллекций зависимостей в БД с использованием методов математической логики.

Исследование ациклических зависимостей продолжено в работе [17]. Представлен формальный контекст для зависимостей соединения, который обобщает подход для анализа многозначных зависимостей. Этот результат является основой для дальнейшего исследования ациклических зависимостей соединения с целью минимизации их множества.

В данной работе рассматриваются только классические [8] зависимости соединения, которые не дополнены какими-либо ограничениями или расширениями. С точки зрения автора статьи, только такой вид зависимостей является значимым для практики.

2. Правила вывода зависимостей соединения

Пусть R – отношение, на множестве атрибутов $U = \{A_1, A_2, \dots, A_n\}$.

Определение 1. *Зависимость соединения* $\bowtie (X_1, X_2, \dots, X_k)$ *выполнена, если для любой реализации отношения* R *выполнено:*

$$R[X] = R[X_1] \bowtie R[X_2] \bowtie \dots \bowtie R[X_k],$$

где $R[X_i]$ *– проекция [1] отношения* R *на множество атрибутов* X_i, $X = X_1 \cup X_2 \cup \dots \cup X_k$.

В [2] сформулирован очевидный признак наличия зависимости соединения.

Утверждение 1. *Зависимость* $\bowtie (X_1, X_2, \dots, X_k)$ *выполнена в* R, *если для любой реализации* R *наличие кортежей* $t_1, t_2, \dots, t_k$, *не обязательно различных, гарантирует наличие кортежа* $t \in R$: $t[X_i] = t_i[X_i]$, $i = \overline{1, k}$, *для которых* $t_i[X_i \cap X_j] = t_j[X_i \cap X_j]$ *если* $X_i \cap X_j \neq \varnothing$.

Анализ утверждения 1 показывает, что разделение зависимостей соединения на полные (full) и встроенные является искусственным. В отношении R все полные зависимости соединения на практике могут оказаться встроенными в процессе эксплуатации БД. Это не должно стать причиной перестроения декомпозиции отношений. Произвольная зависимость $\bowtie (X_1, X_2, \dots, X_k)$ является полной в проекции $R[X]$, где $X = X_1 \cup X_2 \cup \dots \cup X_k$, и в общем случае встроенной в отношение R. В данном случае уместно говорить об области определения зависимости.

Определение 2. *Областью определения зависимости соединения $\bowtie (X_1, X_2, \dots, X_k)$ в отношение R будем называть множество $X = X_1 \cup X_2 \cup \dots \cup X_k$.*

В работе [18] было введено понятие области определения функциональной зависимости с целью разделения значений “не определено (Null)” и “не существует”. Также была разработана формальная теория, которая является естественным обобщением классической теории и позволяет избавиться от значений “не существует” на этапе проектирования схемы БД. Областью определения зависимости в [18] является множество атрибутов, что согласуется с определением 2. Таким образом, рассматриваемые далее зависимости являются полными, но в собственной области определения. Причем, выводимые зависимости могут иметь большую или меньшую область определения, чем исходные зависимости.

Рассмотрение формальных теорий в БД позволяет ответить на главные проблемы проектирования БД:

- является ли проект БД, построенный по множеству зависимостей, оптимальным и правильным относительно некоторых критериев (наличие аномалий, минимальное покрытие ...),
- сохраняет ли декомпозиция $(R_1, R_2, \dots, R_k)$ зависимости в качестве ограничений целостности на данные.

В основе решения указанных проблем лежат правила вывода зависимостей. Рассмотрим множество правил для зависимостей соединения.

P0) $\emptyset \vDash \bowtie (X), X \subseteq U$.

P1) $\bowtie (X_1, X_2, \dots, X_k) \vDash \bowtie (Y_1, Y_2, \dots, Y_m)$, если:

  p11) для любого X_i существует Y_j: $X_i \subseteq Y_j$;

  p12) для любого Y_j выполнено $Y_j \subseteq X_1 \cup X_2 \cup \dots \cup X_k$.

P2) $\bowtie (X_1, X_2, \dots, X_k) \vDash \bowtie (Y_1, Y_2, \dots, Y_k)$, если:

  p21) $X_i \cap (X_1 \cup X_2 \cup \dots \cup X_{i-1} \cup X_{i+1} \dots \cup X_k) \subseteq Y_i$;

  p22) $Y_i \subseteq X_i$.

P3) $\bowtie (X_1, X_2, \dots, X_k) \vDash \bowtie (Y_1, Y_2, \dots, Y_k)$, если:

  p31) $Y_i = X_i \cup Z$.

P4) $\bowtie (X_1, X_2, \dots, X_k, V), \bowtie (Y_1, Y_2, \dots, Y_m) \vDash \bowtie (X_1, X_2, \dots, X_k, V \cap Y_1, V \cap Y_2, \dots, V \cap Y_m)$, если:

  p41) $Y_i \cap Y_j \subseteq V, i \neq j$;

  p42) $V \cap (X_1 \cup X_2 \cup \ ldots \cup X_k) \subseteq V \cap (Y_1 \cup Y_2 \cup \dots \cup Y_m)$.

Рассмотрим соответствие представленных в данной работе правил.

Утверждение 2. *Правила A0–A4 являются следствием правил P0–P4.*

Доказательство. Очевидно, что правила A0 и P0 эквивалентны. Правило A1 является следствием P1 на множестве атрибутов $X = X_1 \cup X_2 \cup \dots \cup X_k$. Действительно, условие p12 правила P1 совпадает с условием $Y \subseteq attr(\mathrm{S})$ правила A1, а условие p11 правила P1 выполнено по построению правила A1. Поскольку P1 соответствует более общему случаю, то A1 является следствием P1 на множестве атрибутов $X = X_1 \cup X_2 \cup \dots \cup X_k$. Не сложно проверить, что $X = Y_1 \cup Y_2 \cup \dots \cup Y_k$. Правило A2 очевидное следствие правила P1. Правило A3 является частным случаем правила P4. Действительно, если $V = Y_1 \cup Y_2 \cup \dots \cup Y_m$, то условия p41 и p42 будут всегда выполнены, а само правило P4 будет совпадать с правилом A3. Правило A4 становится очевидным следствием правила P2, если сделать отождествления $S = (X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_k)$ и $YA = X_i$. □

Утверждение 3. *Правила B1–B5 являются следствием правил P0–P4.*

Доказательство. Очевидно, что правило P3 является обобщением правила B1. Правило B2 выводимо с использованием правил P4 и P1. Действительно, запишем левую часть правила B2 в следующем виде $\bowtie (X, V)$, $\bowtie (Y, X \cap V)$, необходимо получить вывод правила $\bowtie (X, V \cap Y)$. Подставив левую часть в правило P4 получим выводимость $\bowtie (X, V \cap Y, V \cap X)$, условия p41 и p42 выполнены. Затем применим к полученному выражению правило P1, получим выводимость $\bowtie (X, V \cap Y)$, условия p11 и p12 выполнены, что и требовалось показать. Правило B3 эквивалентно правилу P0. Правило B4 является частным случаем правила P1, поскольку компонент [12] является двухреберным гиперграфом, полученным объединением ребер исходного гиперграфа. Правило B5 является частным случаем правила P4. Действительно, условие $Y_1 \cap Y_2 \subseteq X$ гарантирует выполнение условия p41 при $X = V$, а условие $X \cap Xi \subseteq Y_1$ или $X \cap X_i \subseteq Y_2$, $i = \overline{1, k}$, гарантирует выполнение условия p42. Поскольку порядок аргументов зависимости соединения не имеет значения, то правило B5 совпадает с правилом P4, при $m = 2$. □

3. Анализ надежности правил вывода

Для предложенных в данной работе правил P0–P4 необходимо доказать надежность (непротиворечивость). Другими словами, полученные за счет правил P0–P4 зависимости будут удовлетворять условию логического следствия. Это очень важная задача с практической точки зрения, поскольку гарантирует сохранение целостности данных при проектировании БД.

Теорема 1. *Правила P0–P4 являются надежными.*

Доказательство. Надежность правила P0 очевидна.

Для доказательства правила P1 обозначим: $R_X = R[X_1] \bowtie R[X_2] \bowtie \dots \bowtie R[X_k]$ и $R_Y = R[Y_1] \bowtie R[Y_2] \bowtie \dots \bowtie R[Y_m]$, где $X = X_1 \cup X_2 \cup \dots \cup X_k$ и $Y = Y_1 \cup Y_2 \cup \dots \cup Y_m$. Из леммы 5.5 [1] следует, что $R[X] \subseteq R_X$ и $R[Y] \subseteq R_Y$. Необходимо показать, что если $R_X \subseteq R[X]$, тогда $R_Y \subseteq R[Y]$. Рассмотрим произвольный кортеж $t_Y \in R_Y$, надо показать, что $t_Y \in R[Y]$. Для выполнения условия $t_Y \in R_Y$ необходимо наличие, не обязательно различных, кортежей $t_1, t_2, \dots, t_m$ в отношении R (утверждение 1), для которых выполнено $t_p[Y_p] = t_Y[Y_p]$, $p = \overline{1, m}$. Рассмотрим любую пару множеств атрибутов X_i и X_j. Из условия p11 следует, что существуют p и q: $X_i \subseteq Y_p$ и $X_j \subseteq Y_q$. Если $X_i \cap X_j \neq \varnothing$, то $t_p[X_i \cap X_j] = t_q[X_i \cap X_j]$, поскольку они образованы из одного кортежа t_Y. Обозначим $t_{pi} = t_p[X_i]$. Из условия p11 следует, что кортеж t_{pi} существует для любого i: $1 \leq i \leq k$. Из условия p12 следует, что не существует атрибута $A_l \in Y_p$, значение $t_Y[A_l]$ которого не принадлежит какому-либо кортежу t_{pi}. Естественное соединение кортежей t_{pi} совпадает с кортежем t_Y. Следовательно, $t_Y \in R_X$, заметим, что из условий p11 и p12 следует равенство $X = Y$. По предположению $R_X = R[X]$, а значит $t_Y \in R[Y]$. Правило доказано.

Для доказательства правила P2 используем аналогичные обозначения: $R_X = R[X_1] \bowtie R[X_2] \bowtie \dots \bowtie R[X_k]$ и $R_Y = R[Y_1] \bowtie R[Y_2] \bowtie \dots \bowtie R[Y_k]$, где $X = X_1 \cup X_2 \cup \dots \cup X_k$ и $Y = Y_1 \cup Y_2 \cup \dots \cup Y_k$. Из леммы 5.5 [1] следует, что $R[X] \subseteq R_X$ и $R[Y] \subseteq R_Y$. Необходимо показать, что если $R_X \subseteq R[X]$, тогда $R_Y \subseteq R[Y]$. Рассмотрим произвольный кортеж $t_Y \in R_Y$, достаточно показать, что $t_Y \in R[Y]$. Для выполнения условия $t_Y \in R_Y$ необходимо наличие, не обязательно различных, кортежей $t_i \in R$, для которых выполнено $t_i[Y_i] = t_Y[Y_i]$, $i = \overline{1, k}$. Рассмотрим произвольную пару множеств X_i и X_j, где $i \neq j$. Выполнено равенство $X_i \cap X_j \subseteq X_i \cap (X_1 \cup X_2 \cup \dots \cup X_{i-1} \cup X_{i+1} \cup \dots \cup X_k)$, из условия p21 следует $X_i \cap X_j \subseteq Y_i$. С другой стороны, $X_i \cap X_j \subseteq X_i \cap (X_1 \cup X_2 \cup \dots \cup X_{j-1} \cup X_{j+1} \cup \dots \cup X_k)$. Из условия p21 следует $X_i \cap X_j \subseteq Y_j$. В итоге получаем $X_i \cap X_j \subseteq Y_i \cap Y_j$ и $t_i[X_i \cap X_j] = t_j[X_i \cap X_j]$. По свойству операции естественного соединения существует кортеж $t_X = t_1[X_1] \bowtie t_2[X_2] \bowtie \dots \bowtie t_k[X_k] \in R_X$. При пустом множестве $X_i \cap X_j$ отсутствует препятствие соединению кортежей. Из условия p22 следует $t_Y = t_X[Y]$. Поскольку $R_X = R[X]$ и $Y \subseteq X$, то $t_Y \in R[Y]$, что и требовалось доказать.

При доказательстве P3 используем прежние обозначения для R_X, R_Y, X, и Y. Как и прежде, $R[X] \subseteq R_X$ и $R[Y] \subseteq R_Y$. Необходимо показать, что если $R_X \subseteq R[X]$, тогда $R_Y \subseteq R[Y]$. Рассмотрим произвольный кортеж $t_Y \in R_Y$, необходимо показать, что $t_Y \in R[Y]$. Выполнение условия $t_Y \in R_Y$ гарантирует наличие, не обязательно различных, кортежей $t_i \in R$, для которых выполнено $t_i[Y_i] = t_Y[Y_i]$, $i = \overline{1,k}$. Эти кортежи удовлетворяют следующим условиям: $t_i[Y_i \cap Y_j] = t_j[Y_i \cap Y_j]$, $i = \overline{1,k}$. Поскольку $X_i \cap X_j \subseteq Y_i \cap Y_j$ и $Z \subseteq Y_i \cap Y_j$, $i,j = \overline{1,k}$, то кортеж $t_X = t_1[X_1] \bowtie t_2[X_2] \bowtie \ldots \bowtie t_k[X_k]$ принадлежит R_X. Из условия $R_X \subseteq R[X]$ следует, что $t_X \in R[X]$ и существует кортеж $t \in R$, для которого $t_X = t[X]$. По совокупности рассмотренных равенств имеем $t_Y = t[Y]$, то есть $t_Y \in R[Y]$. Следовательно, $R_Y \subseteq R[Y]$, что и требовалось доказать.

Докажем надежность правила P4. Пусть $W_j = V \cap Y_j$, $j = \overline{1,m}$, $W = W_1 \cup W_2 \cup \ldots \cup W_m$. Покажем, что $\bowtie (Y_1, Y_2, \ldots, Y_m) \vDash \bowtie (W_1, W_2, \ldots, W_m)$ выводимо. Для этого используем правило вывода P2. Условие p41 гарантирует выполнение условия $Y_i \cap Y_j \subseteq V \cap Y_j$, $i \neq j$, что, в свою очередь, гарантирует выполнение условия p21: $Y_j \cap (Y_1 \cup Y_2 \cup \ldots \cup Y_{j-1} \cup Y_{j+1} \ldots \cup Y_k) \subseteq W_j = V \cap Y_j$. Условие p22 $W_j \subseteq Y_j$ выполнено по построению. Далее покажем, что $\bowtie (X_1, X_2, \ldots, X_k, V) \vDash \bowtie (X_1, X_2, \ldots, X_k, W)$ выводимо. Для этого также воспользуемся правилом P2. Из условия p42 следует условие p21: $V \cap (X_1 \cup X_2 \cup \ldots \cup X_k) \subseteq W_j = V \cap (Y_1 \cup Y_2 \cup \ldots \cup Y_m)$. Условие p22 $W_j \subseteq V$ выполнено по построению. Поскольку надежность правила P2 уже доказана, то выводимые с его помощью зависимости будут выполнимы. Осталось показать, что выполнима будет зависимость в правой части правила P4. По предположению текущая реализация отношения R удовлетворяет зависимостям $\bowtie (X_1, X_2, \ldots, X_k, V)$ и $\bowtie (Y_1, Y_2, \ldots, Y_m)$. Следовательно, R удовлетворяет зависимостям $\bowtie (X_1, X_2, \ldots, X_k, W)$ и $\bowtie (W_1, W_2, \ldots, W_m)$. Это означает, что:

$$R[X_1 \cup X_2 \cup \ldots \cup X_k \cup W] = R[X_1] \bowtie$$

$$\bowtie R[X_2] \bowtie \ldots \bowtie R[X_k] \bowtie R[W]$$

и

$$R[W] = R[W_1] \bowtie R[W_2] \bowtie \ldots \bowtie R[W_m],$$

подставим W и $R[W]$ в предыдущую формулу:

$$R[X_1 \cup X_2 \cup \ldots \cup X_k \cup W_1 \cup W_2 \cup \ldots \cup W_m] = R[X_1] \bowtie R[X_2] \bowtie$$

$$\ldots \bowtie R[X_k] \bowtie R[W_1] \bowtie R[W_2] \bowtie \ldots \bowtie R[W_m],$$

что соответствует выполнимости зависимости $\bowtie (X_1, X_2, \ldots, X_k, V \cap Y_1, V \cap Y_2, \ldots, V \cap Y_m)$. □

Замечание. При доказательстве теоремы не потребовалось предположение об ацикличности гиперграфов, соответствующих зависимостям соединения.

Для превращения множества правил в систему аксиом необходимо доказать полноту этой системы. К сожалению, все усилия исследователей в этом направлении пока не увенчались успехом.

Заключение

Полученная в работе система правил для зависимостей соединения имеет теоретическое значение с точки зрения развития теории проектирования баз данных. Для анализа полноты системы аксиом традиционно используется процедура прогонки [2], которая не дала удовлетворительных результатов как в этой работе, так и во всех предыдущих работах. Это означает, что поиск новых правил и обобщение известных правил должны быть продолжены. Полученные правила имеют ограниченное практическое значение при проектировании схемы БД. Действительно, после выделения функционально определяемых атрибутов в отдельные отношения при декомпозиции отношения R, либо при формировании обобщенного ключа (суперключа) при синтезе отношений, всякий раз приходится иметь дело с отношением, в котором нет функциональных зависимостей. Однако это отношение не является интерпретируемым. Использование зависимостей соединения позволяет найти корректную декомпозицию и получить интерпретируемые отношения и гарантировать выполнение свойства соединения без потерь информации.

References

[1] J. Ullman, *Principles of Database Systems.* Stanford University: Computer Science Press, 1980.

[2] D. Maier, *The Theory of Relational Databases.* Rockville: Computer Science Press, 1983.

[3] S. Abiteboul, R. Hull, and V. Vianu, *Foundations of Databases: The Logical Level.* Addison-Wesley, 1994.

[4] M. Casanova, R. Fagin, and C. Papadimitriou, "Inclusion dependencies and their interaction with functional dependencies", *J. Comput. Syst. Sci.*, vol. 28, no. 1, pp. 29–59, 1984. DOI: 10.1016/0022-0000(84)90075-8.

[5] M. Levene and M. Vincent, "Justification for inclusion dependency normal form", *IEEE Trans. Knowl. Data Eng.*, vol. 12, no. 2, pp. 281–291, 2000. DOI: 10.1109/69.842267.

[6] H. Köhler and S. Link, "Inclusion dependencies reloaded", in *The 24th ACM International on Conference on Information and Knowledge Management (CIKM '15)*, 2015, pp. 1361–1370. DOI: 10.1145/2806416.2806539.

[7] V. Zykin and S. Zykin, "Analysis of Typed Inclusion Dependences with Null Values", *Automatic Control and Computer Sciences*, vol. 52, no. 7, pp. 638–646, 2018. DOI: 10.3103/S0146411618070258.

[8] J. Rissanen, "Theory of relations for databases – a tutorial survey", in *Mathematical Foundations of Computer Science 1978*, J. Winkowski, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 1978, pp. 536–551.

[9] E. Sciore, "A Complete Axiomatization of Full Join Dependencies", *J. ACM*, vol. 29, no. 2, pp. 373–393, 1982. DOI: 10.1145/322307.322313.

[10] C. Beeri and M. Vardi, "Formal systems for join dependencies", *Theoretical Computer Science*, vol. 38, pp. 99–116, 1985. DOI: 10.1016/0304-3975(85)90212-9.

[11] M. Gyssens, "On the complexity of join dependencies", *ACM Transactions on Database Systems*, vol. 11, no. 1, pp. 81–108, 1986. DOI: 10.1145/5236.5237.

[12] F. Malvestuto, "A Complete Axiomatization of Full Acyclic Join Dependencies", *Inf. Process. Lett.*, vol. 68, no. 3, pp. 133–139, 1998.

[13] I. Düntsch and S. Mikulás, "Cylindric structures and dependencies in relational databases", *Theoretical Computer Science*, vol. 269, pp. 451–468, 2001. DOI: 10.1016/S0304-3975(01)00016-0.

[14] S. Hartmann, H. Köhler, and S. Link, "Full hierarchical dependencies in fixed and undetermined universes", *Annals of Mathematics and Artificial Intelligence*, vol. 50, pp. 195–226, 2007. DOI: 10.1007/s10472-007-9067-0.

[15] J. Biskup and S. Link, "Appropriate inferences of data dependencies in relational databases", *Annals of Mathematics and Artificial Intelligence*, vol. 63, no. 3–4, pp. 213–255, 2011.

[16] M. Hannula and J. Kontinen, "A finite axiomatization of conditional independence and inclusion dependencies", *Information and Computation*, vol. 249, pp. 121–137, 2016. DOI: 10.1016/j.ic.2016.04.001.

[17] J. Baixeries, "A Formal Context for Acyclic Join Dependencies", *Lecture Notes in Computer Science*, vol. 10352, pp. 563–572, 2017. DOI: 10.1007/978-3-319-60438-1_55.

[18] S. Zykin, "Domains of functional dependences in databases", *Trudy Inst. Mat. i Mekh. UrO RAN*, vol. 22, no. 3, pp. 117–129, 2016.