УДК 004.89:004.4

РАЗРАБОТКА РЕЛЯЦИОННЫХ СРЕДСТВ ДЕКОМПОЗИЦИИ ПРЕДИКАТОВ

М.Н. Рудометкина, В.Г. Спицын, В.А. Эттель*

Томский политехнический университет
*Карагандинский государственный технический университет
E-mail: monika_7272@mail.ru

Статья посвящена разработке новых средств декомпозиции предикатов на основе утверждений о зависимостях из области нормализации реляционных отношений. Дано точное определение процедуре декомпозиции по отношению к реляционным отношениям и к предикатам. Рассмотрены утверждения о зависимостях, позволяющих проводить декомпозицию реляционных отношений; предложены их обобщения. Опираясь на связь между реляционной алгеброй и алгеброй предикатов, сформулированы утверждения о зависимостях, преобразованных в новые средства декомпозиции предикатов.

Ключевые слова:
Алгебра предикатов, декомпозиция, реляционная алгебра, зависимость соединения.
Key words:
Predicate algebra, decomposition, relational algebra, dependence of the connection.

Одна из идей алгебраизации математической логики привела к алгебре предикатов, которая активно используется как формульный аппарат описания дискретных отношений. Рассматривается разработка формальных средств построения моделей логических сетей. Эти средства основаны на теоремах из раздела нормализации теории реляционных баз данных и позволяют преобразовывать произвольные уравнения алгебры предикатов в модели логических сетей. Аппаратная реализация логических сетей на программируемых логических интегральных схемах позволит создать новые высокопроизводительные процессоры, обрабатывающие модели естественного языка.

Предикат в базисе дизъюнктивно-конъюнктивной алгебры предикатов (ДКАП) обычно рассматривается как аналитическая запись отношения. Поэтому декомпозиция предиката – это декомпозиция отношения на аналитическом уровне (попросту манипуляция с формулами), когда обрабатываются не сами отношения, а соответствующие им формулы предикатов.

Уточним понятие декомпозиции, как оно понимается в реляционных базах данных. Для этого приведем понятие зависимости соединения, которое используется в теории зависимостей.

Пусть $\Gamma=\{X_1,...,X_l\}$ – семейство подмножеств множества H. Если $H=\bigcup_{i=1}^{l} X_i$, то Γ называется покрытием множества H. Покрытие Γ называется разбиением множества H, если множества $X_1,...,X_l$ попарно не пересекаются.

Пусть R – некоторое отношение, а семейство множеств $\{X_1,...,X_l\}$ образует покрытие множества $\mathrm{at}(R)$. Отношение R удовлетворяет зависимости соединения $^\circ[X_1,...,X_l]$, если

$$R=R[X_1]\circ R[X_2]\circ...\circ R[X_l]. \qquad (1)$$

Выражение (1) является так называемым алгебраическим ограничением целостности [1].

Другие виды зависимостей, которые используются в теории зависимостей – функциональные и многозначные – являются частными случаями зависимости соединения. Всякая зависимость, выполненная в $R\in\mathrm{Re}\ (T)$, позволяет представить R в виде естественного соединения своих проекций (1). В теории реляционных баз данных обычно именно эту процедуру и называют декомпозицией R. Исходя из выражения (1) видно, что непосредственно декомпозиция выполняется путем проектирования исходного отношения по разным группам его атрибутов. Это говорит о том, что оператором декомпозиции служит операция проекции, в то время как оператором композиции (обратной сборки) является естественное соединение.

Однако разбиение отношения может осуществляться разными способами: наряду с проекцией в качестве оператора декомпозиции может быть использована выборка. Тогда оператором композиции уже будет не естественное соединение, а операция объединения. Когда в теории баз данных идет речь о декомпозиции отношений, то практически всегда имеется в виду декомпозиция вида (1). Другие виды практически не используются. Тем не менее, для точности декомпозицию отношений в виде (1) иногда будем называть проекционно-соединительной декомпозицией.

Декомпозиция предиката P – это получение некоторых предикатов $P_1,P_2,...,P_l$ на основе предиката P с возможностью его обратного восстановления из этих предикатов, которое называется композицией предиката P. Важным частным случаем декомпозиции является так называемая бинарная декомпозиция предиката P, характеризующаяся тем, что каждый из предикатов $P_1,P_2,...,P_l$, полученных в результате декомпозиции предиката P, имеет в точности два существенных аргумента. Если оператором композиции (восстановления) служит конъюнкция, то декомпозицию называют конъюнктивной.

Предикатными аналогами реляционных операций проекции являются кванторы существования, аналогом естественного соединения – конъюнкция. Тогда предикатным аналогом проекционно-соединительной декомпозиции будет кванторно-конъюнктивный вид декомпозиции. Его определение заключается в следующем:

Пусть $T \subseteq V \times U$ – остов с конечным множеством имен переменных $V=\{x_1, x_2, ..., x_n\}$, $S=\mathrm{Dom}x_1 \times \mathrm{Dom}x_2 \times ... \times \mathrm{Dom}x_n$. Пусть P – некоторый предикат из Pre (S), а семейство множеств $\{X_1, X_2, ..., X_l\}$ образует покрытие множества V. Будем говорить, что предикат P удовлетворяет зависимости конъюнкции $\wedge\{X_1, X_2, ..., X_l\}$, если

$$\exists Y_1(P) \wedge \exists Y_2(P) \wedge ... \wedge \exists Y_l(P) = P, \qquad (2)$$

где $Y_i = V \setminus X_i$, $i=1, ..., l$.

Представление предиката P в виде (2) будем называть кванторно-конъюнктивной декомпозицией.

Некоторые методы декомпозиции предикатов сводятся к тому, что исходный предикат P заменяется некоторым вспомогательным предикатом P', который строится на основе исходного P с помощью введения одной или нескольких вспомогательных переменных $y_1, y_2, ..., y_k$. Этот предикат связан с исходным предикатом P равенством

$$P = \exists y_1 \exists y_2 ... \exists y_k (P'). \qquad (3)$$

Хотя фактически производится кванторно-конъюнктивная декомпозиция вспомогательного предиката P', но исходный предикат P восстанавливается с помощью равенства (3). Принято считать, что произведена декомпозиция предиката P.

Таким образом, декомпозиция предиката может выполняться без помощи вспомогательных переменных, либо с их помощью. В первом случае декомпозиция называется эквивалентной, во втором – неэквивалентной [2. С. 84]. Так, декомпозиция предиката P' является эквивалентной, а декомпозиция исходного P – неэквивалентной. Кванторно-конъюнктивная декомпозиция является эквивалентной по определению.

Рассмотрим понятия зависимостей и некоторые утверждения относительно них. Последние позволяют для заданного отношения разрешить вопрос о допустимости той или иной проекционно-соединительной декомпозиции. Рассмотрим частные случаи: функциональные и многозначные зависимости, определения которых взяты из [1].

Сначала рассмотрим понятие функциональной зависимости [1. С. 59].

Пусть задано отношение $R \in \mathrm{Re}(T)$, а X и Y – произвольные подмножества множества at(R). Говорят, что Y функционально зависит от X, если для любых двух функциональных данных $f_1, f_2 \in R$ из равенства $f_1[X]=f_2[X]$ следует равенство $f_1[Y]=f_2[Y]$. Если Y функционально зависит от X, то пишут $X \to Y$ и говорят, что отношение R удовлетворяет функциональной зависимости $X \to Y$ или, что функциональная зависимость $X \to Y$ выполняется в отношении R.

Непосредственно из определения следует, что если зависимость $X \to Y$ выполняется в R, то она выполняется в любой его проекции, содержащей множество атрибутов $X \cup Y$.

Если $R \in \mathrm{Re}(T)$ – произвольное отношение, то множество пар (X, Y) таких, что $X \to Y$ в R, обозначается через FD(R). Оно называется структурой функциональных зависимостей отношения R [1. С. 205].

Пусть R – произвольное отношение. В [1. С. 204, 205] доказано, что структура функциональных зависимостей FD(R) обладает следующими ниже свойствами (X, Y, Z, W – подмножества множества at (R)).

Если $X \supseteq Y$, то $X \to Y$. (4)

Если $X \to Y$ и $Y \cup Z \to W$, то $X \cup Z \to W$. (5)

Для дедуктивной системы, заданной правилами (4), (5), алгоритмически разрешима проблема выводимости или логического следования, что означает, что анализ структур функциональных зависимостей может осуществляться автоматически [1. § 5.1]. В частности, существуют алгоритмы, позволяющие для заданного множества функциональных зависимостей Σ найти множество всех зависимостей $\varphi(\Sigma)$, которые логически следуют из множества Σ по правилам (4), (5). Кроме того, имеется алгоритм, позволяющий определить, является некоторая функциональная зависимость Σ следствием множества зависимостей Σ, или нет. Для анализа отношений с функциональными зависимостями все эти алгоритмы основаны на аппарате булевых функций [1. § 5.3].

Следующее ниже утверждение является простым средством декомпозиции отношений. В теории зависимостей оно называется теоремой Хеза [3. С. 293], ее доказательство приводится в [3. С. 316].

Утверждение 1. Теорема Хеза. Пусть R – некоторое отношение из Re (T), а семейство множеств $\{X, Y, Z\}$ образует разбиение множества at(R). Если $X \to Y$ в R, то $R=R[X \cup Y] \circ R[X \cup Z]$. Иными словами Теорема Хеза говорит о том, что функциональная зависимость $X \to Y$ в отношении $R\{X, Y, Z\}$ влечет для него зависимость соединения $^\circ[X \cup Y, X \cup Z]$. И далее вытекает следующее следствие 2:

пусть $R \in \mathrm{Re}(T)$, at(R)$=X \cup Y$, $Y=\{y_1, ..., y_k\}$. Если $X \to Y$ для R, то R удовлетворяет зависимости соединения $^\circ[X \cup y_1, X \cup y_2, ..., X \cup y_k]$.

Чтобы показать справедливость указанного выше следствия достаточно указать на множество зависимостей $X \to y_1, X \to y_2, ..., X \to y_k$, которые следуют из зависимости $X \to Y$ в силу свойств (4), (5).

Рассмотрим понятие многозначной зависимости [1. С. 211].

Пусть задано отношение $R \in \mathrm{Re}(T)$, а X и Y – произвольные подмножества множества at(R). Говорят, что X многозначно определяет Y, если $R=R[X \cup Y] \circ R[X \cup Z]$, где $Z=\mathrm{at}(R) \setminus (X \cup Y)$. Если X многозначно определяет Y, то пишут $X \to > Y$ и говорят, что отношение R удовлетворяет многозначной зависимости $X \to > Y$ или, что многозначная зависимость $X \to > Y$ выполняется в отношении R.

Приведенное определение показывает, что многозначная зависимость является частным случаем зависимости соединения и так же определяется через алгебраическое ограничение целостности [1].

Если $R \in \mathrm{Re}(T)$ – произвольное отношение, то множество пар (X, Y) таких, что $X \to\!> Y$ в R, обозначается через MVD (R). Оно называется структурой многозначных зависимостей отношения R [1. С. 212].

Пусть R – произвольное отношение. В [1. С. 163, 212] доказано, что структура многозначных зависимостей MVD (R) обладает следующими ниже свойствами (X, Y, Z, W – подмножества множества at (R)).

Если $X \supseteq Y$, то $X \to\!> Y$. (6)

Если $X \to\!> Y$ и $Y \cup Z \to\!> W$, то $X \cup Z \to\!> W \setminus (Y \cup Z)$. (7)

Если $X \to\!> Y$ и $X \to\!> Z$, то $X \to\!> Y \cup Z$. (8)

Если $X \to\!> Y$, то $X \to\!> \mathrm{at}\ (R) \setminus (X \cup Y)$. (9)

Свойство (9) говорит о том, что многозначные зависимости всегда образуют связные пары, поэтому для отношения R с $\mathrm{at}(R)=\{X, Y, Z\}$ и зависимостью $X \to\!> Y$ используется обозначение $X \to\!> Y|Z$. Как и для функциональных зависимостей, для дедуктивной системы, заданной правилами (6)–(9), алгоритмически разрешима проблема выводимости [1. § 5.2], поэтому анализ структур многозначных зависимостей также может осуществляться автоматически с применением булевых функций [1. § 5.3].

Согласно [1. С. 163] справедливо следующее утверждение.

Утверждение 3. Пусть $R \in \mathrm{Re}\ (T)$ – произвольное отношение, а X и Y – произвольные подмножества множества $\mathrm{at}(R)$, $Z=\mathrm{at}(R) \setminus (X \cup Y)$. Отношение R удовлетворяет зависимости $X \to\!> Y$ тогда и только тогда, когда

$$\sigma_{(X=A)\wedge(Z=C)}(R)[Y]=\sigma_{X=A}(R)[Y] \qquad (10)$$

для каждого $\{X{:}A\} \in R[X]$ и для каждого $\{Z{:}C\} \in \sigma_{X=A}(R)[Z]$.

Равенство (10) показывает, что для каждого $\{X{:}A\}$ из $R[X]$ множество $\sigma_{(X=A)\wedge(Z=C)}(R)[Y]$ не зависит от $\{Z{:}\ C\} \in \sigma_{X=A}(R)[Z]$. Это означает, что многозначная зависимость $X \to\!> Y$ в R определяет отображение $\{X{:}A\} \to \sigma_{(X=A)\wedge(Z=C)}(R)[Y]$ из $R[X]$ в булеан $\mathbf{P}(R[Y])$, которое можно понимать как многозначную функцию из $R[X]$ в $R[Y]$.

Утверждение 3 в менее формальной форме выбрано в качестве определения многозначной зависимости у Дейта [3. С. 329]. За исключением обозначений это определение выглядит следующим образом.

Пусть имеется некоторое отношение $R \in \mathrm{Re}(T)$, $X, Y, Z \subseteq \mathrm{at}(R)$ и $X \cup Y \cup Z=\mathrm{at}(R)$, причем множества X, Y, Z попарно не пересекаются. Y многозначно зависит от X тогда и только тогда, когда множество значений Y, соответствующее заданной паре (значение X, значение Z) отношения R зависит только от X, но не зависит от Z. В то же время принятое в данной статье определение многозначной зависимости К.Дж. Дейт приводит как теорему Фейгина [3. С. 330], которая формулируется в терминах «тогда и только тогда».

В [1. С. 163] доказывается еще одно утверждение.

Утверждение 4. Пусть $R \in \mathrm{Re}(T)$ – произвольное отношение, а X и Y – произвольные подмножества множества at (R), $Z=\mathrm{at}(R) \setminus (X \cup Y)$. Отношение R удовлетворяет зависимости $X \to\!> Y$ тогда и только тогда, когда

$$\sigma_{X=A}(R)[Y \cup Z]=\sigma_{X=A}(R)[Y] \circ \sigma_{X=A}(R)[Z] \qquad (11)$$

для каждого $\{X{:}\ A\} \in R[X]$.

Заметим, что естественное соединение двух отношений, не имеющих общих атрибутов, можно рассматривать как расширенную версию декартова произведения. Равенство (11) выражает именно такой случай. По сути, утверждение 4 выражает еще один способ, которым можно определить многозначную зависимость. Такую форму определения многозначной зависимости можно найти в [4. С. 92].

Согласно утверждению 4 наличие многозначной зависимости $X \to\!> Y$ в отношении $R\{X, Y, Z\}$ означает, что его атрибуты Y и Z являются взаимно независимыми: множество значений каждого из них зависит только от значения атрибута X. Иными словами, многозначные зависимости $X \to\!> Y|Z$ «появляются тогда, когда для осмысленной группы характеристик X множества значений характеристик из групп Y и Z обязаны встречаться в любых сочетаниях при фиксированном значении X-компоненты» [1. С. 216].

Доказано [1. С. 163, 164], что каждая функциональная зависимость является многозначной зависимостью, т. е. если $X \to Y$ в R, то $X \to\!> Y$ в R. Поэтому $\mathrm{FD}(R) \subseteq \mathrm{MVD}(R)$. Таким образом, многозначные зависимости являются обобщением функциональных зависимостей.

Однако между функциональными и многозначными зависимостями имеется существенное различие. Согласно (9) зависимости $X \to\!> Y$ и $X \to\!> \mathrm{at}(R) \setminus (X \cup Y)$ всегда образуют связанные пары, что обозначается как $X \to\!> Y|Z$, где $Z=\mathrm{at}(R) \setminus (X \cup Y)$. Свойство (9) не имеет аналога для функциональных зависимостей, поэтому в общем случае они не образуют связанных пар. Например, зависимость $X \to Y$ влечет $X \to\!> \mathrm{at}\ (R) \setminus (X \cup Y)$; однако это совсем не означает, что зависимость $X \to\!> \mathrm{at}\ (R) \setminus (X \cup Y)$ тоже является функциональной.

Следующее ниже утверждение 5 обобщает утверждение 4. На этот раз используется понятие зависимости соединения.

Утверждение 5. Пусть R – некоторое отношение из Re (T), а семейство множеств $\{X, Z_1, Z_2, \ldots, Z_m\}$ образует разбиение множества at (R). Отношение R удовлетворяет зависимости соединения $^\circ[X \cup Z_1, X \cup Z_2, \ldots, X \cup Z_m]$ тогда и только тогда, когда

$$\sigma_{X=A}(R)[Z_1 \cup Z_2 \ldots \cup Z_m]= =\sigma_{X=A}(R)[Z_1] \circ \sigma_{X=A}(R)[Z_2] \circ \ldots \circ \sigma_{X=A}(R)[Z_m] \qquad (12)$$

для каждого $\{X{:}A\} \in R[X]$.

Утверждение 6. Обобщение теоремы Фейгина. Пусть R – некоторое отношение из Re (T), а семейство множеств $\{X, Z_1, Z_2, \ldots, Z_m\}$ образует разбиение множества at (R). Отношение R удовлетворяет зависимости соединения $^\circ[X \cup Z_1, X \cup Z_2, \ldots, X \cup Z_m]$ тогда и только тогда, когда оно одновременно удовлетворяет многозначным зависимостям $X \to\!> Z_1, X \to\!> Z_2, \ldots, X \to\!> Z_m$.

Доказательство. Необходимость. Для доказательства воспользуемся следующим вспомогательным понятием. Пусть семейство множеств $\{X_1,...,X_m\}$ является покрытием множества U, т. е. $U=\bigcup_{i=1}^{m} = X$. На основе множеств $X_1,...,X_m$ образуем новые множества $Y_1,..., Y_l$ по формуле

$$Y_j = \bigcup_{s=1}^{l_j} = X_{i_s}, \quad j=1,...,l, \quad i=1,...,m.$$

Если каждое множество X_i входит хотя бы в одно множество Y_j, то $U=\bigcup_{j=1}^{l} Y_j$. Покрытие $\{Y_1,..., Y_l\}$ называется укрупнением покрытия $\{X_1,...,X_m\}$ [1. С. 232].

Рассмотрим одно из свойств зависимостей соединения: из $^\circ[X_1,...,X_m]$ следует $^\circ[Y_1,..., Y_l]$ для любого укрупнения покрытия $\{X_1,..., X_m\}$.

Пусть $Z=\bigcup_{k=1}^{m} Z_k$. Согласно определению многозначная зависимость $X\to>Z_k$, выполненная в R, равносильна зависимости соединения $^\circ[X\cup Z_k, X\cup(Z\backslash Z_k)]$. Очевидно, что для любого $k=1,..., m$ покрытие $\{X\cup Z_k, X\cup(Z\backslash Z_k)\}$ является укрупнением покрытия $\{X\cup Z_1, X\cup Z_2,...,X\cup Z_m\}$. Поэтому для отношения R любая зависимость $^\circ[X\cup Z_k, X\cup(Z\backslash Z_k)]$ является следствием зависимости $^\circ[X\cup Z_1, X\cup Z_2,...,X\cup Z_m]$.

Достаточность. Пусть $Z=\bigcup_{k=1}^{m} Z_k$, $Q_A=\sigma_{X=A}(R)[Z]$, $A\in R[X]$. Тогда с учетом свойств операции проекции равенство (13) можно представить в виде:

$$Q_A=Q_A[Z_1]\circ Q_A[Z_2]\circ...\circ Q_A[Z_m]. \quad (13)$$

В силу утверждения 5 выполнение (14) при каждом $A\in R\,[X]$ равносильно наличию в R зависимости $^\circ[X\cup Z_1, X\cup Z_2,...,X\cup Z_m]$.

Учитывая введенные обозначения, утверждение 4 гласит, что зависимость $X\to>Z_k$ $(k=1,2,..., m)$ в отношении R равносильна выполнению равенства

$$Q_A=Q_A[Z_k]\circ Q_A[Z\backslash Z_k] \quad (14)$$

при любом $A\in R[X]$. Тогда множество зависимостей $X\to>Z_1$, $X\to>Z_2$,..., $X\to>Z_m$ выполненных в R, равносильно тому, что для любого $A\in R\,[X]$ справедлива система, которая состоит из равенств (14) с различными значениями $k=1,2,...,m$.

Для доказательства достаточности данного утверждения необходимо показать, что множество зависимостей $X\to>Z_1$, $X\to>Z_2$,..., $X\to>Z_m$, имеющихся в R, влечет зависимость $^\circ[X\cup Z_1, X\cup Z_2,...,X\cup Z_m]$. Иными словами, требуется доказать, что справедливость системы из m равенств

$$\begin{cases} Q_A = Q_A[Z_1]\circ Q_A[Z\setminus Z_1] \\ Q_A = Q_A[Z_2]\circ Q_A[Z\setminus Z_2] \\ ... \\ Q_A = Q_A[Z_m]\circ Q_A[Z\setminus Z_m] \end{cases}$$

при каждом $A\in R[X]$ влечет справедливость равенства (14) при каждом $A\in R[X]$. Покажем это. Так как $\text{at}\,(Q_A)=Z_1\cup Z_2\cup...\cup Z_m$, то включение

$$Q_A \subseteq Q_A[Z_1]\circ Q_A[Z_2]\circ...\circ Q_A[Z_m]$$

выполняется всегда. Чтобы доказать равенство (14), надо показать, что обратное включение тоже справедливо.

Пусть

$t=\{Z_1{:}C_1, Z_2{:}C_2,...,Z_m{:}C_m\}\in Q_A[Z_1]\circ Q_A[Z_2]\circ...\circ Q_A[Z_m]$.

Значит, отношение Q_A содержит кортежи

$$t'=\{Z_1{:}\ C_1,\ Z_2{:}\ C'_2,...,\ Z_m{:}\ C'_m\},$$
$$t''=\{Z_1{:}\ C''_1,\ Z_2{:}\ C_2,...,\ Z_m{:}\ C''_m\},$$
$$...$$
$$t^{(m)}=\{Z_1{:}\ C_1^{(m)},\ Z_2{:}\ C_2^{(m)},...,\ Z_m{:}\ C_m\}.$$

Пусть $t_1=t'$. Последовательно образуем кортежи $t_2,t_3,...,t_m$ по формуле

$$t_k=t^{(k)}=[Z_k]\circ t_{k-1}[Z\backslash Z_k],\ i=2,..., m. \quad (15)$$

Если $t_{k-1}\in Q_A$, то, согласно (14), t_k тоже принадлежит Q_A. Кортеж $t_1\in Q_A$ по условию, поэтому все последующие кортежи – t_2, t_3,...,t_m – тоже принадлежат отношению Q_A. В соответствии с (15) имеем

$$t_2=\{Z_1{:}\ C_1,\ Z_2{:}\ C_2,\ Z_3{:}\ C'_3,...,\ Z_m{:}\ C'_m\},$$
$$...$$
$$t_k=\{Z_1:C_1,\dots,Z_k:C_k,Z_{k+1}:C_{k+1}^{(k-1)},\dots,Z_m:C_m^{k-1}\},$$
$$...$$
$$t_m=\{Z_1{:}\ C_1,\ Z_2{:}\ C_2,...,\ Z_m{:}\ C_m\}.$$

Так как $t=t_m$, то $t_1\in Q_A$. Следовательно, $Q_A[Z_1]\circ Q_A[Z_2]\circ...\circ Q_A[Z_m]\subseteq Q_A$ и равенство (13) установлено. Поскольку выбор $A\in R[X]$ никак не влияет на ход вышеизложенных рассуждений, то равенство (14) справедливо при любом $A\in R[X]$. Утверждение доказано.

Множество многозначных зависимостей $X\to>Z_1, X\to>Z_2,...,X\to>Z_m$, которым удовлетворяет R с $\text{at}(R)=X\cup Z_1\cup...\cup Z_m$, как и равносильную ему одну зависимость соединения $^\circ[X\cup Z_1, X\cup Z_2,..., X\cup Z_m]$, для удобства иногда будем обозначать через $X\to>Z_1|Z_2|...|Z_m$.

Свойство многозначных зависимостей (9) позволяет представить зависимость $X\to>Z_1|Z_2|...|Z_m$ в виде.

$$\left.\begin{array}{l} X\to> Z_1\,|\,Z_2\cup Z_3\cup...\cup Z_m, \\ X\to> Z_2\,|\,Z_1\cup Z_3\cup...\cup Z_m, \\ ... \\ X\to> Z_m\,|\,Z_1\cup Z_2\cup...\cup Z_{m-1}. \end{array}\right\} \quad (16)$$

Если $m=2$, то множество (16) состоит из пары многозначных зависимостей; если $m>2$, – то из $2m$ зависимостей. На практике возможны как две, так и более многозначных зависимостей в выражении (16), что обеспечивает практическую ценность утверждению 6. Как и теорему Фейгина это утверждение можно использовать для приведения схемы отношения к четвертой нормальной форме при проектировании реляционной базы данных [3].

Согласно утверждению 4 многозначная зависимость $X\to>Y|Z$, выполненная в отношении R с $\text{at}(R)=X\cup Y\cup Z$, выражает взаимную независимость атрибутов Y и Z в отношении R, притом что множество значений каждого из них определяется значением атрибута X. Точно так же утверждение 5 позволяет рассматривать зависимость $X\to>Z_1|Z_2|...|Z_m$, выполненную в R, $\text{at}(R)=X\cup Z_1\cup Z_m$: она выражает взаимную независимость атрибутов $Z_1,Z_2,...,Z_m$ друг от друга, притом что множество значений каждого из них определяется значением атрибута X.

Согласно теореме Хеза функциональная зависимость выступает в роли достаточного условия для декомпозиции заданного отношения на две своих проекции. Утверждения 3 и 4 можно рассматривать как критерии наличия определенной многозначной зависимости в заданном отношении. Утверждение 5 является критерием наличия зависимости соединения определенного вида, а утверждение 6 устанавливает равносильность такой зависимости группе многозначных зависимостей с одинаковым детерминантом. Благодаря этому утверждение 5 можно рассматривать как обобщение утверждения 4 на случай группы многозначных зависимостей.

Сформулируем и докажем некоторые из рассмотренных утверждений в терминах подстановочной дизъюнктивно-конъюнктивной алгебры предикатов (ПДКАП) – для данной цели базиса этой алгебры достаточно. Для этого сначала необходимо сформулировать понятия функциональной, многозначной зависимости и зависимости соединения для предикатов.

Согласно [1. С. 162] функциональная зависимость аргумента x_n от аргументов $x_1,\ldots,x_{n-1}$ в предикате $P(x_1,\ldots,x_n)$ выражается следующей формулой:

$$\forall x_1 \forall x_2 \ldots \forall x_n \forall x'_n (P(x_1,\ldots,x_{n-1},x_n) \wedge \wedge P(x_1,\ldots,x_{n-1},x'_n) \Rightarrow x_n = x'_n. \qquad (17)$$

Любая функциональная зависимость одной группы переменных предиката $P(x_1,\ldots,x_n)$ от другой группы может быть задана либо несколькими формулами вида (17), либо одной такой формулой, в которой после импликации будет стоять конъюнкция нескольких предикатов равенства. Если X и Y являются подмножествами переменных предиката $P(x_1,\ldots,x_n)$, и Y функционально зависит от X, то будем писать $X \to\!> Y$ и говорить, что предикат $P(x_1,\ldots,x_n)$ удовлетворяет функциональной зависимости $X \to\!> Y$.

Зависимость конъюнкции для предикатов была определена выражением (2). Она является аналогом зависимости соединения отношений. Многозначную зависимость для предикатов можно выразить через зависимость конъюнкции. Будем говорить, что предикат $P \in \mathrm{Pre}(S)$ удовлетворяет многозначной зависимости $X \to\!\!\to\!> Y$, если он удовлетворяет зависимости конъюнкции $\wedge\{X \cup Y, X \cup Z\}$, где $Z = V \setminus (X \cup Y)$. Ясно, что в этом случае так же будет выполняться зависимость $X \to\!\!\to\!> Z$, что можно обозначить как $X \to\!\!\to\!> Y|Z$.

Как и для реляционных отношений, для предикатов также можно задать структуры зависимостей. Пусть $P \in \mathrm{Pre}(S)$, $X, Y \subseteq V$. Тогда множество пар (X, Y), таких, что $X \to Y$ в P, обозначим через $\mathrm{FD}(P)$ и назовем структурой функциональных зависимостей предиката P. Аналогично множество пар (X, Y), таких, что $X \to\!\!\to\!> Y$ в P, обозначим через $\mathrm{MVD}(P)$ и назовем структурой многозначных зависимостей предиката P.

Судя по всему, множества $\mathrm{FD}(P)$ и $\mathrm{MVD}(P)$ будут обладать точно такими же свойствами, как структуры зависимостей для отношений. Но для проверки этого предположения требуется дополнительное исследование, которое выходит за рамки данной статьи.

Переходим к разработке предикатных утверждений о зависимостях.

Прежде всего, рассмотрим теорему Хеза, следствие 2 и утверждение 6. Так как в них отсутствуют алгебраические выражения, то их не нужно формулировать отдельно в терминах подстановочной дизъюнктивно-конъюнктивной алгебры предикатов. Так как практически очевидно, что зависимости в предикатах подчиняются тем же свойствам, что и зависимости в реляционных отношениях, то указанные три утверждения можно применять как к отношениям, так и к предикатам. Если доказать, что указанные свойства совпадают, то весь аппарат теории зависимостей можно будет обоснованно применять для анализа зависимостей в предикатах, что существенно расширит возможности декомпозиции предикатов.

Для определенности уточним, что означают теорема Хеза, следствие 2 и утверждение 6 по отношению к предикатам.

Рассмотрим теорему Хеза. Пусть $P \in \mathrm{Pre}\ (S)$, $V = X \cup Y \cup Z$. Если P удовлетворяет функциональной зависимости $X \to\!> Y$, то для него выполняется зависимость конъюнкции $\wedge\{X \cup Y, X \cup Z\}$.

Следствие 2 по отношению к предикатам означает следующее. Пусть $P \in \mathrm{Pre}(S)$, $V = X \cup Y$, $Y = \{y_1,\ldots,y_k\}$. Если $X \to\!> Y$ для P, то P удовлетворяет зависимости конъюнкции $\wedge\{X \cup y_1, X \cup y_2, \ldots, X \cup y_k\}$.

Пусть теперь $V = X_1 \cup \ldots \cup X_m \cup Y$. Тогда на основании утверждения 6 зависимость конъюнкции $\wedge\{X_1 \cup Y, X_2 \cup Y, \ldots, X_m \cup Y\}$ в предикате $P \in \mathrm{Pre}\ (S)$ будем отождествлять с множеством многозначных зависимостей $Y \to\!\!\to\!> X_1$, $Y \to\!\!\to\!> X_2, \ldots,$ $Y \to\!\!\to\!> X_m$ в этом предикате и условно обозначать $Y \to\!\!\to\!> X_1|X_2|\ldots|X_m$.

Утверждение 5 содержит алгебраическое выражение (12), поэтому его следует переформулировать в терминах ПДКАП и доказать.

Для удобства в дальнейшем изменим обозначения в условии утверждения 5 и сформулируем его следующим образом. Пусть семейство множеств $\{X_1, X_2, \ldots, X_m, Y\}$ образует разбиение множества at (R) для некоторого $R \in \mathrm{Re}(T)$. Отношение R удовлетворяет зависимости соединения $^{\circ}[X_1 \cup Y, X_2 \cup Y, \ldots, X_m \cup Y]$ тогда и только тогда, когда

$$\sigma_{Y=A} = (R)[X_1 \cup X_2 \ldots \cup X_m] = \sigma_{Y=A} = (R)[X_1] \circ \sigma_{Y=A} = (R)[X_2] \circ \ldots \circ \sigma_{Y=A} = (R)[X_m] \qquad (18)$$

для каждого $\{Y{:}\ A\} \in R\ [Y]$.

Введем несколько удобных обозначений. Пусть задано произвольное разбиение $\Gamma = \{X_1, X_2, \ldots, X_k\}$ множества $V = \{x_1, x_2, \ldots, x_n\}$, $k \le n$. Иногда вместо каждой переменной в отдельности $x_1, x_2, \ldots, x_n$ бывает удобно рассматривать непересекающиеся группы переменных $X_1, X_2, \ldots, X_k$ – слои разбиения множества V. В таких случаях вместо обычной записи $P(x_1, x_2, \ldots, x_n)$ будет применяться условное обозначение $P(X_1, X_2, \ldots, X_k)$.

Пусть $T\subseteq V\times U$ – остов с конечным множеством имен переменных V. Пространство S задано как декартово произведение областей изменения переменных из V. Если $X=\{x_1,\dots,x_k\}\subseteq V$, то через $\mathrm{Dom}X$ будет обозначаться произведение $\mathrm{Dom}x_1\times\dots\times\mathrm{Dom}x_k$ [1. С. 131]. Пусть P – произвольный предикат из Pre (S).

Переходим к утверждению 5. Зависимость соединения $^\circ[X_1\cup Y,X_2\cup Y,\dots,X_m\cup Y]$ перейдет в зависимость конъюнкции $^\wedge\{X_1\cup Y,X_2\cup Y,\dots,X_m\cup Y\}$, которая выражается равенством

$$P=\exists V\setminus(X_1\cup Y)(P)\wedge\exists V\setminus(X_2\cup Y)(P)\wedge\dots\wedge$$
$$\wedge\exists V\setminus(X_m\cup Y)(P). \qquad (19)$$

Равенство (18) примет вид:

$$\exists Y(P\wedge Y^A)=\exists V\setminus X_1(P\wedge Y^A)\wedge$$
$$\wedge\exists V\setminus X_2(P\wedge Y^A)\wedge\dots\wedge\exists V\setminus X_m(P\wedge Y^A). \qquad (20)$$

Представим выражения (19) и (20) в более удобной форме. Пусть X_1, $X_2,\dots X_k\subseteq V$, введем обозначение

$$\exists(X_1\cup X_2\cup\dots\cup X_m)(P)=\exists X_1\exists X_2\dots\exists X_m(P),$$
$$P\in\mathrm{Pre}(S).$$

Перепишем сомножители из (20) с помощью операции подстановки. Пользуясь тождеством $Y/A(P)\equiv\exists Y(P\wedge Y^A)$, $Y\subseteq V$, $Y\subseteq V$, $P\in\mathrm{Pre}(S)$, имеем:

$$\exists V\setminus X_i(P\wedge Y^A)=$$
$$=\exists X_1\dots\exists X_{i-1}\exists X_{i+1}\dots\exists X_m\exists Y(P\wedge Y^A)=$$
$$=\exists X_1\dots\exists X_{i-1}\exists X_{i+1}\dots\exists X_m(\exists Y(P\wedge Y^A))=$$
$$=\exists X_1\dots\exists X_{i-1}\exists X_{i+1}\dots\exists X_m(Y/A(P)),\ i=1,\dots,m.$$

Теперь сформулируем и докажем утверждение 5 для предикатов.

Утверждение 7. Пусть семейство множеств $\{X_1,X_2,\dots,X_m,Y\}$ образует разбиение множества V. Предикат $P\in\mathrm{Pre}(S)$ удовлетворяет зависимости конъюнкции $^\wedge\{X_1\cup Y, X_2\cup Y,\dots, X_m\cup Y\}$, т. е.

$$P=(\exists X_2\exists X_3\dots\exists X_m(P))\wedge$$
$$\wedge(\exists X_1\exists X_3\dots\exists X_m(P))\wedge\dots\wedge(\exists X_1\exists X_2\dots\exists X_{m-1}(P)) \qquad (21)$$

тогда и только тогда, когда

$$Y/A(P)=(\exists X_2\exists X_3\dots\exists X_m(Y/A(P)))\wedge$$
$$\wedge(\exists X_1\exists X_3\exists X_m(Y/A(P)))\wedge\dots\wedge$$
$$\wedge(\exists X_1\exists X_2\dots\exists X_{m-1}(Y/A(P))), \qquad (22)$$

для каждого $A\in P_T[Y]\subseteq\mathrm{Dom}\,Y$, где P_T – область истинности предиката P.

Тогда условие (22) примет вид:

$$Y/A(P)=Y/A((\exists X_2\exists X_3\dots\exists X_m(P))\wedge\dots\wedge$$
$$\wedge(\exists X_1\exists X_2\dots\exists X_{m-1}(P))),$$

это равенство справедливо при любом $A\in P_T[Y]$, если выполняется (21).

Утверждение 8. Пусть семейство множеств $\{X_1,X_2,\dots,X_m Y\}$ образует разбиение множества V. Предикат $P\in\mathrm{Pre}(S)$ удовлетворяет зависимости $^\wedge\{X_1\cup Y,X_2\cup Y,\dots,X_m\cup Y\}$ тогда и только тогда, когда для каждого $A\in P_T[Y]$ (P_T – область истинности предиката P) найдутся такие предикаты $P_1^A(X_1),\dots P_m^A(X_m)\in\mathrm{Pre}(S)$, что подстановку $Y/A(P)=P(X_1,X_2,\dots X_m,A)$ можно представить конъюнкцией

$$P(X_1,\dots,X_m,A)=$$
$$=P_1^A(X_1)\wedge P_2^A(X_2)\wedge\dots\wedge P_m^A(X_m). \qquad (23)$$

Чтобы с помощью утверждения 8 проверить, удовлетворяет ли предикат зависимости $^\wedge\{X_1\cup Y,X_2\cup Y,\dots,X_m\cup Y\}$, выполним дизъюнктивное разложение предиката P по переменным Y:

$$P(X_1,\dots,X_m,A)=\bigvee_{A\in P_T[Y]}Y^A P(X_1,\dots,X_m,A). \qquad (24)$$

Формула любого предиката P $(X_1,\dots,X_m,A)$, $A\in P_T[Y]$ обычно намного меньше, чем формула исходного предиката P $(X_1,\dots,X_m,Y)$. Поэтому после дизъюнктивного разложения легко проверить: можно представить каждый предикат P $(X_1,\dots,X_m,A)$ конъюнкцией (23) или нет. Для сравнительно небольших формул это довольно легко определить по их внешнему виду.

Следует заметить, что равенство (24) не является дизъюнктивным разложением предиката в том виде, который предложен в [2. С. 28], и поэтому правомерность такого разложения еще надо доказать.

Утверждение 9. Модифицированная теорема о дизъюнктивном разложении предиката. Любой конечный предикат PPre(S) может быть представлен в виде

$$P(x_1,x_2,\dots,x_n)=$$
$$=\bigvee_{(a_1,a_2,\dots a_m)\in N}x_1^{a_1}x_2^{a_2}\dots x_m^{a_m}P(a_1,a_2,\dots,a_m,x_{m+1},x_{m+2},\dots x_n). \qquad (25)$$

Таким образом, для правильного представления предиката PPre(S) с помощью дизъюнктивного разложения логическое суммирование можно проводить не по всей области M, а по ее подмножеству N. В дальнейшем при выполнении дизъюнктивного разложения предикатов всегда будет использоваться выражение (25), а кортежи множества N могут быть проиндексированы (в силу конечности множества $N=P_T[x_1,x_2,\dots,x_m]$, которое вытекает из конечности множества P_T).

Следует заметить, что для конъюнктивного разложения [2. С. 36] подобная модификация будет неверна.

Основным методом познания интеллекта считают алгебраизацию логики, т. е. точное формальное описание той алгебры, которая реализована в интеллекте и, как следствие, проявляет себя в естественном языке. В результате такого описания была разработана алгебра предикатов и алгебра предикатных операций. Первая потенциально может описывать мысли, вторая – действия над мыслями, т. е. мышление. Схемная реализация формул этих алгебр приводит к инженерным решениям, которые называются логическими сетями. Логические сети ориентированы на параллельные вычисления и, по мнению создателей, призваны стать основой мозгоподобных компьютеров.

Однако неадекватность тьюрингоподобной или, другими словами, последовательной природы персональных компьютеров феномену естественного языка была осознана уже давно. В поисках новых аппаратных решений для моделирования естественного языка и других функций интеллекта средствами алгебры конечных предикатов в 1980-е гг. разрабатывались специальные технические средства – переключательные цепи. Эти средства были предложены для эффективного решения уравнений алгебры предикатов. В начале 2000-х гг. на смену переключательным цепям первого и второго рода пришла новая более прогрессивная концепция логических сетей, которая основана на идеях теории категорий [5–7].

Логические сети способны стать той аппаратной базой, которая позволит в реальном режиме времени обрабатывать логические уравнения большого объема, описывающие полноценные модели естественного языка (на что последовательные компьютеры не способны в принципе). Развитие метода декомпозиции предикатов основано на утверждениях о функциональных, многозначных зависимостях и зависимости соединения реляционных отношений, теоремах Хеза и Фейгина и утверждениях о предикатных зависимостях. Метод позволяет обоснованно применить аппарат теории зависимостей для анализа зависимостей в предикатах и существенным образом расширить возможности декомпозиции предикатов. Опираясь на связь между реляционной алгеброй и алгеброй предикатов, утверждения о зависимостях преобразованы в новые средства декомпозиции предикатов. С помощью полученных средств декомпозиции разработан метод бинарной декомпозиции функциональных предикатов, который является мощным инструментом декомпозиции предикатов. Непосредственными операторами декомпозиции в данном методе являются кванторы существования [5].

Выводы

1. Выявлены основные виды зависимостей атрибутов в реляционных отношениях: функциональные, многозначные и зависимости соединения.
2. Рассмотрены утверждения о зависимостях, позволяющие выполнять декомпозицию реляционных отношений.
3. Сформулированы и доказаны обобщающие утверждения о зависимостях соединений особого вида, которые усовершенствовали теорию нормализации отношений.
4. Разработаны новые критерии проверки предикатов на выполнение зависимости конъюнкции, что позволяет выполнять декомпозицию предикатов.
5. Выполнена модификация теоремы о дизъюнктивном разложении предиката на случай бесконечного предметного пространства.

СПИСОК ЛИТЕРАТУРЫ

1. Цаленко М.Ш. Моделирование семантики в базах данных. – М.: Наука, 1989. – 288 с.
2. Шабанов-Кушнаренко Ю.П. Теория интеллекта. Математические средства. – Харьков: Вища шк. Изд-во при Харьк. ун-те, 1984. – 144 с.
3. Дейт К.Дж. Введение в системы баз данных / пер. с англ. и ред. К.А. Птицына. – М.: Издательский дом «Вильямс», 2008. – 1327 с.
4. Мишенин А.И. Теория экономических информационных систем. – М.: Финансы и статистика, 2008. – 237 с.
5. Рудометкина М.Н. Разработка метода бинарной декомпозиции функциональных предикатов // Известия Томского политехнического университета. – 2012. – Т. 320. – № 5. – С. 5–11.
6. Бондаренко М.Ф., Чикина В.А., Шабанов-Кушнаренко Ю.П. Модели языка // Бионика интеллекта. – 2004. – № 1 (61). – С. 27–37.
7. Bondarenko M.F., Hahanova I.V. Logic networks application for computing process organization // Радиоэлектроника и информатика. – 2003. – № 3. – С. 150–156.

Поступила 31.01.2013 г.