

Л. ВЫХАНДУ, Я. ТЕПАНДИ

УДК 681.3.06

МОДЕЛЬ НЕПРОЦЕДУРНОГО ЯЗЫКА ДОСТУПА К СЕТЕВОЙ БАЗЕ ДАННЫХ

(Представил А. Хумал)

Для автоматизации проектирования пакетов программ проблемной ориентации, основанных на системах управления базами данных, необходимы общие и в то же время достаточно эффективные средства доступа проблемных программ к базам данных. В настоящей статье изучаются свойства не процедурного языка отбора данных из сетевой базы данных, подмножество которого реализовано в пакете СХОДИ [1].

Опишем неформально не процедурный язык и поставленные проблемы. Для этого рассмотрим простой запрос к сетевой базе данных: задан набор T с типом В-записей R_1 и типом Ч-записей R_2 ; задано множество M записей типа R_2 (напр., фиксированием значения некоторого элемента). Требуется найти все записи типа R_1 , которые связаны с записями из множества M через набор T . Отметим, что здесь неявно ставится условие: каждая запись, выданная в ответ на запрос, должна иметь по меньшей мере одну Ч-запись из множества M в экземпляре набора T . Ответ на этот запрос однозначен и требует не более одного просмотра записей типа R_1 и R_2 . Обобщение этого и подобных ему запросов можно представить следующим образом. Дано: множество имен типов записей и типов наборов (**схема области поиска**); фиксированные подмножества записей некоторых типов из области поиска (**сужение поиска**); множество имен **требуемых типов записей** из области поиска; **условия согласованности**, которым должен удовлетворять ответ на запрос. Требуется: найти наибольшее множество всех записей требуемых типов, входящих в **область согласованности** — область поиска, удовлетворяющую ограничениям сужения поиска и условиям согласованности. Ответ на такой запрос не является в общем случае однозначным и требует итеративного определения области согласованности, инвариантной к данным условиям согласованности.

Цель настоящей статьи — выделение классов запросов, для которых существует однозначный ответ, или быстрый алгоритм нахождения ответа. Для этого строятся формальные модели состояния базы данных и запроса к базе данных. Выделяется класс запросов, для которых существует однозначный ответ. Из этого класса выделяется подкласс, для которого число проходов одного набора, нужное для выполнения запроса, определяется на уровне схемы и не зависит от содержания базы данных. Формальные построения основываются на понятиях алгебраической системы [2] и смешанного графа [3]. Вводя понятие В-вычисления, мы развиваем подход, предложенный в [4-6].

Модели

Пусть $\mathfrak{A} = \langle A, v^{\mathfrak{A}} \rangle$ — алгебраическая система сигнатуры $\Sigma = \langle R, F, \mu \rangle$ [2], причем A является счетным множеством, а F содержит только символы констант. Конституентой сигнатуры Σ называется формула вида $p(c_1, \dots, c_k)$, где $p \in R$, $\mu(p) = k$, $c_1, \dots, c_k \in F$ [4]. Пусть Z — некоторое множество конституент сигнатуры Σ . Алгебраическая система сигнатуры Σ с носителем A , в котором истинны те и только те конституенты, которые входят в Z , обозначается через \mathfrak{A}_Z и, наоборот, множество всех конституент сигнатуры Σ , истинных на \mathfrak{A} , обозначается через $K(\mathfrak{A})$. Если $N \subseteq R$, тогда множество всех конституент из Z вида $r(c_1, \dots, c_k)$, где $r \in N$, $c_1, \dots, c_k \in F$, обозначается через $Z[N]$.

Определение 1. Упорядоченная пара $G = \langle \Sigma, \mathfrak{A} \rangle$ называется моделью состояния базы данных с сетевой структурой (коротко — **базой данных**), если выполняются следующие условия:

- 1) $\mathfrak{A} = \langle A, v^{\mathfrak{A}} \rangle$ — алгебраическая система сигнатуры $\Sigma = \langle R, F, \mu \rangle$;
- 2) носитель A системы \mathfrak{A} — конечное множество (по смыслу — множество записей);
- 3) $R = \{R_0, R_1, \dots, R_n, T_1, \dots, T_m\}$; $\mu(R_i) = 1$, $i = 0, 1, \dots, n$; $\mu(T_j) = 2$; $j = 1, \dots, m$ (по смыслу: R_0, R_1, \dots, R_n — имена типов записей; R_0 — тип записей SYSTEM; T_1, \dots, T_m — имена типов наборов);
- 4) $F = \{c_0, \dots, c_k\}$, $\mu(c_i) = 0$, $i = 0, \dots, k$ (по смыслу — множество идентификаторов записей);
- 5) на системе \mathfrak{A} истинны следующие предложения, имеющие естественное истолкование в базе данных:

$$(D1) \quad (\forall x) (R_0(x) \rightarrow (x = c_0));$$

$$(D2) \quad (\forall x) (R_j(x) \rightarrow \neg R_l(x)), \quad j \neq l, \quad j, l = 0, \dots, n;$$

$$(D3) \quad (\forall x_1 x_2 y) [(T_i(x_1, y) \& T_i(x_2, y)) \rightarrow (x_1 = x_2)], \quad i = 1, \dots, m.$$

Для каждого T_i , $i = 1, \dots, m$, однозначно определены $R_{i_1}, R_{i_2} \in R$, $i_1 \neq i_2$, такие, что справедливо

$$(D4) \quad (\forall xy) [T_i(x, y) \rightarrow (R_{i_1}(x) \& R_{i_2}(y))].$$

В некоторых случаях, кроме предложений вида (D4), выполняются еще предложения вида (D5) и вида (D6), где i, i_1, i_2 имеют те же значения, что и в (D4):

$$(D5) \quad (\forall y) [R_{i_2}(y) \rightarrow (\exists x) (T_i(x, y) \& R_{i_1}(x))];$$

$$(D6) \quad (\forall x) [R_{i_1}(x) \rightarrow (\exists y) (T_i(x, y) \& R_{i_2}(y))], \quad i_1 \neq 0.$$

Определение 2. Запросом к базе данных $\langle \Sigma, \mathfrak{A} \rangle$ называется четверка $P = \langle X, E, D, U \rangle$, где:

- 1) $X \subseteq R$ (схема области поиска); обозначим $V = X \cap \{R_0, R_1, \dots, R_n\}$, $W = X \cap \{T_1, \dots, T_m\}$;
- 2) E — множество предложений (условий сужения поиска) вида $(\forall x) (R_i(x) \rightarrow S_i(x))$, $R_i \in V$, $i \neq 0$; S_i — некоторые новые предикатные символы;
- 3) D — множество замкнутых формул сигнатуры Σ вида

$$(A) \quad (\forall x_1 \dots x_k) (\varphi \rightarrow \psi),$$

где φ — конъюнкция атомных формул, ψ — формула [2]; предложения из D называются **условиями согласованности**;

4) $U \subseteq V \setminus \{R_0\}$ (множество имен **требуемых типов записей**).

Определение 3. Пусть дан запрос $P = \langle V \cup W, E, D, U \rangle$, где множество D содержит только предложения видов (D4), (D5), (D6). **Графом запроса** P называется смешанный граф $\langle V, W, Q \rangle$, где инцидентор Q [3] истинен на всех тройках типа $\langle R_1, T, R_2 \rangle$, для которых существует предложение вида (D5) или вида (D6) с предикатными символами R_2 в условии и символами T, R_1 в следствии импликации. Запрос называется **сильно ациклическим**, если его граф не содержит замкнутых путей.

Определение 4. **В-вычислением** над множеством конститuent Z_0 по множеству F предложений вида (A) (**по программе F**) называется последовательность $Z_0, Z_1, \dots, Z_k = Z$, где:

1) $Z_{i+1} = Z_i \setminus W_i, i = 0, \dots, k-1$;

2) W_i — непустое множество всех конститuent из Z_i , входящих в условие φ некоторого предложения $(\forall x_1 \dots x_n) (\varphi \rightarrow \psi)$ из F при некоторой интерпретации $\gamma: X \rightarrow A$ множества X свободных переменных формулы $(\varphi \rightarrow \psi)$, для которой $\mathcal{A}_{Z_i} \Rightarrow \neg(\varphi \rightarrow \psi) [\gamma]$. Если \mathcal{A}_Z — модель множества формул F , то говорят, что В-вычисления **закончены** и Z есть **результат В-вычисления**. Выбор различных предложений $\alpha \in F$ и различных интерпретаций γ определяет различный порядок В-вычислений.

Определение 5. Пусть имеется запрос $P = \langle X, E, D, U \rangle$ к базе данных $G = \langle \Sigma, \mathcal{A} \rangle$, причем $\Sigma = \langle R, F, \mu \rangle$. Пусть $S = \{S_i \mid S_i \text{ входит в некоторое предложение из } E\}$, $R_E = R \cup S$, а алгебраическая система \mathcal{A}_E сигнатуры $\Sigma_E = \langle R_E, F, \mu_E \rangle$ есть обогащение системы \mathcal{A} . **Выполнением запроса P над базой данных G** называется В-вычисление над множеством конститuent $Z_0 = K(\mathcal{A}_E)[X \cup S]$ по программе $E \cup D$. **Результатом выполнения запроса P** называется множество $Z[U]$, где Z — результат соответствующего В-вычисления.

Вычисления

Теорема 1. Пусть программа F состоит из замкнутых формул вида (B1), (B2):

$$(B1) \quad (\forall x_1 \dots x_n) (r(x_{i_1}, \dots, x_{i_k}) \rightarrow \psi),$$

$$(B2) \quad (\forall x_1 \dots x_n) (r(x_{i_1}, \dots, x_{i_k}) \rightarrow (Jy_1 \dots y_m) \psi),$$

где $\{i_1, \dots, i_k\} = \{1, \dots, n\}$, $\mu(r) = k$, ψ — конъюнкция атомных формул вида $r(v_{j_1}, \dots, v_{j_l})$, $\mu(p) = l$. Тогда В-вычисление над конечным множеством конститuent Z_0 заканчивается за конечное число шагов, а результат не зависит от порядка вычислений.

Замечание. Если условие некоторого предложения из F есть конъюнкция двух и более атомных формул, то результат В-вычисления может оказаться неоднозначным, причем алгебраические системы, соответствующие различным результатам, могут быть неизоморфными.

Следствие. Пусть D — множество условий согласованности вида (B1) или вида (B2) (заметим, что таковыми являются и условия видов (D4), (D5), (D6)). Выполнение запроса $P = \langle X, E, D, U \rangle$ над базой

данных $\langle \Sigma, \mathfrak{M} \rangle$ заканчивается за конечное число шагов, а результат Z соответствующего В-вычисления и результат выполнения запроса P определяются однозначно. Формулы видов $(D1)$, $(D2)$, $(D3)$, истинные на алгебраической системе \mathfrak{M} , истинны также и на \mathfrak{M}_Z .

В общем случае выполнение запроса представляет собой итерационный процесс определения области согласованности, инвариантной к данным ограничениям. Ниже показывается, что для класса сильно ациклических запросов существует алгоритм, позволяющий избежать итерации.

Алгоритм А.

Вход. Сильно ациклический запрос $P = \langle X, E, D, U \rangle$ к базе данных $G = \langle \Sigma, \mathfrak{M} \rangle$.

Выход. Результат выполнения запроса P над базой данных G .

Метод. К P и G применяется следующая программа:

начало

1. $Y \leftarrow \{R \mid R \text{ входит в некоторое предложение из } E\};$
2. $Z_E \leftarrow \{R(c) \mid (\forall x) (R(x) \rightarrow S(x)) \text{ входит в } E\} \& S(c) \& R(c)\};$
3. $Z \leftarrow Z_E \cup K(\mathfrak{M})[X \setminus Y];$
4. **если** система \mathfrak{M} не является моделью множества формул D **то** $Y \leftarrow V$;

коммент если \mathfrak{M} не является моделью D , то любое условие согласованности может вызывать уменьшение множества конститuent, влияющее на требуемые типы записей;

5. найти множество M путей от всех $y \in Y$ до всех $u \in U$;
6. **пока** множество M не пусто **делать**

начало

7. найти ребро T , которое является начальным в путях $m_1, \dots, m_k \in M, k \geq 1$, и не входит ни в какой путь из $M \setminus \{m_1, \dots, m_k\}$. Пусть R_2 — конец ребра T ;
8. $D_T \leftarrow$ все предложения программы D , содержащие предикатный символ T ;
9. $Z \leftarrow$ результат В-вычисления над множеством Z по программе D_T ;
10. **если** множество конститuent вида $R_2(c)$ уменьшилось при вычислении по программе D_T **то** удалить из путей m_1, \dots, m_k ребро T
11. **иначе** удалить из M пути m_1, \dots, m_k ;
12. удалить из M все пути нулевой длины;

конец

13. $Z \leftarrow Z[U];$

коммент Z — результат выполнения запроса;

конец

Теорема 2. Алгоритм А выдает результат выполнения сильно ациклического запроса P над базой данных G , совершая при этом не более одного В-вычисления по программе D_T для каждого $T \in W$.

ЛИТЕРАТУРА

1. Выханду Л. К., Лучковский Т. Ф., Микли Т. И., Тепанди Я. Я., УСнМ, № 1, 99—102 (1981).
2. Ершов Ю. Л., Палютин Е. А., Математическая логика, М., «Наука», 1979.
3. Мелихов А. Н., Ориентированные графы и конечные автоматы, М., «Наука», 1971.
4. Клещев А. С., Программирование, № 4, 20—29 (1980).
5. Крицкий С. П., В кн.: Автоматизация производства пакетов прикладных программ (автоматизация производства трансляторов), Тезисы докладов, Таллин, Изд. ТПИ, 1980, с. 95—98.
6. Тепанди Я. Я., В кн.: Всесоюзная конференция по методам трансляции (3—5 февраля 1981 г.), Тезисы докладов, Новосибирск, Изд. ВЦ СО АН СССР, 1981, с. 75—77.

Таллинский политехнический
институт

Поступила в редакцию
7/IV 1981

L. VOHANDU, J. TEPANDI

VÖRKSTRUKTUURILISE ANDMEBAASI MITTEPROTSEDUURSE POORDUSKEELE MUDEL

Artikkel käsitleb probleemorienteeritud andmeotsingukeele seostamist võrkstruktuuriliste andmepankade juhtimise süsteemiga. On esitatud andmebaasi, otsingukeeles antud päringu ja päringu täitmise mudelid, kasutades algebraliste süsteemide teooriat. Näidatakse, et päringud, mille kooskõlatingimused omavad kuju

$$\forall x_1 \dots x_n (r(x_{i_1}, \dots, x_{i_k}) \rightarrow \mathbf{J}y_1 \dots y_m \psi),$$

kus r on predikaadi sümbol ja ψ atomaarsete valemite konjunktsioon, vastavad võrkstruktuurilise andmebaasi omadustele ja on ühesed. On iseloomustatud tugevalt atsüklilisi päringuid, mille täitmine ei vii iteratsioonini.

L. VOHANDU, J. TEPANDI

A MODEL OF A NONPROCEDURAL ACCESS LANGUAGE FOR A NETWORK DATA BASE

Unambiguousness and complexity properties of a retrieval language for a network data base (DB) are studied. A query involves: retrieval region schema — a set of names of record types and set types; retrieval restriction — a subset of record occurrences for some record types from retrieval region; a set of names of the requested record types; balancing conditions for the retrieval region. The answer must contain all the records of the requested types, which belong to the balancing region — the greatest retrieval region satisfying the retrieval restrictions and the balancing conditions.

The formal descriptions of DB, query and calculation on the DB, based on the notion of the algebraic system (structure) are presented. A class of queries with balancing conditions of the type $\forall x_1 \dots x_n (r(x_{i_1}, \dots, x_{i_k}) \rightarrow \mathbf{J}y_1 \dots y_m \psi)$ and similar to them, where r is a predicate symbol, ψ — a conjunction of the atomic formulae, is introduced. It is shown that those queries are natural for a network DB and that they lead to a uniquely defined set of records. A subclass of strongly acyclic queries is introduced. An algorithm is given to implement those queries through a noniterative process.