УДК 510.67

РАЗРЕШИМАЯ ТЕОРИЯ БЕЗ ТРАНСЛЯЦИОННОЙ ТЕОРЕМЫ[1]

Дудаков С.М.
Кафедра информатики

Предлагается пример теории, который опровергает гипотезу из [1] о том, что не существует разрешимых теорий без транансляционной теоремы.

We construct a theory which the hypothesis from [1] that there is no decidable theory without the collapse result.

Введение. Типичной моделью базы данных со времён Кодда является реляционная модель, в которой база данных мыслится как собрание конечного числа конечных таблиц (см. Кодд, [4, 5]). Эта модель реализуется в большинстве существующих средств управлениями базами данных и в предлагаемых ими языках запросов. При этом в качестве языка запросов обычно предлагается та или другая стилизация языка логики предикатов первого порядка. Эта традиция тоже восходит к Кодду, который в качестве языка запросов предложил использовать язык реляционных выражений, практически эквивалентный языку логики предикатов первого порядка.

Обычно при этом удобно предполагать, что элементы хранящихся таблиц выбираются из фиксированного линейно упорядоченного множества, называемого универсумом. Например, в качестве такового можно взять множество натуральных чисел, множество всех слов некоторого конечного алфавита или какое-то другое множество. Это множество должно быть бесконечным. Оно может быть снабжено своими собственными отношениями и операциями. Они составляют сигнатуру универсума. Эти отношения и операции обычно по своей природе не могут быть заданы конечными таблицами.

Итак, базы данных предназначены для хранения текущей информации о как-то структурированной предметной области. В каждый момент времени эта информация является конечной и представляет собой конечный набор конечных таблиц ([4, 5]).

Более формально, каждая таблица — это конечноместное конечное отношение, а сама база данных — это конечный набор конечноместных конечных отношений. Для удобства разговора о базе данных каждому ее отношению приписывают некоторое имя с указанием числа аргументов (или местности) этого имени отношения. Схема (или сигнатура) базы данных и есть конечная последовательность этих имен отношений с указанием местности каждого имени. В каждый момент времени именам отношений из этой схемы присвоены некоторые отношения соответствующих местностей. Это — состояние базы данных в данный момент. Состояние называется конечным, если все его отношения конечны (каждая таблица содержит конечное множество строк). Запросами являются формулы логики предикатов первого порядка.

Для линейно упорядоченных универсумов мы будем рассматривать локально генерические запросы. Так называются запросы, которые сохраняются при любых сохраняющих упорядочение отображениях конечных подмножеств универсума в универсум. Это означает, что ответ на такой запрос основывается на хранящейся в базе

[1]Работа выполнена при финансовой поддержке РФФИ, гранты 04-01-00015 и 04-01-00565

данных информации, но не зависит от способа кодировки этой информации при хранении.

При построении запросов можно использовать или только отношения базы данных (включая порядок), или наряду с отношениями базы данных — отношения универсума. В первом случае запросы называются ограниченными, во втором — расширенными. Очевидно, каждый ограниченный запрос является локально генерическим, так как вообще не использует информацию универсума. Утверждение, что каждый локально генерический расширенный запрос эквивалентен для конечных состояний базы данных некоторому ограниченному запросу, называется трансляционной теоремой или трансляционным результатом для рассматриваемого универсума. [2]

Как известно, многие легко вычисляемые запросы не могут быть записаны с помощью отношений базы данных. Например, вопрос о том, является ли количество элементов какого-либо множества четным, в общем случае записан быть не может ([3]). Возникает вопрос: расширяет ли использование знаний универсума выразительные возможности языка запросов?

Известно, что для многих разрешимых теорий это не так — например, для теории действительных чисел или арифметики Пресбургера ([2]). Все известные примеры теорий, которые расширяют выразительные возможности языка запросов, например, арифметика Пеано, были неразрешимы. В [1] высказана гипотеза о том, что разрешимых теорий, расширяющих возможности языков первого порядка, нет.

Данная работа посвящена опровержению этой гипотезы. Мы строим пример теории — обогащение плотного линейного порядка, которая является разрешимой. Вместе с тем, в этой теории существует формула, которая способна из любого конечного множества выделить произвольное подмножество. Таким образом, в данной теории трансляционный результат места не имеет.

1. Конструкция. Рассмотрим множество рациональных чисел $\mathbb{Q}$. Пронумеруем его элементы натуральными числами:

$$\mathbb{Q} = \{q_0, q_1, q_2, \dots\}.$$

Определим на множестве $\mathbb{Q}$ бинарный предикат $\sqsubseteq$.

Определять его будем индукцией по n. Изначально считаем, что у нас есть одно множество $I_{\varnothing}^{\varnothing} = \mathbb{Q}$, а значение предиката $\sqsubseteq$ не определено ни для каких чисел.

Рассмотрим n-ый шаг определения $\sqsubseteq$. По индукции считаем, что значение предиката $\sqsubseteq$ определено на тех парах (q_i, q_j), для которых $i < n$ или $j < n$. При этом построены множества I_L^K для всевозможных $K, L \subseteq \{q_0, \dots, q_{n-1}\}$, которые являются всюду плотными в $\mathbb{Q}$ и образуют разбиение $\mathbb{Q}$. Для всех $i < n$ выполнено $q_i \not\sqsubseteq q_i$. Для всех $i < n$ и $p > i$ таких, что $q_p \in I_L^K$, предикат $\sqsubseteq$ определён так, что

$$q_p \sqsubseteq q_i \Leftrightarrow q_i \in K,$$

$$q_i \sqsubseteq q_p \Leftrightarrow q_i \in L.$$

Итак, начнём определять отношение $\sqsubseteq$ для пар рациональных чисел вида (q_i, q_j), для которых $\min\{i, j\} = n$.

1. Полагаем $q_n \not\sqsubseteq q_n$.

[2] *Англ.* collapse theorem, collapse result.

2. Каждое из существующих множеств I_L^K разобьём на два множества $I_L^{K\cup\{q_n\}}$ и I_L^K, каждое из которых было бы плотно в $\mathbb{Q}$. Для $p > n$ положим $q_p \sqsubseteq q_n$ для $q_p \in I_L^{K\cup\{q_n\}}$ и $q_p \not\sqsubseteq q_n$ для $q_p \in I_L^K$.

3. Каждое из полученных на шаге (2) множеств I_L^K разобьём на два множества $I_{L\cup\{q_n\}}^K$ и I_L^K, каждое из которых было бы плотно в $\mathbb{Q}$. Для $p > n$ положим $q_n \sqsubseteq q_p$ для $q_p \in I_{L\cup\{q_n\}}^K$ и $q_n \not\sqsubseteq q_p$ для $q_p \in I_L^K$.

Очевидно, что построенные таким образом множества снова будут удовлетворять условиям индукционного предположения. Следовательно, этот процесс построения можно продолжать бесконечно, и определить отношение $\sqsubseteq$ для всех рациональных чисел.

2. Результаты. Мы будем рассматривать теорию $T = \mathrm{Th}(\mathbb{Q}, <, \sqsubseteq)$.

Л е м м а 1. *Пусть $a, b, c_1, \ldots, c_k, d_1, \ldots, d_l, e_1, \ldots, e_m, f_1, \ldots, f_n, g_1, \ldots, g_s$ — такие рациональные числа, что*

1) $a < b$;

2) $c_i \neq d_j$ для всех $i \in \{1, \ldots, k\}$ и $j \in \{1, \ldots, l\}$;

3) $e_i \neq f_j$ для всех $i \in \{1, \ldots, m\}$ и $j \in \{1, \ldots, n\}$.

Тогда существует рациональное число x, такое, что

4) $a < x < b$;

5) $c_1 \sqsubseteq x, \ldots, c_k \sqsubseteq x$;

6) $d_1 \not\sqsubseteq x, \ldots, d_l \not\sqsubseteq x$;

7) $x \sqsubseteq e_1, \ldots, x \sqsubseteq e_m$;

8) $x \not\sqsubseteq f_1, \ldots, x \not\sqsubseteq f_n$;

9) $x \notin \{g_1, \ldots, g_s\}$.

Д о к а з а т е л ь с т в о. Все рациональные числа из условия леммы имеют некоторые номера в нумерации $\{q_i\}_i$. Пусть натуральное число S превышает все эти номера. На шаге S построения предиката $\sqsubseteq$ мы получим в том числе такое множество I_L^K, что $K = \{c_1, \ldots, c_k\}$ и $L = \{e_1, \ldots, e_m\}$. Это множество будет плотно в $\mathbb{Q}$. Следовательно, между a и b располагается бесконечно много элементов этого множества. Выберем такое $q_M \in I_L^K$, лежащее между a и b, чтобы $M > S$. Тогда q_M будет удовлетворять всем условиям леммы. ■

С л е д с т в и е 1. *Если $A \subseteq \mathbb{Q}$ — конечное множество и $B \subseteq A$, то существует такой x, что*

$$(\forall a \in A)(a \sqsubseteq x \leftrightarrow a \in B).$$

Т е о р е м а 2. *Теория $\mathrm{Th}(\mathbb{Q}, <, \sqsubseteq)$ не обладает трансляционным результатом.*

Доказательство. Покажем, например, что существует расширенный запрос, истинный в базе данных со схемой, состоящей из имени одного одноместного отношения P, тогда и только тогда, когда количество элементов в P чётно.

Действительно, этот запрос задаётся формулой

$$(\exists p_0, p_m \in P)((\forall p \in P)(p \geqslant p_0 \wedge p \leqslant p_m) \wedge (\exists x)(p_0 \sqsubseteq x \wedge \neg p_m \sqsubseteq x \wedge \\ (\forall p_1, p_2 \in P)((p_1 < p_2 \wedge \neg(\exists p \in P)(p_1 < p < p_2)) \to (\neg p_1 \sqsubseteq x \leftrightarrow p_2 \sqsubseteq x)))).$$

Но как известно, никакой ограниченный запрос чётность числа элементов P выразить не может. ■

Теорема 3. *Теория* $\mathrm{Th}(\mathbb{Q}, <, \sqsubseteq)$ *допускает эффективную элиминацию кванторов.*

Доказательство. Так как $\neg x \sqsubseteq x$ для всех x, то такие формулы можно исключить. В общем случае формула, в которой мы должны удалять квантор, имеет вид:

$$(\exists x)\Big(a < x < b \wedge \Big(\bigwedge_i c_i \sqsubseteq x\Big) \wedge \Big(\bigwedge_i \neg d_i \sqsubseteq x\Big) \wedge \\ \wedge \Big(\bigwedge_i x \sqsubseteq e_i\Big) \wedge \Big(\bigwedge_i \neg x \sqsubseteq f_i\Big) \wedge \Big(\bigwedge_i x \neq g_i\Big)\Big).$$

Если $c_i = d_j$ или $e_i = f_j$ для некоторых i, j, то формула, очевидно, ложна. Также она ложна, если $a \geqslant b$. Иначе по лемме 1 она истинна. Таким образом, эта формула эквивалентна такой:

$$a < b \wedge \Big(\bigwedge_{i \neq j} c_i \neq d_j\Big) \wedge \Big(\bigwedge_{i \neq j} e_i \neq f_j\Big).$$

■

Список литературы

[1] Baldwin J., Benedikt M. Stability theory, permutations of indiscernibles, and embedded finite models. // Trans. Amer. Math. Soc., V.352(11), 2000. P.4937–4969.

[2] Belegradek O.V., Stolboushkin A.P., Taitslin M.A. Extended order-generic queries. // Annals of Pure and Applied Logic, V.97(1–3), 1999. P.85–125.

[3] Benedikt M., Dong G., Libkin L., Wong L. Relational expressive power of constraint query languages. // Proc. 15th ACM Symp. on Principles of Database Systems, 1996. P.5–16.

[4] Codd E.F. A relational model for large shared data banks. // Communications of the ACM, V.13, 1970. P.377–387.

[5] Codd E.F. Relational completeness of data base sublanguages. // Database Systems (ed. Rustin R.), Prentice-Hall, 1972. P.33–64.