



## **МОДЕЛИ СЕМАНТИЧЕСКОЙ ЭКВИВАЛЕНТНОСТИ В БАЗАХ ДАНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ**

**Алисейко Е. В., Бондаревский С. Н., Половин Б. А.**

*Харьковский торгово-экономический институт*

*Киевского национального торгово-экономического университета*

*г. Харьков, пер. Отакара Яроша 8а*

*bondarevskij.s@gmail.com*

Уровень организации доступа к данным совместно со степенью автоматизации их обработки является фундаментом современной информационной индустрии. Данные, определенным образом организованные в структуры, являются носителем семантической (смысловой) информации и должны рассматриваться как жизненно важный ресурс общества.

Решение задачи создания и использования производственного потенциала связано с проблемой оптимального распределения ресурсов (в том числе природных и информационных) на основе применения математических методов и использования вычислительной техники. Успех ее решения определяется, помимо экономико – организационных мер, использованием современных информационных технологий для поддержки концепции взаимоувязанного и согласованного решения задач в больших предметных областях (ПО) с высоким уровнем организации обмена между семантически и лексически многозначными данными пользователей различных специальностей и уровней иерархии.

Решение задачи повышения эффективности использования информационного ресурса связано с разработкой механизма семантической сравнимости структур данных и их элементов базе данных, при этом использование реляционной модели базы данных (РБД), считается перспективной. Разработка моделей семантической эквивалентности данных РБД, критериев и методов сравнимости структур данных и их элементов является важным направлением в решении проблемы совместного доступа к данным. Решение этой проблемы позволит не только эффективно выполнять задачи интеграции данных и поддержания семантических ограничений целостности, но и служить инструментом нормализации лексики языка описания предметной области (ПО) для увязки задач по информации в рамках единого подхода.

Одной из проблем, возникающих при моделировании семантики ПО на уровне логической схемы, является отсутствие средств поддержания смыслового содержания функциональной зависимости, что является следствием многозначности лексического и семантического значения имени атрибута РБД [1].

В малых ПО эта задача решалась путем нормализации лексики языка описания ПО и, как следствие, имен атрибутов. Работа с современными информационными системами требует единой методики описания ПО и построения критериев семантического сравнения имен атрибутов при условии многозначности их лексического значения.

Предлагается методика описания ПО с использованием абстрактных типов данных [2]. При таком подходе ПО состоит из множества  $T$  информационных объектов  $TO_i$ :

$$T = \{TO_1, TO_1, \dots, TO_i, \dots, TO_r\}, i = 1, 2, \dots, r,$$

каждый из которых является параметризованным, или родовым типом данных вида:

$$TO_r(\{RO_1, RO_2, \dots, RO_S\}; \Omega_r),$$

где  $\Omega_r$  - множество операций на  $\{RO_i\}$ .

В свою очередь, каждый из параметров  $RO_i$  на  $k$ -ом объекте может быть представлен множеством своих значений  $\{A_1, A_2, \dots, A_n\}$ .

В этом случае родовой тип данных будет иметь вид:

$$TO_r(\{\langle RO_1\{A_1, \dots, A_{n1}\}\rangle, \langle RO_2\{A_1, \dots, A_{n2}\}\rangle, \dots, \langle RO_S\{A_1, \dots, A_n\}\rangle\}; \Omega_r)$$

Множества  $\{A_1, \dots, A_n\}$  могут, в частности представлять собой множество атрибутов РБД. Свяжем с каждым  $A_i \in A$  конечное множество возможных  $\{a_{ik}\}$  элементов простого типа, называемых значением атрибута с именем  $A_i$ . Пару  $\langle A_i\{a_{ik}\}\rangle$  будем называть атрибутом и обозначать в виде  $[A_i]$ . Пусть задано множество имен доменов  $D = \{D_i\}$  таких, что каждому имени домена  $D_k$  сопоставлено множество его значений  $\{d_{ik}\}$ . Пару  $\langle D_i, \{d_{ik}\}\rangle$  будем называть доменом и обозначать в виде  $[D_i]$ , причем тип домена однозначно определяется его именем, а каждому  $[A_i]$  соответствует один и только один домен, содержащий все допустимые значения  $\{a_{ik}\}$ . Атрибут  $[A_i]$  связан доменом  $[D_i]$  тогда и только тогда, когда  $\{a_i\} \in \{d_i\}$ . Этот факт будем изображать в виде  $[A_i] \rightarrow [D_i]$ . Будем говорить, что множество  $A$  согласовано по типам с множеством  $D$  тогда и только тогда, когда каждый из атрибутов с именем из  $A$  связан ровно с одним доменом из  $D$ . В этом случае множество имен атрибутов реляционной базы данных  $\{A_1, A_2, \dots, A_n\}$  может быть однозначно заменено на множество имен доменов  $\{D_1, D_2, \dots, D_n\}$  и модель родового типа данных будет иметь вид:

$$TO_r(\{\langle RO_1\{D_1, \dots, D_{n1}\}\rangle, \langle RO_2\{D_1, \dots, D_{n2}\}\rangle, \dots, \langle RO_S\{D_1, \dots, D_{nm}\}\rangle\}; \Omega_r)$$

Параметры объектов рассматриваются в виде пар  $\langle RO_i, \{D_i\}\rangle$ , называемых идентификатором  $-[RO_i]$ , где  $RO_i$  - имя идентификатора (смысл на объекте),  $[D_i]$  - множество значений идентификатора. Предлагаемая модель данных позволяет осуществлять семантическую сравнимость атрибутов РБД в случае их лексической и семантической многозначности в соответствии с критериями сравнимости: с точностью до связи с доменом (то есть множества значений



сравниваемых атрибутов являются подмножествами значений одного домена), с точностью до имени идентификатора (атрибуты имеют одинаковый смысл на объекте), с точностью до типа объекта.

Предлагаются следующие критерии сравнимости:

1. Атрибуты сравнимы с точностью до связи с доменом (то есть множество значений атрибутов является подмножеством одного домена)

$$[A_k] \square^D [A_l] \stackrel{def}{\Leftrightarrow} ([A_k] \rightarrow [D_k]) \& ([A_l] \rightarrow [D_k]) \& (A \Delta D),$$

что позволяет производить операцию естественного соединения по этим атрибутам и ведет к ошибке в случае, если множества значений атрибутов имеют различную семантическую интерпретацию.

2. Атрибуты сравнимы с точностью до имени идентификатора:

$$[A_l] \square^{RO} [A_l] \stackrel{def}{\Leftrightarrow} [A_k] \rightarrow [RO_l] \& [A_l] \rightarrow [RO_l],$$

$$[A_l] \rightarrow [RO_l] \stackrel{def}{\Leftrightarrow} (A_l \in \{A_l\}) \& [RO_l] = \langle RO_l \{A_l\} \rangle.$$

Такие атрибуты поддерживают одинаковый смысл, но, возможно, на различных объектах ПО, однако, они могут быть согласованы с различными типами объектов и являться подмножествами различных доменов.

3. Атрибуты сравнимы с точностью до идентификатора:

$$[A_k] \square^{RO_l} [A_l] \stackrel{def}{\Leftrightarrow} \left( [A_l] \square^{RO} [A_l] \right) \& (A \square D) \& \exists D_s \in D ([A_k] \rightarrow [D_s] \& [A_l] \rightarrow [D_s]).$$

В таком случае атрибуты  $A_l$  и  $A_k$  согласованы с одним доменом и, возможно, с различными типами объектов со смыслом  $RO_l$ .

Такие атрибуты имеют один определенный доменом тип данных и поддерживают одинаковый смысл, но возможно, на различных объектах.

4. Атрибуты сравнимы с точностью до типа объекта:

$$[A_k] \square^{TO} [A_l] \stackrel{def}{\Leftrightarrow} [A_k] \rightarrow [RO_k] \& [A_l] \rightarrow [RO_l] \& [RO_k \in TO_l] \& [RO_l \in TO_l].$$

Если атрибуты  $[A_i]$  и  $[A_j]$  удовлетворяют приведенным критериям, то они допускают интеграцию по данным и ограничения на операции реляционной алгебры отсутствуют, в частности, операция естественного соединения не приведет к семантическим аномалиям.

#### Список литературы

1. Мейер Д. Теория реляционных баз данных: Пер. с англ. – М.: Мир, 1987. – 608 с.
2. Замулин А. В. Типы данных в языках программирования и базах данных. – Новосибирск: Наука, 1987. – 148 с.