

УДК 004.652

Универсальная модель данных и ее отличительные особенности

В. И. Есин

Харьковский национальный университет им. В.Н. Каразина, Украина

Формулируется проблема создания стандартной модели данных. В качестве решения предлагается универсальная модель данных, которая позволяет представлять данные любой предметной области в общем (стандартном) и понятном, не только профессионалам в области разработки информационных систем, но и обычным пользователям, виде. Раскрывается структурная, манипуляционная и целостная части универсальной модели данных, как инструмента моделирования любой предметной области в рамках реляционной модели, с приведением ее отличительных особенностей.

Ключевые слова: модель данных, универсальная модель данных, отношение, объект, событие, класс объектов, класс событий.

Формулюється проблема створення стандартної моделі даних. Як рішення пропонується універсальна модель даних, яка дозволяє представляти дані будь-якої наочної області у загальному (стандартному) і зрозумілому, не лише професіоналам в області розробки інформаційних систем, але і звичайним користувачам, вигляді. Розкривається структурна, маніпуляційна і цілісна частини універсальної моделі даних, як інструменту моделювання будь-якої наочної області в рамках реляційної моделі, з приведенням її відмітних особливостей.

Ключові слова: модель даних, універсальна модель даних, відношення, об'єкт, подія, клас об'єктів, клас подій.

The problem of creation database model is formulated. Universal data model is proposed as a solution. Universal data model presents information of any subject domain in (standard) general and understandable, not only for professionals in the area of information systems development, but also the ordinary user form. Structural, manipulated and integrated parts of universal data model, as an instrument of design any subject domain within the relational model are described. The distinctive features of universal data model are given.

Keywords: data model, universal data model, relation, object, event, objects class, events class.

1. Проблема создания стандартной модели данных

Проблема создания стандартной модели данных интересовала мировое сообщество баз данных давно. Еще на международных симпозиумах в 1988 и в 1998 годах [1] поднимался этот вопрос. На них рассматривалась актуальность и перспективность этого направления. Вопрос о необходимости создания универсальной модели данных обсуждался в 2008 г. на 10-ой Всероссийской научной конференции [2] и на Шестой всероссийской открытой ежегодной конференции [3]. Однако и по сегодняшний день эта проблема остается открытой и актуальной.

2. Универсальная модель данных как решение проблемы стандартной модели

В качестве решения проблемы стандартной модели предлагается универсальная модель данных (УМД), которая позволяет представлять данные

любой предметной области в стандартном (универсальном, общем) и понятном, как профессионалам в области разработки информационных систем, так и обычным пользователям, виде и дает возможность осуществить простой переход к компьютерной реализации.

При разработке УМД учитывались основные требования, предъявляемые к моделям данных [4]:

- структурная достоверность (возможность представления реального мира);
- простота (легкость понимания модели как профессионалами в области разработки информационных систем, так и обычными пользователями);
- отсутствие избыточности (исключение излишней информации);
- способность к совместному использованию (отсутствие принадлежности к какому-то особому приложению или технологии и, следовательно, возможность использования модели во многих приложениях и технологиях);
- расширяемость (способность эволюционировать с целью включения новых требований с минимальным влиянием на уже существующих пользователей);
- целостность.

Рассмотрим принципы построения такой универсальной модели данных и ее отличительные особенности.

В работе [5] отмечалось, что в силу схожести математических отношений (являющихся теоретической основой универсальной модели данных) с реляционными, УМД целесообразно реализовывать в рамках реляционной модели. «За» использование реляционной модели, как основной логической модели построения базы данных (БД), говорят много и других фактов, например, следующих. Реляционная модель данных в подавляющем большинстве случаев вполне достаточна для моделирования любых данных [6]. Системы управления реляционными базами данных представляют собой одну из наиболее удачных технологий в компьютерной науке и индустрии [7]. Они имеют строгое математическое обоснование и формализацию. При этом реляционная модель не остается неизменной — с годами она постоянно развивалась и расширялась и эта тенденция сохраняется [8]. В этом отношении она напоминает математику (область математических знаний также не является застывшей, а развивается со временем). Более того, реляционная модель может рассматриваться как малая ветвь математики. Теория, построенная для реляционной модели, может быть перенесена на большинство других моделей [9].

Итак, рассмотрим универсальную модель данных подробнее.

В соответствии с современной трактовкой определения модели данных, в котором выделяется три компонента: структурная, манипуляционная и целостная, вначале рассмотрим структурную.

Структура модели данных

Структурная составляющая модели данных полностью согласуется с математическим определением понятия модели как множества с заданными на нем отношениями, основные формульные зависимости которых приведены в [5]. Структурная схема этих отношений представлена на рис. 1.



Рис. 1. Структурная схема основных отношений УМД

В случае реализации УМД в рамках реляционной модели данных эти математические отношения можно легко представить в виде реляционных отношений или соответствующих им таблицам и связей между ними.

Целостность данных

Поскольку универсальную модель данных решено реализовывать в рамках реляционной модели, то, естественно, ей будут присущи основные реляционные ограничения целостности. А именно, целостность сущностей и ссылочная целостность.

Целостность сущностей заключается в том, что в базовом отношении ни один атрибут первичного ключа не может содержать отсутствующих значений, обозначаемых как NULL. То есть в основных отношениях УМД множества соответствующих условных идентификаторов (идентификатор раздела, идентификатор класса объектов, идентификатор типа объектов, идентификатор экземпляра объектов, идентификатор класса событий, идентификатор экземпляра событий, идентификатор класса параметров объектов и т.д.), а в реляционной модели соответствующих этим идентификаторам атрибутов, не могут содержать значений NULL.

Ссылочная целостность заключается в том, что если в отношении существует внешний ключ, то значение внешнего ключа должно, либо соответствовать значению потенциального ключа некоторого кортежа в его базовом отношении, либо внешний ключ должен полностью состоять из значений NULL.

Механизм реализации ссылочной целостности в универсальной модели данных следующий. Имеется отношение экземпляров объектов – O , которое связано с отношениями типов объектов – T , классов объектов – C , разделов – R , предметных областей – M . И чтобы все эти отношения находились в непротиворечивом состоянии в УМД, значение атрибута $t_i \in T_i$ i -го экземпляра объектов отношения O соответствует одному из значений атрибута идентификатора типа объекта (T_i) отношения T ; значение атрибута $c_i \in C_i$ i -го экземпляра объектов отношения O соответствует одному из значений атрибута идентификатора класса объектов (C_i) отношения C ; значение атрибута $r_i \in R_i$ i -го экземпляра объектов отношения O соответствует одному из значений атрибута идентификатора раздела (R_i) отношения R и значение атрибута $m_i \in M_i$ i -го экземпляра объектов отношения O соответствует одному из значений атрибута идентификатора предметной области (M_i) отношения M .

Отношение экземпляров событий – B связано с отношениями классов событий – E , экземпляров объектов – O , предметных областей – M . И чтобы все эти отношения также находились в непротиворечивом состоянии, осуществляется поддержка ссылочной целостности. А именно: значение атрибута $e_i \in E_i$ i -го экземпляра события отношения B соответствует одному из значений атрибута идентификатора класса события (E_i) отношения E ; значение атрибута $o_i \in O_i$ i -го экземпляра события отношения B соответствует одному из значений атрибута идентификатора экземпляра объекта (O_i) отношения O и значение атрибута $m_i \in M_i$ i -го экземпляра события отношения B соответствует одному из значений атрибута идентификатора предметной области (M_i) отношения M .

И так далее для остальных отношений УМД.

Так как универсальная модель данных основывается на семантической модели «объект-событие» [10], то в ней также как и в модели «объект-событие» уникальная идентификация объектов (целостность сущностей) достигается иерархическим именем объекта в пределах рассматриваемой предметной области. При этом также обеспечивается и ссылочная целостность. Так как для поддержания такого иерархического имени требуется, чтобы для каждой ссылки в отношениях существовало значение атрибута, на которое она ссылается.

Кроме того, поскольку каждый атрибут (элемент кортежа) отношения УМД связан с некоторым множеством (считай доменом), то для множества допустимых значений каждого атрибута отношения можно говорить о так называемых ограничениях домена, которые можно сформулировать в виде:

- ограничений на допустимые перечисляемые (списочные) значения доменов характеристик объектов, событий, параметров объектов (например, множество типов значений характеристик объектов – H_3 и множество признаков характеристик экземпляров объектов – H_4 отношения H состоят из следующих элементов: $H_3 = \{boolean, char, number, date\}$, $H_4 = \{\gamma_1, \gamma_2\}$, где

$\gamma_i = \{ 'Y', 'N' \}$ – множество признаков принадлежности к списочным характеристикам в отношении; множество признаков характеристик параметров объектов – J_i отношения J состоит из следующих элементов: $J_i = \{ \gamma_i, \lambda_i \}$, где $\lambda_i = \{ 'H', 'D', 'M', 'Q', 'Y', null \}$ – множество признаков интервала измерения параметра объекта: 'H' – час, 'D' – день (сутки), 'M' – месяц, 'Q' – квартал, 'Y' – год.);

- ограниченный на единицы измерений, рассматриваемых характеристик объектов, событий, параметров объектов.

В УМД также поддерживается ограничения на события: с одним экземпляром объекта в один и то же момент (интервал) времени может происходить только одно событие одного класса. В этом случае отношение B можно определить в соответствии с выражением (1).

$$\begin{aligned}
 B = \{ & (b_1, b_2, b_3, b_4, b_5, e_1, e_3, o_1, m_1) \mid b_1 \in B_1 \ \& \ b_2 \in B_2 \ \& \ b_3 \in B_3 \ \& \ b_4 \in B_4 \ \& \ e_1 \in E_1 \ \& \\
 & b_5 \in (B_1 \cup \{null\}) \ \& \ e_3 \in (E_1 \cup \{null\}) \ \& \ o_1 \in O_1 \ \& \ m_1 \in M_1 \ \& \ ((\exists i = 1..n_1) \ \& \\
 & (\exists j = 1..n_2) \ \& \ (\exists l = 1..|B_3|) \ \& \ (\exists k = 1..|B_4|) \ \& \ (\exists \lambda = 1..(n_1+1)) \ \& \ (\exists v = 1..n_3) \ \& \\
 & (\exists \alpha = 1..n_4) \ \& \ (\exists \mu = (1..(n_3+1))) \ \& \ (\exists \beta = 1..n_5)) \ \exists b_1^i \in B_1 \ \& \ \exists b_2^j \in B_2 \ \& \ \exists b_3^l \in B_3 \ \& \\
 & \exists b_4^k \in B_4 \ \& \ \exists b_5^\lambda \in (B_1 \cup \{null\}) \ \& \ e_1^v \in E_1 \ \& \ e_3^\mu \in (E_1 \cup \{null\}) \ \& \ o_1^\alpha \in O_1 \ \& \\
 & m_1^\beta \in M_1 \ \& \ \exists (b_1^i, b_2^j, b_3^l, b_4^k, b_5^\lambda, e_1^v, e_3^\mu, o_1^\alpha, m_1^\beta) \in B \ \& \ ((\forall o_1 \in O_1) \ \& \ ((\forall \tau, \chi = 1..n_{18}) \ \& \\
 & (\tau \neq \chi) \ \& \ ((\forall \vartheta, \omega = 1..|B_3|) \ \& \ (\vartheta \neq \omega)) \ \& \ ((\forall \delta, \varepsilon = 1..|B_4|) \ \& \ (\delta \neq \varepsilon)) \\
 & \neg \exists ((b_1^i, b_2, b_3^g, b_4^\delta, b_5, e_1, e_3, o_1, m_1) = (b_1^\chi, b_2, b_3^\omega, b_4^\varepsilon, b_5, e_1, e_3, o_1, m_1)) \ \} \},
 \end{aligned} \tag{1}$$

где B_1 – множество условных идентификаторов экземпляров событий;

B_2 – множество условных имен экземпляров событий;

B_3 – множество времен начала событий;

B_4 – множество времен окончания событий;

n_1, n_2, n_3, n_4, n_5 – мощность множеств B_1, B_2, E_1, O_1 и M_1 соответственно;

O_1 – множество условных идентификаторов экземпляров объектов;

E_1 – множество условных идентификаторов классов событий;

M_1 – множество условных идентификаторов предметных областей.

Средства манипулирования данными

Средством манипулирования данными в универсальной модели данных является специальный непроцедурный (декларативный) язык модели данных (ЯМД) [11], который в случае использования реляционной модели данных позволяет определять и манипулировать данными подобно языку SQL-запросов, но в терминологии близкой к естественному языку, не привязанной к структуре данных и, не прибегая к контекстному поиску.

Отличительные особенности универсальной модели данных состоят в том, что в ней:

- любая предметная область описывается с помощью одних и тех же общих стандартных отношений (рис. 1);
- связи описываются, как необходимость отражения реальных процессов объективного мира, в виде иерархий реальных объектов и событий, которые с ними происходят в определенный момент или интервал времени, в

отличие от прописываемых связей, необходимых для определения значений соответствующих атрибутов при традиционном подходе;
- уникальная идентификация объектов в модели достигается иерархическим именем объекта в пределах рассматриваемой предметной области, которое содержит раздел с подразделами (если таковые имеются), класс объекта и экземпляр объекта. Если имеется иерархия классов объектов и экземпляров объектов, то и они указывается в иерархической последовательности.

3. Выводы

1. С точки зрения структурной составляющей определения модели данных, универсальная модель данных – это набор стандартных отношений, которые используются для описания данных и связей между ними.

2. УМД присущи следующие ограничения целостности: целостность сущностей и ссылочная целостность, которые свойственны как реляционным моделям, так и семантической модели «объект-событие». Для последней уникальная идентификация объектов (целостность сущностей) достигается иерархическим именем объекта в пределах рассматриваемой предметной области, а ссылочная целостность, как необходимость поддержания иерархического имени (для каждой ссылки в отношениях существует значение атрибута, на которое она ссылается).

3. Средством манипулирования данными в универсальной модели данных является специально разработанный непроцедурный язык модели данных.

4. В целом универсальная модель данных - это инструмент моделирования любой предметной области, который легко реализуется в рамках реляционной модели данных и может быть использован, в том числе, и для построения модели базы данных.

ЛИТЕРАТУРА

1. Кузнецов С. Д. Будущие направления исследований в области баз данных: десять лет спустя, 1999 [Электронный ресурс] – режим доступа : http://www.citforum.ru/database/articles/future_01.shtml
2. Вязилов Е.Д. Унификация структур данных в области изучения, освоения и использования ресурсов Мирового океана. 10-ая Всероссийская научная конференция ["Электронные библиотеки: перспективные методы и технологии, электронные коллекции"], (Дубна, Россия, 2008) [Электронный ресурс] / Е.Д. Вязилов, А.А. Федорцов, А.Е. Кобелев – режим доступа : http://rcdl2008.jinr.ru/presentation/conf_hall/08_10_2008/Vyazilov.ppt
3. Вязилов Е. Д. Универсальная модель хранения данных с учетом жизненного цикла объектов. Шестая всероссийская открытая ежегодная конференция ["Современные проблемы дистанционного зондирования Земли из космоса"], (Москва, ИКИ РАН, 10-14 ноября 2008 г. Сборник тезисов конференции) [Электронный ресурс] / Е.Д. Вязилов, А.А. Федорцов – режим доступа : <http://d902.iki.rssi.ru/theses-cgi/thesis.pl?id=1197>
4. Коннолли Т. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. 3-е издание. : Пер. с англ. / Т. Коннолли, К. Бегг. – М. : Издательский дом "Вильямс", 2003. – 1440 с. : ил.

5. Есин В. И. Универсальная модель данных и ее математические основы / В. И. Есин // Системы обработки информации. – Х. : Харьковский университет Повітряних Сил, 2011. – № – С. ...
6. Кузнецов С. Д. Концептуальное проектирование реляционных баз данных с использованием языка UML [Электронный ресурс] – режим доступа : <http://www.citforum.ru/database/articles/umlbases.shtml>
7. Donald D. Chamberlin. Anatomy of an Object-Relational Database. DB2 Online Magazine, Winter 1996.
8. Райордан Р. Основы реляционных баз данных / Пер, с англ. – М.: Издательско-торговый дом «Русская Редакция», 2001. – 384 с.: ил.
9. Тальхайм Б. Обзор семантических ограничений для моделей баз данных [Электронный ресурс] – режим доступа : [http://www.intsys.msu.ru/magazine/archive/v3\(3.../thalheim-307-351.pdf](http://www.intsys.msu.ru/magazine/archive/v3(3.../thalheim-307-351.pdf)
10. Есин В. И. Семантическая модель данных "объект-событие" / В. И. Есин // Вісник Харківського національного університету. – Х. : Харьковский национальный университет им. В.Н. Каразина, 2010. – № 925. – С. 65-73 – (Серія : Математичне моделювання. Інформаційні технології. Автоматизовані системи управління" ; вип. 14).
11. Есин В. И. Язык описания и манипулирования данными, хранящимися в БД с УМД / В. И. Есин, М. В. Есина // Компьютерное моделирование в наукоемких технологиях (КМНТ-2010) : междунар. науч.-техн. конф., 18-21 мая 2010 г. : тезисы докл. – Х. : Харьковский национальный университет им. В.Н. Каразина, 2010. – Часть 2. — С. 104-108.