

УДК 004.652

В.И. Есин

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

УНИВЕРСАЛЬНАЯ МОДЕЛЬ ДАННЫХ И ЕЕ МАТЕМАТИЧЕСКИЕ ОСНОВЫ

Описывается проблема создания модели базы данных и существующий подход к ее решению, а также его недостатки. Предлагается решение проблемы с помощью универсальной (стандартной) модели данных, которая основывается на семантической модели данных «объект-событие», теории множеств и логическом исчислении. Приводятся математические выражения, послужившие теоретической основой создания универсальной модели данных, являющейся инструментом моделирования любой предметной области, который легко реализуется в рамках реляционной модели.

Ключевые слова: модель данных, универсальная модель данных, множество, отношение, n -арное отношение, объект, событие, класс объектов, класс событий.

1. Постановка задачи

Понятие «модель данных» является одним из фундаментальных в области баз данных, хотя даже сегодня в известной литературе нередко дается его неоднозначная трактовка. В современном понимании *модель данных* — это не результат, а инструмент моделирования, то есть совокупность правил структурирования данных, допустимых операций над ними и видов ограничений целостности, которым они должны удовлетворять [1].

В соответствии с трехуровневой архитектурой ANSI/X3/SPARC модели данных рассматриваются по отношению к каждому ее уровню: внешнему, концептуальному и внутреннему. При этом функции моделей существенно различаются. Кроме того, различаются они и объектами моделирования, с которыми связаны. С этой точки зрения, различаются модели предметной области (ПрО) информационной системы и модели базы данных (БД). Хотя в процессе проектирования систем баз данных, как правило, используются обе эти модели.

В связи с этим становится интересным рассмотрение вопросов, посвященных разработке модели базы данных с учетом модели предметной области, когда сама структура модели базы данных (ее схема) не зависит от рассматриваемой предметной области.

2. Существующий подход к решению проблемы создания модели базы данных и его недостатки

Многие организаций или отдельные проектировщики при разработке модели базы данных в незначительной степени полагаются на какие-либо методологии. В большей степени их действия определяются опытом их работы, осведомленностью, имеющимися «под рукой» средствами и некоторыми

другими субъективными факторами. И именно поэтому такие действия часто считаются основной причиной неудач при разработке моделей баз данных и информационных систем в целом. А если, к тому же, корпорация, организация, предприятие и т.п. занимается несколькими значительными и объемными видами деятельности, такими как кадровая, финансовая, технологическая, маркетинговая и т.д. (а это достаточно распространенная сегодня ситуация), то в этом случае необходимо иметь несколько моделей БД. Как быть в этом случае? Каждый раз разрабатывать и сопровождать новую модель? На это потратится значительное время и средства, и тем большее, чем больше и сложнее эти виды деятельности.

3. Предлагаемый подход к решению проблемы

Проанализировав недостатки, присущие вышеизложенному подходу, было принято решение о поиске возможной универсальной (стандартной) для любой предметной области модели данных. *Цель ее построения* – представление данных, вне зависимости от рассматриваемой предметной области, в общем (стандартном) и понятном как профессионалам в области разработки информационных систем, так и обычным пользователям виде.

Тема создания стандартной модели данных обсуждалась на международных симпозиумах еще в 1988 и в 1998 годах [2]. На них рассматривалась ее актуальность и перспективность. Вопрос о необходимости создания универсальной модели данных обсуждался в 2008 г. на 10-й Всероссийской научной конференции [3] и на Шестой всероссийской открытой ежегодной конференции "Современные проблемы дистанционного зондирования Земли из космоса" [4]. Однако и по сегодняшний день эта проблема остается открытой и актуальной.

Работы в этом направлении в Харькове для газовой промышленности начали активно проводиться примерно 10 лет назад [5]. Затем предметные области исследований стали расширяться [6, 7, 8], и позволили приблизиться к разработке универсальной модели данных (УМД), позволяющей в определенной мере решить проблему стандартной модели. Созданию УМД способствовал также подход, учитывающий не только объекты, но и события с ними происходящие, впоследствии приведший к созданию семантической модели данных «объект-событие» [9].

4. Математические основы универсальной модели данных

Итак, универсальная (стандартная) модель данных основывается на семантической модели данных «объект-событие», теории множеств и логическом исчислении. В модели «объект-событие» все объекты, процессы, явления любой предметной области описываются с помощью мета-онтологий (мета-онтологии выступают как система понятий модели, ориентированная на описание элементов любой предметной области). А, как известно, основными компонентами онтологии являются отношения, которые представляют тип взаимодействия между понятиями предметной области. Формально n -арные отношения определяются как подмножество прямого (декартова) произведения n множеств. Сформируем такие отношения.

Так для описания классов объектов (как одной из основных мета-онтологий семантической модели данных «объект-событие») воспользуемся следующим набором множеств:

- C_1 – множество условных идентификаторов классов объектов;
- C_2 – множество условных имен классов объектов

для определения отношения C .

Множества C_1 и C_2 равноможны: $|C_1|=|C_2|=n_1$ (для конечных множеств мощность множества – это число элементов).

Известно, что классы объектов могут находиться в иерархической зависимости. Кроме того, классы объектов могут быть связаны с различными рассматриваемыми предметными областями.

Тогда отношение C , содержащее данные о классах объектов и их иерархии для всех рассматриваемых предметных областей, можно определить как подмножество декартова произведения множеств: C_1, C_2 , объединенного множества условных идентификаторов классов объектов C_1 и множества, содержащего единственный элемент – $null$ ($C_1 \cup \{null\}$) (характеризует иерархию событий) и M_1 (множество условных идентификаторов предметных областей, мощность множества $M_1 - |M_1|=n_2$) ($C \subset C_1 \times C_2 \times (C_1 \cup \{null\}) \times M_1$). А именно:

$$C = \{(c_1, c_2, c_3, m_1) \mid c_1 \in C_1 \ \& \ c_2 \in C_2 \ \& \ c_3 \in (C_1 \cup \{null\}) \ \& \ m_1 \in M_1 \ \& \ ((\exists i, j = 1..n_1) \ \& \ (\exists l = 1..(n_1+1)) \ \& \ (\exists k = 1..n_2) \ \exists c_1^i \in C_1 \ \& \ \exists c_2^j \in C_2 \ \& \ \exists c_3^l \in (C_1 \cup \{null\}) \ \& \ \exists m_1^k \in M_1 \ \& \ (\exists (c_1^i, c_2^j, c_3^l, m_1^k) \in C) \ \& \ ((\forall i, j = 1..n_1 \ \& \ i \neq j) \ \exists c_1^i \in C_1 \ \& \ \exists c_2^j \in C_2 \ \& \ \neg \exists (c_1^i, c_2^j) \in C_{1-2})\},$$

где $C_{1-2} \subset C_1 \times C_2$.

Отношение C является отношением порядка 4. Мощность множества C – это суммарное количество четверок элементов, характеризующих классы объектов всех рассматриваемых предметных областей.

Для экземпляров объектов

Пусть для конкретной предметной области имеется набор множеств, описывающих экземпляры объектов:

- O_1 – множество условных идентификаторов экземпляров объектов;
- O_2 – множество условных имен экземпляров объектов.

Число рассматриваемых условных идентификаторов и условных имен экземпляров объектов в общем случае неодинаково. Количество элементов множества O_1 равно n_3 ($|O_1|=n_3$), а множества O_2 равно n_4 ($|O_2|=n_4$), причем $n_3 \geq n_4$.

Тогда, учитывая тот факт, что экземпляры объектов принадлежат определенным разделам, типам и классам объектов рассматриваемой предметной области, и что экземпляры объектов могут находиться в определенной иерархической зависимости [9], по аналогии с вышеизложенным определим отношение O , как подмножество декартова произведения множеств O_1, O_2, R_1 (множество условных идентификаторов разделов предметной области, $|R_1|=n_5$), C_1, T_1 (множество условных идентификаторов типов объектов, $|T_1|=n_6$) и M_1 ($O \subset O_1 \times O_2 \times (O_1 \cup \{null\}) \times R_1 \times C_1 \times T_1 \times M_1$). Отношение O содержит данные обо всех экземплярах объектов и их иерархиях для всех предметных областей. Выражение (2) – есть формализованное представление отношения O .

Отношение O является отношением порядка 7. Мощность O – это суммарное количество семерок элементов, характеризующих экземпляры объектов для всех рассматриваемых ПрО.

$$O = \{(o_1, o_2, o_3, r_1, c_1, t_1, m_1) \mid o_1 \in O_1 \ \& \ o_2 \in O_2 \ \& \ o_3 \in (O_1 \cup \{null\}) \ \& \ r_1 \in R_1 \ \& \ c_1 \in C_1 \ \& \ t_1 \in T_1 \ \& \ m_1 \in M_1 \ \& \ ((\exists i = 1..n_3) \ \& \ (\exists j = 1..n_4) \ \& \ (\exists l = 1..(n_3+1)) \ \& \ (\exists \lambda = 1..n_5) \ \& \ (\exists v = 1..n_1) \ \& \ (\exists \alpha = 1..n_6) \ \& \ (\exists k = 1..n_2)) \ \& \ \exists o_1^i \in O_1 \ \& \ \exists o_2^j \in O_2 \ \& \ \exists o_3^l \in (O_1 \cup \{null\}) \ \& \ \exists r_1^\lambda \in R_1 \ \& \ \exists c_1^\alpha \in C_1 \ \& \ \exists t_1^\alpha \in T_1 \ \& \ \exists m_1^k \in M_1 \ \& \ \exists (o_1^i, o_2^j, o_3^l, r_1^\lambda, c_1^\alpha, t_1^\alpha, m_1^k) \in O\}.$$

Для классов событий

Пусть для конкретной ПрО имеется набор множеств, описывающих классы событий:

- E_1 – множество условных идентификаторов классов событий;
- E_2 – множество условных имен классов событий.

Число рассматриваемых условных идентификаторов и условных имен классов событий в общем случае неодинаково.

Для множества E_1 мощность $|E_1| = n_7$, а для множества E_2 – $|E_2| = n_8$, причем $n_7 \geq n_8$.

Известно, что классы событий могут находиться в иерархической зависимости. А так как классы событий к тому же связаны с определенным классом объектов конкретной предметной области (события происходят с объектами), и таких ПрО может быть несколько, то отношение E , содержащее данные о классах событий и их иерархии для всех рассматриваемых предметных областей, есть подмножество декартова произведения множеств E_1, E_2, C_1, M_1 и объединенного множества условных идентификаторов классов событий E_1 и множества, содержащего единственный элемент – null ($E_1 \cup \{null\}$) ($E \subseteq E_1 \times E_2 \times (E_1 \cup \{null\}) \times C_1 \times M_1$).

А именно:

$$E = \{(e_1, e_2, e_3, c_1, m_1) \mid e_1 \in E_1 \ \& \ e_2 \in E_2 \ \& \ e_3 \in (E_1 \cup \{null\}) \ \& \ c_1 \in C_1 \ \& \ m_1 \in M_1 \ \& \ ((\exists i = 1..n_7) \ \& \ (\exists j = 1..n_8) \ \& \ (\exists l = 1..(n_7+1)) \ \& \ (\exists \lambda = 1..n_1) \ \& \ (\exists k = 1..n_2) \ \& \ (\exists e_1^i \in E_1 \ \& \ \exists e_2^j \in E_2 \ \& \ \exists e_3^l \in (E_1 \cup \{null\}) \ \& \ \exists c_1^\lambda \in C_1 \ \& \ \exists m_1^k \in M_1 \ \& \ \exists (e_1^i, e_2^j, e_3^l, c_1^\lambda, m_1^k) \in E)\} \quad (3)$$

Мощность множества E – это суммарное количество пятюрок элементов, характеризующих классы событий для всех рассматриваемых ПрО.

Для экземпляров событий

Пусть для конкретной ПрО имеется набор множеств, описывающих экземпляры событий:

- V_1 – множество условных идентификаторов экземпляров событий;
- V_2 – множество условных имен экземпляров событий;
- V_3 – множество времен начала событий;
- V_4 – множество времен окончания событий.

Мощность множества V_1 равна n_9 ($|V_1| = n_9$). Множество $V_2 \subseteq E_2$. Так как не для всех экземпляров событий обязательно время его окончания, то $|V_4| \leq |V_3|$.

Известно, что экземпляры событий могут находиться в иерархической зависимости. Кроме того, экземпляры событий связаны с определенным клас-

сом событий (находящимся в определенной иерархической зависимости) и экземпляром объекта конкретной ПрО.

Тогда, воспользовавшись известным приемом, определим отношение B , содержащее данные об экземплярах событий и их иерархии для всех ПрО, как подмножество декартова произведения $B \subseteq V_1 \times V_2 \times V_3 \times V_4 \times (V_1 \cup \{null\}) \times E_1 \times (E_1 \cup \{null\}) \times O_1 \times M_1$.

$$B = \{(b_1, b_2, b_3, b_4, b_5, e_1, e_3, o_1, m_1) \mid b_1 \in V_1 \ \& \ b_2 \in V_2 \ \& \ b_3 \in V_3 \ \& \ b_4 \in V_4 \ \& \ e_1 \in E_1 \ \& \ b_5 \in (V_1 \cup \{null\}) \ \& \ e_3 \in (E_1 \cup \{null\}) \ \& \ o_1 \in O_1 \ \& \ m_1 \in M_1 \ \& \ ((\exists i = 1..n_9) \ \& \ (\exists j = 1..n_8) \ \& \ (\exists l = 1..|V_3|) \ \& \ (\exists k = 1..|V_4|) \ \& \ (\exists \lambda = 1..(n_9+1)) \ \& \ (\exists v = 1..n_7) \ \& \ (\exists \mu = (1..(n_7+1))) \ \& \ (\exists \alpha = 1..n_3) \ \& \ (\exists \beta = 1..n_2) \ \& \ \exists b_1^i \in V_1 \ \& \ \exists b_2^j \in V_2 \ \& \ \exists b_3^\lambda \in V_3 \ \& \ \exists b_4^k \in V_4 \ \& \ \exists b_5^\mu \in (V_1 \cup \{null\}) \ \& \ e_1^v \in E_1 \ \& \ e_3^\mu \in (E_1 \cup \{null\}) \ \& \ o_1^\alpha \in O_1 \ \& \ m_1^\beta \in M_1 \ \& \ \exists (b_1^i, b_2^j, b_3^\lambda, b_4^k, b_5^\mu, e_1^v, e_3^\mu, o_1^\alpha, m_1^\beta) \in B)\}.$$

Мощность множества B – это суммарное количество девяток элементов, характеризующих экземпляры событий, которые произошли с объектами для всех рассматриваемых ПрО.

Аналогичным образом, с помощью математических отношений, описываются все остальные мета-онтологии семантической модели данных «объект-событие».

Такие отношения и составляют основу УМД. С помощью них можно описать любую предметную область.

И теперь, когда встает проблема построения модели базы данных, последняя может быть достаточно просто решена.

А именно: даже поверхностный анализ математических выражений (1)...(4) позволяет сделать вывод о том, что УМД можно легко реализовать в рамках реляционной модели.

Так как представленные выше математические отношения естественно преобразуются в реляционные, ввиду того, что большая часть свойств реляционных отношений происходит от свойств математических отношений.

В реляционной модели отношение можно представить как произвольное подмножество декартова произведения доменов атрибутов. Поэтому если рассматривать множества в декартовых произведениях математических отношений (которые составляют основу УМД) как домены атрибутов реляционного отношения, то мы придем именно к описанию реляционных отношений известной и хорошо себя зарекомендовавшей реляционной модели данных, в рамках которой достаточно просто реализуется модель базы данных любой ПрО.

Выводы

1. Приведенные математические отношения послужили теоретической основой создания универсальной модели данных.

2. Универсальная модель данных, с одной стороны – это набор стандартных математических отношений, которые используются для описания данных, связей между ними и ограничений, накладываемых на данные любой предметной области. С другой стороны, согласно определению модели данных и выбранного объекта моделирования – предметной области информационной системы, 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.

Список литературы

1. Коголовский М.Р. Перспективные технологии информационных систем / М.Р. Коголовский. – М.: ДМК Пресс; М.: Компания АйТи, 2003. – 288 с. (Серия «ИТ-Экономика»).

2. Кузнецов С.Д. Будущие направления исследований в области баз данных: десять лет спустя [Электронный ресурс] / С.Д. Кузнецов. – Режим доступа к ресурсу: http://www.citforum.ru/database/articles/future_01.shtml.

3. Вязилов Е.Д. Унификация структур данных в области изучения, освоения и использования ресурсов Мирового океана [Электронный ресурс] / Е.Д. Вязилов, А.А. Федорцов, А.Е. Кобелев // 10-я Всероссийская научная конференция "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", Дубна, Россия, 2008г. – Режим доступа к ресурсу: http://rcdl2008.jinr.ru/presentation/conf_hall/08_10_2008/Vyazilov.ppt.

4. Вязилов Е.Д. Универсальная модель хранения данных с учетом жизненного цикла объектов [Электронный ресурс] / Е.Д. Вязилов, А.А. Федорцов // Шестая всероссийская открытая ежегодная конференция

"Современные проблемы дистанционного зондирования Земли из космоса", Москва, ИКИ РАН, 10-14 ноября 2008 г.: Сборник тезисов конференции. – режим доступа к ресурсу: <http://d902.iki.rssi.ru/theses/cgi/thesis.pl?id=1197>.

5. Пергаменцев Ю.А. Информационные, функциональные модели и техническое задание на сегмент ОБД "Основные фонды" / Ю.А. Пергаменцев // Материалы Научно-технического совета ОАО "Газпром" по проекту "Отраслевой банк данных (ОБД)". – Санкт-Петербург, ноябрь 2000 г., Москва 2001. – С. 66.

6. Есин В.И. Разработка информационных систем с использованием универсальной модели данных / В.И. Есин // I Наукова конференція молодих вчених ХВУ: наук. конф., 13-14 листопада 2002 р.: тези докл. – Х., 2002. – С. 41.

7. Можливі шляхи створення єдиної бази даних діяльності ХВУ на основі універсальної моделі даних / [Єсін В.І. Варакута В.П., Смірнов Є.Б., Шматков С.І.]. – Х.: ХВУ. Науково-методичний збірник ХВУ, 2003. – № 3(89). – С. 3-12.

8. Есин В.И. Технология проектирования модели предприятия на основе универсальной модели данных, 2008. [Электронный ресурс] / В.И. Есин, Ю.А. Пергаменцев. – Режим доступа к ресурсу: <http://www.citforum.ru/database/articles/udm/>

9. Есин В.И. Семантическая модель данных "объект-событие" / В.И. Есин // Вісник Харківського національного університету. – Х.: Харківський національний університет ім. В.Н. Каразіна, 2010. – № 925. – С. 65-73 – (Серія : Математичне моделювання. Інформаційні технології. Автоматизовані системи управління" ; вып. 14).

Поступила в редколлегию 20.01.2011

Рецензент: д-р техн. наук, проф. А.А. Кузнецов, Харьковский университет Воздушных Сил им. И. Кожедуба, Харьков.

УНІВЕРСАЛЬНА МОДЕЛЬ ДАНИХ І ЇЇ МАТЕМАТИЧНІ ОСНОВИ

В.І. Єсін

Описується проблема створення моделі бази даних і існуючий підхід до її рішення, а також його недоліки. Пропонується рішення проблеми за допомогою універсальної (стандартної) моделі даних, яка ґрунтується на семантичній моделі даних «об'єкт-подія», теорії множин і логічному численні. Наводяться математичні вирази, що послужили теоретичною основою створення універсальної моделі даних, яка є інструментом моделювання будь-якої наочної області, і яка легко реалізується в рамках реляційної моделі.

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

UNIVERSAL DATA MODEL AND ITS MATHEMATICAL BASES

V.I. Yesin

The problem of creation database model is described. Existing approaches for creation database model are described. The shortcomings of existing approaches are described. An approach to solving problems by using a universal (standard) data model is given. Universal data model is based on the semantic data model "object-event", the theory of sets and logical calculus. The mathematical expressions are given. The mathematical expressions are served as the theoretical basis of creation of universal data model that is tool for modeling any subject domain, which easily will be realized within the relational model.

Keywords: data model, universal data model, set, relation, n-ary relation, object, event, objects class, events class.