

УДК 004.652: 004.655

Семантическая модель данных "объект-событие"

В. И. Есин

Харьковский национальный университет имени В.Н. Каразина, Украина

Описываются основные положения разработанной семантической модели данных "объект-событие", позволяющей существенно расширить (в сравнении с существующими семантическими моделями данных) необходимые для проектирования БД возможности по адекватному отображению реального мира, как динамической системы.

Ключевые слова: *семантическая модель данных, объект, событие, класс объектов, класс событий, характеристика объекта, характеристика события.*

Описуються основні положення розробленої семантичної моделі даних "об'єкт-подія", що дозволяє істотно поширити (порівняно з існуючими семантичними моделями даних) необхідні для проектування БД можливості по адекватному відображенню реального світу, як динамічної системи.

Ключові слова: *семантична модель даних, об'єкт, подія, клас об'єктів, клас подій, характеристика об'єкту, характеристика події.*

Bases of semantic data model are described "object-event", which allows substantially extending (by comparison to the existent semantic data models) possibilities on the adequate reflection of the real world, as a dynamic system, necessary for DB designing.

Key words: *semantic data model, object, event, objects class, events class, property of object, property of event.*

1. Общая постановка проблемы построения БД и ее актуальность

Одна из наиболее сложных проблем проектирования баз данных (БД) состоит в том, что специалисты предметной области, разработчики баз данных и программного обеспечения, а также конечные пользователи, как правило, рассматривают данные и их назначение по-разному. Поэтому, чтобы добиться полного их взаимопонимания в процессе обмена информацией, необходимо использовать понятную общую модель данных, которая не усложнена техническими подробностями и двояким толкованием. Потребность в более удобных и мощных средствах моделирования предметной области вызвала к жизни направление семантических моделей данных.

Семантическое моделирование стало предметом интенсивных исследований с конца 70-х годов прошлого столетия. Основная причина активизации таких исследований крылась в том, что системы баз данных тогда (да и сейчас это не редкость) обладали весьма ограниченными сведениями о смысле хранящихся в них данных. Чаще всего они знали, как обращаться с данными определенных простых типов, и понимали некоторые простейшие ограничения целостности, наложенные на эти данные. Но любая более сложная интерпретация возлагалась на пользователя.

Хорошо было бы, если, например, язык *SQL* (или ему подобный) мог понимать, что вес изделия и поставляемое их количество являются, хотя и числовыми, но все же семантически разными величинами. Это позволило бы

системе подвергнуть сомнению или даже вовсе отвергнуть запрос, в котором, например, сравнивался вес изделия с их количеством.

Общий подход к проблеме семантического моделирования характеризуется четырьмя основными этапами [1]:

- 1) Выявление некоторого множества семантических понятий, которые используются для описания "реального мира".
- 2) Определение набора соответствующих символических (формальных) объектов, которые могут использоваться для представления описанных выше семантических понятий.
- 3) Определение набора формальных общих правил целостности, предназначенных для работы с такими формальными объектами.
- 4) Определение набора формальных операторов, предназначенных для манипулирования этими формальными объектами.

2. Существующие подходы к решению проблемы и их недостатки

Сегодня существует достаточное количество семантических моделей. Это и хорошо известная модель "сущность-связь" в различных ее нотациях (нотация Чена, нотация Мартина, нотация IDEF1X, нотация Баркера, нотация языка моделирования UML), и расширенная модель "сущность-связь" (*Enhanced Entity-Relationship* — *EER*), и модель "объект-роль", и объектно-ориентированная модель данных, и менее известная модель "объект-качество", а также некоторые другие модели.

Теоретически упомянутые и неупомянутые выше модели данных, равносильны в том смысле, что все, выразимое в одной из них, выразимо в остальных. Их отличает только то, насколько удобно использовать ту или иную модель разработчику при работе с реальными задачами, и то, насколько эффективно можно, используя их, реализовать в будущем логическую и физическую модели на компьютере (если это вообще возможно).

Известно, что информация о реальном мире (конкретной предметной области) дается через восприятие. При этом само восприятие достаточно сложно и состоит из множества взаимосвязанных фактов. Эти факты можно описать на естественном языке с помощью объектов, событий, происходящих с этими объектами, свойств (характеристик) объектов и событий, времени наступления события.

Формально факт можно определить как кортеж (O, E, C, T) , где O – набор объектов, E – набор событий, C – свойства объектов (или событий), T – время наступления события.

То есть можно говорить о том, что с *объектом* произошло *когда-то* (*тогда-то*) и *где-то* *какое-то* событие. С точки зрения причинно-следственного подхода, как общефилософского метода познания, эта формулировка является корректной.

Таким образом, в конкретный момент времени предметная область – это выделенная совокупность объектов, событий и их характеристик (свойств), называемая состоянием ПрО. Предметную область целесообразно рассматривать как некую динамическую систему, которая состоит из определенной последовательности состояний. Конечность множества состояний такой

системы в каждый момент времени обуславливается необходимостью и возможностями конкретной информационной системы.

В перечисленных выше моделях данных не предусмотрен явный учет времени. Причина отсутствия в них поддержки концепции времени, как правило, объясняется "трудностью практической реализации". Поэтому последовательность возникновения событий в большинстве моделей представляется с помощью упорядочения. Да и сам термин "*событие*" в них имеет совершенно иной смысл, чем рассматриваемый ниже. Так, например, в модели данных "сущность-связь" понятие событие упоминается в качестве примера сущности.

Ни одна из указанных выше моделей не оперирует понятием "*событие*" как фактом или действием, которое происходит с некоторым элементом предметной области в определенный момент или интервал времени, хотя все процессы реального мира протекают именно в пространстве и времени.

Существование объекта, как правило, связано с такими событиями, как возникновение, исчезновение и изменение. Объект возникает, когда субъект (человек, система) начинает проявлять к нему интерес, и исчезает, когда этот интерес утрачивается. При этом в модели данных объект может существовать независимо от того, определены или нет его свойства и связи с другими объектами. Единственное свойство, с которым следует соотносить существование объекта, - это время его возникновения и исчезновения [2]. Из чего можно заключить, что время является важным аспектом любого факта. Поэтому включение в модель данных понятия "*событие*", напрямую связанного со временем (именно в том контексте, который был определен выше), существенно расширяет возможности модели по адекватному отображению реального мира, и обеспечивает сохранение сведений о свойствах и связях, которые либо являются актуальными и достоверными на текущий момент (состояние ПрО), либо их утратили.

Поэтому на основе анализа известных семантических моделей данных (выявления и обобщения их лучших сторон), а также использования подхода, при котором аппарат моделирования, являясь независимым от конкретной предметной области, опирается на мета-онтологии [4], была получена модель "объект-событие", лишенная указанного выше недостатка.

3. Предлагаемый подход к решению проблемы

В соответствии с современной трактовкой понятия модели данных, в которой явно вычерчиваются три компоненты: структура данных, операции над данными, ограничения целостности данных, рассмотрим модель "объект-событие".

Известно, что одним из основных способов структуризации данных является использование абстракций. *Абстракция* предполагает, что несущественные аспекты сущности должны быть опущены, а внимание должно быть сконцентрировано на ее основных общих свойствах. Абстракция широко применяется как средство улучшения понимания сложных предметов [2].

В моделировании данных абстракция используется для образования категорий данных.

И если в модели "сущность-связь" ключевыми являются категории сущность и связь, в модели "объект-роль" – объект и роль, в объектной модели – объект, в модели "объект-качество" – объект и качество, то в модели "объект-событие" – такими ключевыми мета-онтологиями являются – объект и событие.

Существует два типа абстракции: обобщение и агрегация.

Обобщение – это абстракция, которая позволяет обобщенно представлять себе некоторый класс индивидуальных объектов как единый именованный объект. При этом обобщение, является, вероятно, наиболее важным механизмом для концептуализации реального мира.

Механизм обобщения в модели "объект-событие" поддерживается с помощью введенных мета-онтологий: *класс объектов, класс событий, тип объекта*.

Тип объекта – это некоторая совокупность схожих по нескольким значительным качественным признакам экземпляров объектов (или просто объектов). Тип объекта поддерживает обобщение на уровне экземпляров объектов (в *семiotике* это обобщение называют обобщением на уровне знаков).

Класс объектов – это совокупность типов объектов, объединяющих соответствующие однородные экземпляры объектов. Идентифицируется общим названием.

Класс объектов может включать один или несколько типов объектов.

Например, класс объектов "Автомобиль" может включать следующие типы объектов: "Вольво", "Мерседес", "ГАЗ-3110", "ВАЗ-21310", "КАМАЗ-43101" и т.д., при этом каждый из типов имеет собственные экземпляры объектов.

Класс событий – это совокупность *событий* (экземпляров событий), выделенных по какому-либо принципу, которые могут происходить с некоторыми объектами в определенный момент или интервал времени, и идентифицируемая общим названием.

Примеры классов событий: "Прием на работу", "Проведение технического обслуживания", "Встреча", "Изменение технического состояния", "Установка двигателя", "Поставка газа" и т.д. Так, если субъекты Иванов, Петренко, Сидорова получили водительские права, то чтобы этот факт отобразить в модели необходимо использовать, например, класс события "Получение водительских прав".

Экземпляр события – это факт или действие, которое происходит с некоторым объектом в определенный момент или интервал времени. Экземпляр события идентифицируется временем и объектом, принадлежит некоторому классу событий. С одним экземпляром объекта в один и то же момент (интервал) времени может происходить только одно событие одного класса (но несколько событий разных классов).

Уточним предыдущий пример. Пусть Иванов получил права 10.04.2000 г., Петренко – 11.03.1990 г., а Сидорова – 01.07.2009 г. В этом случае экземпляры события будут представлять собой уточнения класса события конкретным временем (если же событие требовало бы использование интервала времени, то вместо конкретного значения времени использовался бы интервал времени).

Таким образом, обобщение позволяет соотнести совокупность подобных экземпляров объектов с типом объекта, один или несколько типов объектов - с

одним классом объектов и совокупность подобных экземпляров событий - с классом событий.

Каждый тип объектов, класс объектов и класс событий обладают определенными свойствами (качествами, признаками). Для того чтобы их каким-то образом идентифицировать и использовать, в модели вводятся следующие мета-онтологии: *фактическая характеристика объекта, паспортная характеристика объекта, характеристика события*.

Фактическая характеристика объекта – это один поименованный признак (качество, свойство) из всей совокупности признаков, характеризующих экземпляры объектов определенного класса.

Паспортная характеристика объекта – это один поименованный признак (качество, свойство) из всей совокупности признаков, характеризующих тип объекта определенного класса.

Характеристика события – это один поименованный признак (качество, свойство) из всей совокупности признаков, характеризующих событие определенного класса.

Существует достаточное количество предметных областей реального мира, в которых необходимо проводить всевозможные измерения некоторых величин, называемых параметрами, которые в дальнейшем необходимо иметь в модели данных. *Параметр* – это величина, характеризующая однотипные процессы, состояния в конкретный момент времени.

Поэтому в модели "объект-событие" предлагается использовать мета-онтологию *характеристика параметра объекта*.

Характеристика параметра объекта – это изменяемый во времени один поименованный признак (качество) из всей совокупности признаков, характеризующих экземпляры объектов определенного класса.

Для уточнения описания характеристики параметра объекта, например, характеристика "давление", может требовать уточнения: "внутреннее" или "внешнее", "подтвержденное" или "неподтвержденное", используется понятие тип значения параметра объекта (что-то подобное имеет место в модели "объект-качество", когда речь идет о том, что любому атрибуту объекта ставиться в соответствие два имени: "качественное" и "структурное"). *Тип значения параметра объекта* – это один поименованный признак (качество) из всей совокупности признаков, позволяющий различать некоторые одинаковые по названию, но разные по существу характеристики параметров объектов.

Характеристика параметра объекта связана с определенным классом объекта и, так называемым классом параметров объекта. *Класс параметров объекта* – это совокупность параметров, выделенных по какому-либо принципу и идентифицируемая общим названием.

Каждая характеристика имеет свое значение. Хотя ранее говорилось о том, что модель данных является средством абстракции, которое дает возможность увидеть информационное содержание данных, а не конкретные значения данных, но для обозначения самого факта присвоения конкретного значения данных для некоторых понятий модели "объект-событие" были введены мета-онтологии: значение характеристики объекта, значение характеристики события, значение характеристики параметра объекта.

Каждая из рассмотренных выше характеристик, как было отмечено, имеет свое значение. Но эти значения могут определяться для каждой характеристики произвольно, а могут задаваться с помощью заранее сформированного списка значений. Фактически такой перечисляемый тип значений является *доменом* (набором допустимых значений) соответствующей характеристики (свойства объекта, события, параметра объекта), подобно доменам, имеющим место в модели "сущность-связь".

Следующий тип абстракции, используемый в модели - агрегация. В данном случае агрегация предполагает построение классов объектов и классов событий путем их конструирования из других классов объектов и событий. При этом могут агрегироваться и экземпляры объектов из других экземпляров объектов.

Например, конкретный экземпляр телевизора может состоять из конкретных экземпляров радиокомпонент. Точно также класс объектов "Телевизор" может включать в свой состав классы объектов: "Резистор", "Конденсатор", "Транзистор", "Микросхема" и т.д. А для класса объектов "Транспортное средство" может иметь место класс событий "Работа транспортного средства", который может включать классы событий "Заправка транспортного средства", "Расход топлива", "Учет пробега по трассе", "Учет пробега по бездорожью", "Транспортная работа" и т.д.

Как известно, абстракция может быть многоуровневой (иерархичной). В связи с чем, в модели имеется возможность строить различные иерархии. При этом иерархии могут быть получены не только для классов объектов и событий, но и для классов параметров объектов и экземпляров объектов и событий. При этом *классом-владельцем объектов* называется класс объектов, которому может принадлежать рассматриваемый класс. А *классом-владельцем события* - класс событий, которому может принадлежать рассматриваемый класс.

Подчиненные классы объектов и событий могут наследовать свойства своих соответствующих классов-владельцев. А можно запретить наследование определенных свойств.

Владельцем класса параметров объектов называется класс параметров объектов, которому принадлежит рассматриваемый класс.

Владельцем экземпляра объекта называется экземпляр объекта, которому принадлежит рассматриваемый объект. *Владельцем экземпляра события* называется экземпляр события, которому принадлежит рассматриваемое событие.

Связь, в смысле ассоциации, устанавливаемой между сущностями, в модели "объект-событие" проявляется следующим образом:

- между классами объектов - через их иерархию;
- между классами событий - через их иерархию;
- между классами параметров объектов - через их иерархию;
- между экземплярами объектов - через их иерархию и через экземпляры событий;
- между экземплярами событий - через их иерархию.

Если для описания свойств экземпляра объекта, типа объекта, класса объекта, класса события, класса параметров объекта и т. д. необходимо расширенное, понятное объяснение и толкование, а с помощью соответствующих

характеристик этих свойств сделать это невозможно, то для решения этой проблемы в модели "объект-событие" введено понятие "документ". Под *документом* в модели понимается структурированный или неструктурированный объект, назначение которого содержательное описание, графическое отображение, возможная демонстрация, необходимые для детализации или уточнения основных мета-онтологий модели. Документ идентифицируется именем. Если таких документов несколько, то они могут объединяться, в так называемые, *пакеты* документов, имеющих свою иерархию.

Кроме того, для лучшей структуризации предметной области, в модели предусмотрена возможность (при необходимости) разбивать ее на несколько частей с помощью, так называемых, разделов. *Раздел* – это некоторая выделенная и поименованная часть предметной области. Разделы тоже могут иметь свою иерархию.

Таким образом, структурная часть модели данных "объект-событие" определяется мета-онтологиями справедливыми для любой предметной области.

В процессе создания концептуальной модели конкретной ПрО мета-онтологии уточняются онтологиями рассматриваемой предметной области, которые, в свою очередь, представляют собой спецификации на формальном языке договоренностей групп специалистов о том, что и как называется в их области и каким свойствам (соотношениям) удовлетворяет. При этом онтологии должны быть удобны для понимания специалистами и интерпретироваться системами при использовании [3].

Исследования показали, что графическое представление позволяет пользователям лучше и быстрее усваивать, а также запоминать структуру данных [2]. Поэтому для схематического изображения схем модели "объект-событие" используются достаточно простые и наглядные диаграммы (несколько сходные с ER-диаграммами модели "сущность-связь"), являющиеся к тому же своего рода документацией, которая необходима на всех этапах жизненного цикла базы данных.

Сделаем ряд общих замечаний о значении графических представлений и связанных с ними проблемах реализации на компьютере. Зрительный образ на экранах мониторов в компактной и обобщенной форме эффективно воспроизводит основное содержание текста, описывающего ПрО. А размещение достаточно громоздких диаграмм на экране может оказаться либо вообще невозможным, либо нецелесообразным из-за большого числа элементов, связей между ними, пересечений, переносов на другие экраны и т.д. Поэтому кроме диаграммного изображения ПрО в модели предусмотрено представление данных и метаданных в виде связанного иерархического дерева объектов и событий, их иерархий, наборов характеристик, их значений и т. д., которое поддерживается с помощью специального программного средства.

Далее, в соответствии с современной трактовкой понятия "модель данных", в которой явно выделяются три компоненты, первый из которых уже был рассмотрен, перейдем к рассмотрению второго - *целостность данных*.

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.

Например, имеется организационная структура "Национальная акционерная компания", в которую входят различные экземпляры объектов различных типов класса "Субъект". Экземпляр объекта "Харьковгаз" типа «ОАО» подчинен экземпляру объекта "Газ Украины" типа "ДК". Тогда иерархическое имя экземпляра объекта "Харьковгаз" можно схематично отобразить следующим образом:

<Раздел>=Национальная акционерная компания; / <КлассО>=Субъект;
<ТипО>=ДК; <ЭкзО>=Газ Украины; → <КлассО>=Субъект; <ТипО>=ОАО;
<ЭкзО>=Харьковгаз;

Может показаться, что такая идентификация достаточно громоздка, но она всегда информативна и имеет семантический смысл в отличие от ничего не значащего уникального *OID (Object Identifier - объектный идентификатор)* в объектной модели и модели "объект-качество", а также первичных ключей в модели "сущность-связь". Кроме того, такое иерархическое имя объекта позволяет контролировать иерархию классов объектов, экземпляров объектов этих классов, тем самым, обеспечивая целостность данных, в том числе и ссылочную, подобно объектной модели.

Подобного вида записи формируются для мета-данных и данных любой рассматриваемой предметной области, что позволяет контролировать, а, следовательно, и обеспечивать целостность данных (объектов, событий, характеристик и т.д.) на семантическом уровне, понятном не только специалистам информационных технологий, но и специалистам ПрО и конечным пользователям.

В модели "объект-событие" поддерживаются также следующие ограничения: ограничения на допустимые перечисляемые (списочные) значения доменов характеристик объектов, событий, параметров объектов; ограничения на единицы измерений, рассматриваемых характеристик объектов, событий, параметров объектов; ограничения на события: с одним экземпляром объекта в один и то же момент (интервал) времени может происходить только одно событие одного класса.

Третий компонент модели – это *средство манипулирования данными*. Средством манипулирования данными для модели "объект-событие" является специальный непроцедурный (декларативный) язык модели данных (ЯМД). В случае использования реляционной модели БД, с помощью ЯМД можно манипулировать данными такой БД подобно языку SQL-запросов, но в терминологии близкой к естественному языку, не привязанной к структуре данных и, не прибегая к контекстному поиску.

ЯМД могут использовать различные специалисты и пользователи на протяжении всего жизненного цикла БД.

4. Выводы

В сравнении с существующими семантическими моделями данных, модель данных "объект-событие" существенно расширяет, так необходимые при

проектировании БД, возможности по адекватному отображению реального мира, как динамической системы, и обеспечивает сохранение сведений о свойствах и связях, которые либо являются актуальными и достоверными на текущий момент (состояние ПрО), либо их утратили.

ЛИТЕРАТУРА

1. Дейт К. Введение в системы баз данных, 7-е издание.: Пер. с англ. — М.: Издательский дом "Вильямс", 2001. — 1072с.: ил.
2. Цикритзис Д., Лоховски Ф. Модели данных /Пер. с англ. – М.: Финансы и статистика, 1985. – 344с.
3. Бениаминов Е. М. Некоторые проблемы широкого внедрения онтологий в ИТ и направления их решений // <http://beniaminov.rsuh.ru/BeniaminovOntoNew.pdf>
4. Когаловский М. Р. Концептуальное моделирование в технологиях баз данных и онтологические модели. Доклад на Симпозиуме «Онтологическое моделирование: состояние, направления исследований и применения», Звенигород, 20-21 мая 2008 г. // <http://www.cemi.rssi.ru/mei/articles/koga08-1.pdf>.