

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

УДК 004.652.8

А.М. Бабанов

ДВА СОВРЕМЕННЫХ ПОДХОДА К СЕМАНТИЧЕСКОМУ МОДЕЛИРОВАНИЮ – ORM И ERM

Используя схему Дейта для синтеза семантической модели [1] в качестве методики анализа, автор сравнивает два современных подхода к семантическому моделированию – модель Халпина «Объект – Роль» (OR-модель) [2] и собственную модель «Сущность – Связь – Отображение» (ERM-модель) [3]. Обе эти модели отличает от предшествующих семантических моделей (ER-модели и ее модификаций) более богатый набор структурных и ограничивающих понятий, позволяющий представить в схеме данных практически всю семантику моделируемого мира. Проведенный анализ показывает, что возможности определения структур данных и ограничений целостности ERM-модели, по крайней мере, не хуже, чем у OR-модели, а в чем-то даже превосходят их.

Ключевые слова: семантическая модель данных; OR-модель, ERM-модель, неформальные семантические концепции; формальные понятия; формальные правила; формальные операторы.

Своим возникновением семантическое моделирование обязано П. Чену и его модели «Entity – Relationship», или, кратко, ER-модели [4]. В указанной статье вместе с представлением новой модели заложил основы семантической методики проектирования схем БД. Эта методика блестяще дополняет классическую методику нормализации, снимая такие проблемы последней, как нетрадиционный для большинства людей и весьма нетривиальный способ восприятия и формализации предметной области (ПрО), практическая неприменимость для сложных ПрО, неоднозначность решения проблемы проектирования, приводящая к прямому перебору многочисленных вариантов схемы в поисках наиболее подходящего. Успех применения семантической методики проектирования БД, помимо знаний и умений человека, определяют два фактора: мощность выразительных средств используемой семантической модели и детальность анализа семантической схемы в применяемых правилах ее трансляции в модель СУБД [5].

Именно поэтому представляет интерес синтез максимально выразительной семантической модели, в рамках которой на этапе первоначальной формализации описаний ПрО возможна фиксация в схеме всех основных понятий ПрО и всех закономерностей взаимоотношений между ними. Идеальная семантическая схема ПрО должна содержать полное формальное описание ПрО с тем, чтобы все последующие межмодельные преобразования схемы носили чисто синтаксический характер и не требовали повторного анализа семантики ПрО и работы с текстами на естественном языке, в которых предварительно зафиксированы выявленные бизнес-правила. Возможность построения таких семантических схем откроет перспективу полной автоматизации процесса проектирования БД, включая генерацию таких элементов реализации, как представления, триггеры и т.д.

Далее в статье будет представлен анализ двух наиболее выразительных семантических моделей – OR-модели и ERM-модели.

1. Подход Дейта к проблеме семантического моделирования

В своей монографии К. Дейт определяет процесс синтеза семантической модели следующим образом. «Общий подход к проблеме семантического моделирования характеризуется четырьмя основными этапами.

1. Прежде всего, попытаемся выявить некоторое множество семантических концепций (понятий), которые могут быть полезны при неформальном обсуждении реального мира.
2. Далее попытаемся определить набор соответствующих символических (т.е. формальных) объектов, которые могут использоваться для представления упомянутых выше семантических концепций.
3. Кроме того, следует определить набор формальных общих правил целостности, предназначенных для работы с такими формальными объектами.
4. Наконец, необходимо также определить набор формальных операторов, предназначенных для манипулирования этими формальными объектами» [1].

2. Неформальные семантические концепции

В поисках неформальных семантических концепций авторы обоих анализируемых моделей обратились к самой мощной семиотической системе – естественному языку. Вот как толкуют их современные логики. Все значащие выражения языка, т.е. выражения, являющиеся знаками, могут быть разбиты на некоторую совокупность классов, называемых семантическими категориями. Семантическая категория – это класс выражений с однотипными предметными значениями, при этом включающий все выражения с предметным значением данного типа.

Имена – слова и словосочетания, являющиеся знаками предметов. По существу, предметом мысли является все то, что обозначено каким-либо именем. Имена бывают единичные, их предметными значениями являются отдельные предметы, и общие. Общее имя является знаком произвольного, любого предмета некоторого класса (определенного вида предметов) и в силу этого как знак представляет в нашем мышлении именно этот класс, который и считается предметным значением общего имени.

Предикаторы – выражения языка (слова и словосочетания), предметными значениями которых являются свойства (одноместные предикаторы) или отношения (многоместные предикаторы). При этом имеются в виду свойства и отношения, которые употребляются как характеристики предметов познания, т.е. наличие или отсутствие чего у предметов мы утверждаем в наших высказываниях.

Предметные функторы – это знаки так называемых предметных функций. Наряду с математическими функциями сюда относят такие характеристики предметов, как скорость, плотность, возраст, пол и т.п. Такие характеристики иногда называют свойствами, хотя это весьма неточно; свойства – это, например, «жидкий», «электропроводный» и т.п. В отличие от свойств предметные функции могут сопоставлять предметам значения из произвольных множеств (числа, строки, даты), а не только истинностные значения Истина и Ложь.

Логические термины – знаки логических отношений (операций), с помощью которых в языке строятся сложные понятия и высказывания. Обобщенно все они могут трактоваться как знаки специфических – логических – функций.

Повествовательные предложения – знаковые формы суждений. Возможными предметными значениями повествовательных предложений являются Истина и Ложь. Смыслом повествовательного предложения является суждение. Суждение, взятое вместе со знаковой формой, в логике принято называть высказыванием.

В логическом анализе языка с целью придания этому анализу большей точности и достижения при этом некоторых обобщений применяется разработанная в логике функциональная трактовка некоторых выражений языка. В логике осуществлено значительное обобщение математического понятия «функция», в силу которого все значимые выражения языка, кроме предложений, единичных имен и их аналогов – переменных, могут трактоваться как функции [6].

Методология мировосприятия каждого человека в той или иной степени совпадает с мнением логики, но, тем не менее, каждый подходит к нему по-своему. Вот мнение Т. Халпина [2, 7], автора ОР-модели, по поводу лежащих в ее основе концепций восприятия моделируемого мира. Поскольку люди естественно обращаются со словами, картинками и примерами, лучший способ достичь ясного описания ПрО – это использовать естественный язык, интуитивно понятные диаграммы и примеры.

По аналогии с термином UML мы называем эти примеры вариантами использования данных, поскольку они являются образцами данных, используемых в системе.

ORM началось в начале 1970-х гг. как семантический подход к моделированию, который воспринимает мир просто как объекты, которые играют роли. Элементарный факт – это простое утверждение или атомарное суждение о том, что конкретный объект играет конкретную роль. Слово «факт» указывает, что утверждение принято истинным пользователями этой ПрО. Мы можем думать о ПрО как о множестве объектов, играющих роли.

Роль, играемую объектом, иногда называют свойством объекта. Тогда элементарный факт утверждает, что некоторый объект имеет некоторое свойство. Это также называют односторонней взаимосвязью, поскольку только одна роль вовлечена в факт [2].

А каковы же семантические концепции, лежащие в основе ERM-моделирования?

Моделируемый мир составляют объекты. Основной информационной (мыслительной) единицей является суждение, которое может быть выражено в языковой форме – высказывании. Причем главную роль (по крайней мере, при проектировании и формировании БД) играют единичные атрибутивные суждения и единичные суждения об отношениях, утверждающие наличие у конкретного объекта определенного свойства или отношения к другим объектам.

Характерным для единичных суждений является использование в качестве логических подлежащих их высказываний единичных имен, предметными значениями которых являются отдельные предметы или объекты (эмпирические или идеальные в зависимости от реальности или идеальности моделируемого мира). Таким образом, имеем одну из основных семантических концепций ERM-модели – объект.

Логическое сказуемое в таких высказываниях может быть задано с использованием одиночного общего имени, предикатора или предметного функтора, либо более сложного логического выражения, включающего эти знаки и логические термины. Все указанные семантические категории могут быть выражены через предметные функции. Последние и представляют вторую основную семантическую концепцию ERM-модели, которую мы будем называть более подходящим термином «отображение». Он, во-первых, не несет математического, количественного смысла образов и прообразов, традиционного для понятия «функция». Во-вторых, с функцией, как правило, ассоциируется требование единственности образа, чего нам также хотелось бы избежать.

Итак, по мнению автора ERM-модели (которое солидаризуется с мнением логиков), для выражения любых суждений о моделируемом мире достаточно двух семантических концепций – объект и отображение. Объект ПрО есть уникальная целостность, которую человек в процессе мировосприятия и мышления способен отличить от всего того, что не является данной целостностью. Отображение ПрО – это некоторый закон предметной области, по которому каждому объекту моделируемого мира (прообразу) может быть поставлен в соответствие (а может быть, и нет) один или более объектов (образов).

Отметим, что пока речь идет о конкретных понятиях «объект» и «отображение», элементами объема которых являются эмпирические или идеальные явления ПрО. Синтез на их основе формальных понятий ERM-модели еще впереди.

3. Формальные понятия

Базовые неформальные семантические концепции играют настолько важную роль, что они часто составляют название модели. Так произошло и с анализируемыми моделями (ERM-модель первоначально называлась OM-моделью – модель «Object – Mapping»). Эти же концепции нашли однозначное отражение в базовых формальных понятиях обеих моделей. Отметим, что базовых понятий OR-модели «объект» и «роль» недостаточно для представления большей части информации. Так, зная только то, что Джон играет роль Муж, а Мэри – роль Жена, нельзя установить факт их супружества. Необходимы дополнительные формальные понятия, например, позволяющие представить тот факт, что «два или более объектов участвуют вместе во взаимосвязи» [2]. А значит, помимо «одно-

местных взаимосвязей», необходимы, по крайней мере, еще и взаимосвязи порядка больше одного. В ERM-модели же понятия «объект» и «отображение» обеспечивают всю выразительную силу языка а остальные формальные понятия являются производными от них. Этому потенциалу они обязаны функциональной трактовке семантических категорий естественного языка.

Итак, сначала рассмотрим систему формальных понятий Халпина [2, 8, 9].

Элементарные факты утверждают, что объекты играют роли. Объекты являются или значениями, или сущностями. Значение – константа, которая самоидентифицируется в смысле, что когда вы видите константу, записанную в некотором контексте, вы всегда знаете то, на что она указывает. В результате на значения можно сослаться непосредственно, без необходимости идентифицировать их описанием.

Концептуально на сущность (например, конкретного человека или автомобиль) ссылаются в информационной системе посредством определенного описания. Например, кенгуру прыгают по сущности, идентифицированной как «Страна по имени 'Австралия'». В отличие от значений, некоторые, а в действительности – большинство объектов могут изменяться со временем. Сущность может быть материальным объектом (например, «Город по имени 'Париж'») или абстрактным объектом (например, «Курс с кодом 'CS 114'»). Мы рассматриваем и сущности и значения как объекты, которые существуют в ПрО.

Теперь, когда мы знаем, как указать объекты, поговорим о том, как мы определяем роли, которые они играют. Мы используем логические предикаты. В логике предикат – это аналог повествовательного предложения с пустотами для объектов в нем. Чтобы окончательно сформировать предложение, надо пустоты для объектов заполнить терминами объектов.

Предикат можно представить как «...болеет», используя многоточие для обозначения пустоты или «ниши» для экземпляра объекта. Мы продемонстрировали одноместный предикат или предложение с одной нишей для объекта. Его можно также назвать свойством или одноместным типом взаимосвязей.

Бинарный предикат – это предложение с двумя нишами для объектов. Большинство предикатов в информационных моделях являются бинарными. Такой предикат можно показать как «...разработал...» с двумя нишами для объектов. Заметим, что порядок, в соответствии с которым объекты будут помещены в него, важен.

Тернарный предикат – предложение с тремя нишами для объектов. N -арный предикат – предложение с n нишами для объектов ($n > 0$). Так как порядок существен, заполненный n -арный предикат связан с последовательностью из n терминов объектов, необязательно различных. Значение n – арность или степень предиката. Элементарный факт утверждает суждение в форме: $R O_1 \dots O_n$, где R – предикат арности n и $O_1 \dots O_n$ – n терминов объектов, не обязательно различных.

При изменении порядка терминов мы полностью меняем предикат. Мы говорим, что предикат «был разработан» – инверсия предиката «разработал». Хотя семантически мы могли бы расценить два предложения, построенные с их участием, как выражения одного и того же факта, синтаксически они различны. Большинство логиков описали бы это как случай двух различных предложений, выражающих одно и то же суждение. Важно не рассмотреть оба предложения как различные, несвязанные факты. Наш подход к бинарным типам фактов состоит в том, чтобы выбрать один первичный способ объявления предиката и опционально позволить инверсное прочтение, например: «The Person with surname 'Wirth' *designed* / *was designed* by the Language named 'Pascal'».

Здесь предикат слева от «/» используется для чтения слева направо. Предикат справа от «/» используется для инверсного чтения. Наличие различных способов говорить о типе фактов может помочь взаимодействию и улучшить спецификацию ограничений. N -арный тип фактов имеет много возможных упорядочений, но обычно показывается только один из них.

В отличие от ER-модели, ORM не использует атрибуты. Все факты представляются в терминах объектов (сущностей или значений), играющих роли. Хотя это часто приводит к большим диаграммам, подход без использования атрибутов имеет преимущества для концептуального анализа, включающие простоту, семантическую стабильность и легкость изменения.

Другая особенность ORM – ее гибкая поддержка специализации, включая множественное наследование, основанная на формальных определениях подтипов [2].

Основные базовые понятия ERM-модели – объект, класс и отображение, взаимно-однозначно соответствующие нашим семантическим концепциям, обеспечивают большую часть выразительной мощи модели. Неудивительно, что именно эти понятия представлены в синтаксисе и аксиоматике формальной системы теории семантически значимых отображений, являющейся математической основой ERM-модели. Однако человек не всегда использует этот весьма абстрактный уровень мышления. Для простоты работы со схемой выделяются частные виды объектов, классов и отображений, образующие множество производных понятий модели.

Объекты, мыслимые в высказываниях о ПрО как предметы, представляют собой сущности, а классы таких объектов есть не что иное, как множества сущностей. Идеальные объекты, такие как числа, даты, строки символов, являются значениями. Они не обладают свойствами, характеристиками и не вступают в отношения с другими объектами, кроме того что являются значениями характеристик объектов. Их объединяют во множества значений по синтаксическим особенностям. Объекты, подпадающие под конкретные понятия об n -ках предметов, представляют собой связи, а классы объектов, составляющих объемы таких понятий, есть не что иное, как множества связей. Каждый объект в n -ке связи играет определенную роль, характеризующую его функцию в этой связи.

Отображения, определяемые множествами связей, которые в качестве областей определения и значений имеют одиночные множества сущностей или их Декартовы произведения, называются реляционными. Роли объектов-прообразов и объектов-образов этих отображений такие же, как и роли сущностей в связях. Общее количество реляционных отображений, определяемых одним множеством связей степени n , равно $2^n - 2$.

Отображение, ставящее в соответствие объекту истинностное значение, называется отображением-свойством. Если в качестве области значений в отображении используется произвольное множество значений, такое отображение будем называть отображением-характеристикой. Отображения-характеристики являются не чем иным, как атрибутивными отображениями или просто атрибутами.

Таким образом, нам удалось связать основные базовые понятия ERM-модели со структурными понятиями модели «Сущность – Связь». Все вместе они составляют понятийный базис ERM-модели. Отметим, что понятие «объект» является обобщением понятий «сущность», «связь» и «значение». Соответственно, понятие «класс» является обобщением понятий «множество сущностей», «множество связей» и «множество значений». А понятия «реляционное отображение» и «атрибутивное отображение» суть специальные виды отображений. Очевидно, в отличие от первых двух последняя специализация неполна.

Следует отметить, что наряду с основными структурными базовыми понятиями – «класс» и «отображение» – для более полного выражения семантики ПрО используются дополнительные базовые понятия – специализация и категоризация, роли объектов-прообразов и объектов-образов в экземплярах отображений, а также операции и отношения (непересекаемости, включения и равенства – для классов, несовместности, следствия и эквивалентности – для отображений), определенные на множествах классов и отображений.

Предлагаемый подход к выбору структурных понятий модели (абстрактных базовых и более простых производных) позволяет при формализации ПрО оперировать в основном знакомыми понятиями («сущность», «связь»), прибегая к использованию новых форм («отображение») лишь в случае недостаточной выразительности первых. В некоторых случаях проектирования схем данных использования понятий «класс» и «отображение» удастся избежать вовсе.

Рассматривая недостаточно выразительные семантические концепции – объект и роль, – Халпин уже на этапе определения формальных понятий своей модели вынужден ввести действительно необходимое ему понятие «предикат» для представления таких важных явлений, как взаимосвязи между объектами. Но тем самым модель теряет индивидуальность и оригинальность, приближаясь по своим базовым понятиям к ER-модели и реляционной модели.

Понимая важность различных упорядочиваний объектов в предикатах, Халпин вводит их в свою модель в виде прочтений предикатов. При этом он ограничивается ролью этих прочтений исключительно для вербализации схемы и данных – наличие различных способов говорить о типе фактов может помочь взаимодействию с экспертом и улучшить спецификацию ограничений. Лежащих в основе этого отображений он не видит, а они открывают дополнительные возможности для формализации бизнес-правил.

4. Формальные правила для задания ограничений целостности

Говоря о типах ограничений целостности, предусмотренных в модели Халпина, отмечается: ER-нотация менее выразительна, чем ORM, по охвату ограничений и бизнес-правил. Отказ ER- и UML-диаграмм включать стандартные нотации для многих важных ORM-ограничений затрудняет создание всесторонней модели или выполнение преобразований модели [2].

Вот какие дополнительные «стандартные» ограничения целостности Халпин предлагает определять на OR-диаграмме. Внутренние ограничения уникальности указывают, какие роли или комбинации ролей одной ассоциации должны иметь уникальные вхождения. Внешние ограничения уникальности действуют аналогично, но применяются к ролям различных предикатов, концептуально соединенных друг с другом.

Ограничение обязательности роли означает, что каждый экземпляр класса, для которого она определена, должен играть эту роль. Роли, для которых отсутствуют такие ограничения, называются необязательными или опциональными.

Ограничение включающего «или» указывает, что дизъюнкция двух или более ролей является обязательной. Ограничение исключения запрещает одновременное включение ролей ассоциаций. Можно определить два ограничения одновременно – ограничение обязательности и взаимной исключительности ролей. Оба эти ограничения вместе обеспечивают так называемое хог-ограничение (исключающее «или»).

Ограничения сравнения множеств ограничивают способ экземпляризации одной роли или последовательности ролей, касающийся экземпляризации других ролей. Рассматриваются следующие отношения, которыми два множества могут быть связаны: подмножество, равенство, исключение [2].

Следует отметить, что указанные ограничения могут применяться как к одной роли, так и к группе ролей или всей ассоциации.

Из оставшихся нерассмотренными OR-ограничений упомянем ограничения на специализации, ограничения частоты появления объекта в роли и кольцевые ограничения.

Отличительной особенностью аппарата ограничений целостности модели Халпина является, то, что для любой полезной особенности данных в нем изобретены свои специальные термин и графический примитив, что приводит к громоздкости определения модели и сложности ее применения.

Теперь посмотрим, как обстоят дела с правилами задания ограничений целостности в ERM-модели.

Первую группу ограничений целостности составляют правила, определяющие области значений отображений-характеристик. Область значений такого отображения можно задать определением синтаксиса составляющих его значений, перечислением допустимых значений (экстенсионалом), операцией сравнения, логическим выражением.

Вторая группа ограничений целостности определяет количественные характеристики (глобальное минимальное кардинальное число, реальное минимальное кардинальное число, максимальное кардинальное число) и тип (неограниченное, полностью определенное, частичное функциональное, полное функциональное) каждого отображения-характеристики и обратного ему отображения.

Третья группа ограничений целостности определяет количественные характеристики и тип реляционных отображений.

Четвертая группа ограничений целостности определяет количественные характеристики и тип отображений, сопутствующих атрибутным отображениям множеств связей.

Пятая группа ограничений целостности определяет получаемые отображения (образуемые с помощью операций над отображениями), а также их количественные характеристики и тип.

Шестая группа ограничений целостности определяет взаимоотношения между классами и между отображениями.

Последняя группа ограничений целостности связана со специализациями. Для каждой из них необходимо указать, является она полной или частичной, пересекающейся или непересекающейся, а также определить отображения, которые являются основаниями выделения подклассов.

Отличительной особенностью аппарата ограничений целостности ERM-модели является то, что, в отличие от OR-модели, в ней предлагается ограниченный универсальный набор конструкций, с помощью которых можно представить многие (если не все) законы взаимоотношения данных. По крайней мере, все типы ограничений целостности OR-модели можно выразить предлагаемыми в ERM-модели средствами.

5. Формальные операторы

Если Халпин вообще не рассматривает операции над данными (по крайней мере, в монографии [2]), то в ERM-модели они играют важную роль еще при построении схемы, поскольку с их помощью определяются производные отображения, которые, так же как и базовые отображения, являются объектом ограничений целостности. Все операции над отображениями составляют алгебру отображений.

Отображение φ^{-1} является инверсией отображения φ тогда и только тогда, когда $\forall x \forall y (y = \varphi(x) \leftrightarrow x = \varphi^{-1}(y))$.

Отображение $\psi(\varphi())$ является композицией отображений φ и ψ тогда и только тогда, когда $\forall x \forall z (z = \psi(\varphi(x)) \leftrightarrow \exists y (y = \varphi(x) \wedge z = \psi(y)))$.

Отображение $\varphi \cup \psi$ является объединением отображений φ и ψ тогда и только тогда, когда $\forall x \forall y (y = (\varphi \cup \psi)(x) \leftrightarrow y = \varphi(x) \vee y = \psi(x))$.

Отображение $\varphi \cap \psi$ является пересечением отображений φ и ψ тогда и только тогда, когда $\forall x \forall y (y = (\varphi \cap \psi)(x) \leftrightarrow y = \varphi(x) \wedge y = \psi(x))$.

Отображение $\varphi - \psi$ является разностью отображений φ и ψ тогда и только тогда, когда $\forall x \forall y (y = (\varphi - \psi)(x) \leftrightarrow y = \varphi(x) \wedge \neg y = \psi(x))$.

Отображение $\varphi[Y]$ является проекцией отображения φ на определяемую роль (или группу ролей) Y тогда и только тогда, когда $\forall x \forall y (y = \varphi[Y](x) \leftrightarrow \exists z (< y, z > = \varphi(x))$. Здесь для упрощения записи мы опустили имена ролей и сложный объект-образ представили упорядоченным кортежем $< y, z >$.

Отображение $\varphi \diamond \psi$ является агрегатом отображений φ и ψ тогда и только тогда, когда $\forall x \forall y \forall z (< y, z > = (\varphi \diamond \psi)(x) \leftrightarrow y = \varphi(x) \wedge z = \psi(x))$.

Из-за ограниченности объема статьи здесь представлена не вся алгебра отображений, и приведенные операции даются в варианте для простых отображений (один объект – прообраз и один объект – образ).

6. Метасхемы моделей

Понятийный базис модели наглядно представляется ее метасхемой. Если обычная схема данных описывает моделируемый мир, то метасхема представляет структуру схемы ПрО, в ней данными являются элементы этой схемы.

Из-за разнообразия специфических базовых понятий полная метасхема OR-модели Халпиным не приводится. Ее фрагменты из [2, 10, 11] представлена на рис. 1 и 2.

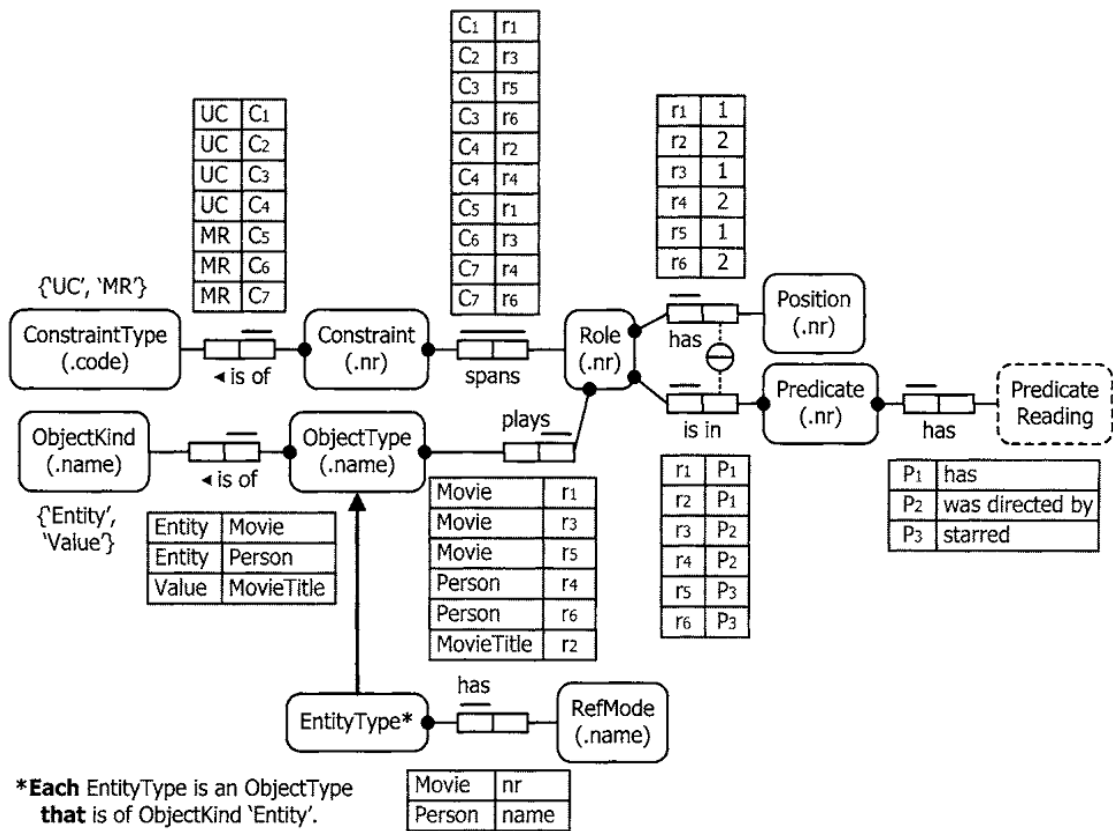


Рис. 1. Фрагмент метасхемы OR-модели (объекты, предикаты и внутренние ограничения целостности); таблицы представляют экстенционалы ассоциаций

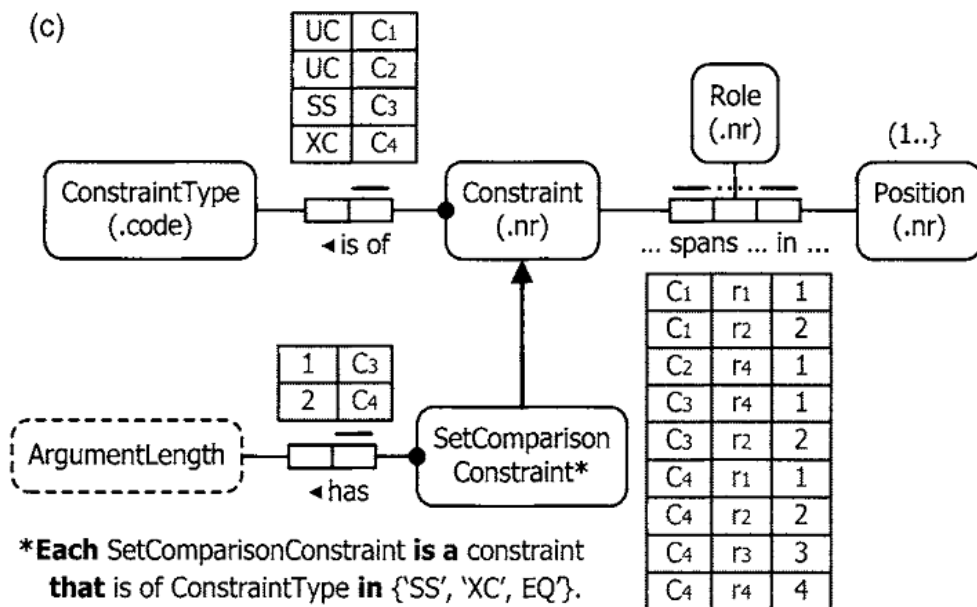


Рис. 2. Фрагмент метасхемы OR-модели (ограничения сравнения множеств)

На рис. 3 представлена полная текущая метасхема ERM-модели. Поскольку модель находится в развитии, постоянно пересматривается и ее метасхема. Серым цветом показаны базовые концепции, белым – производные. Для компактности рисунка опущены атрибуты множеств сущностей и множеств связей метасхемы.

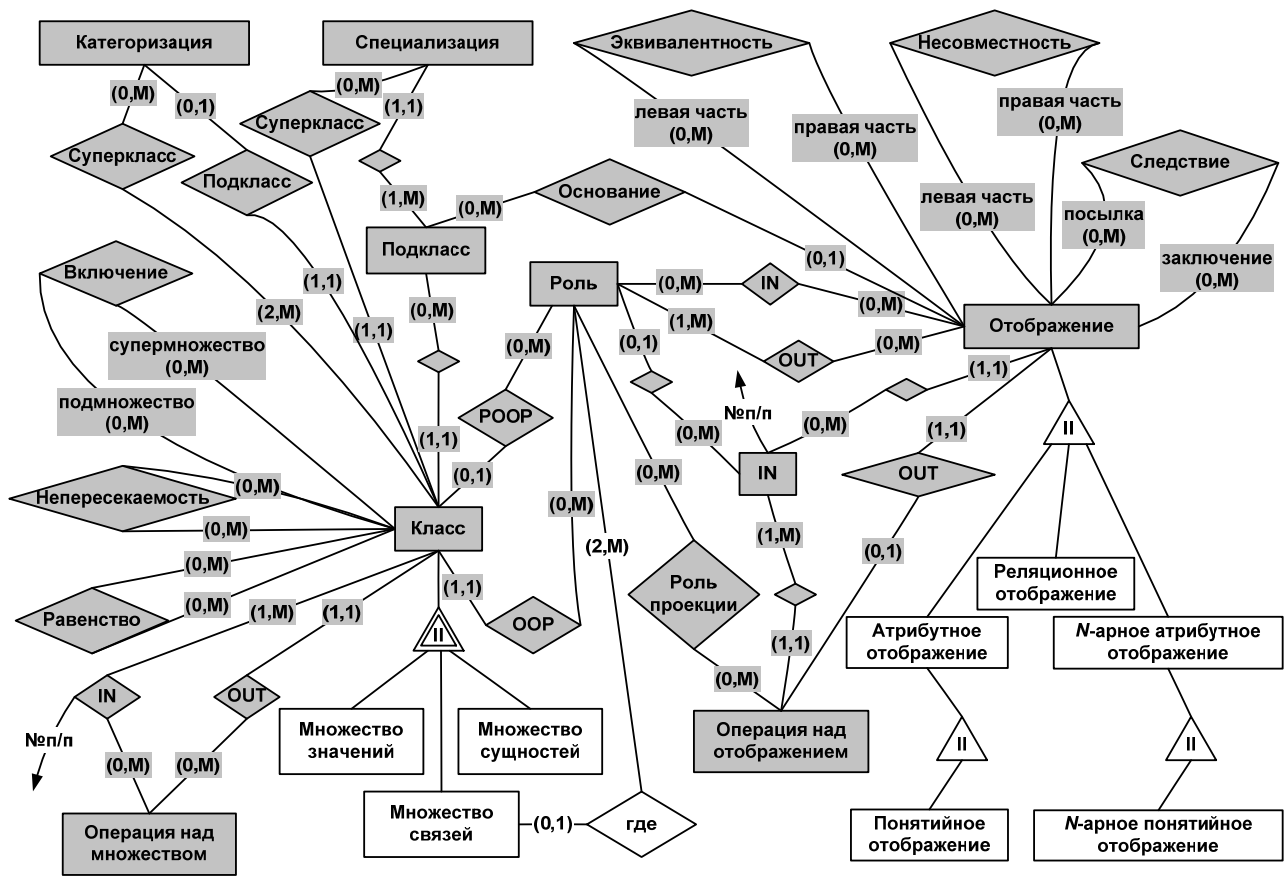


Рис. 3. Полная метасхема ERM-модели

На рис. 3 хорошо видны универсальные структурные и ограничивающие базовые понятия, заимствованные из логики и математики, но это не мешает (или помогает?) ER-модели по своим выразительным возможностям приближаться к естественному языку.

Заключение

Проведенный анализ показывает, что, понимая важность синтеза мощной семантической модели и рассматривая в качестве эталона выразительности естественный язык, авторы ORM- и ERMM-моделей идут различными путями. Если Халпин декомпозирует предикат (или взаимосвязь) степени n на множество n ролей, агрегация экстенционалов которых не восстанавливает экстенционал предиката, то в ERMM-модели вместо типа связей степени n рассматривается $2^n - 2$ типов отображений, каждый из которых эквивалентен одной связи. Вместо n ролей мы имеем $2^n - 2$ отображений, характеристики которых полнее описывают семантику ПрО. Если к этому добавляется возможность определения производных отображений с использованием алгебры отображений, то описательные способности ERM-модели становятся практически безграничными, что, кстати, подтверждается реализацией в ней всех более-менее значащих примеров OR-схем, приведенных в [2].

Богатые выразительные возможности ERM-модели – не результат механистического изобретения специальных конструкций для каждой характерной особенности данных (как это делает Халпин). Основу ERM-модели составляют универсальные понятия логики и математики, сочетая которые, как и в естественном языке, достигается вся ее выразительная мощь.

ЛИТЕРАТУРА

1. Дейт К. Введение в системы баз данных. 7-е изд. : пер. с англ. М. : Вильямс, 2001. 1072 с.

2. Halpin T., Morgan T. Information Modeling and Relational Databases, Second Edition. Morgan Kaufman, 2008. 943 p.
3. Бабанов А.М. Семантическая модель «Сущность – Связь – Отображение» // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2007. № 1. С. 77–91.
4. Chen P.P.-S. The Entity-Relationship Model – Toward a Unified View of Data // ACM Transactions on Database Systems. 1976. V. 1, No. 1. P. 9–36.
5. Бабанов А.М. Семантическая методика проектирования БД и ее перспективы, открывающиеся с применением ERM-модели данных // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2011. № 3(16). С. 58–66.
6. Войцивилло Е.К., Дегтярев М.Г. Логика как часть теории познания и научной методологии (фундаментальный курс) : в 2 кн. М. : Наука, 1994.
7. Halpin T. What is an elementary fact? // Proceedings of First NIAM-ISDM Conference. 1993. Utrecht. P. 1–11.
8. Halpin T. Object-Role Modeling: Principles and Benefits // International Journal of Information Systems Modeling and Design. 2010. No. 1(1). P. 32–54.
9. Halpin T. Object-Role Modeling // Encyclopedia of Database Systems. Springer. 2009. P. 1941–1946.
10. Cuyler D., Halpin T. Two Meta-Models for Object-Role Modeling // Information Modeling Methods and Methodologies. Idea Publishing Group, Hershey PA, USA, 2005. P. 17–42.
11. Halpin T. Comparing Metamodels for ER, ORM and UML Data Models // Advanced Topics in Database Research. Idea Publishing Group, Hershey PA, USA, 2004. V. 3, pt. II. P. 23–44.

Бабанов Алексей Михайлович, канд. техн. наук. E-mail: babanov2000@mail.ru
Томский государственный университет

Поступила в редакцию 4 марта 2014 г.

Babanov Alexey M. (Tomsk State University, Tomsk, Russian Federation).

Two modern approaches to semantic modeling: ORM and ERMM.

Keywords: semantic data model; OR-model; ERM-model; informal semantic concepts; formal structural concepts; formal integrity rules; formal data operators.

Semantic modeling occurrence is obliged to P. Chen and his model «Entity – Relationship». He laid the bases for a semantic technique of the DB schemes designing in his initial article together with the representation of a new model. This technique excellent supplements a classical technique of normalization. Two factors determine a success of semantic technique application (besides designer knowledge and skills): used semantic model expressiveness and minuteness of the semantic scheme analysis in used rules of its translation in the DBMS model.

For this reason, synthesis of the maximum expressive semantic model is of interest. Thus, fixing in the scheme of all basic concepts and all laws of mutual relations between them will be possible within the framework of this model at a stage of initial data domain formalization. The ideal data domain semantic scheme should contain the full formal data domain description, so that all subsequent intermodeling transformations of the scheme had be only syntactic and did not demand the repeated analysis of data domain semantics and works with texts in a natural language, in which the revealed business rules are preliminary fixed. The opportunity of construction of such semantic schemes will open prospect of full automation of DB designing process, including generation of such realization elements as views, triggers, etc.

Using Date's pattern for synthesis of semantic model as a technique of the analysis, the author compares two modern approaches to semantic modeling: Halpin's model «Object – Role» (OR-model) and own model «Entity – Relationship – Mapping» (ERM-model). Both these models are distinguished from previous semantic models with the richer set of the structural and constraining concepts, allowing to present in the data scheme practically all semantics of the modeled world.

In the ERM-model concepts «object» and «mapping» provide all expressive force of language, and other formal concepts are derivatives from them. For this potential they are obliged to functional treatment of natural language semantic categories.

Semantic concepts of OR-model «object» and «role» are not enough for representation of the most part of the information. For this purpose, Halpin already at a stage of model formal concepts definition entered the concept «predicate». But thus the model loses the individuality and originality, coming nearer on the base concepts to the ER-model and relational model.

Understanding the importance of various objects orderings in predicates, Halpin enters them in the model as the perusals of predicates. But he does not go beyond usage of these perusals for verbalization of the scheme and the data. He does not see mappings underlying it, and they open additional opportunities for the formalization of business rules.

Distinctive feature of Halpin's model integrity constraints is, that he is invented the special term and graphic primitive for any useful feature of the data. That leads to the unhandiness of the model definition and complexity of its application. As against the OR-model and ERM-model, we offer the limited universal set of constructions. The basis of ERM-model composes the generalized concepts of logic and mathematics, combined in use, as well as in a natural language, provides all its expressive power.

If Halpin does not consider data operation at all, in ERM-model they play the important role even at the construction of the scheme as with their help derivative mappings are defined, and they are the object of integrity constraints as well as the base mappings.

Understanding the importance of powerful semantic model synthesis and considering as the standard of expressiveness a natural language, authors of the ORM and ERMM go by various ways. If Halpin decomposes the predicate (or relationship) of degree n on

the set of n roles, which extensions aggregation does not restore predicate extension, in the ERMM instead of the relationship type of degree n is considered as 2^n-2 types of mappings, which each copy is equivalent to the whole relationship in information sense. Instead of n roles we have 2^n-2 mappings which characteristics more full describe semantics. If to this, the opportunity of determining the derivatives of mappings, using algebra maps, is added, then descriptive abilities of ERM-model become practically boundless.

The carried out analysis shows that the opportunities of data structures and integrity constraints definition of the ERM-model are not worse, than ones of the OR-model, and sometimes even surpass them.

REFERENCES

1. Date C.J. *An introduction to database systems*. Addison-Wesley Longman Inc., 2000. 1072 p.
2. Halpin T., Morgan T. *Information modeling and relational databases*. Morgan Kaufman, 2008. 943 p.
3. Babanov A.M. Semantic model “Entity – Relationship – Mapping”. *Tomsk State University Journal of Control and Computer Science*, 2007, no. 1(1), pp. 77-91. (In Russian).
4. Chen P.P-S. The entity-relationship model – toward a unified view of data. *ACM Transactions on Database Systems*, 1976, vol. 1, no. 1, pp. 9-36. DOI: 10.1145/320434.320440.
5. Babanov A.M. Semantic method of database designing and its prospects opening with application of the ERM data model. *Tomsk State University Journal of Control and Computer Science*, 2011, no. 3(16), pp. 58-66. (In Russian).
6. Vojshvillo E.K., Degtyarev M.G. *Logika kak chast' teorii poznaniya i nauchnoy metodologii (fundamental'nyy kurs)* [Logic as a part of the theory of knowledge and scientific methodology]. Moscow: Nauka Publ., 1994. (In Russian).
7. Halpin T. What is an elementary fact? *Proc. of First NIAM-ISDM Conference*. Utrecht, 1993, pp. 1-11.
8. Halpin T. Object-Role modeling: Principles and Benefits. *International Journal of Information Systems Modeling and Design*, 2010, no 1(1), pp. 32-54. DOI: 10.4018/jismd.2010092302.
9. Halpin T. Object-Role modeling. *Encyclopedia of Database Systems*, Springer, 2009, pp. 1941-1946. DOI: 10.4018/jismd.2010092302.
10. Cuyler D., Halpin T. *Two meta-models for object-role modeling*. In: Krogstie J., Halpin T., Siau K. (eds.) *Information Modeling Methods and Methodologies*. Idea Publishing Group, Hershey PA, USA, 2005, pp. 17-42.
11. Halpin T. *Comparing Metamodels for ER, ORM and UML Data Models*. In: Siau K. (ed.) *Advanced Topics in Database Research*. Idea Publishing Group, Hershey PA, USA, 2004, pt. II, vol. 3, pp. 23-44.