

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

УДК 004.652.8

DOI: 10.17223/19988605/34/8

А.М. Бабанов, Е.С. Квач

«IS-THE»-ОТНОШЕНИЯ В СЕМАНТИЧЕСКИХ МОДЕЛЯХ ДАННЫХ: ОСНОВНЫЕ ПОНЯТИЯ И РАЗНОВИДНОСТИ

Статья открывает серию публикаций, посвященных исследованиям и результатам, которые получены в такой редко упоминаемой области моделирования данных, как «IS-A»-отношения. В первой статье вводится понятие «IS-THE»-отношения, более подходящее для работы с данными, даются основополагающие определения и рассматриваются различные виды «IS-THE»-отношений. Предлагаемый подход отличается следованием логическим канонам, но без ущерба функциональности, исторически сложившейся в этом разделе моделирования данных.

Ключевые слова: «IS-THE»-отношение; «IS-THE»-кластер; «IS-THE»-категоризация; семантическая модель данных; ERM-модель.

«IS-A»-отношения (отношения «ЕСТЬ-НЕКОТОРЫЙ») являются важным элементом человеческой умственной деятельности (не только научной). И на бытовом уровне мы часто используем обобщенные и специализированные понятия: «Пушкин А.С. есть некоторый поэт», «Поэт есть некоторый человек».

Что же касается науки, то в ней операции обобщения и специализации являются важными мыслительными инструментами, позволяющими переходить от частных явлений к их абстракциям различных уровней и наоборот. Многочисленные классификации в различных отраслях науки тому подтверждение. Это важное для человека явление давно изучается исследователями, правда, задачи конкретной науки накладывают свой отпечаток на представление об этом предмете [1, 2]. Соответственно, используются и различные термины для обозначения «IS-A»-отношений: собственно «IS-A»-отношение, «IS-IN»-отношение (семантические сети), отношение наследования (системы представления знаний и системы объектно-ориентированного программирования), отношения принадлежности и включения (теория множеств), операции обобщения и ограничения понятий (логика), отношения «суперкласс-подкласс», «супертип-подтип» и т.п. Интуитивно все эти термины имеют близкие смыслы, но есть и существенные отличия.

В [2, 3] выделяют две группы смысловых значений «IS-A»-отношений – класс / объект (англ. generic / individual) и класс / класс (англ. generic / generic). Первую группу составляют трактовки этого отношения, близкие теоретико-множественному отношению принадлежности элемента множеству. Вторую группу составляют трактовки «IS-A»-отношения, близкие теоретико-множественному отношению включения одного множества в другое.

В моделировании данных эта естественная для людей форма представления информации была предложена не сразу. В семантических моделях она появилась вместе с расширенной моделью «Сущность–Связь» (Enhanced / Extended Entity-Relationship Model – EERM) в 80-х гг. прошлого века [4, 5]. В этой модели для представления «IS-A»-отношений введены такие структурные понятия, как специализация и категоризация.

Однако даже в моделировании данных нет единого мнения во взглядах на «IS-A»-структуры, что уж говорить об отличии этих представлений от логических основ. В статье будет предложено новое по-

нятие «IS-THE»-отношения (отношения «ЕСТЬ-ЭТОТ»), ориентированного на моделирование данных, и предпринята попытка синтеза логически непротиворечивого понятийного базиса рассматриваемого явления. Далее в работе будут определены разновидности структур данных для представления «IS-THE»-отношений при моделировании данных. Среди них будут и уже традиционные специализация и категоризация.

Приведенные определения положены в основу правил структуризации данных и определения ограничений целостности семантической модели «Сущность–Связь–Отображение» (Entity–Relationship–Mapping Model – ERMM) [6], касающихся представления в ней «IS-THE»-отношений.

1. Базовые понятия «IS-A»-отношений

Чтобы получить полную и непротиворечивую картину представлений об «IS-A»-отношении и окружающих его понятиях, начнем рассмотрение с позиции современной логики по этому вопросу.

«Обобщение некоторого понятия есть операция образования из этого понятия некоторого нового с более широким объемом. Обратная операция перехода от некоторого понятия к понятию с меньшим объемом называется ограничением понятия. Обобщение понятия осуществляется за счет определенного ослабления содержания первого. Ограничение понятия осуществляется за счет расширения содержания исходного. Каждый акт обобщения должен быть переходом от вида к некоторому ближайшему роду, а ограничения – от рода к некоторому ближайшему виду.

Деление понятий – это операция разбиения объема понятия на подклассы, представляющие собой виды предметов, мыслимых в этом понятии. То, что мы называем точкой зрения или аспектом различения предметов, называют основанием деления понятий. Цель деления понятия состоит в том, чтобы выделить все возможные виды предметов каждый раз по некоторому определенному основанию. В составе каждого деления можно выделить делимое понятие, основание деления и члены деления.

Деление является правильным, если оно удовлетворяет следующим условиям:

1. Деление должно происходить по одному определенному основанию.
2. Полученные при делении понятия должны быть попарно несовместимы.
3. Члены деления как классы должны исчерпывать объем исходного понятия.
4. Никакой из членов деления не должен быть пустым классом.
5. Деление должно быть непрерывным, т.е. все его члены являются ближайшими видами исходного понятия, выделяемыми по выбранному основанию» [7].

В.А. Бочаров и В.И. Маркин дополняют эту картину рассмотрением возможных вариантов деления понятий.

«С каждым делением связывается основание деления – такая характеристика предметов, входящих в объем делимого понятия, модификация которой и порождает систему членов деления. В зависимости от того, что берется в качестве основания деления, различают два их вида – дихотомическое деление и деление по видоизменению основания. В случае дихотомического деления родового понятия основанием деления является некоторый признак, присущий лишь части предметов, входящих в объем исходного понятия. Деление осуществляется по наличию или отсутствию этого признака у предметов делимого понятия. В случае деления по видоизменению основания в качестве основания деления используются предметно-функциональные характеристики элементов объема делимого понятия. Необходимо только, чтобы предметная функция, берущаяся в качестве основания, была всюду определенной на этом объеме. При таком делении объемы видовых понятий представляют собой собрания тех и только тех предметов a и b , для которых выполняется условие $f(a) = f(b)$.

Всякое деление должно осуществляться по одному основанию» [8].

Насколько представления логиков, формировавшиеся веками, точны и понятны, настолько размыты и неглубоки определения, которыми пользуются в моделировании данных. Там принято то, что логики называют делением понятия, обозначать термином «специализация». Он же используется для именованного нисходящего процесса образования «IS-A»-иерархий с полным наследованием признаков. Обратный процесс, при котором эти «IS-A»-иерархии образуются с помощью восходящего процесса, принято называть генерализацией.

По мнению специалистов по моделированию данных, специализация – это:

- «процесс выделения подклассов в суперклассе, который основан на группировке исключительных характеристик и связей подклассов» [9];
- «процесс выявления различий между экземплярами типа сущностей» [10];
- «процесс определения набора подклассов для типа сущностей на основе отличительных характеристик» [11];
- «процедура, в ходе которой вводятся подклассы более общих типов сущностей, чтобы объявить, что определенные роли зарегистрированы только для этих подклассов» [12];
- «процесс концептуального уточнения суперкласса путем выделения подклассов» [13].

Как видим, если логики обращают внимание, прежде всего, на содержание (признак) понятий деления, то специалисты по моделированию данных видят преимущественно структурные и ограничительные особенности классов данных. Главное назначение специализаций в моделировании данных – возможность группирования признаков (атрибутов и типов связей): обобщенные поднимаются на уровень суперкласса, специализированные опускаются до соответствующего подкласса. Т. Халпин, кроме этого, замечает, что специализации позволяют делать опциональные роли суперкласса обязательными для подкласса и за счет этого вводить дополнительные ограничения целостности [12. Р. 238].

И только Дж. Смит и Д. Смит подходят к концепции специализации аналогично логикам.

«Наш метод представления родовой иерархии требует, чтобы непосредственные потомки любого узла были разбиты на группы. Каждая группа должна содержать родовые объекты, классы которых являются взаимно исключающими. На практике такое группирование обычно может быть весьма легко сделано из семантических соображений. Например, потомки класса Транспорт могли бы быть сгруппированы следующим образом: {Транспортные средства, приводимые в движение ветром; Моторизованные транспортные средства; Транспортные средства, приводимые в движение человеком} и {Воздушный транспорт; Водный транспорт; Наземный транспорт}. Первая группа содержит взаимно-исключающие классы, которые соответствуют альтернативным типам “Системы приведения в движение”. Вторая группа, в свою очередь, содержит взаимно исключающие классы, которые соответствуют различным типам “Среды передвижения”».

Назовем взаимно исключающие группы родовых объектов, имеющие общие родительские объекты, кластерами. Будем говорить, что кластер относится к его родительскому родовому объекту. Например, мы можем говорить о двух кластерах, относящихся к типу объектов Транспорт. Кластер, относящийся к “Моторизованным транспортным средствам”, и кластер, относящийся к “Воздушному транспорту”, могут иметь общий элемент – Вертолет» [14].

Дж. Смит и Д. Смит предлагают, как и логики, сводить иерархии обобщения к отдельным правильным делениям родительских понятий, правда, они их называют кластерами. Но по-прежнему налицо одно основание деления и непересекаемость подклассов одного кластера.

2. Дополнительные понятия «IS-A»-отношений

Важными понятиями, сопутствующими рассмотрению «IS-A»-отношений, являются понятия, применяемые к элементам этих отношений.

«Если экземпляры типа объектов классифицируются в более специфичный тип, то этот специализированный тип называется подтипом» [12. Р. 260]. «Супертип – это обобщенная форма его подтипов, в таком случае подтип – специализированная форма супертима» [Там же].

«Суперкласс – тип сущностей, включающий одну или несколько различных вспомогательных группировок его экземпляров, которые должны быть представлены в модели данных. Подкласс – различимая вспомогательная группировка экземпляров типа сущностей, которая должна быть представлена в модели данных» [4. С. 430]. «Связь между суперклассом и подклассом является связью “один-к-одному” (1 : 1) и называется связью суперкласс / подкласс» [Там же. С. 431].

Одни авторы (Халпин) предпочитают рассматривать иерархии обобщения типов данных, другие, такие как Т. Коннолли и К. Бэгг, строят «IS-A»-иерархии на классах данных. Логика же говорит об

иерархиях понятий. И эта позиция выглядит более универсальной по сравнению с узкими трактовками специалистов по моделированию данных.

Действительно, «понятие – это мысль, которая посредством указания на некоторый признак (содержание понятия) выделяет из универсума и собирает в класс (обобщает) все предметы, обладающие этим признаком (объем понятия) [8]». И далее: «Содержание представляет собой интенциональную характеристику понятия; экстенциональной характеристикой понятия является его объем» [8]. Таким образом, говоря об иерархии понятий, мы предполагаем и иерархию признаков, и иерархию объемов. Оба эти аспекта исчерпывают вариации других авторов.

В таком случае, рассматривая двухуровневую «IS-A»-иерархию, корректнее говорить о родовом (аналог супертипа или суперкласса) и видовых (аналоги подтипов или подклассов) понятиях, точнее, системе видовых понятий. Содержание видового понятия отличает от содержания родового понятия дополнительное условие (видовое отличие), которое истинно для объектов объема этого видового понятия. Таким образом, спуск по «IS-A»-иерархии сопровождается усложнением содержания понятий. Что касается объемов понятий, то чем выше уровень понятия в «IS-A»-иерархии, тем его объем шире, и, наоборот, опускаясь вниз, мы переходим от объемов родовых понятий к их подмножествам – объемам видовых понятий.

В моделировании данных, исходя из практических потребностей, в одну специализацию суперкласса могут собирать подклассы, образованные по разным основаниям. Это нарушает логическую стройность схемы данных и часто приводит к проблемам при реализации этой схемы. Предпочтительнее вслед за логиками и Смитами рассматривать только специализации с одним основанием деления суперкласса на подклассы.

А вот в чем следует отступить от канонов логики, так это второе и третье условия правильного деления – требования попарной несовместимости видовых понятий и покрытия их объемами объема родительского понятия. Необходимость этого в моделировании данных доказала практика использования специализаций. В некоторых моделях по этому поводу введены даже специальные типы специализаций: полные непересекающиеся (аналог правильного деления в логике), полные пересекающиеся, частичные непересекающиеся, частичные пересекающиеся.

Таким образом, приходим к необходимости расширенной, по сравнению с логиками, трактовки операции деления понятий, учитывающей все особенности специализаций в моделировании данных.

3. Понятие «IS-THE»-отношения

Подход логиков в вопросе «IS-A»-отношений кажется более основательным, и его нельзя обойти стороной при синтезе понятий, описывающих эти отношения, но ориентированных на моделирование данных. Вторым краеугольным камнем этой системы понятий являются отображения и теория семантически значимых отображений [15]. Анализ отображений, сопутствующих «IS-A»-отношениям, позволяет глубже рассмотреть этот объект исследований.

Прежде чем говорить о представлении «IS-A»-отношений в моделях данных, введем понятие «IS-THE»-отношения.

Два элемента объемов разных понятий связаны «IS-THE»-отношением (отношением «ЕСТЬ-ЭТОТ») тогда и только тогда, когда они представляют один и тот же объект предметной области (ПрО) (отсюда использование определенного артикля «THE»). Например, фраза «Пациент Иванов есть этот врач Иванов» описывает ситуацию с заболевшим врачом. «IS-THE»-отношение – это бинарное 1 : 1-отношение, определяющее пару взаимобратных функциональных «IS-THE»-отображений. В этом отношении и отображениях участвуют непосредственно объекты ПрО (когда речь идет о законе соответствия объектов) или их абстракции – данные (когда речь идет об аналогичном законе соответствия данных, представляющих объекты).

Структуры данных (множества сущностей, отношения) предназначены для представления данных об объектах ПрО, и каждая из них соответствует определенному понятию (о предметах, *n*-ках предметов), экстенционал которого и представлен данными, принадлежащими этой структуре. Каждый объект ПрО определен в ней своей абстракцией (сущностью, кортежем) в контексте этого понятия. Таким об-

разом, один и тот же объект может быть представлен в нескольких различных структурах данных. Именно между этими элементами-абстракциями одного и того же объекта и устанавливается «IS-THE»-отношение в данных. Если известные трактовки «IS-A»-отношения связывают класс с объектом или класс с классом, то в «IS-THE»-отношения вступают объекты.

В максимально обобщенном понимании «IS-THE»-отношение позволяет указывать в каком-то смысле «близость» понятий: понятия близки, если найдется хотя бы один объект, подпадающий под оба эти понятия. «IS-THE»-отношение в таком варианте определяет два частичных функциональных отображения (действительно, соответствия между элементами объемов произвольных понятий может и не быть).

«IS-THE»-отношения, определенные на объемах понятий, представляются в схеме данных специальными универсальными взаимосвязями между соответствующими структурными компонентами – классами. Если представить классы в виде вершин, а взаимосвязи – в виде ребер, то получится в общем случае граф. Назовем его «IS-THE»-графом.

4. Разновидности «IS-THE»-отношений и дополнительные понятия

Если рассматривать два понятия, одно из которых является обобщением другого (а то, в свою очередь, является ограничением первого), то в этом случае «IS-THE»-отношение становится родовидовым отношением. «IS-THE»-отображение из объема родового понятия (суперкласса) в объем видового (подкласса) по-прежнему частичное функциональное (объект подпадет под видовое понятие только в случае удовлетворения содержания этого понятия). А вот «IS-THE»-отображение из подкласса в суперкласс стало полным функциональным (объект видового понятия всегда подпадет под родовое понятие). Это становится очевидным, если представить, как выглядит содержание любого видового понятия. В нем к родовому признаку конъюнктивно добавлено условие видового отличия. Понятно, что каждый элемент каждого подкласса удовлетворяет обоим признакам – и родовому и видовому.

«IS-THE»-кластер (или, по аналогии со Смитами, просто кластер) – это система классов, построенная для ограничений (возможно, единственного ограничения) понятия одного и того же суперкласса, такая, что все понятия подклассов используют в своих основаниях выделения одно и то же отображение (предметную функцию). Основание выделения подкласса – это видовое отличие объектов этого подкласса, которое представляет собой дополнительное условие содержания понятия подкласса. Единство отображения в основаниях выделения подклассов означает для дихотомического деления разбиение на подклассы в зависимости от наличия и отсутствия образов у элементов суперкласса при одном и том же отображении, для деления по видоизменению основания – объединение в подклассы элементов суперкласса, имеющих совпадающие образы при одном и том же отображении.

Такие отображения можно назвать дискриминирующими, ведь именно эти признаки объектов суперкласса определяют их попадание в подклассы. В этом качестве могут использоваться любые признаки – свойства, характеристики объектов, отношения с другими объектами. Роль прообразов таких отображений играют объекты суперкласса или их абстракции, а роль образов – истинностные значения, произвольные значения и другие объекты (абстракции) соответственно.

Дуальными к экземплярам этих отображений выступают экземпляры отображений между суперклассом и объединением подклассов кластера. С понятийной точки зрения это объединение представляет объем понятия, содержанием которого является дизъюнкция условий всех видовых понятий. По сути эти и упомянутые ранее отображения являются двумя альтернативными формами представления одной и той же информации о ПрО.

Для определения кластеров в схеме данных могут использоваться ориентированные гиперребра – ребра, соединяющие в общем случае несколько вершин: начальные вершины, представляющие подклассы, и конечную вершину, представляющую суперкласс.

Понятие «IS-THE»-кластера совпадает по смыслу с понятием двухуровневой специализации, подклассы которой образованы в соответствии с одним и тем же основанием. Например, рассмотрим специализацию людей на мужчин и женщин (рис. 1). Видовое отличие мужчин – «Пол = 'мужской'», женщин – «Пол = 'женский'». И в том и в другом случае используется одно и то же отображение (Пол). От-

личаются объекты подклассов только своими образами при этом отображении. С точки зрения логики имеем деление по видоизменению основания.

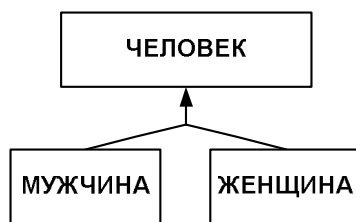


Рис. 1. «IS-THE»-кластер

Последовательно применяя операцию деления к полученным на предыдущем шаге видовым понятиям, можно построить иерархию кластеров (когда нет классов, входящих в качестве подклассов в разные кластеры) или граф кластеров (в противном случае).

Многие исследователи семантических моделей и авторы, повествующие о них, используют термин «специализация» и для обозначения процесса порождения подклассов, и для обозначения его результата – двухуровневой иерархии. Так, Т. Халпин определяет специализацию и как «процедуру, в ходе которой вводятся подклассы более общих типов сущностей» [12. Р. 260], и как «граф подтипов, который может возникнуть нисходящим способом, специализируя тип объекта в подтипы» [Там же. Р. 244]. На наш взгляд, корректнее использовать этот термин только для процесса (и тут все авторы солидарны), а результат именовать «IS-THE»-кластером (в случае двух уровней) или графом кластеров (в случае нескольких уровней). Такой подход позволит к тому же непротиворечиво ввести типизацию структур, некоторые элементы которой не имеют ничего общего со специализациями.

Подавляющее большинство «IS-THE»-отношений, представляющих интерес при построении БД, определяется «IS-THE»-кластерами и подразумевает полное наследование родительских признаков (атрибутов и связей) объектами дочерних понятий. То есть эти признаки на видовом уровне не декларируются, а предполагаются.

В ERM-модели значимые объекты ПрО представляются, так же как и в ER-модели, сущностями. Каждая сущность является абстракцией объекта, построенной в соответствии с некоторым понятием о предметах. Один и тот же объект ПрО, таким образом, может быть представлен несколькими сущностями разных типов. После введения таких представлений можно утверждать, что каждый «IS-THE»-кластер определяет полное функциональное «IS-THE»-отображение между объединением множеств сущностей видовых понятий и множеством сущностей родового понятия. Это является отличительной особенностью «IS-THE»-отношений с полным наследованием. Также действует полное функциональное «IS-THE»-отображение между каждым из множеств сущностей видовых понятий и множеством сущностей родового понятия.

В том случае, когда множество сущностей является подклассом только в одном «IS-THE»-кластере (соответствующее видовое понятие связано только с одним родовым понятием), говорят, что осуществляется полное единичное наследование. Когда множество сущностей входит подклассом в несколько «IS-THE»-кластеров (соответствующее видовое понятие связано с несколькими родовыми понятиями), говорят, что осуществляется полное множественное наследование. Но и в этом случае сущности подкласса наследуют признаки всех своих родителей. В обоих случаях полного наследования «IS-THE»-отношение представляет собой широко известное отношение наследования. Направление гиперссылки «IS-THE»-кластера указывает на родовое понятие, определяющее наследуемые признаки.

В моделировании данных введен особый подвид двухуровневых «IS-A»-иерархий – категоризации, определяющие особые «IS-THE»-отношения между объединением множеств сущностей-суперклассов и множеством сущностей-подклассом, в этом случае называемым категорией [4. С. 209]. Чтобы лучше всего понять суть категоризации, надо представить содержание понятия, объем которого представлен множеством сущностей-подклассов. Первая часть его логического выражения – дизъюнкция условий содержаний множеств сущностей-суперклассов, а вторая часть – условие, которому долж-

ны удовлетворять сущности этих множеств, чтобы быть представленными в категории. Такие двухуровневые «IS-A»-иерархии будем называть «IS-THE»-категоризациями (или просто категоризациями).

В определении категоризации подчеркивается, что суперклассы не содержат абстракции-сущности одних и тех же объектов и настолько различны, что их не связывают общие классы-предки. Другими словами, понятия суперклассов попарно несовместимы. В таком случае элемент категории всегда наследует признаки только одного суперкласса, того, где содержится связанный с ним родительский элемент. Поэтому говорят, что такие двухуровневые «IS-A»-иерархии предполагают выборочное наследование.

По аналогии с «IS-THE»-кластерами можно утверждать, что с каждой «IS-THE»-категоризацией связано полное функциональное отображение между множеством сущностей-категорией и объединением множеств сущностей-суперклассов. А вот между множеством сущностей-категорией и каждым из множеств сущностей-суперклассов имеется частичное функциональное отображение. Что касается обратных отображений, то в зависимости от семантики ПрО они могут быть как полностью определенными, так и нет.

Для представления «IS-THE»-категоризаций также можно использовать ориентированные гиперребра, только в отличие от «IS-THE»-кластеров они соединяют начальную вершину, представляющую подкласс (категорию), и в общем случае несколько конечных вершин, представляющих суперклассы.

Типичным примером «IS-THE»-категоризации являются «IS-THE»-отношения между понятием «субъект права» и понятиями, объекты которых могут выступать в роли владельцев прав, – физическое лицо, юридическое лицо и административно-территориальное образование (рис. 2). Действительно, каждый субъект права – это или гражданин, или организация, или территория, и в зависимости от этого наследуются признаки соответствующего суперкласса.

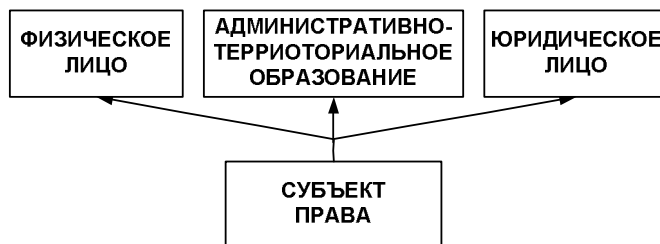


Рис. 2. «IS-THE»-категоризация

Теперь для «IS-THE»-кластеров и «IS-THE»-категоризаций (и только для них) можно ввести обобщенные понятия «суперкласс» и «подкласс». Суперкласс определяет наследуемые признаки, а элементы подклассов их наследуют. В сторону суперклассов ориентированы гиперребра, представляющие эти структуры.

В зависимости от применяемого при образовании «IS-THE»-кластера принципа и основания деления суперкласса на подклассы выделяют кластеры с динамичным и статичным набором подклассов. Если в ходе эволюции данных возможно появление нового подкласса – кластер динамичен, если нет – статичен. При использовании метода дихотомии всегда образуются два статичных подкласса: для объектов одного признак деления истинен, для объектов другого – ложен. Чаще всего второй подкласс не создается, а его экстенционал определяется как разность экстенционала суперкласса и экстенционала другого подкласса. В случае деления по видоизменению основания статичность и динамичность кластера определяет то, под все возможные образы дискриминирующего отображения созданы подклассы или нет.

Преыдущие типы «IS-THE»-кластеров характеризовали интенционал данных, т.е. определялись на основании статичности или динамичности схемы данных. Статичность или динамичность самих данных (экстенционала) приводит к кластерам со стабильными или мигрирующими экземплярами подклассов. Т. Халпин называет эти разновидности кластеров строгими (rigid) и ролевыми (role) соответственно [12. Р. 251].

Во многих моделях данных (в частности, в реляционной модели, модели «Объект–Роль» Халпина, ERMM) в схеме данных выделяются так называемые хранимые и получаемые структуры. Т. Халпин говорит даже о трех состояниях данных – хранимые, получаемые и частично получаемые.

«Получаемый (англ. derived) факт – это факт, который выводится из других фактов математическими вычислениями или логическим выводом. Факт, который нельзя вывести из других фактов, называется хранимым или утверждаемым пользователем (англ. asserted) фактом. Для каждого получаемого факта в схеме данных задается правило его получения» [12. Р. 33]. «Частично получаемый (англ. semi-derived) тип фактов определяется в том случае, когда ряд фактов этого типа можно вывести, а другие факты будут заданы пользователем» [12. Р. 99].

Что касается «IS-THE»-кластеров, то, как уже отмечалось, альтернативной формой представления информации о принадлежности объекта подклассу является наличие соответствующего экземпляра дискриминирующего отображения. Именно между этими формами данных и осуществляет выбор проектировщик схемы, в каждом конкретном случае предпочитая либо хранение экстенсиналов подклассов (хранимый кластер), либо хранение экземпляров отображения (получаемый кластер), либо их смесь (частично получаемый кластер).

Кстати, именно характер дискриминирующего отображения определяет еще одну типизацию «IS-THE»-кластеров, которую упоминают многие авторы [4, 12]. Это деление специализаций на полные и частичные, пересекающиеся и непересекающиеся. Некоторые говорят об этих видах специализаций как об особом типе ограничений целостности на специализации [4. С. 208].

Вот как определяются эти типы «IS-THE»-кластеров в ERM-модели в том случае, когда при делении по видоизменению основания создаются подклассы для всех образов. Если дискриминирующее отображение – полное функциональное, кластер является полным непересекающимся. Если дискриминирующее отображение – частичное функциональное, кластер является частичным непересекающимся. Если дискриминирующее отображение – полное нефункциональное, кластер является полным пересекающимся. Если дискриминирующее отображение – неограниченное, кластер является частичным пересекающимся.

Аналогичные высказывания можно сделать и по поводу отображений между суперклассом и объединением подклассов, дуальных дискриминирующим отображениям. Причем указанное отображение в любом случае определяет характер кластера, чего не скажешь о дискриминирующем отображении. Так, если для некоторых образов дискриминирующего отображения подклассы не определяются, кластер будет частичным независимо от свойств отображения. Функциональность этого отображения гарантирует непересечение подклассов. Но даже нефункциональное дискриминирующее отображение может в этом случае привести к непересекающемуся кластеру.

Если кластер образуется методом дихотомии и включает два подкласса, он всегда будет полным непересекающимся. Если же в таком кластере создается только один подкласс, кластер будет частичным непересекающимся.

Для «IS-THE»-категоризаций существует похожее деление на полные и частичные [4. С. 210]. Если каждый объект всех суперклассов присутствует в подклассе-категории, категоризация считается полной, если есть объекты суперклассов, не отображенные в категории, категоризация частичная. Или на языке отображений: если отображение из объединения суперклассов в категорию – полное функциональное, категоризация полная, если оно – частичное функциональное, категоризация частичная.

И, наконец, в ERM-модели можно строить «IS-THE»-графы на классах одного, но любого типа – множествах сущностей, множествах связей или множествах значений. Это имеет практическое значение и хорошо соотносится с логикой: там наряду с понятиями о предметах (сущностях и значениях) рассматривают и понятия об *n*-ках предметов (связях). Обобщение множеств сущностей и множеств значений (типов данных) стало уже стандартом во многих областях информатики. Потребность в построении «IS-A»-иерархий множеств связей встречалась до сих пор только в системах представления знаний. Теперь такое возможно и в системах баз данных.

Заключение

Значение «IS-A»-иерархий в моделировании данных многими недооценивается. А ведь с их помощью, помимо прочего, еще и восстанавливается единство объекта, рассыпанного в базе данных в виде своих абстракций-сущностей по несвязанным между собой множествам. В семантической модели

ERM предпринята попытка непротиворечивого включения «IS-A»-иерархий в правила построения структур и ограничений целостности. Основным принципом является следование безупречным логическим канонам при сохранении той функциональности этих структур, которая сложилась в моделировании данных. Настоящая статья закладывает основы правил использования «IS-THE»-отношений в виде определений базовых понятий и рассмотрения системы типизации структур, предназначенных для представления этих отношений в данных. В последующих статьях авторы постараются полнее раскрыть феномен «IS-THE»-отношений и их использование в моделировании данных.

ЛИТЕРАТУРА

1. Wang P. From Inheritance Relation to Nonaxiomatic Logic // International Journal of Approximate Reasoning. 1994. No. 11. P. 281–319.
2. Palomäki J., Kangassalo H. That IS-IN Isn't IS-A: A Further Analysis of Taxonomic Links in Conceptual Modelling // Advances in Knowledge Representation. 2012. P. 3–18.
3. Brachman R.J. What IS-A Is and Isn't: An Analysis of Taxonomic Links in Semantic Networks // IEEE Computer. 1983. V. 16, No. 10. P. 30–36.
4. Коннолли Т., Бегг К. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. 2-е изд. М. : Вильямс, 2000. 1120 с.
5. Teorey T., Yang D., Fry J. A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model // Computing Surveys. 1986. V. 18, No. 2. P. 197–222.
6. Бабанов А.М. Семантическая модель «Сущность – Связь – Отображение» // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2007. № 1. С. 77–91.
7. Войшвилло Е.К., Дегтярев М.Г. Логика как часть теории познания и научной методологии (фундаментальный курс) : в 2 кн. М. : Наука, 1994.
8. Бачаров В.А., Маркин В.И. Основы логики. М. : Инфра-М, 1998. 297 с.
9. Coronel C., Morris S., Rob P. Database Systems: Design, Implementation, and Management. 10th ed. Course Technology, 2013. 1054 p.
10. Connolly T., Begg C. Database Systems: A Practical Approach to Design, Implementation, and Management. 4th ed. Addison-Wesley, 2010. 1400 p.
11. Elmasri R., Navathe S. Fundamentals of Database Systems. 6th ed. Addison-Wesley, 2010. 1200 p.
12. Halpin T., Morgan T. Information Modeling and Relational Databases. 2nd ed. Morgan Kaufman, 2008. 943 p.
13. Ferragine V., Doorn J., Rivero L. Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends. Information Science Reference, 2009. 1124 p.
14. Smith J., Smith D. Database Abstractions: Aggregation and Generalization // ACM Transactions on Database Systems. 1977. V. 2, No. 2. P. 105–133.
15. Бабанов А.М. Теория семантически значимых отображений // Вестник Томского государственного университета. Сер. Математика. Кибернетика. Информатика. 2003. № 280. С. 239–248.

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

Квач Елена Сергеевна. E-mail: kvachelena93@gmail.com
Томский государственный университет

Поступила в редакцию 1 декабря 2015 г.

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

«IS-THE»-relations in semantic data models: basic notions and variants.

Keywords: «IS-THE»-relation; «IS-THE»-cluster; «IS-THE»-categorization; semantic data model; ERM-model.

DOI: 10.17223/19988605/34/8

«IS-A»-relations are important intellectual tools allowing to pass from the individual phenomena to their abstractions of the different levels and on the contrary. Importance of these relations in data modeling is underestimated. In fact, these relations in addition recover the unity of an object scattered in a database as unconnected abstract entities. In semantic model ERM (Entity-Relationship-Mapping) attempt of consistent «IS-A»-relation inclusion to rules of construction of structures and integrity constraints is undertaken. Main principle is maintenance of faultless logic canons with preservation of that functionality of these structures which has developed in data modeling.

There this natural for people form of the information representation has been offered not at once. In semantic models it has appeared together with the Enhanced Entity-Relationship Model in 80th years of the last century. In this model such structural concepts as specialization and categorization are entered for representation of «IS-A»-relations. But even in data modeling there is no common opinion about «IS-A»-structures so it is no wonder that these representations are different from logic bases.

The logician approach to the «IS-A»-relation problem looks more thorough, and it cannot be bypassed at the definition of the concepts describing these relations, but focused on data modeling. The second corner stone of this system of concepts are mappings and the

semantically significant mapping theory. The analysis of the mappings accompanying «IS-A»-relations allows to consider this subject of inquiry more deeply.

Two elements of scopes of the different concepts are connected by «IS-THE»-relation if and only if they represent the same application domain (AD) object. «IS-THE»-relation is the binary 1:1-relation determining pair of inverse functional «IS-THE»-mappings. AD objects or their abstractions (data) participate in these relation and mappings. If the known treatments of «IS-A»-relations connect a class with an object or a class with a class, «IS-THE»-relations connect objects.

«IS-THE»-relation in maximum generalized understanding allows specifying «nearness» of concepts in such sense: concepts are near if there is at least one object falling under both these concepts. «IS-THE»-relation in such variant determines two partial functional mappings.

If two concepts, one of which is generalization of another, are considered, then «IS-THE»-relation becomes a subsumption relation. «IS-THE»-mapping from scope of generic concept (superclass) to scope of specific concept (subclass) is still partial functional. And here, «IS-THE»-mapping from a subclass to a superclass becomes full functional.

«IS-THE»-cluster (or, simply, cluster) is a system of classes constructed for restrictions of the same superclass concept; such, that all subclass concepts use the same mapping in the bases of them subtyping. Such mapping can be named discriminating; in fact it is this characteristic of superclass objects that determines their hits in subclasses. When an entity set is a subclass only in one «IS-THE»-cluster, it is said that full single inheritance takes place. When an entity set enters as a subclass in more than one «IS-THE»-cluster, it is said that full multiple inheritance is carried out. In both cases entities of subclass inherit characteristics of all their parent classes.

In data modeling the special subspecies of two-level «IS-A»-hierarchies are entered. This is a categorization determining special «IS-THE»-relations between the union of superclass entity sets and subclass entity set (in this case named a category). To understand core of categorization it is necessary to present the content of the concept which scope is represented by subclass entity set. The first part of its logic expression is a disjunction of content conditions of superclass entity sets, and the second part is a condition to which should satisfy entities of these sets to be presented in a category. Such two-level «IS-A»-hierarchy we name «IS-THE»- categorization (or, simply, categorization).

In categorization definition it is emphasized that superclasses do not contain entities been abstractions of the same objects and are so various that they are not connected with the common ancestor. In other words, concepts of superclasses are incompatible in pairs. In that case the element of a category always inherits characteristics only one superclass, where the parental element connected to it contains. Therefore it is said that such two-level «IS-A»-hierarchies assume selective inheritance.

The definitions submitted in this article are based ERMM rules of data structurization and integrity constraint definition concerning representation of «IS-THE»- relations.

REFERENCES

1. Wang, P. (1994) From Inheritance Relation to Nonaxiomatic Logic. *International Journal of Approximate Reasoning*. 11. pp. 281-319. DOI: 10.1016/0888-613X(94)90021-3
2. Palomäki, J. & Kangassalo, H. (2012) That IS-IN Isn't IS-A: A Further Analysis of Taxonomic Links in Conceptual Modelling. In: Gutiérrez, C.R. (ed.) *Advances in Knowledge Representation*. Publisher: InTech. pp. 3-18. DOI: 10.5772/36484
3. Brachman, R.J. (1983) What IS-A Is and Isn't: An Analysis of Taxonomic Links in Semantic Networks. *IEEE Computer*. 16(10). pp. 30-36. DOI: 10.1109/MC.1983.1654194
4. Connolly, T. & Begg, C. (2000) *Базы данных. Проектирование, реализация и сопровождение. Теория и практика* [Database Systems. A Practical Approach to Design, Implementation, and Management]. Translated from English by R. Imamutdinova, K. Ptitsyn. 2nd ed. Moscow: Williams.
5. Teorey, T., Yang, D. & Fry, J. (1986) A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model. *Computing Surveys*. 18(2). pp. 197-222. DOI: 10.1145/7474.7475
6. Babanov, A.M. (2007) Semantic model "Entity – Relationship – Mapping". *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(1). pp. 77-91. (In Russian).
7. Voyshvillo, E.K. & Degtyarev, M.G. (1994) *Логика как часть теории познания и научной методологии (фундаментальный курс): в 2 кн.* [Logic as a part of the theory of knowledge and scientific methodology]. Moscow: Nauka.
8. Bacharov, V.A. & Markin, V.I. (1998) *Основа логики* [Foundation of logic]. Moscow: Infra-M.
9. Coronel C., Morris S. & Rob, P. (2013) *Database Systems: Design, Implementation, and Management*. 10th ed. Course Technology.
10. Connolly, T. & Begg, C. (2010) *Database Systems: A Practical Approach to Design, Implementation, and Management*. 4th ed. Addison-Wesley.
11. Elmasri, R. & Navathe, S. (2010) *Fundamentals of Database Systems*. 6th ed. Addison-Wesley.
12. Halpin, T. & Morgan, T. (2008) *Information Modeling and Relational Databases*. 2nd ed. Morgan Kaufman.
13. Ferraggine, V., Doorn, J. & Rivero, L. (2009) *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends*. Information Science Reference.
14. Smith, J. & Smith, D. (1977) Database Abstractions: Aggregation and Generalization. *ACM Transactions on Database Systems*. 2(2). pp. 105-133. DOI: 10.1145/320544.320546
15. Babanov, A.M. (2003) Theory of semantically significant mapping. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 280. pp. 239-248. (In Russian).