

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

УДК 004.652.8

DOI: 10.17223/19988605/36/8

А.М. Бабанов, Е.С. Квач**«IS-THE»-ОТНОШЕНИЯ В СЕМАНТИЧЕСКИХ МОДЕЛЯХ ДАННЫХ:
ПРЕДПОСЫЛКИ ИСПОЛЬЗОВАНИЯ И ПРАВИЛА ПОСТРОЕНИЯ
«IS-THE»-КЛАСТЕРОВ**

Статья продолжает серию публикаций, посвященных представлению «IS-THE»-отношений в модели данных «Сущность – Связь – Отображение». Она повествует о наиболее часто встречающейся при моделировании данных разновидности этих отношений – родовидовых отношениях – и об образуемых из них «IS-THE»-кластерах. Освещаются условия, при которых их необходимо использовать, и предлагаются две методики проектирования кластеров в семантических моделях данных – сверху вниз (методом специализации) и снизу вверх (методом генерализации).

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

Важным элементом семантики данных являются так называемые «IS-A»-отношения (отношения «ЕСТЬ-НЕКОТОРЫЙ»). Они традиционно применялись в системах управления знаниями [1]. С появлением в конце прошлого века семантической модели Enhanced/Extended Entity-Relationship Model (EERM) [2] эти отношения стало возможным представлять и при моделировании данных. Во время синтеза EERM была предпринята попытка переосмыслить этот феномен [3]. Основным принципом являлось следование безупречным логическим канонам при сохранении той функциональности, которая востребована в моделировании данных.

В ходе исследований было введено понятие «IS-THE»-отношения (отношения «ЕСТЬ-ЭТОТ»), более подходящее к моделированию данных. Рассмотрение частных случаев этого отношения привело к образованию производных понятий «IS-THE»-кластера и «IS-THE»-категоризации, являющихся аналогами известных структур данных – специализации и категоризации [4].

Настоящая статья посвящена подробному описанию «IS-THE»-кластеров. После краткого напоминания об этом понятии будет проведен анализ предпосылок их использования при моделировании данных. Во второй части статьи предлагаются две методики проектирования «IS-THE»-кластеров: сверху вниз (метод специализации) и снизу вверх (метод генерализации). Реализация этих методик предполагает определение как самих структур данных, так и ограничений целостности, накладываемых на них.

1. Понятие «IS-THE»-кластера

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). «IS-THE»-отношение – это бинарное 1:1-отношение, определяющее пару взаимнообратных функциональных «IS-THE»-отображений. В этом отношении (и отображениях) участвуют либо непосредственно объекты ПрО, либо их абстракции – данные [4]. Если известные трактовки «IS-A»-отношения связывают класс с объектом или класс с классом [5], 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 отоб-

ражения. Если рассматривать два понятия, одно из которых является обобщением другого, то «IS-THE»-отношение становится родовидовым отношением. «IS-THE»-отображение из объема родового понятия (суперкласса) в объем видового (подкласса) по-прежнему частичное функциональное. А вот «IS-THE»-отображение из подкласса в суперкласс стало полным функциональным.

«IS-THE»-кластер (или 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. 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. Прообразами таких отображений являются объекты суперкласса или их абстракции, а роль образов играют истинностные значения, произвольные значения и другие объекты (абстракции) соответственно.

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). Отличаются объекты подклассов только своими образами при этом отображении.

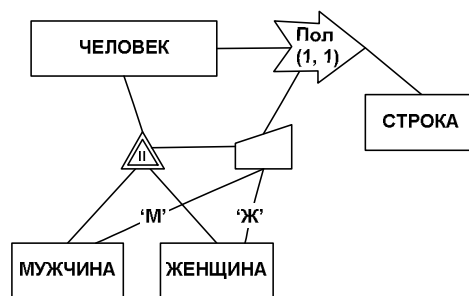


Рис. 1. «IS-THE»-кластер в расширенной графической нотации ERMM

Несколько слов о расширенной графической нотации ERMM. «IS-THE»-кластер обозначается с помощью треугольника, с вершиной которого соединяется суперкласс кластера, а с основанием – подклассы.

Основание деления обозначается в виде прямоугольной трапеции, от которой отходят как минимум три ребра: одно – к отображению (вершина в виде стрелки), на основании которого делится суперкласс, второе – к треугольнику «IS-THE»-кластера, третье ребро – к подклассу. В зависимости от количества подклассов кластера, выделяемых по данному основанию, будет меняться и количество ребер, инцидентных трапеции. Также каждое ребро к подклассу имеет пометку: в случае дихотомического деления – exists (образ существует) или not exists (образ отсутствует), в случае деления по видоизменению основания – значение образа отображения-основания для соответствующего подкласса.

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). Далее в этой статье речь пойдет в основном лишь о двухуровневом варианте графа, состоящем из одного «IS-THE»-кластера. В случае необходимости предлагаемые для этого случая процедуры следует многократно повторить, что приведет к построению произвольного графа кластеров.

Каждый «IS-THE»-кластер определяет полное функциональное «IS-THE»-отображение между объединением объемов видовых понятий и объемом родового понятия. Также действует полное функциональное «IS-THE»-отображение между каждым из классов видовых понятий и классом родового понятия. 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.

Характер отображения между объемом родового понятия и объединением объемов видовых понятий определяет ставшую традиционной типизацию «IS-THE»-кластеров. Если это отображение – полное функциональное, кластер является полным непересекающимся. Если оно – частичное функциональное, кластер является частичным непересекающимся. Если отображение между объемом родового понятия и объединением объемов видовых понятий – полное нефункциональное, кластер является полным пересекающимся. Если это отображение – неограниченное, кластер является частичным пересекающимся.

Ограничения целостности, связанные с кластерами, графически задаются следующим образом. Пересекающиеся кластеры помечаются двумя перекрещивающимися отрезками внутри треугольника, непересекающиеся – двумя параллельными отрезками. В случае полного кластера граница треугольника двойная, иначе – одинарная.

В нашем примере (см. рис. 1) кластер, который делит множество всех людей на мужчин и женщин, будет полным непересекающимся.

2. Предпосылки использования «IS-THE»-кластеров

Повествуя о специализациях (более широко распространенных аналогах «IS-THE»-кластеров), известные авторы выделяют многочисленные доводы в пользу их создания.

Дж. Смит и Д. Смит считают, что «польза от такой дисциплины структурирования заключается в следующем:

- могут быть эффективно интегрированы абстракции, имеющие отношение к различным пользователям базы данных, и они могут поддерживаться совместно;
- при различных видах эволюционных изменений может обеспечиваться стабильность моделей (иногда называемая независимостью данных);
- без значительной утраты интеллектуальной управляемости могут поддерживаться сильно структурированные модели;
- может быть разработан более семантический подход к проектированию баз данных;
- становятся возможными более эффективные реализации, поскольку может быть сделано больше предположений относительно структуры верхнего уровня» [6. С. 107].

Введение специализаций обеспечивает повышенную стабильность схемы в процессе развития базы данных. Изменения атрибутов или множеств связей множества сущностей суперкласса не приведут к изменению множеств сущностей подклассов. Так же как и изменения во множестве сущностей одного или нескольких подклассов не повлияют на множество сущностей суперкласса.

Схема может иметь слишком много атрибутов и множеств связей в рамках некоторого множества сущностей. В подобных случаях можно обеспечить понимание схемы путем введения специализации. Специализация позволяет вынести часть атрибутов и множеств связей в подклассы без их дублирования.

По мнению Т. Халпина и Т. Моргана, «основная причина использования специализации – это определение связанного с ней набора ограничений целостности. Кроме этого, благодаря введению специализации, можно показать, что атрибуты или множества связей уникальны или обязательны для конкретного подкласса» [7. С. 238].

Ну и, наконец, «специализация приводит к исключению большого количества неопределённых значений в атрибутах при занесении данных в БД» [8. С. 148].

При построении схемы данных люди используют два разных подхода – сверху вниз и снизу вверх. При первом подходе проектировщик выделяет в ПрО сначала сложные высокоуровневые типы объектов, а затем методами декомпозиции и специализации порождает более простые типы, представляющие интерес с точки зрения бизнес-процессов ПрО. Сторонники второго подхода идут от конкретных явлений ПрО, классифицируют их в типы данных, строят агрегаты знаков и типов и таким образом тоже порождают структуру данных ПрО.

Неудивительно, что и для кластеров предусмотрен как метод специализации, так и метод генерализации. Первый метод отталкивается от наличия структуры для суперкласса и порождает необходимый набор структур подклассов, второй, имея изначально структуры подклассов, строит структуру для суперкласса. Естественно, что для обоих случаев существуют свои предпосылки создания «IS-ТНЕ»-кластера.

Когда спроектирована структура только для суперкласса, определить потребность в его специализации и выделении отдельных структур для подклассов можно по следующим признакам:

- суперкласс «раздут» от множества опциональных атрибутов и связей;
- часть опциональных атрибутов и связей задается только при соответствии экземпляра класса одному условию из фиксированного набора условий;
- при несоответствии экземпляра суперкласса ни одному из условий задаются только обязательные атрибуты и связи.

Необходимым условием проектирования «IS-ТНЕ»-кластера снизу вверх является наличие в схеме данных классов, чьи понятия близки (в соответствии с данным ранее определением). И далее возникает вопрос, стоит ли объединить эти классы в один или несколько кластеров или оставить их несвязанными.

Создание «IS-ТНЕ»-кластеров в этом случае оправдано, если:

- в бизнес-процессах ПрО есть потребность в объединении экземпляров различных, но близких понятий;
- один и тот же объект ПрО представлен данными в нескольких классах, и необходимо восстанавливать единство этого объекта.

Следует отметить, что при построении семантической схемы не принимается никаких решений по повышению эффективности реализации БД. Это задача последующих этапов проектирования. А вот обеспечение соответствующих методик и механизмов полной информацией для принятия адекватного решения – задача этапа первоначальной фиксации семантики ПрО. Поэтому если в ней имеются «IS-ТНЕ»-отношения, которые могут понадобиться в моделируемых бизнес-процессах, смело проектируйте соответствующие структуры. Способы их реализации в БД могут быть самыми разными, включая и максимально эффективный.

3. Правила проведения специализаций

Возьмем в качестве примера проведения специализации ситуацию в академической ПрО. К моменту принятия решения о проектировании «IS-ТНЕ»-кластера текущих контрольных точек (КТ) был создан фрагмент ERM-схемы, приведенный на рис. 2.

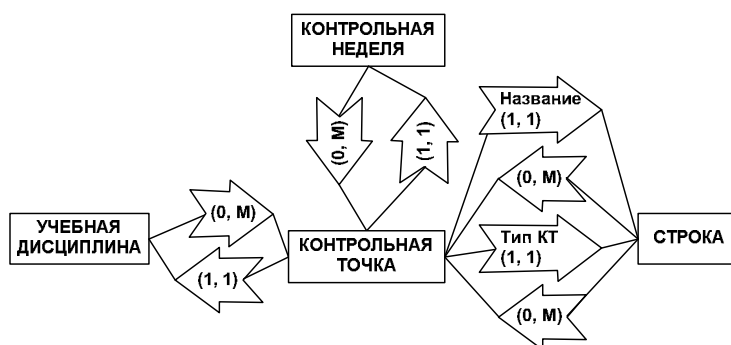


Рис. 2. ERM-схема до специализации

Эта схема отражает структуры данных, необходимые для планирования текущих КТ по учебным дисциплинам. Эти КТ проводятся в специально отведенные для этого контрольные недели. При проектировании структур под необходимую для проведения КТ информацию было обнаружено, что КТ одного типа нуждаются в одних структурных компонентах, КТ второго типа – в других, а КТ третьего типа – в третьих. Так, для лабораторной работы требуется определить задание на работу, для контрольной работы – формулировку запроса на поиск данных, а для тестов – указать до 10 тестовых заданий с текстом вопроса и альтернативами ответов к каждому заданию.

Если все эти потребности реализовать как отображения класса КОНТРОЛЬНАЯ ТОЧКА, будет много редко используемых структур, по большей части заполняемых в БД неопределенными значениями. Налицо первые два признака, указывающие на необходимость специализации класса КОНТРОЛЬНАЯ ТОЧКА.

Если при проектировании ERM-схемы Про принято решение, что специализацию использовать необходимо, то при ее проведении проектировщик может придерживаться следующей методики.

1. Выбрать класс, который будет являться суперклассом «IS-THE»-кластера, определить все его отображения в другие классы.

В нашем примере суперклассом является класс КОНТРОЛЬНАЯ ТОЧКА. Кроме показанных на рис. 2 полностью определенных отображений Учебная дисциплина, Контрольная неделя, Название и Тип КТ у этого класса есть опциональные отображения Задание на лабораторную работу, Формулировка запроса и Тестовое задание (для упрощения рисунка они не показаны).

2. Сгруппировать отображения по совпадению их областей определения: в одну группу попадут отображения, определенные на всем суперклассе, остальные группы отображений будут определены на некоторых подмножествах суперкласса.

Некоторые авторы считают, что критерием разбиения атрибутов на группы является обязательность значения для элементов подмножества. Но ведь до какого-то момента у объекта может не быть образа, позже он может появиться. Таким образом, атрибут остается опциональным, но приписывается к группе, в подмножество которой входит указанный объект. Решающими становятся потенциальная возможность наличия образа для объектов подмножества и отсутствие такой возможности у объектов, не принадлежащих этому подмножеству. А это фактически и есть нахождение области определения отображения (ООО).

Полностью определенные отображения Учебная дисциплина, Контрольная неделя, Название и Тип КТ естественно сформируют группу отображений с ООО в виде всего суперкласса КОНТРОЛЬНАЯ ТОЧКА. ООО Задание на лабораторную работу – множество КТ типа сдача лабораторной работы; ООО Формулировка запроса – множество КТ типа написание контрольной работы; ООО Тестовое задание – множество КТ типа сдача теста.

3. Указать для каждого подмножества вид выделения (дихотомическое или по видоизменению основания). Определить, экземпляры какого отображения суперкласса порождают каждое подмножество. Это отображение является дискриминирующим для создаваемого «IS-THE»-кластера. В случае деления по видоизменению основания каждому подмножеству сопоставить характерный для него образ дискриминирующего отображения.

Подмножества могут образовываться по принципу дихотомии (у объектов одного подмножества образ есть, у объектов другого – нет) или по совпадению образов. В первом случае выделяется один (когда подмножество не имеющих образа объектов не создается) или два подмножества, во втором – несколько. При наличии в кластере одного подкласса он фактически представляет операцию ограничения понятия.

Если все подмножества, сформированные на втором шаге, порождаются одним и тем же отображением, продолжаем выполнять шаги методики, в результате будет построен один «IS-THE»-кластер. В противном случае для представления всех имеющихся «IS-THE»-отношений необходимо построить несколько «IS-THE»-кластеров, расположенных друг относительно друга либо горизонтально (у них один и тот же суперкласс), либо вертикально (подкласс одного кластера является суперклассом другого), либо горизонтально и вертикально одновременно. Обсуждение этого случая отложим до конца методики.

В нашем примере все подмножества выделяются в соответствии с одним дискриминирующим отображением Тип КТ. Значит, в ходе проектирования будет образован один «IS-THE»-кластер. Трём сформированным подмножествам соответствуют образы дискриминирующего отображения Тип КТ в виде строк «ЛР», «КР», «ТЕСТ».

4. У суперкласса остаются только отображения его группы. Каждое выделенное на втором шаге подмножество становится подклассом построенного «IS-THE»-кластера. Этот подкласс является ООО всех отображений соответствующей группы.

Отображения Учебная дисциплина, Контрольная неделя, Название и Тип КТ остаются на уровне суперкласса КОНТРОЛЬНАЯ ТОЧКА. Подкласс СДАЧА ЛАБОРАТОРНОЙ РАБОТЫ будет определять отображение Задание на лабораторную работу, подкласс НАПИСАНИЕ КОНТРОЛЬНОЙ РАБОТЫ – отображение Формулировка запроса, а подкласс СДАЧА ТЕСТА – отображение Тестовое задание.

5. Определить ограничения целостности для построенного «IS-THE»-кластера и отображений подклассов.

Если дискриминирующее отображение – полностью определенное и в кластере сформированы подклассы для всех возможных его образов, кластер будет полным; если это отображение к тому же функционально, кластер будет полным непересекающимся, если не функционально, кластер будет полным пересекающимся. Если дискриминирующее отображение – не полностью определенное или у него есть образы без подклассов, кластер будет частичным; если это отображение к тому же функционально, кластер будет частичным непересекающимся. При нефункциональном дискриминирующем отображении в ряде случаев также возможно непересечение подклассов, но чаще кластер будет пересекающимся.

В примере отображение Тип КТ – полное функциональное, и все его образы порождают подклассы, следовательно, получившийся кластер полный и непересекающийся, что и отражено в ERM-схеме на рис. 3. Как и ранее, в схеме не нашли отражения отображения подклассов. Но следует отметить, что опциональные на уровне суперкласса КОНТРОЛЬНАЯ ТОЧКА отображения Задание на лабораторную работу, Формулировка запроса и Тестовое задание стали обязательными на уровне его подклассов.

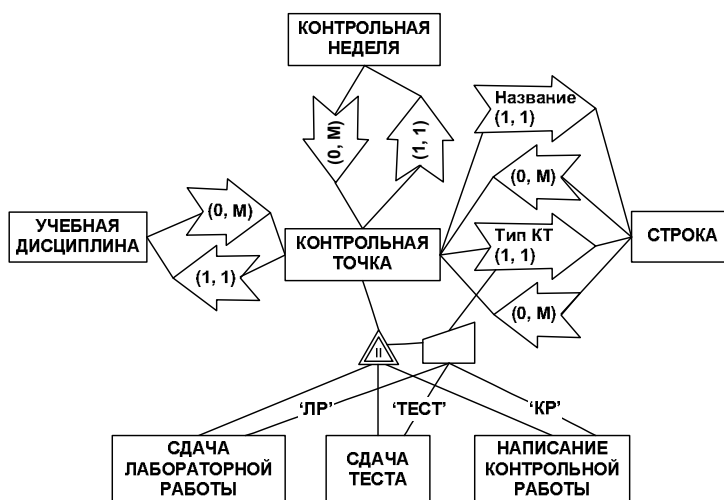


Рис. 3. ERM-схема после специализации

Ситуацию с несколькими различными основаниями выделения подмножеств суперкласса поясним на примере различных ролей, которые играют люди в академической ПроО. В этой ПроО все люди делятся на тех, кто ведет занятия, тех, кто посещает эти занятия, и тех, кто делает и то и другое. Первых назовем преподавателями, вторых – студентами, а третьих – преподающими студентами.

При проведении специализации всех людей вуза (суперкласс ЧЕЛОВЕК) в первую группу отображений (уровень суперкласса) попадут такие общие для всех людей отображения, как Фамилия, Имя, Дата рождения, Пол и т.д. Подмножество преподавателей будет ООО таких отображений, как Дата принятия на работу, Должность, Степень, Звание, Педагогический стаж. Подмножество студентов будет ООО таких отображений, как Дата поступления, Академическая специальность, Группа, Средний кре-

дитный балл. Подмножество преподающих студентов специфических отображений не имеет (хотя, в принципе, они могли бы быть) и представляет собой пересечение подмножеств преподавателей и студентов.

Подмножество преподавателей выделено методом дихотомии на основании наличия у людей образа отображения Дата принятия на работу. Подмножество студентов выделено методом дихотомии на основании наличия у людей образа отображения Дата поступления. Объекты третьего подмножества имеют образы обоих отображений. Различия оснований выделения говорят о том, что нельзя обойтись одним «IS-THE»-кластером.

В этом случае необходимо построить четыре кластера, каждый из которых фактически будет представлять ограничение: понятия ЧЕЛОВЕКА – до понятия ПРЕПОДАВАТЕЛЯ, понятия ЧЕЛОВЕКА – до понятия СТУДЕНТА, понятия ПРЕПОДАВАТЕЛЯ – до понятия ПРЕПОДАЮЩЕГО СТУДЕНТА и понятия СТУДЕНТА – до понятия ПРЕПОДАЮЩЕГО СТУДЕНТА. Полученная «ромбовидная» структура отражает граф кластеров, так как подкласс ПРЕПОДАЮЩИЙ СТУДЕНТ является подклассом двух кластеров.

4. Правила проведения генерализаций

Пример для проведения генерализации возьмем из ПрО функционирования поликлиники. В схеме данных этой ПрО, помимо прочего, есть класс ПАЦИЕНТ с отображениями Фамилия, Пол, Регистрация в поликлинике, Консультация у врача и класс ВРАЧ с отображениями Фамилия, Пол, Врачебная специальность, Консультация пациента (рис. 4). Для большинства бизнес-процессов поликлиники потребности в наличии «IS-THE»-отношения между сущностями типов ПАЦИЕНТ и ВРАЧ не возникает. Предположим, что иногда встает задача определения тех врачей, к которым в случае болезни обращаются сами врачи. Для решения этой задачи необходимо связать сущности типов ПАЦИЕНТ и ВРАЧ «IS-THE»-отношением. Можно предусмотреть для этого частное решение в виде соответствующего бинарного множества связей. Но если аналогичные задачи приходится решать еще и с классами ОБСЛУЖИВАЮЩИЙ ПЕРСОНАЛ и АДМИНИСТРАТОР, имеет смысл обобщить все объекты этих типов до класса ЧЕЛОВЕК и представить исходные классы как его подклассы. Это и позволяет проделать генерализацию. При этом мы восстанавливаем единство объектов ПрО, разбитых на свои многочисленные абстракции-сущности разных типов. Другим мотивом генерализации в нашем примере может быть потребность в предоставлении общего списка сотрудников поликлиники независимо от должности.

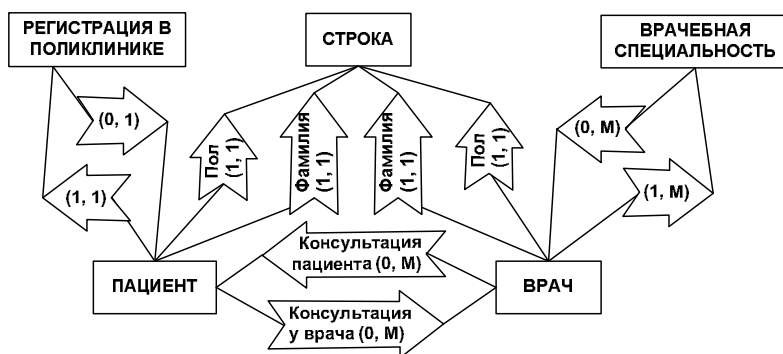


Рис. 4. ERM-схема до генерализации

Если при проектировании ERM-схемы ПрО принято решение о необходимости обобщения некоторых классов близких понятий до их общего класса-предка, то для проведения генерализации проектировщик может придерживаться следующей методики.

1. Найти ближайшее общее понятие-предок. В нашем случае таким обобщающим пациентов и врачей понятием будет понятие ЧЕЛОВЕК.
2. Определить видовое отличие каждого исходного понятия-потомка и участвующее в нем отображение.

Если во всех видовых отличиях фигурирует одно и то же отображение и в случае дихотомии для одного понятия-потомка его образ присутствует, а для другого – отсутствует, а в случае видоизменения основания понятия-потомки отличаются образами, это отображение является дискриминирующим и создается один «IS-THE»-кластер. Для этого варианта продолжаем выполнение методики. Если у понятий-потомков в видовых отличиях участвуют разные отображения, для представления всех имеющихся «IS-THE»-отношений необходимо построить несколько «IS-THE»-кластеров, расположенных друг относительно друга либо горизонтально, либо вертикально, либо горизонтально и вертикально одновременно. При этом рассуждать надо примерно так же, как и в аналогичном случае методики специализации.

Характерным для пациентов признаком является наличие регистрации в поликлинике, отличительная особенность врачей – одна или несколько врачебных специальностей. Первое понятие-потомок образуется методом дихотомии с помощью отображения Регистрация в поликлинике, второе – аналогичным методом с использованием отображения Врачебная специальность. Соответственно, в нашем случае необходимо построить два «IS-THE»-кластера суперкласса ЧЕЛОВЕК, в каждом из которых по одному подклассу – ПАЦИЕНТ и ВРАЧ.

3. Если у дискриминирующего отображения не все варианты образов соответствуют исходным понятиям-потомкам, необходимо подумать, не добавить ли в схему соответствующие подклассы.

Для нашего примера отсутствующие подклассы людей, не являющихся пациентами, и людей, не являющихся врачами, с одной стороны, не представляют интереса для задач ПрО, с другой стороны, легко получаются с помощью операции разности.

4. Создать суперкласс и перенести на уровень суперкласса все общие отображения подклассов, включая дискриминирующее отображение.

Суперкласс ЧЕЛОВЕК будет являться ООО отображений Фамилия, Пол, Регистрация в поликлинике и Врачебная специальность. Подкласс ПАЦИЕНТ будет определять отображение Консультация врача, а подкласс ВРАЧ – отображение Консультация пациента (см. рис. 5).

5. Определить ограничения целостности для построенного «IS-THE»-кластера и для отображений суперкласса в соответствии с шагом 5 методики специализации.

Кластер, построенный методом дихотомии с отсутствующим одним из подклассов, всегда частичный и непересекающийся. В нашем примере такими будут оба построенных «IS-THE»-кластера (рис. 5). Дискриминирующие отображения Регистрация в поликлинике и Врачебная специальность, бывшие обязательными для подклассов, становятся опциональными при переносе их на уровень суперкласса.

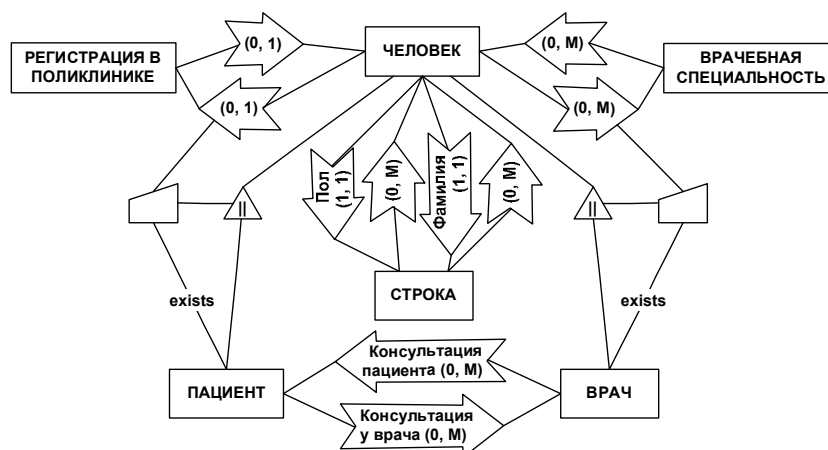


Рис. 5. ERM-схема после генерализации

Мы рассмотрели две методики создания «IS-THE»-кластеров – метод специализации, когда имеется суперкласс и создаются подклассы, и метод генерализации, при котором, наоборот, исходными являются подклассы и производится суперкласс. Редко, но встречается ситуация, когда все классы уже спроектированы и необходимо создать лишь кластер. Ее можно рассматривать как частный случай специализации или генерализации и применить одну из приведенных методик.

Заключение

Семантическая методика проектирования схем БД предусматривает на первом этапе максимально полную формализацию информации о ПрО в рамках той или иной семантической модели. Без фиксации в должной степени «IS-THE»-отношений такое представление вряд ли может обойтись. Вот почему любая семантическая модель данных должна предусматривать соответствующие средства для определения структур и ограничений целостности. ERM-модель в этом плане не является исключением. В ней помимо обычных 1:1-множеств связей для представления «IS-THE»-отношений имеются такие структуры, как «IS-THE»-кластеры и «IS-THE»-категоризации.

В статье подробно обсуждаются вопросы проектирования «IS-THE»-кластеров: базовые определения, предпосылки и методики создания. Процедуры проектирования снабжены подробными примерами. Материал статьи позволит проектировщикам схем БД разобраться с такими явлениями ПрО, как иерархии обобщения, и научит использовать соответствующие правила и инструменты структуризации, что приведет в конечном счете к более эффективным и адекватным схемам БД.

ЛИТЕРАТУРА

1. 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.
2. 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.
3. Бабанов А.М. Семантическая модель «Сущность – Связь – Отображение» // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2007. № 1. С. 77–91.
4. Бабанов А.М., Квач Е.С. «IS-THE»-отношения в семантических моделях данных: основные понятия и разновидности // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2016. № 1(34). С. 69–78.
5. 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.
6. Smith J., Smith D. Database Abstractions: Aggregation and Generalization // ACM Transactions on Database Systems. 1977. V. 2, No. 2. P. 105–133.
7. Halpin T. and Morgan T. Information Modeling and Relational Databases. Second Edition. Morgan Kaufman, 2008.
8. Coronel C., Morris S., Rob P. Database Systems: Design, Implementation, and Management. 10th edition. Course Technology. 2013.

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

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

Томский государственный университет

Поступила в редакцию 29 января 2016 г.

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

«IS-THE»-relations in semantic data models: prerequisites and construction rules of «IS-THE»-clusters.

Key words: «IS-THE»-relation; «IS-THE»-cluster; specialization; generalization; semantic data model; ERM-model.

DOI: 10.17223/19988605/36/8

«IS-THE»-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.

At data schema construction people use two different approaches – from top to down and from below upwards. So it is provided both a method of specialization, and a method of generalization for clusters. The first method makes a start from presence of superclass structure and generates a necessary set of subclass structures, the second, having initially structures of subclasses, builds structure for a superclass.

When only the structure for a superclass was designed, one can determine need for its specialization and detachment of separate structures for subclasses on the basis of the following signs:

- superclass is swelled from optional attributes and relationships;
- part of optional attributes and relationships are fixed only if superclass instance satisfies one condition from the fixed set of conditions;
- if superclass instance does not satisfy any condition from the fixed set of conditions superclass mandatory attributes and relationships are fixed only.

Necessary condition for «IS-THE»-cluster designing from below upwards is presence of classes, whose concepts are close. And further there is a question whether it is necessary to unite these classes in one or several clusters or to leave their untied.

«IS-THE»-clusters creation in this case is justified, if:

- any business process needs for union of instances of various, but close concepts;
- an object of application domain is represented by data in several classes, and it is necessary to restore unity of this object.

If at ERM-schema designing one makes a resolve that specialization use is necessary, the designer can adhere to the following technique.

1. To choose a class which will be a superclass of «IS-THE»-cluster, to determine all its mappings to other classes.
2. To group mappings according to their domain: one group is formed from mappings defined on the superclass, other groups of mappings will be defined on some subsets of the superclass.
3. To specify for each subset a kind of detachment (dichotomizing or by basis modification). To determine mapping, which instances form each subset. This mapping is discriminating for created «IS-THE»-cluster. In case of detachment by basis modification should establish a correspondence between each subset and discriminating mapping image.
4. The superclass has still only mappings of its group. Each subset distinguished on the second step is a subclass of constructed «IS-THE»-cluster. This subclass is a domain of all mappings of corresponding group.
5. To define integrity constraints for constructed «IS-THE»-cluster and subclass mappings.

If at ERM-schema designing one makes a resolve that generalization use is necessary, the designer can adhere to the following technique.

1. To find the nearest general concept-ancestor to given classes with close concepts.
2. To determine specific difference of each initial concept-descendant and mapping participating in it.

If the same mapping appears in all specific differences, and in the case of dichotomizing for one concept-descendant its image is present, and for another concept-descendant is absent, and in case of basis modification concepts-descendants differ in the images, this mapping is discriminating, and one «IS-THE»-cluster is created.

3. If not all variants of discriminating mapping images correspond to initial concepts-descendants, it is necessary to think whether to add to the schema corresponding subclasses.

4. To create a superclass and to transfer on a level of the superclass all common mappings of subclasses, including discriminating mapping.

5. To define integrity constraints for constructed «IS-THE»-cluster and superclass mappings.

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

REFERENCES

1. Brachman, R.J. (1983) What IS-A Is and Isn't: An Analysis of Taxonomic Links in Semantic Networks. *IEEE Computer*. 16(10). pp. 3036. DOI: 10.1109/MC.1983.1654194
2. 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
3. Babanov, A. M. (2007) Semanticheskaya model' "Sushchnost' – Svyaz' – Otobrazhenie" [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.
4. Babanov, A. M. & Kvach, E.S. (2016) "IS-THE"-relations in semantic data models: basic notions and variants. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(34). pp. 6978. (In Russian). DOI: 10.17223/19988605/34/8
5. Palomäki, J. & Kangassalo, H. (2012) That IS-IN Isn't IS-A: A Further Analysis of Taxonomic Links in Conceptual Modelling. *Advances in Knowledge Representation*. pp. 3-18. DOI: 10.5772/36484
6. 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
7. Halpin, T. & Morgan, T. (2008) *Information Modeling and Relational Databases*. 2nd ed. Morgan Kaufman.
8. Coronel, C., Morris, S. & Rob, P. (2013) *Database Systems: Design, Implementation, and Management*. 10th ed. Course Technology.