# БАЗЫ ДАННЫХ
# DATABASE

УДК 004.652.8 DOI: 10.17587/it.26.460-471

**А. Н. Родионов,** д-р техн. наук, вед. науч. сотр., ran@newmail.ru,
Вычислительный центр ДВО РАН, г. Хабаровск

# Концептуальное и логическое моделирование данных: обнаружение и конфигурирование *R*-типов и *n*-арных отношений в предметных областях

*Решаются задачи обнаружения и конфигурирования комплексов высокоуровневых отношений баз данных (отношений с арностью 3 и выше) и представления части последних посредством R-типов. R-тип — одна из категорий синтетических объектных типов, экземпляры которых ассоциируются с отношениями, возникающими между объектами. Такие объекты-отношения обладают собственными свойствами и в состоянии вступать во взаимодействие с другими объектами. Показано, что многозначные функциональные зависимости, которые могут присутствовать среди атрибутов внешних ключей заголовочных документальных типов, однозначно идентифицируют и сами n-арные отношения, и соответствующие им R-типы. Вводится понятие комплекса бинарных n-арных отношений-типов, которые объединяет наличие общих исходных взаимодействующих типов. На основе предлагаемой p/a-характеристики бинарных отношений разрабатывается метод установления первичных (подлежащих хранению) и производных (выводимых) их первичных отношений, образующих комплекс.*

***Ключевые слова:*** *высокоуровневые отношения, R-типы, p/a-характеристика, многозначные функциональные зависимости*

## Введение

Выявление типов бинарных отношений, в которых могут состоять, будут состоять и состоят объекты предметных областей, — тривиальная, рутинная процедура концептуального и логического моделирования данных. Проблемы моделирования возникают, когда в отношения оказываются вовлеченными одновременно более двух объектов. (На отношения с арностью 3 и выше будем ссылаться далее как на *n*-арные (высокоуровневые) отношения, где *n* — степень отношения.) Перечислим и раскроем суть каждой из проблем.

В первую очередь, *n*-арные отношения должны быть обнаружены в предметной области или, если это приемлемо, логически выведены из уже ранее установленных бинарных отношений. Существующие технологии и отдельные методы концептуального моделирования в состоянии описать структуру и ограничения кардинальности отношений произвольной арности [1—3], но не включают методы идентификации последних.

Вторая проблема — проблема аномальности обновлений высокоуровневых отношений — достаточно хорошо изучена. Специально разработанная и непрерывно совершенствующаяся теория нормальных форм [4—8] нацелена на поиск и исключение аномалий в предварительно полученных отношениях, но ограничивается зачастую формальными аспектами, делая акцент на достоверность и постоянство имеющейся в распоряжении аналитиков семантики, что не всегда соответствует действительности. Семантика отношений, которая выражается в форматах функциональных зависимостей и кардинальности связей, иногда бывает подвержена изменениям, а это, в свою очередь, требует получения таких моделирующих конструкций, которые были бы нечувствительны к этим изменениям. Для обеспечения устойчивости логических схем было предложено несколько решений — от использования различных систем классификации отношений [9—12] до специально разработанных шаблонов [3, 10, 11], "предвосхищающих" потенциальные изменения.

Отношения между отношениями составляют суть следующей актуальной проблемы моделирования данных. Показательна работа [13], в которой можно найти несколько специализированных паттернов моделирования, по-

зволяющих представить подобные отношения на концептуальном уровне. Правда, при этом не указываются и не унифицируются формальные условия, приводящие к этим паттернам.

Отправной точкой для решения всех перечисленных задач может стать обнаружение и (или) вывод *n*-арных отношений из бинарных отношений. Некоторые из высокоуровневых отношений находятся достаточно просто. В предметных областях присутствует специфический класс высокоуровневых отношений, экземпляры которого естественным образом воспринимаются в качестве объектных типов. Некоторые из таких типов у всех "на слуху". Например, спортивные матчи — ассоциации между двумя командами, бухгалтерские проводки, связывающие два счета, смс-сообщения, включающие источники и приемники сообщений. Список может продолжен.

Одной из первых работ, в которой *n*-арные отношения трактовались как сущности, была работа [14]. В современных технологиях моделирования подобные отношения называются по-разному. Например, в ORM они именуется как "*objectified associations"* [13], а в Onto UML — как "*relators"* [15]. Из соображения лаконичности будем пользоваться последней нотацией.

Формально *relator* (*R*-тип, рилейтор) — это "слабая сущность", каждый экземпляр которой идентифицируется составным суррогатным ключом (рис. 1).

Как видно из схемы (рис. 1), ассоциация с кардинальностью *M:M* может иметь две альтернативные формы представления: либо в виде "слабой сущности" (*W*-типа) с составным суррогатным ключом<Id_1*, Id_2*>, либо в виде "*relator*" с единственным ключевым суррогатным атрибутомId_r*. В последнем случае уникальность пары <Id_1, Id_2> в составе *R*-типа становится обязательным ограничением целостности. (Надстрочный индекс, которому на рис. 1 присвоен символ "*", указывает на ключевой атрибут.)

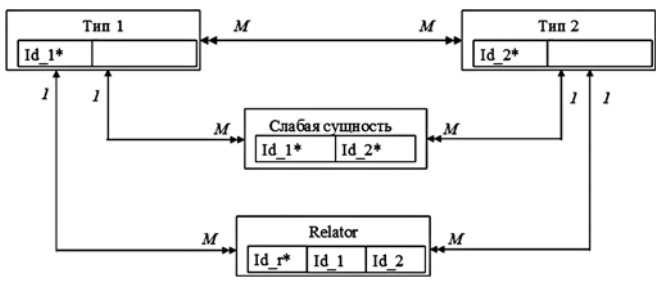


**Рис. 1. Источники и структура *R*-типа**

Поскольку *R*-типы и *W*-типы — две альтернативные формы представления высокоуровневых ассоциаций, то идентификация *R*-типа будет автоматически приводить к обнаружению искомой ассоциации.

Альтернативный, дополняющий подход к обнаружению *n*-арных отношений может строиться на основе анализа грамматических конструкций естественных языков [9, 12, 16—18], в частности, их глагольных форм. В статье также используются модальные и немодальные формы глаголов для классификации высокоуровневых отношений в целях локализации исходных и производных классов отношений.

Настоящая работа нацелена на решение нескольких тесно связанных друг с другом задач, среди которых: обнаружение и конфигурирование отношений с арностью 3 и выше; установление признаков, на основании которых *n*-арные отношение должны быть приведены к объектным типам (*R*-типам); разработка рамочной основы для построения, по возможности, полной и устойчивой системы отношений в моделируемой предметной области.

Статья организована следующим образом. В первой части приводятся задействованные в работе понятия и нотации, рассматривается "природа"*R*-типов и излагается процедура идентификации последних, одновременно приводящая к идентификации и *n*-арных отношений. Формальные схемы отношений, приводящие к получению 3-компонентных рилейторов, и возникающие при этом неопределенности исследуются во второй части статьи. В третьем разделе работы на примере конфигурирования 4-компонентных отношений раскрывается суть и значимость *p/a* характеристики бинарных связей между объектами для построения полной системы отношений. В четвертой части статьи рассматривается один из вариантов подсхем организации данных, увязывающий *R*-типы с подтипами сущностных типов, образующих кластер.

## 1. Используемая нотация и процедура идентификации простейших рилейторных типов

Претендентами на рилейторы, если исходить из изложенного выше, могут быть только слабые сущности. Рассмотрим формальные ассоциации, которые могут возникать между типами в моделях данных, и выделим те из них, которые могут приводить к появлению *R*-типов.

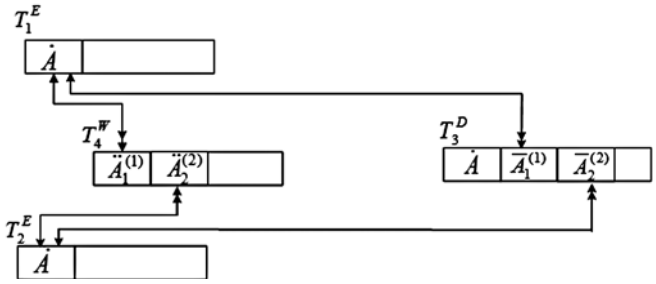

**Рис. 2. Виды ассоциаций между типами и используемая нотация**

Но прежде всего, определимся с терминами, обозначениями и моделирующими конструкциями, которыми далее будем пользоваться.

Среди множества типов, входящих в состав баз данных, будем различать сущностные типы, документальные заголовочные типы и слабые сущности. (Документальный заголовочный тип — обязательный корневой тип документального кластера, атрибуты которого, согласно работе [19], соотносятся с однозначными реквизитами моделируемого документа.) В качестве базовой моделирующей конструкции остановимся на диаграмме базы данных (рис. 2), которую, в отличие от концептуальной и логической схем, несущих схожую функцию, отличает полный набор элементов, достаточных и для представления объектов и отношений реального мира, и для однозначного преобразования последних в физические структуры хранения.

Будем также полагать, что, по умолчанию, каждый тип — это одновременно и переменная отношения в реляционной модели, в которой для задания типа достаточно перечислить имена атрибутов, входящих в его состав, и области определения (домены) этих атрибутов. (Ссылки на домены как несущественные для задач, которые решаются в работе, далее будем опускать, ограничившись только перечнями атрибутов.) Соответственно реализации типа — не что иное, как тело отношения (или просто отношение), представленное совокупностью кортежей.

В обозначении типов, позволяющих их однозначно идентифицировать в диаграмме базы данных, будем придерживаться следующей символики. Пусть $T_n^{Z_m}$ ссылается на некоторый тип. Верхний индекс $Z$, там, где это требуется в работе, принимает одно из трех значений: $E$ — сущностный тип ($E$-тип), $D$ — документальный тип ($D$-тип), $W$ — слабая сущность ($W$-тип), $R$ — рилейтер ($R$-тип). Нижний индекс $n$ — порядковый номер этого типа. В некоторых примерах $n$ заменяется на название типа. Еще один индекс — $m$ в $Z_m$ указывает категорию сущностного типа: *Prototype, Sample* или *Instance*.

Согласно работе [20] в базах данных экземпляры сущностных типов могут нести различную функциональную нагрузку, что выражается в их соответствующей функциональной дифференциации. Экземпляры *Prorotype*-типов — это прообразы, модели сущностей. В свою очередь, экземпляры реальных сущностей распределяются среди *Sample*- и *Instance*-типов. Основное различие между экземплярами *Sample*- и *Instance*-типов — принятый в конкретной предметной области способ учета этих экземпляров. *Instance* образован единичными экземплярами, каждый из которых ассоциируется с конкретной сущностью. В отличие от *Instance*-экземпляров, экземпляры *Sample*-типов ссылаются одновременно на несколько конкретных сущностей, характеризующихся одинаковыми значениями их подлинных свойств. Согласно последней особенности *Sample*-тип должен включать обязательный количественный атрибут.

Сущностные категории *Prototype, Sample* и *Instance* связаны друг с другом предопределенными отношениями и совместно образуют сущностный кластер (рис. 3).

В тексте статьи ссылки на ассоциации, связывающие типы, даются в виде: $T_i \leftrightarrow T_j$ — связь с кардинальностью 1:1, $T_i \leftarrow\!\!\!\rightarrow T_j$ — связь с кардинальностью 1:M, $T_i \leftarrow\!\!\!\rightarrow T_j$ — связь с кардинальностью M:M.

Отношения между типами представим в виде отношений между атрибутами простых первичных ключей $\dot{A}$, атрибутами составных первичных ключей $\ddot{A}$ и атрибутами внешних ключей $\bar{A}$. За каждым конкретным атрибутом "скрывается" множество, образованное значениями, принимаемыми этим атрибутом. Привяжем атрибуты к типам с помощью индексов. В итоге обозначение атрибута примет следующий вид: $A_m^{n(n)}$.

Принадлежность атрибута конкретному (собственному) типу отражена посредством

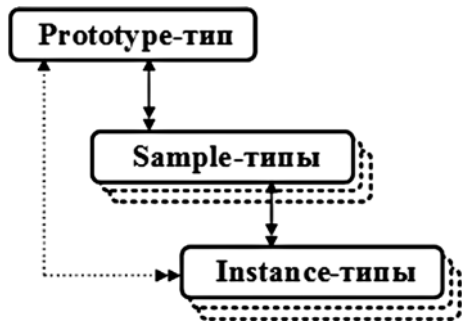


**Рис. 3. Конфигурация унифицированного сущностного кластера**

первого компонента верхнего индекса. Другой компонент верхнего индекса (заключенный в круглые скобки)указывает на родительский тип и, по очевидной причине, используется только для $\dot{A}$- и $\bar{A}$- атрибутов. Порядковые номера атрибутов внешнего и составного первичного ключей в типе задаются с помощью нижнего индекса $m$.

Опишем области определения перечисленных разновидностей ключевых атрибутов. Условимся, что в целом соответствует повседневной практике, что для любого $\dot{A}$ справедливо: $\dot{A} \subset \mathbb{Z}$. Другими словами, $\dot{A}$ в нашем случае — это всегда суррогатный ключ.

Для атрибутов внешних ключей, а также отдельных атрибутов, входящих в состав составных ключей, действуют следующие ограничения:

$$\bar{A}_m^{n(n)} \subseteq \dot{A}^n; \quad (1)$$

$$\ddot{A}_m^{n(k\neq n)} \subseteq \dot{A}^{k\neq n}. \quad (2)$$

Первое ограничение (1) определяет множество значений, которые может принимать $\bar{A}_m^{n(n)}$. При этом ссылки на собственный и родительский типы $n(n)$, ввиду допустимости рекурсий в сущностных типах, могут совпадать.

Для атрибутов составных первичных ключей, входящих в состав слабых сущностей, их области определения также формируются на основании множества значений, принимаемых первичными ключами сущностных типов. Форма записи ограничения (2) для $\ddot{A}$ несколько отличается от ограничения (1). Запись $k \neq n$ в обозначении верхнего индекса свидетельствует о недопустимости присутствия рекурсий в $W$-типах.

По определению, составной первичный ключ (СПК), который, в том числе, является отличительным признаком $W$-типа, образован несколькими упорядоченными атрибутами. Область его значений $\ddot{A}$ — ни что иное, как подмножество множества упорядоченных кортежей, являющихся результатом декартова произведения определенной совокупности $\bar{A}$- множеств. В общем виде:

$$\ddot{A}^n \subseteq \dot{A}_1^{k\neq n} \times \ldots \times \dot{A}_M^{k\neq n} = \left\{ \left\langle a_1, \ldots, a_m, \ldots, a_M \right\rangle \right\}, \quad (3)$$

где $M$ —степень составного первичного ключа, $M \geqslant 2$; $\dot{A}_m^{k\neq n}$ — первичный ключ $k$-го сущностного типа; $a_m$ — значение $m$-го атрибута СПК.

Основная особенность ассоциаций, возникающих между сущностными и документальными типами, показанными на рис. 2, заключается в том, что тело отношения, соответствующее $T_3^D$, может определяться не только на основании соотношения (1).

Пусть $\overline{\overline{A^n}}$, по аналогии с СПК, — составной внешний ключ (СВК), который так же, как и СПК, представляет собой подмножество множества упорядоченных кортежей и определяется согласно соотношению (3). Предположим, что в формировании $\overline{\overline{A^n}}$ участвуют те же самые атрибуты, что образуют СПК некоторой слабой сущности $T^W$. В предметных областях нередки ситуации, когда СПК и СВК связаны друг с другом соотношением:

$$\overline{\overline{A^n}} \subseteq \ddot{A}^{(k\neq n)}. \quad (4)$$

Записанное ограничение иллюстрируется примерами (рис. 4). Из отношения, соответствующего $T_3^D$, вычеркнуты кортежи, которых нет в $T_4^W$.

Действие ограничения (4) ведет к появлению аномалий вставки и модификации кортежей в $T_3^D$. Действительно, чтобы добавить новый кортеж в $T_3^D$, потребуется первоначально проверить наличие аналогичного кортежа в $T_4^W$. В свою очередь, удаление кортежа в $T_4^W$ должно сопровождаться удалением группы кортежей в $T_3^D$, содержащих те же значения упорядоченных атрибутов внешних ключей, что и значения ключевых атрибутов в удаляемом кортеже $T_4^W$.

Исключить указанные аномалии можно, преобразовав исходную схему (см. рис. 2). В представленном варианте (рис. 5) "слабая сущность" (рис. 5, *б*) заменена на рилейторный тип (рис. 5, *в*), а вместо связей $T_1^E \longleftrightarrow\!\!\!\!\rightarrow T_3^D$ и $T_2^E \longleftrightarrow\!\!\!\!\rightarrow T_3^D$ задействована связь $T^R \longleftrightarrow\!\!\!\!\rightarrow T_3^D$.

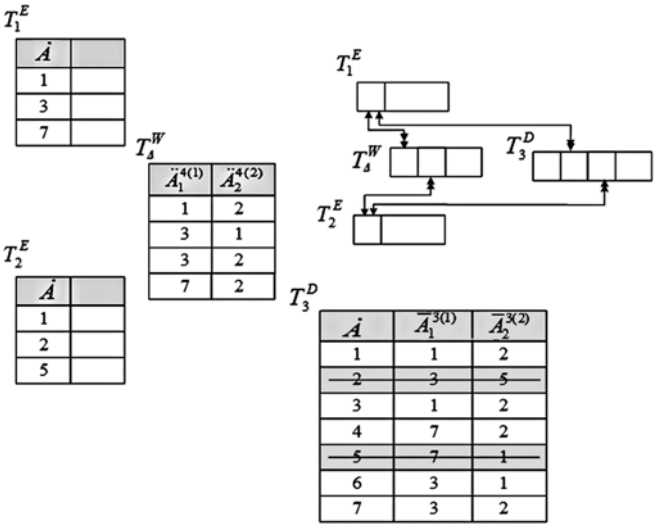


**Рис. 4. Ассоциации между типами, которые могут приводить к появлению *R*-типа**

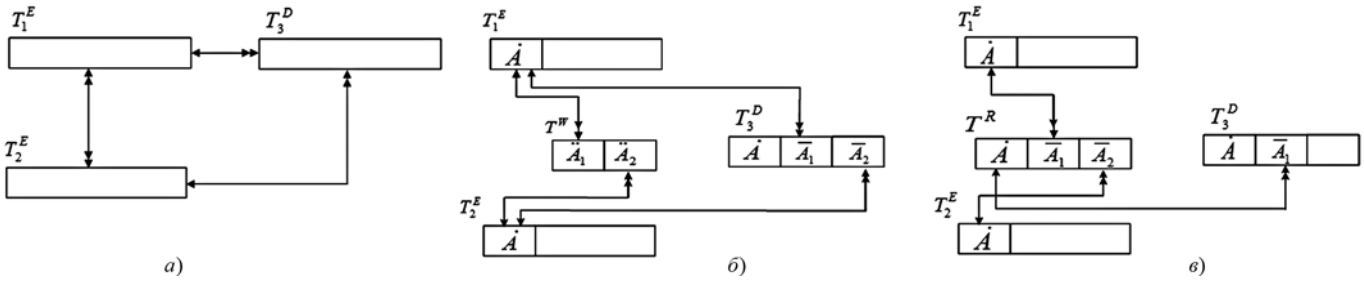

**Рис. 5. Замена *W*-типа на *R*-тип**

Основываясь на вышесказанном, сформулируем первое из двух условий появления *R*-типов в базах данных.

Заметим, что атрибуты внешних ключей $T_3^D$ связаны друг с другом многозначной функциональной зависимостью. Отсюда следует, что если между произвольными атрибутами внешних ключей некоторого заголовочного документального типа обнаружена многозначная функциональная зависимость, то в схеме организации данных должен присутствовать *W*-тип, который следует преобразовать в *R*-тип с последующей заменой пары атрибутов внешних ключей исходного документального типа на атрибут внешнего ключа, наследуемого от первичного ключа *R*-типа.

Нетрудно оттолкнуться и от обратного. Если какие-либо два типа связаны друг с другом ассоциацией, кардинальность которой *M:M,* и эти же типы одновременно ссылаются на третий тип и кардинальность этих связей составляет *1:M* или *1:1* (*1:1* здесь трактуется как частный случай *1:M*), то *W*-тип должен быть заменен на *R*-тип для исключения многозначной функциональной зависимости между наследуемыми внешними ключами третьего типа.

Второе условие, которое должно выполняться, чтобы состоялась замена *W*-типа на *R*-тип, предполагает идентичность всех ролевых подмножеств, участвующих во взаимодействиях между сущностными типами и между сущностными и документальным типом. Поясним сказанное на примере той же самой формальной подсхемы, изображенной на рис. 5, снабдив ее некоторой семантикой (рис. 6), которая указывает на содержание сущностных типов и состав ролевых, взаимодействующих множеств, являющихся определенными подмножествами сущностных типов.

В двух взаимодействиях между типами "Личности" и "Подразделения" разрешено участвовать разным подмножествам "Личностей". В одном случае — это будут "*Студенты*", в другом — "*Преподаватели*". В то же время, домен атрибута $\bar{A}_1$ ($\bar{A}_1$ — внешний ключ $T_3^D$), — это "*Преподаватели*". Поэтому только один из *W*-типов, "связывающий" типы "Личности" и "Подразделения" с соответствующими ролями "*Преподаватели*" и "*Кафедры*", может быть преобразован в *R*-тип.

Способы задания ролевых подмножеств подробно описаны в работе [21] и поэтому в данной статье не рассматриваются.

## 2. Многокомпозитные рилейторы

Только что рассмотренный случай, приводящий к появлению *R*-типов, можно считать одновременно и показательным, и простым, потому что *R*-тип синтезирован из отношения, в котором одновременно "участвуют" только два типа и "слабая сущность" обнаруживает-

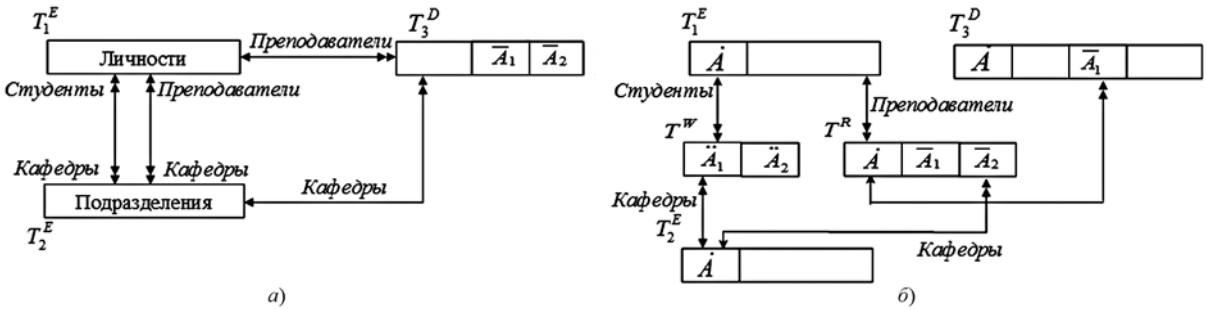


**Рис. 6. Замена *W*-типа на *R*-тип с учетом семантики связей**

ся достаточно просто — как необходимость в представлении ассоциации *M:M* в качестве самостоятельного типа. Но в орбиту *R*-типов могут оказаться вовлечены и другие типы, непосредственно соприкасающиеся с источниками *R*-типов. Установим признаки, на основании которых атрибутный состав рилейторов подлежит расширению.

Добавим в анализируемую конструкцию (см. рис. 2) еще один тип $T_5$ и исследуем влияние ассоциаций, связывающих $T_1$, $T_2$ и $T_5$, и кардинальность этих ассоциаций на необходимость ввода рилейторного типа и его атрибутный состав. Попрежнему $T_1$, $T_2$ и $T_5$ ссылаются на документальный тип $T_3$.

Несколько упростим систему обозначений ключевых атрибутов, помечая их буквами английского алфавита и латинскими цифрами, и откажемся от индексов, которые несущественны для содержимого текущего раздела.

Для перечисленных на рис. 7 сочетаний сущностных типов только две комбинации влекут за собой появление *R*-типов. В первом случае (рис. 7, *г*) *R*-тип — следствие ассоциации с кардинальностью *M:M* между $T_1$ и $T_2$.

Во втором (рис. 7, *в*) причиной его возникновения является наличие двух функциональных зависимостей: $\overline{II}^{1(2)} = f_1(\dot{I})$ и $\overline{II}^{5(2)} = f_2(I\dot{I}I)$, а также очевидное ограничение, в соответствии с которым для каждого кортежа отношения с заголовком $R\{\dot{R}, \bar{I}, \overline{III}\}$ должно выполняться условие $f_1(\dot{I}) = f_2(I\dot{I}I)$. Заметим, что в последнем случае *R*-тип никак не связан с *M:M*-ассоциацией. Таким образом, имеет место комбинация функциональных зависимостей между атрибутами внешних ключей документальных типов, которая также приводит к необходимости формирования *R*-типов.

Показанный на рис. 8 пример отношений, соответствующих подсхеме на рис. 7, *в*, иллюстрирует описанную ситуацию. Кортежи $t_3$ и $t_4$, которые ввиду действия представленного ограничения не могут присутствовать в отношении ***D***, вычеркнуты из последнего. На другой части рисунка показано то же самое отношение ***D***, содержащее атрибут внешнего ключа, ссылающегося уже на *R*-тип.

Две другие подсхемы (рис. 7, *а* и рис. 7, *б*) не влекут за собой появление *R*-типов.

Следующие схемы (рис. 9) включают в себя обязательную ассоциацию с кардинальностью *M:M* между $T_1$ и $T_2$ и дополнительные ассоциации с кардинальностями *1:M* противоположной направленности между $T_2$ и $T_5$. Показа-

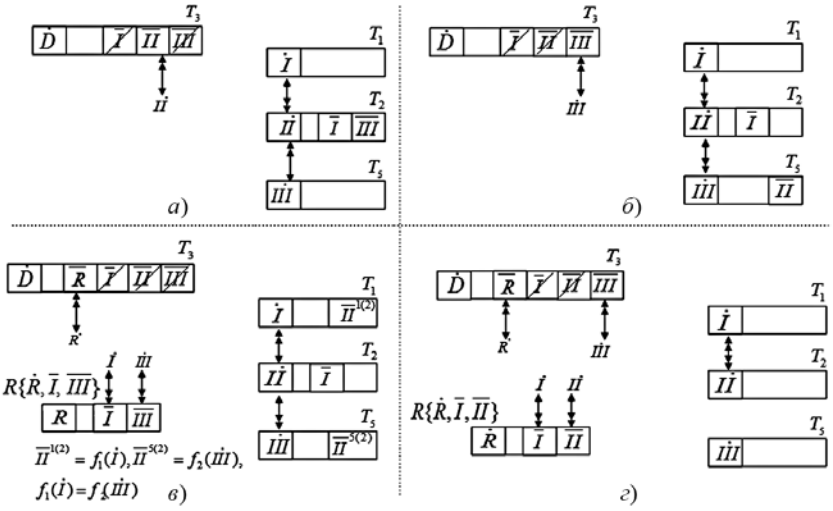


**Рис. 7. Простейшие ассоциации, приводящие к *R*-типам**

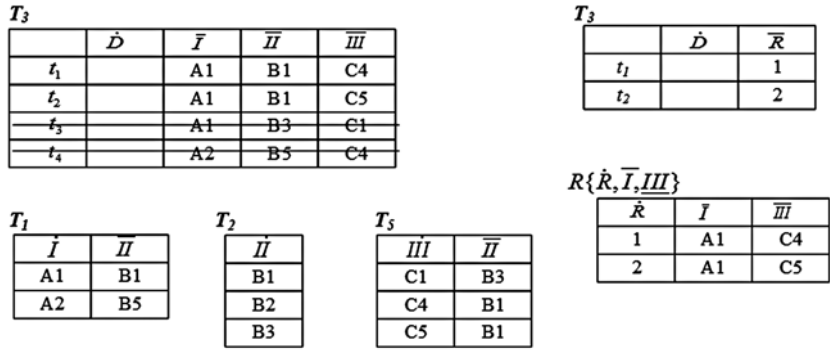

**Рис. 8. Состав и содержание *R*-типа при отсутствии многозначных функциональных зависимостей**

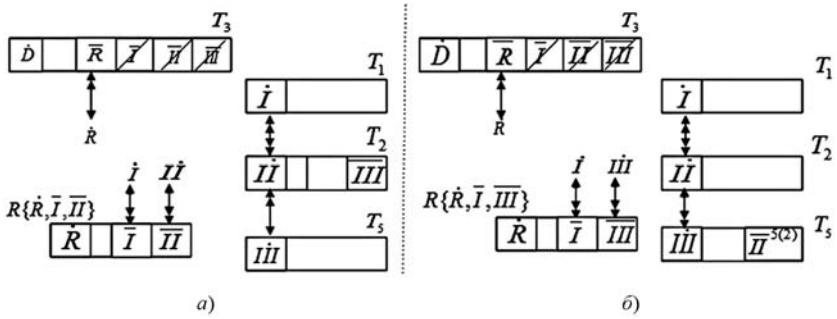


**Рис. 9. Атрибутный состав и ограничения *R*-типов при наличии *M:M*-ассоциаций**

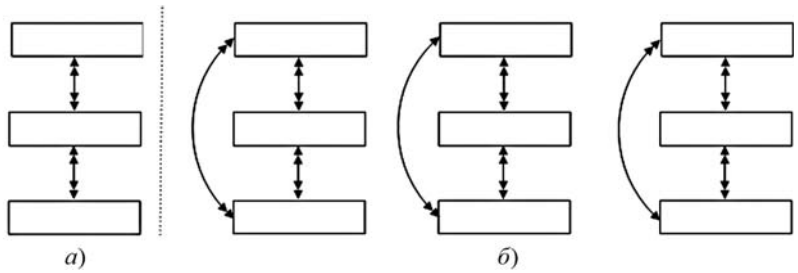


**Рис. 10. Варианты концептуальных схем с множественными *M:M*-ассоциациями**

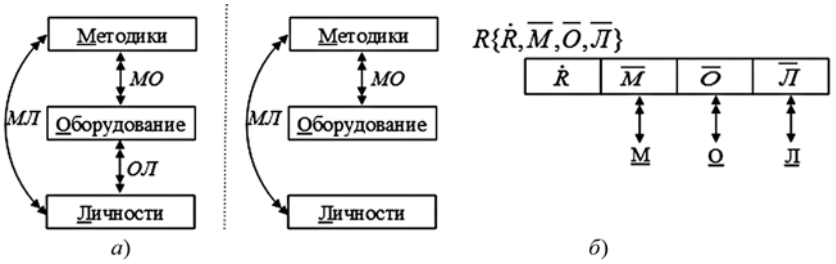


**Рис. 11. Фрагмент концептуальной модели предметной области "Поверка средств измерений"**

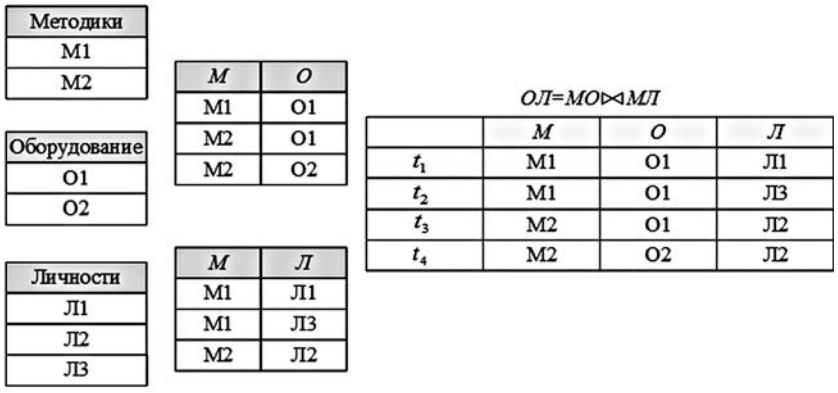


**Рис. 12. Фрагмент выводимых отношений логической модели**

тельна конструкция на рис. 9, *б*, в которой значение внешнего ключа $\overline{II}$ $T_3$-типа может быть найдено на основании функциональной зависимости $\overline{II}^{5(2)} = f(I\dot{I}I)$. Заметим, что нулевые или пустые значения атрибута $\overline{II}^{5(2)}$, если таковые допускаются, делают текущую конфигурацию *R*-типа некорректной и предполагают переход к схеме $R\{\dot{R}, \overline{I}, \overline{II}\}$ с отменой запрета на присутствие $\overline{III}$ в $T_3$. При этом вступает в силу ограничение $\overline{III}^3 = f(\overline{R}^3)$.

Алгоритм идентификации *R*-типов для схем, содержащих две и более ассоциации с кардинальностями *M:M* (рис. 9), не столь очевиден и однозначен. Он требует, помимо учета рассматриваемых структурных конфигураций (формальных по своей сути), отслеживания и ряда семантических аспектов взаимодействий. Но прежде чем перейти к обоснованию и изложению такого алгоритма, поставим вопрос о возможной эквивалентности двух подсхем, изображенных на рис. 10, *а* и *б*. Возьмем в качестве примера объекты и ассоциации, присутствующие в области поверки средств измерений и связывающие друг с другом такие типы сущностей, как "Методики", "Оборудование" и "Личности".

Соответствующий бизнес-процесс, содержащий перечисленные объекты и отношения между ними, представим в виде сценария.

*Сценарий*. "Организация выполняет поверку приборов, используя для этого специализированное оборудование. Одни те же методики поверки могут поддерживаться разными видами оборудования и наоборот. Личности, допущенные к поверке приборов на конкретных видах оборудования, должны владеть соответствующими методиками. По результатам поверки заказчику выдается свидетельство о поверке, в заголовке которого указывается вид оборудования, методика поверки и персона, проводившая поверку".

Концептуальная схема, содержащая объекты и отношения в представленной предметной области, может быть организована двояким образом (рис. 11).

Покажем, что отношение *ОЛ—ОЛ* {Оборудование, Личность} — выводимо из отношений *МО* {Методики, Личность} и *ЛО*. Действительно, ссылка на то, что приборы могут быть проверены только личностями, владеющими определенными методиками, приводит к тому, что $ОЛ = МО \bowtie МЛ$ (рис. 12). Отсюда, в частности, следует, что *R*-тип — $R\{\dot{R}, \overline{M}, \overline{O}, \overline{Л}\}$, как и подобная ему слабая сущность, являются лишними в логической модели, так как дублируют данные двух других *W*-типов: *МО* и *МЛ*.

Несмотря на это, подсхема, присутствующая на рис. 11, *а*, также имеет право на существование. Достаточно снять ограничение, согласно которому "Личности, допущенные к поверке на конкретных видах оборудования, должны владеть соответствующими методиками." Тогда отношение *ОЛ* не может рассматриваться как результат естественного соединения *МО* и *МЛ*.

### 3. Процедура конфигурирования комплексов *n*-арных отношений и *R*-типов

Выдвинем гипотезу, согласно которой любое отношение с арностью 3 и выше является подмножеством отношения, полученного в результате естественного соединения ряда бинарных отношений. Под комплексом (комплектом) отношений будем понимать совокупность исходных бинарных отношений, участвующих в синтезе 3- и более арных отношений.

Из содержимого предыдущего раздела следует, что кроме семантики еще и невыводимость некоторых отношений задают потребность в построении 3-арных отношений. Между тем, имеется еще несколько факторов, которые также следует учитывать, когда принимается решение о конфигурировании высокоуровневых отношений. Далее, на конкретном примере (рис. 13), приводящем к конфигурированию 3- и 4-арных отношений, рассмотрим полный спектр вопросов, которые могут возникнуть в процессе построения отношений с арностью 3 и выше.

Пусть исходный набор отношений образован слабыми сущностями, устанавливающими следующие бинарные факты-отношения:

$ГД^{A}$ — список дисциплин, изучаемых студентами в составе учебной группы;

$ДЗ^{P}$ — виды занятий, предусмотренные дисциплиной;

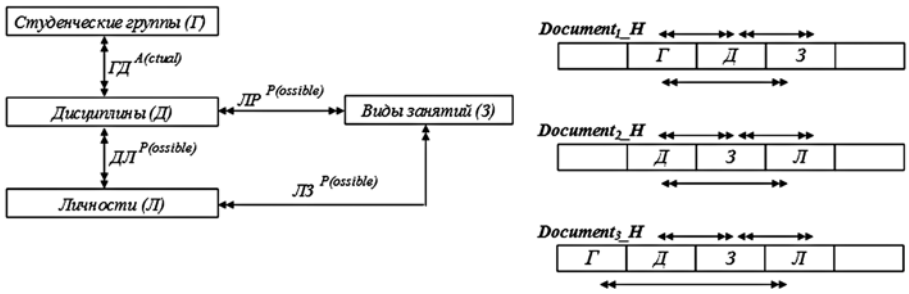


**Рис. 13. Типы и ассоциации, участвующие в синтезе высокоуровневых отношений**

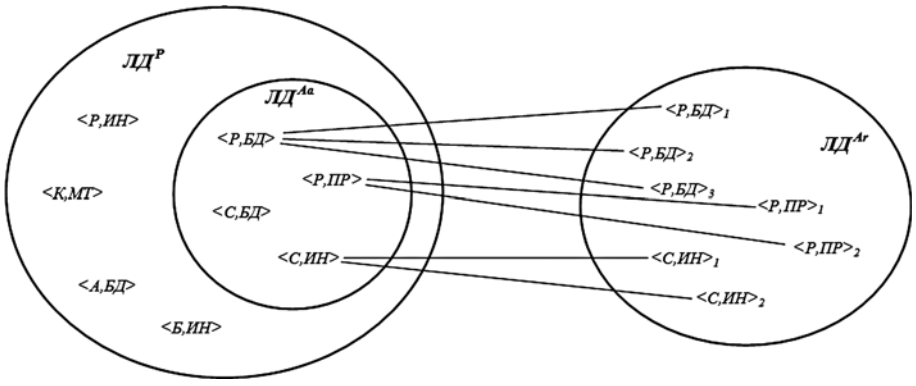


**Рис. 14. Отношения между множествами потенциальных, разрешенных и состоявшихся фактов**

$ЛД^{P}$ — преподаватели, которые могут вести те или иные дисциплины;

$ЛЗ^{P}$ — виды занятий, которые могут проводить преподаватели.

(Так как в данном конкретном случае в расчет принимается семантика отношений, то удобно отдельные кортежи отношений интерпретировать в качестве фактов.)

Среди перечисленных групп (типов) фактов неявно можно выделить факты по меньшей мере двух разных форматов: "потенциальные" (*Potential*) — *P*-факты и фактические (*Actual*) — *A*-факты. Если первые указывают на возможность существования отношений определенного типа,то вторые — это уже реальные факты, являющиеся подмножествами первых: $F_j^A \subseteq F_j^P$. Здесь $F_j^A$ и $F_j^P$ — множества соответственно фактических и потенциальных фактов *j*-го вида.

Если подходить к указанным различиям в форматах фактов более строго, то среди $F_j^A$ можно обнаружить подмножества двух видов: $F_j^{A_a}$ и $F_j^{A_r}$, экземпляры которых связаны друг с другом функциональной зависимостью $F_j^{A_a} = f(F_j^{A_r})$. Элементы состоявшихся взаимодействий образуют подмножество $F_j^{A_r}$, элементы разрешенных взаимодействий — подмножество $F_j^{A_a}$.

Например, применительно к отношению *ЛД* элементы $ЛД^{P}$ будут включать пары "личность—дисциплины", показывающие, какие дисциплины *может* вести преподаватель. (В работе мы не рассматриваем процедуры, приводящие к возникновению этих отношений, несмотря на то, что такая задача существует.) Дисциплины, которые *будет* проводить преподаватель, составят $ЛД^{A_a}$. Факты, касающиеся дисциплин, которые ведет в настоящее время преподаватель или уже провел, будут зафиксированы в $ЛД^{A_r}$. Поэтому между $ЛД^{A_a}$ и $ЛД^{A_r}$ установится взаимно однозначное соответствие — $ЛД^{A_a} = f(ЛД^{A_r})$. Очевидно, что и $ЛД^{A_a} \subseteq ЛД^{P}$.

Для наглядности элементы упомянутых множеств и отношения между множествами показаны на рис. 14.

Далее, чтобы сосредоточиться только на вопросах

конфигурирования $n$-арных отношений, не будем проводить различий между $F_j^{A_a}$ и $F_j^{A_r}$, полагая, что $F_j^{A_a} = F_j^A$. (Конфигурирование $F_j^{A_r}$ автоматически влечет за собой учет фактора времени. Например, в контексте $ДЗ^{A_r}$ это будет означать, что преподаватель, скорее всего, не ограничится одним занятием по одной дисциплине.)

Определимся с необходимостью поддержки 3- и 4-арных отношений, воспользовавшись в качестве маркеров таких отношений документальными типами.

Пусть имеются документы: $Document_1$, $Document_2$ и $Document_3$, часть атрибутов заголовочных структур которых ссылаются на объектные типы (см. рис. 13). Между атрибутами соответствующих внешних ключей, обозначенных как *Г*, *Д*, *З* и *Л*, присутствуют многозначные функциональные зависимости, которые предполагают синтез рилейтерных типов: *ГДЗ*, *ДЗЛ* и *ГДЗЛ*.

*ГДЗ* может быть получен в результате естественного соединения *ГД* и *ДЗ*: $ГДЗ = ГД \bowtie ДЗ$. *В* свою очередь, *ДЗЛ* определен как: $ДЗЛ = ДЗ \bowtie ДЛ \bowtie ЛЗ$. Возникает вопрос, а к каким форматам (*Р* или *А*) следует отнести полученные отношения? Исходный набор бинарных отношений, что очевидно, может состоять или только из *Р*-отношений, или *А*-отношений, или их комбинации.

Понятно, что если все исходные отношения относятся к *Р*-формату, то и результирующее отношение также будет *Р*-отношением.

Для остальных двух комбинаций сделать столь однозначное заключение, не прибегая к специальному анализу результирующих отношений, будет затруднительно. Покажем, почему это так. Воспользовавшись тем же самым примером, сконфигурируем два варианта *ГДЗ*:

$$ГДЗ^I = ГД^A \bowtie ДЗ^A \text{ и}$$
$$ГДЗ^{II} = ГД^A \bowtie ДЗ^P.$$

Если исходить из того, что $ДЗ^A \subseteq ДЗ^P$, то $M(ДЗ^P) \geqslant M(ДЗ^A)$, где $M$ — мощность отношения. Отсюда следует, что по мощности $ГДЗ^{II}$ будет всегда сопоставимо или превышать $ГДЗ^I$, $M(ГДЗ^{\mathrm{I}}) \geqslant M(ГДЗ^{\mathrm{II}})$. На этом основании $ГДЗ^{II}$ не может быть отнесено к *А*-формату. Следовательно, наличие хотя бы одного *Р*-отношения в исходном бинарном наборе делает результирующее отношение *Р*-отношением.

В отличие от $ГДЗ^{II}$ $ГДЗ^I$ может быть присвоен как *А*, так и *Р*-формат, потому что может оказаться, что и $ГД^A = ГД^P$ и (или) $ДЗ^A = ДЗ^P$.

Предположим, что $ГДЗ^A = ГДЗ^I$. Но $ГДЗ^I$ по определению не является исходным, первичным отношением. Следовательно, могут существовать множества, являющиеся подмножествами $ГДЗ^I$, что, собственно говоря, подтверждается практикой. Применительно к *ГДЗ* такие подмножества без проблем обнаруживаются в соответствующих предметных областях.

Поэтому, если исходный набор отношений представлен только *А*-форматными отношениями, то формат результирующего отношения может быть определен только экспертом.

Рассмотрев 3-арные отношения, перейдем к отношениям следующего порядка. Используя операции естественного соединения и исходный набор отношений, получим *ГДЗЛ:*

$$ГДЗЛ^P = ГД^A \bowtie ДЗ^P \bowtie ДЛ^P \bowtie ЛЗ^P,$$

который будет иметь *Р*-формат.

Содержимое $ГДЗЛ^P$ будет показывать, какие дисциплины и какие виды занятий в них в состоянии вести преподаватели.

Поскольку и $ГДЗ^P$, и $ДЗЛ^P$, и $ГДЗЛ^P$ были получены из исходных отношений: *ГД*, *ДЗ*, *ДЛ* и *ЛЗ*, осталось выяснить, не связаны ли каким-то образом $ГДЗ^P$, $ДЗЛ^P$ и $ГДЗЛ^P$ друг с другом.

Семантически — это все разные факты. В то же время, $ГДЗЛ^P$ может быть найдено и как $ГДЗ^P \bowtie ДЗЛ^P$, что не существенно, поскольку дает тот же результат, что и $ГД^A \bowtie ДЗ^P \bowtie ДЛ^P \bowtie ЛЗ^P$.

Но что будет, если формат *ГДЗ* и *ДЗЛ* изменится с *Р* на *А*? Результатом естественного соединения $ГДЗ^A$ и $ДЗЛ^A$, как видно из рис. 15, станет *ГДЗЛ*.

Если исключить из *ГДЗЛ* кортеж $t_1$ или $t_2$ и (или) удалить один из двух кортежей — $t_4$ или $t_5$, то две проекции $ГДЗЛ$ — $Пр_{ДЗЛ}(ГДЗЛ)$ и $Пр_{ГДЗ}(ГДЗЛ)$ — дадут соответственно $ДЗЛ^A$ и $ГДЗ^A$. Следовательно, $ГДЗЛ^A$ первично по отношению к $ДЗЛ^A$ и $ГДЗ^A$, а $ДЗЛ^A$ и $ГДЗ^A$ выводимы из $ГДЗЛ^A$.

С учетом всего вышеизложенного можно прийти к итоговой схеме отношений, увязывающей комплекс исходных бинарных и производных от них 3- и 4-арных отношений (рис. 16).

Пунктирной линией очерчены производные отношения, сплошной — единственное первичное отношение $ГДЗЛ^A$, которое подлежит хранению. Отсюда, в частности, следует, что если в $Document_3\_H$, *ГДЗЛ* принадлежит *А*-формату, то оно должно быть преобразовано в *R*-тип. В противном случае, если *ГДЗЛ* помечено как *Р*-формат, то *R*-тип, замещающий $ГДЗЛ^A$, не формируется. Но остается от-

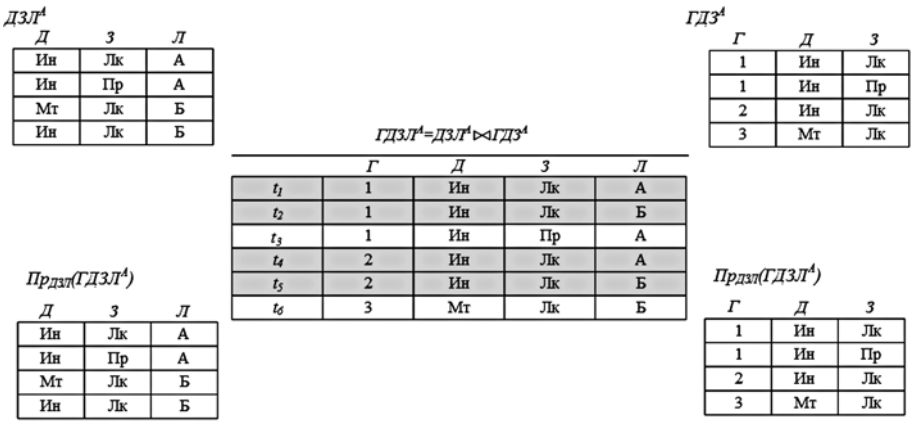


**Рис. 15. Результаты естественного соединения исходных отношений и последующих проекций полученного результирующего отношения**

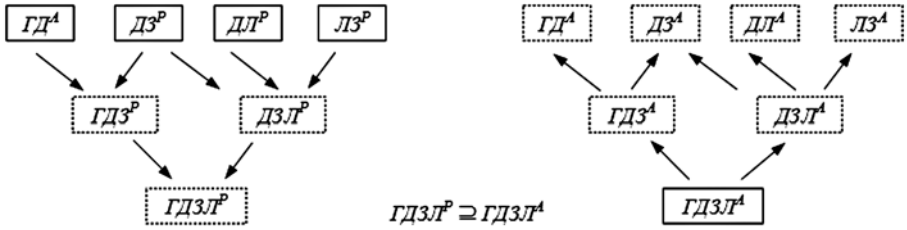


**Рис. 16. Первичные и производные отношения логической модели**

крытым вопрос, касающийся потенциальных *R*-типов, объединяющих внешние ключи двух других заголовочных структур: *Document*$_1$*_H* и *Document*$_2$*_H*. *ГДЗ* и *ДЗЛ* могут относиться и (или) к *P*- или к *A*-форматам.

Воспользуемся той же самой логикой, чтобы решить, какие из отношений (*ГДЗ* и (или) *ДЗЛ*) следует представлять в явном виде, а какие могут быть выведены из других отношений. (В явном виде означает, что они приобретают *R*-форму и сохраняются в базе данных.)

Если *ГДЗ* и (или) *ДЗЛ* имеют *P*-формат, то они выводимы, и необходимость в их сохранении отпадает. В противном случае, одно из них или оба отношения (будучи семантически не связаны друг с другом) приобретают *R*-форму и включаются в логическую модель.

Подведем промежуточные итоги. Вопросы, рассмотренные в настоящем разделе, позволяют решить одновременно несколько задач:

— идентифицировать и определять формат 3- и более арных отношений, в формировании которых участвует комплект бинарных отношений;

— устанавливать выводимые 3- и более арные отношения, а также первичные отношения (включая и бинарные), подлежащие хранению;

— задавать ограничения по составу сохраняемых типов.

## 4. Сцепленные рилейтеры и актуальные задачи моделирования сцепленных типов

Ранее было отмечено, что для исчерпывающего, унифицированного представления сущностей требуется кластер, образованный *Prototype (P), Sample (S)* и *Instance (I)*-подтипами, функционально связанными друг с другом. Покажем, в каких ассоциациях окажутся рилейтеры, сконфигурированные из одних и тех же типов, но различающиеся "собственными" подтипами — *P*-, *S*- или *I*-подтипами, являющимися родительскими по отношению к *R*-типам. Воспользуемся тем же примером рилейтора (см. рис. 11), который ранее объединял "методики", "оборудование" и "личности".

Пусть имеются два заголовочных документальных типа: *Document*$_1$*_H* и *Document*$_2$*_H*. Один из внешних ключей первого типа — $O^P$ — ссылается на прототип (модель, марку) оборудования, а другой — $O^I$ — на конкретный экземпляр оборудования, который принадлежит этой модели (рис. 17).

Один из вариантов замещения возникающих 3-арных отношений $W_1\{M, O^P, Л\}$ и $W_2\{M, O^I, Л\}$ приведет к появлению двух рилейторных типов: $R_1$ и $R_2$, между атрибутами внешних ключей которых — $\bar{O}^I$ и $\bar{O}^P$ — установится функциональная зависимость $f: \bar{O}^I \to \bar{O}^P$. В этом плане об $R_1$ и $R_2$ можно говорить как о сцепленных *R*-типах.

Если проанализировать содержимое отношений $R_1$ и $R_2$, то в них обнаружится большое число избыточных, дублирующих друг друга данных, обусловленное тем, что домены $\bar{M}^I$ и $\bar{Л}^I$, соответствующие $R_1$ и $R_2$, будут пересекаться.

Представленная на рис. 17 конфигурация *R*-кластера — не единственная. Атрибутный состав и наполнение $R_1$- и $R_2$-типов, скорее всего, изменится для отдельных сочетаний исходных *p*- и *a*-фактов.

Перечисленное может составить предмет дальнейших исследований, результатом которых должно стать получение универсальных шаблонных подсхем для представления

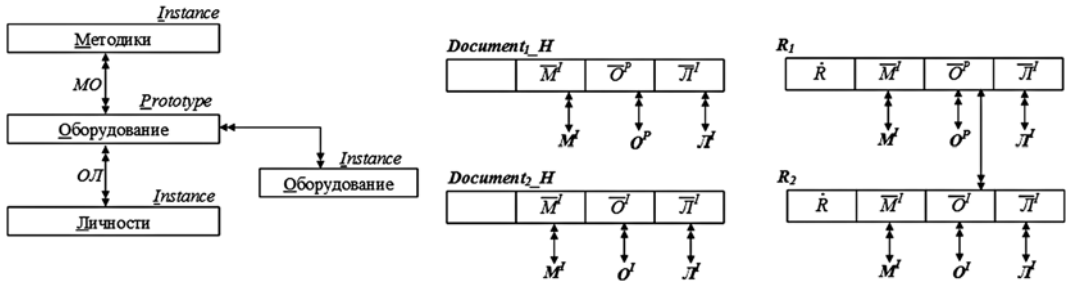


**Рис. 17. Сцепленные рилейтеры**

*n*-арных отношений и соответствующих им *R*-типов в логических конструкциях.

## Заключение

Альтернативные формы представления некоторых классов ассоциаций между сущностями в качестве самодостаточных объектов известны давно, равно как и проблема обнаружения и последующего моделирования высокоуровневых отношений. Мы обратили внимание на тесную взаимосвязь указанных задач. Оказалось, что такие объекты, как документы, в частности, в своей заголовочной части, содержат исчерпывающую информацию, указывающую на присутствие *n*-арных отношений.

Разработанная технология анализа и конфигурирования высокоуровневых типов строится с привлечением некоторых глагольных форм естественного языка, позволяющих различать в рамках одного класса отношений потенциальные и фактические отношения-факты. В работе приводится обоснование первичности фактических и вторичности потенциальных *n*-арных фактов по отношению к аналогичным бинарным фактам, которые всегда устанавливаются на начальных фазах моделирования вне зависимости от применяемого метода моделирования. Тем самым появляется возможность выявить первичный набор отношений в предметной области, из которого впоследствии с помощью операций проекций или естественного соединения могут быть получены все другие актуальные для предметной области отношения. Параллельно решается задача выбора формы представления первичных отношений — в виде или *W*-типов, или *R*-типов.

В заключение очертим некоторые направления перспективных исследований, которые могут быть выполнены, если отталкиваться от результатов настоящей работы.

Сформированный каркас первичных отношений может служить базисом для конфигурирования еще одного подкласса отношений, которые в работе обозначены как "состоявшиеся" отношения. Кроме упомянутых выше универсальных шаблонов, вследствие того, что предметные области — динамические системы с изменяемой структурой отношений, не менее актуальны решения по сохранности уже имеющихся данных при переходах от одних шаблонных схем к другим с увеличением или уменьшением арности *p*-отношений.

### Список литературы


1. **McAllister A.** Complete rules for n-ary relationship cardinality constraints // Data and knowledge engineering. 1998. No. 27. P. 255—288.
2. **Jones T. H., Song Y.** Analysis of binary/ternary combinations in entity-relationship modeling // Data and knowledge engineering. 1995. No. 19. P. 39—64.
3. **Krishna P. R., Khandekar A., Karlapalem K.** Modeling dynamic types for subsets of entity type instances and across entity types // Information systems. 2016. V. 60. P. 114—126.
4. **Armstrong W. W.** Dependency structures of database relations // Proceedings of IFIP congress. 1974. P. 580—583.
5. Fagin R. Multivalued dependences and new normal form for relational databases // ACM transactions database systems. 1977. No. 2(3). P. 262—278.
6. **Мейер Д.** Теория реляционных баз данных. М.: Мир, 1987. 608 с.
7. **Kohler Hp. R., Link S.** SQL schema design: foundation, normal forms and normalization // Information systems. 2018. V. 60. P. 88—113.
8. **Ling T.** A normal form for entity-relationship diagrams // Proceedings 4th International Conference on ER Approach. 1985. P. 24—35.
9. **Storey V. C.** Comparing relationships in conceptual modeling: mapping to semantic classifications // IEEE Transactions on knowledge and data engineering. 2005. Vol. 17, N. 11. P. 1478—1489.
10. **Guarino N., Guizzardi G.** We need to discuss the relationship: revisiting relationships as modeling constructs // International Conference on Advanced Information Systems Engineering. 2015. P. 279—294.
11. **Fonseca C. M., Porello D., Guizzardi G., Almeida J. P. A., Guariono N.** Relations in ontology-driven conceptual modeling // Conceptual modeling. 38-th international conference, ER 2019. P. 28—42.
12. **Olive A.** Conceptual modeling of information systems. Springer-Verlag Berlin Heidelberg, 2007. P. 471.
13. **Halpin T., Morgan T.** Information modeling and relational databases. Morgan Kaufmann Publishers, 2008. 970 p.
14. **Martin J.** Information engineering. Planning & Analysis. Book 2. Prentice Hall, 1990.
15. **Guizzardi G.** Ontological foundations for structural conceptual models. Center for Telematics and Information Technology, University of Twente, The Nethelands, 2005. 441 p.

16. **Hartman S., Link S.** English sentences structures and ERR modeling // The fourth Asia-pacific conference on conceptual modeling. Conferences in research and practice in information technology. 2007. Vol. 67. P. 27—35.

17. **Chen P. P.** English sentence structure and entity-relationships diagrams // Information science. 29. P. 127—149.

18. **Ghash S., Mukherjee P., Chakraborty B., Bashar R.** Automated Generation of E-R Diagram from a Given Text in Natural Language // 2018 International Conference on Machine Learning and Data Engineering (iCMLDE).

19. **Родионов А. Н.** "Мигрирующие" объекты моделей данных: "слабые сущности" и документы // Вестник ХГАЭП. 2011. № 1. С. 40—65.

20. **Родионов А. Н.** Семантическая идентификация, конфигурирование и моделирование типов сущностей в моделях данных // Вестник НГУ. Сер. Информационные технологии. 2014. Т. 12, Вып. 1. С. 64—78.

21. **Родионов А. Н.** Абстрактные роли и примитивы ролевого моделирования сущностей в системе "концептуальная—логическая—физическая модели данных"// Информационные технологии. 2019. Т. 25. № 4, С. 451—466.

**A. N. Rodionov,** Dr. of Tech. Sc.,
Computer Centre of Far-Eastern Branch of RAS, e-mail: ran@newmail.ru

# Conceptual and Logical Data Modeling: Detection and Configuration of *R*-types and *n*-ary Relationships in Domains

*The paper addresses the problems of detecting and configuring complexes of high-level database relationships (relationships with arity 3 and above) and representing part of the latter by means of R-types. R-type is one of the categories of synthetic object types whose instances are associated with links that occur between objects. Such types can come with their own properties and can interact with other objects. It is pointed out that multi-valued functional dependencies, which can be present among the foreign key attributes of header document types, uniquely identify both the n-ary relations themselves and the corresponding R-types. R-types that replace n-ary relations eliminates the occurrence of tuple update anomalies in the header documentary types. We study the influence of different relationship cardinality constraints that involve multiple entity types on the attribute composition of the corresponding R-types. It is introduced the concept of a binary relations hipscomplex that subsume initial interacting types. On the basis of the proposed p/a criteria for binary relations, which allows to distinguish potential and actual relations of a complex, it is developed a method for ascertaining primary relations that are subject to storage, and secondary relations that derive from primary ones. It is concluded that all high-level p-relations can be deduced (by means of natural join operation) from potential binary relations, and all binary a-relations can be obtained by applying projection operations to some source n-ary a-relation, if one exists in the domain. In conclusion, an example of linked R-types that may appear in entity type clusters is given, and current modeling problems are listed.*

***Keywords:*** *high-level relationships, R-types, p/a-binary relations, multi-valued dependences, PSI entity cluster*

DOI: 10.17587/it.26.460-471

**References**

1. **McAllister A.** Complete rules for n-ary relationship cardinality constraints,*Data and Knowledge Engineering*, 1998, no. 27, pp. 255—288.

2. **Jones T. H., Song Y.** Analysis of binary/ternary combinations in entity-relationship modeling, *Data and Knowledge Engineering,* 1995, no. 19. pp. 39—64.

3. **Krishna P. R., Khandekar A., Karlapalem K.** Modeling dynamic types for subsets of entity type instances and across entity types, *Information Systems*, 2016, vol. 60, pp. 114—126.

4. **Armstrong W. W.** Dependency structures of database relations, *Proceedings of IFIP congress*, 1974, pp. 580—583.

5. **Fagin R.** Multivalued dependences and new normal form for relational databases, *ACM transactions database systems*, 1977, no. 2(3), pp. 262—278.

6. **Maier D.** The theory of relational databases, Moscow, Mir, 1987, 608 p. (in Russian).

7. **Kohler Hp. R., Link S.** SQL schema design: foundation, normal forms and normalization, Information systems, 2018, vol. 60, pp. 88—113.

8. **Ling T.** A normal form for entity-relationship diagrams, *Proceedings 4th International Conference on ER Approach*, 1985, pp. 24—35.

9. **Storey V. C.** Comparing relationships in conceptual modeling: mapping to semantic classifications, *IEEE Transactions on knowledge and data engineering*, 2005, vol. 17, no. 11, pp. 1478—1489.

10. **Guarino N., Guizzardi G.** We need to discuss the relationship: revisiting relationships as modeling constructs, *International conference on advanced information systems engineering*, 2015, pp. 279—294.

11. **Fonseca C. M., Porello D., Guizzardi G., Almeida J. P. A., Guariono N.** Relations in ontology-driven conceptual modeling, *Conceptual modeling. 38-th international conference*, ER 2019, 2019, pp. 28—42.

12. **Olive A.** Conceptual modeling of information systems, Springer-Verlag Berlin Heidelberg, 2007, p. 471.

13. **Halpin T., Morgan T.** Information modeling and relational databases, Morgan Kaufmann Publishers, 2008, 970 p.

14. **Martin J.** Information engineering, Planning & Analysis. Book 2, Prentice Hall, 1990, 497 p.

15. **Guizzardi G.** Ontological foundations for structural conceptual models, Center for Telematics and Information Technology, University of Twente, The Nethelands, 2005, 441 p.

16. **Hartman S., Link S.** English sentences structures and ERR modeling, *The fourth Asia-pacific conference on conceptual modeling. Conferences in research and practice in information technology*, 2007, vol. 67, pp. 27—35.

17. **Chen P. P.** English sentence structure and entity-relationships diagrams, *Information science,* 29, pp. 127—149.

18. **Ghash S., Mukherjee P., Chakraborty B., Bashar R.** Automated generation of E-R diagram from a given text in natural language, *International Conference on machine learning and data engineering (iCMLDE),* 2018, pp. 112—117.

19. **Rodionov A. N.** Migrating objects of data models: weak entities and documents, *Vestnik KSAEL,* 2011, no. 1, pp. 40—65 (in Russian).

20. **Rodionov A. N.** Semantic identification, configuration and entities types modeling for the data model engineering, *Vestnik NSU. Series: Information Technologies*, 2014. vol. 12, no. 1, pp. 64—78 (in Russian).

21. **Rodionov A. N.** The abstract roles and the primitives of role modeling in the conceptual, logical, and physical data models system, *Information Technologies,* 2019, vol. 25, no. 8, pp. 451—466 (in Russian).