

В.А. Филатов, С.С. Танянский, Е.Б. Чапланова, А.И. Сизов

## Модель «сущность–связь» в задачах представления объектно-реляционных свойств предметной области

Исследована семантическая модель «сущность–связь». Предложены расширения структурной компоненты для отображения объектно-реляционных свойств предметной области средствами ER-модели. Дополнена и формально описана операционная спецификация данной модели. Рассмотрен пример.

The «entity-relationship» semantic model is investigated. The expansions of the structural component are suggested to display the object-relational properties of the subject domain by means of the ER-model. The operating specification of the given model is completed and formally described. An example is considered.

Досліджено семантичну модель «сутність–зв'язок». Запропоновано розширення структурної компоненти для відображення об'єктно-реляційних властивостей предметної області засобами ER-моделі. Довповнено та формально описано операційну специфікацію даної моделі. Розглянуто приклад.

**Введение.** В контексте проектирования автоматизированных информационных систем модели предметной области применяются на стадиях предпроектного обследования и системного анализа. На разных уровнях проектирования модели базы данных (БД) эффективность информационных процессов понимается по-разному. Кроме логического и физического уровня представления данных, существует более высокий «семантический» уровень. Традиционной архитектурой систем баз данных он не поддерживается, однако предполагается как неотъемлемый элемент создания эффективной модели БД. На уровне семантического моделирования данных решается основной вопрос – сохранение свойств предметной области.

Инфологические модели данных как способ семантического моделирования отображают данные на более высоком уровне абстракции в сравнении с даталогическими моделями. Поэтому семантическая концептуальная модель БД не связана с инструментальными методами обработки данных, т.е. не зависит от выбранной для реализации проекта системы управления базой данных (СУБД). Открытым остается вопрос, к какой структуре БД привести семантическую модель. Традиционно для этих целей используется реляционное описание данных,

основанное на теории множеств, для которой основной структурной компонентой есть плоская (реляционная) таблица. С другой стороны, известен факт ограниченности реляционной модели в вопросах поддержки семантических свойств предметной области [1, 2].

Современные информационные технологии дают возможность широко использовать объектный подход к описанию и обработке данных. Сохраняя основные свойства реляционной модели, объектная технология дает возможность использовать дополнительные типы данных, обеспечивая построение и поддержку типизированных иерархических связей, что, в свою очередь, позволяет в схеме данных описать больше семантических свойств предметной области.

Обзор основных решений задачи семантического моделирования приведен в [3]. Кроме этого, вопросы соответствия моделей данных рассматриваются в монографии Л.А. Калинин [4]. В работах [5, 6] исследуются ограничения модели «сущность–связь» и механизм приведения этой модели к реляционному виду.

Таким образом, вопросы представления объектных свойств различных предметных областей средствами модели «сущность–связь», с помощью которых можно описывать на фор-

мальном уровне концептуальные схемы предметных областей, актуальны.

### **Цель проводимых исследований**

Статья посвящена исследованию известной модели «сущность–связь». Формально определяются ее структурная компонента, ограничения целостности, операционная спецификация, а также особенности отображения в объектно-реляционную модель данных (ОРМД).

### **Разработка и исследование методов отображения объектно-реляционных свойств предметной области средствами модели «сущность–связь»**

#### ***ER – модель в задачах семантического моделирования***

Термин «инфологическая модель» может применяться как к моделированию данных, так и к моделированию предметной области. В информационно-логических моделях акцент делается на структуру предметной области и логику функционирования. При этом информационные аспекты вспомогательны. Инфологические модели данных предназначены для решения задач построения эффективной совокупности сохраненных данных и процедур их обработки.

Семантической моделью БД назовем концептуальное представление непосредственно интерпретирующее знаковые и функциональные особенности предметной области в терминах объектов данных. Семантические модели характеризуются достаточно высокой степенью формализации, но в общем случае не обеспечивают возможности однозначного перехода к концептуальной схеме БД.

Один из вариантов семантического моделирования – модель «сущность–связь» или *ER*-модель, предложенная в 1976 году П. Ченом (*P. Chen*) [3]. Модель была неоднократно модифицирована и пересмотрена. Сегодня можно говорить о *ER*-подходе к семантическому моделированию данных.

Модель «сущность–связь» можно рассматривать как семантическое расширение реляционной модели данных. Основные структуры данных (множество сущностей и связей) могут

быть представлены отношениями реляционной БД. Но это предположение не означает, что на логическом уровне *ER*-модель должна быть обязательно реализована средствами реляционной СУБД.

Самым распространенным видом представления схемы БД в виде *ER*-модели является *ER*-диаграмма (диаграмма «сущность–связь»). Она представляет собой ориентированный граф  $G(X, Y)$ , где  $X = \{x_1, \dots, x_n\}$  – множество вершин;  $Y = \{y_1, \dots, y_m\}$ , где  $y_k = \{\langle x_i, x_j \rangle \mid i \neq j\}$  – множество упорядоченных пар вершин (ребро).

Диаграмма «сущность–связь» может содержать три вида вершин: типы сущностей, типы связей, домены (множество значений). Ребро графа  $X$  означает соответствующую агрегацию типов вершин, если ребро поименованное, то имя ребра определяет роль сущности в связи. Расширение *ER*-диаграммы также может быть представлено в виде графа  $G$ . При этом каждому типу сущности или типу связи соответствует таблица, которая в общем случае может быть и нереляционной.

#### ***Исследование структурных компонентов модели «сущность–связь»***

В данной модели выделяются два вида объектов предметной области – сущности  $E$  и связи  $R$ . Объекты характеризуются свойствами, при этом сущности рассматриваются как агрегаты свойств, а связи – как агрегаты сущностей и, может быть, собственных свойств.

Из этого следует, что объекты можно разделить на две категории: объект–сущность и объект–связь. Каждая сущность обладает определенным набором свойств. Можно провести аналогию с реляционной моделью данных, где свойствам объекта реального мира отвечают атрибуты отношения.

Объект–связь может быть определен как ассоциация между несколькими объектами–сущностями (чаще – между двумя сущностями). Такой объект может обладать собственным набором свойств.

Граница между категориями сущностей нечеткая. Один тип может рассматриваться и как свойство объекта и быть самим объектом. Каждое свойство объекта ассоциируется с типом

сущности или типом связи путем задания отображения от объекта к множеству значений.

Для определения формальной структурной спецификации *ER*-модели введем множество сущностей *E* вида:

$$E = \{ E_1, \dots, E_m \}, \quad (1)$$

где каждый элемент *E* представляет собой множество:

$$E_j = \{ e_{j1}, \dots, e_{jm} \}, \quad j = \overline{1, n}. \quad (2)$$

Элементы множества *E* необязательно должны быть различны с точки зрения семантического смысла или роли, которые несет в себе каждая сущность. Они могут пересекаться или являться подмножеством других элементов.

Для обозначения типов связей введем множество *R*:

$$R = \{ R_1, \dots, R_p \}. \quad (3)$$

Пусть существует  $(E_1, \dots, E_n)$  – совокупность сущностей, которые агрегированы в *n*-арную связь  $R_j$ . Множество ролей каждого типа сущности  $E_k$  связи  $R_j$  обозначим как множество вида:

$$L_j = \{ l_{j1}, \dots, l_{jm} \}, \quad j = \overline{1, n}. \quad (4)$$

Тогда множество связей можно представить отношением:

$$R_j = \{ (l_{j1} / e_1, \dots, l_{jk} / e_k) \mid e_1 \in E_1, \dots, e_k \in E_m \}, \quad j = \overline{1, p}, \quad (5)$$

где  $l_{jk} / e_k$  – это сущность  $e_k$ , именованная ролью  $l_{jk}$ .

Порядок следования сущностей и связей не имеет значения, так как сущности в связи сопровождаются своей ролью. Множества *R* и *E* конечны.

Пусть *V* – множество значений агрегированных простых свойств сущности:

$$V = \{ V_1, \dots, V_n \}. \quad (6)$$

Свойства объекта–сущности или объекта–связи определены как отображение  $\varphi$ :

$$\varphi_i : E_i \rightarrow V_1 \times \dots \times V_n \text{ или } \varphi_j : R_j \rightarrow V_1 \times \dots \times V_n, \quad i = \overline{1, m}, \quad j = \overline{1, p}. \quad (7)$$

Если агрегатное свойство объекта многозначно, изменим отображение  $\varphi_i$ , которое примет вид:

$$\varphi_i : E_i \rightarrow (V_1 \times \dots \times V_k)^q \text{ или } \varphi_j : R_j \rightarrow (V_1 \times \dots \times V_k)^q, \quad i = \overline{1, m}, \quad j = \overline{1, p}, \quad (8)$$

где *k* – количество агрегированных простых свойств; *q* – количество повторений агрегата.

При  $k=1$  получаем простое свойство;  $q=1$  – свойство однозначно, при  $q=0$  – по определению содержит единственный элемент – пустой кортеж. Так как объект–сущность или объект–связь могут содержать многозначные и однозначные свойства, то тогда  $\varphi_i$  примет вид:

$$\begin{aligned} \varphi_i : E_i &\rightarrow (V_1 \times \dots \times V_k)^q \times V_l \times \dots \times V_n \text{ или} \\ \varphi_j : R_j &\rightarrow (V_1 \times \dots \times V_k)^q \times V_l \times \dots \times V_n \\ i = \overline{1, n}, \quad j = \overline{1, p}, \quad \{V_l, \dots, V_n\} &\subseteq \\ &\subseteq \{V_1, \dots, V_k\} \text{ и } l < n, \quad k < n, \quad l < k. \end{aligned} \quad (9)$$

При этом в (8) свойство ассоциативности не выполняется.

При построении семантической модели БД следует учитывать такие правила структуризации в *ER*-модели:

- свойство объекта всегда связано с одним определенным типом объектов. Это означает, что одно и то же свойство не может характеризовать несколько типов, если эти типы не подчинены одному общему типу;
- один и тот же объект–сущность может быть агрегирован в неограниченное число объектов–связей;
- объекты–связи могут агрегировать неопределенное количество сущностей.
- объекты–связи не могут агрегировать другие объекты–связи.

Число сущностей, агрегированных в связь, называется степенью связи. Если степень связи равна *n*, связь называется *n*-арной.

Во многих случаях схема предметной области, построенная с использованием *ER*-модели, находит свое отображение в концептуальной схеме реляционной БД.

Выражения (7) и (8) дают возможность получить схему БД в рамках объектно-реляцион-

ной модели данных [7]. В этом случае каждое многозначное свойство сущности может быть представлено составным объектом БД, если оно не представлено в виде ограничения целостности, т.е. не является ключом будущего отношения.

### **Исследование ограничений целостности модели «сущность–связь»**

Связи модели «сущность–связь» моделируют отображения объектов–сущностей. Основная характеристика таких отображений – ограничения целостности.

Рассмотрим отображение  $\mu$ , которое является строгим соответствием между двумя абстрактными множествами сущностей  $E_1$  и  $E_2$ :

$$\mu \subseteq E_1 \times E_2 \text{ или } \mu : E_1 \rightarrow E_2. \quad (10)$$

Отображение такого вида характеризуется кардинальными числами, показывающими, какое число элементов одного множества отвечает одному элементу другого множества. Можно описать  $\mu : E_1 \rightarrow E_2$  соотношением

$$\mu[(M_1, N_1) : (M_2, N_2)]. \quad (11)$$

где  $M_1$  – минимальное допустимое число элементов множества  $E_1$ , связанных отображением с одним элементом множества  $E_2$ ;  $N_1$  – максимальное допустимое число элементов множества  $E_1$ , связанных с одним элементом множества  $E_2$ ,  $M_2$  – минимальное допустимое число элементов множества  $E_2$ , связанных отображением с одним элементом множества  $E_1$ ;  $N_2$  – максимальное допустимое число элементов множества  $E_2$ , связанных с одним элементом множества  $E_1$ .

Например, пусть  $E_1$  – сущность *БЛЮДА* в рамках предметной области «Кулинарные рецепты»,  $E_2$  – сущность *ПРОДУКТЫ*,  $R$  – связь, *СОСТАВ*. Необходимо охарактеризовать эту связь. Пусть из свойств предметной области известно, что один и тот же продукт не может одновременно встречаться более чем в одном блюде; некоторые продукты не используются ни в одном блюде; готовое блюдо может состоять не менее чем из пяти, но не более чем из 30 продуктов. Тогда соотношение кардинальных чисел будет иметь вид:  $[(5, 30) : (0, 1)]$ .

В процессе моделирования зачастую неизвестно точное максимальное кардинальное число. Поэтому, если это число больше единицы, то соотношение будет иметь вид « $N$ » или « $\infty$ ». Это правило может распространяться и на минимальное кардинальное число. Если связь характеризуется только минимальными кардинальными числами, в этом случае она будет иметь вид  $(N_1 : N_2)$ .

Множество сущностей и связей может быть представлено отношениями. В *ER*-модели существуют понятия ключа сущности и ключа связи. Они аналогичны понятию ключа в отношении РМД. Если заданы ключи типов сущностей, агрегированных в некоторую связь, тогда в представлении экземпляра связи нет необходимости включать все экземпляры сущностей. Достаточно использовать только значение ключа. Соответственно, ключом связи в общем случае является совокупность ключей, агрегированных сущностей, и, возможно, – собственных атрибутов связи.

В случаях избыточности концептуальной модели возможно возникновение аномалий добавления, изменения, удаления. Для устранения аномалий используется алгоритм нормализации отношений баз данных.

### **Операционная спецификация ER-модели**

В инфологических моделях операции отображения, обеспечивающие работу с типами, приоритетны в отличие от операций над данными, которые являются вспомогательными [5].

Объекты–сущности и связи представлены в виде множеств, значит, на них определены базовые и производные операции реляционной алгебры  $\Psi$ , представленные кортежем:

$$\Psi = \langle \{ E \}, \{ R \}, \cup, \times, -, \sigma, \pi, \cap, \div, \theta \rangle. \quad (12)$$

Для отображения семантики структурирования данных в моделях данных используется два основных типа абстракций – обобщение и агрегация.

**Определение 1.** Операция  $\phi_i$  агрегирования выполняется над свойствами объектов и может быть представлена как отображение:

$$\phi_i : V_i \rightarrow (V_1 \times \dots \times V_k)^q, \quad i = \overline{1, N}, \quad (13)$$

где  $k$  – число агрегированных простых свойств; показатель декартовой степени  $q$  – число повторений агрегата.

Процесс, обратный агрегации, называется пошаговой детализацией.

**Определение 2.** Обобщение – это вид абстракции, позволяющей соотнести множество объектов данных с одним общим для них типом или множеством типов – с супертипом.

Операция обобщения выполняется над объектами и может быть супертипом для объектов, свойства которых участвовали в операции обобщения.

Для множества сущностей  $E$ , представленным выражением (1), можно определить объект  $T$  как  $\tau$ :

$$\tau: T \mapsto [E_n \chi E_k \chi \dots \chi E_m]. \quad (14)$$

Операция  $\chi$  может быть представлена как множество операций:

$$\chi = \{ \cup, \cap, -, \sigma, \pi \}, \quad (15)$$

где « $\cup$ » – объединение свойств объектов; « $\cap$ » – пересечение свойств объектов; « $-$ » – разность; « $\sigma$ » – операция селекции; « $\pi$ » – проекция свойств.

**Определение 3.** Действие, обратное обобщению, определяется как операция специализации  $\bar{\tau}$  – разделение объекта  $\bar{T}$  на множество других объектов  $\{ \bar{T}_i \}$ :

$$\bar{\tau}: \chi [ \bar{T} ] \mapsto \{ \bar{T}_1, \bar{T}_2, \dots, \bar{T}_n \}. \quad (16)$$

В этом случае операция  $\chi$  определена одним из элементов множества:

$$\chi = \{ \sigma, \pi \}, \quad (17)$$

где  $\sigma$  – операция селекции,  $\pi$  – проекция свойств.

Множество объектов с одним общим типом может отображаться в концептуальную схему БД в рамках объектно-реляционной модели данных. Это обусловлено тем, что супертип есть составной объект.

**Пример представления расширенной модели «сущность–связь»**

Для иллюстрации проводимых исследований рассмотрим известный пример построения

инфологической модели базы данных «Кулинарные рецепты».

Необходимо отобразить взаимосвязь продуктов, из которых готовят различные блюда, и их рецептов.

Предметная область характеризуется следующим набором сущностей  $E = \{ E_1, E_2, E_3 \}$ , где  $E_1$  – Блюда;  $E_2$  – Продукты;  $E_3$  – Рецепты и свойства сущностей  $E_1$  и  $E_2$  пересекаются –  $E_2 \cap E_1$ , т.е. элементы множества  $E_1$  могут быть подмножеством множества  $E_2$ .

Каждый кортеж сущности  $E_1$  будет определять отдельное блюдо. Количество элементов  $E_1$  определяется числом блюд рассматриваемой предметной области, например  $E_1 = \{ e_1, e_2, e_3, e_4 \}$ . Значит, сущность Блюда будет содержать четыре кортежа.

Пусть  $E_1$  – блюда,  $E_2$  – продукты агрегированы в бинарную связь *состав*. Тогда, согласно (3), определим множество связей как  $R = \{ R_1 \}$ , где  $R_1 = \{ \text{состав} \}$ .

В соответствии с (4) роль сущности в связи обозначим как  $L_1 = \{ l_{11}, l_{12}, l_{13}, l_{14} \}$ , где  $L_1 = \{ \text{вес}(\bar{a}) \}$ . Следовательно, согласно (5), получаем:

$$R_1 = \{ (l_{11}/e_1, l_{12}/e_2, l_{13}/e_3, l_{14}/e_4) | e_1 \in E_1, e_2 \in E_1, e_3 \in E_1, e_4 \in E_1 \},$$

где  $l_{11}/e_1$  – вес, определенный для блюда  $e_1$  сущности  $E_1$ ;  $l_{12}/e_2$  – вес, определенный для блюда  $e_2$  сущности  $E_1$ , аналогично для остальных кортежей сущности  $E_1$ .

Набор сущностей  $E = \{ E_1, E_2, E_3 \}$  обладает следующим набором свойств:

*Блюда* (БЛ, Название\_БЛ\_к, Название\_БЛ\_н, Вид);

*Продукты* (ПР, Продукт, Калорийность);

*Рецепты* (БЛ, Рецепт),

где БЛ – номер блюда в кулинарной книге или меню, ПР – номер продукта,

*Название\_БЛ\_к* – наименование блюда по кулинарной книге, *Название\_БЛ\_н* – общепринятое или краткое наименование блюда.

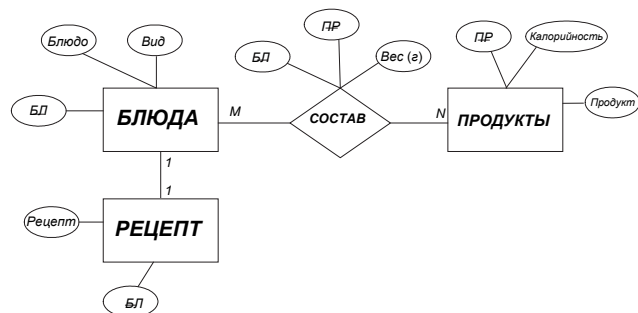
Согласно (6–7) элементы {*Название\_БЛ\_к*, *Название\_БЛ\_н*} могут быть агрегированы в свойство *Блюдо*.

Пусть  $V_1 = \text{Название\_БЛ\_к}$ ;  $V_2 = \text{Название\_БЛ\_н}$ ,  $k = 2$ , так как число агрегированных простых свойств равно 2;

$q = 2$ , так как декартова степень также равна 2.

Следовательно, получаем, что элементы множества  $E_1$  – *Блюда* – {*Название\_БЛ\_к*, *Название\_БЛ\_н*} могут быть агрегированы в многозначное свойство *Блюдо*.

Фрагмент схемы БД предметной области «Кулинарные рецепты» представлен на рисунке.



Фрагмент схемы БД «Кулинарные рецепты»

**Заключение.** Проектирование схемы базы данных всегда начинается с анализа предметной области и разработки концептуальной модели. От степени адекватности модели предметной области зависит в дальнейшем и качество реляционной модели данных. Таким образом, к функциональным характеристикам концептуальных моделей предъявляются повышенные требования.

В статье исследована модель «сущность–связь». Расширена структурная компонента модели. Дополнена операционная спецификация семантической модели. Формально определены многозначные свойства объектов, операции агрегирования, обобщения и детализации. Предложено использовать отображение логического представления в концептуальную схему объектно-реляционной модели данных. Приведен пример построения модели «сущность–связь», иллюстрирующий предлагаемый в статье подход.

1. Цаленко М.Ш. Моделирование семантики в базах данных. – М.: Наука, 1988. – 288 с.
2. Цикритзис Д., Лоховски Ф. Модели данных. М.: Финансы и статистика, 1985. – 344 с.
3. Чен П. Модель «сущность–связь» – шаг к единому представлению о данных // СУБД. – 1995. – № 3. – С. 137–158.
4. Калиниченко Л.А. Методы и средства интеграции неоднородных баз данных. – М.: Наука, 1983. – 424 с.
5. Дейт К. Введение в системы баз данных. – М.: Изд. дом «Вильямс», 2001. – 1072 с.
6. Когаловский М.Р. Энциклопедия технологий баз данных. – М.: Финансы и статистика, 2003. – 800 с.
7. Чапланова Е.Б. Об одном подходе к построению объектно-реляционной модели данных // Зб. наук. пр. Військового ін-ту Київського нац. ун-ту імені Тараса Шевченка. – К.: ВІКНУ, 2009. – 20. – С. 141–146.

Поступила 09.10.2010  
Тел. для справок: (057) 702-1354, 702-1432 (Харьков)  
E-mail: tanyanskyss@yahoo.com  
© В.А. Филатов, С.С. Таянский, Е.Б. Чапланова,  
А.И. Сизов, 2011