

МЕТОДОЛОГИЯ ОПИСАНИЯ ПРЕДМЕТНОЙ ОБЛАСТИ ПРИ ПРОЕКТИРОВАНИИ БАЗ ДАННЫХ СЛОЖНЫХ СИСТЕМ

А.А. Карпук

Расчетный центр Национального банка Республики Беларусь

Минск, Республика Беларусь

E-mail: Anatoly_Karpuk@bisc.by

Изложена разработанная автором методология построения информационных моделей предметной области сложных автоматизированных информационных систем. Представлены методики построения инфологических моделей фрагментов предметной области, отображения этих моделей в канонические модели фрагментов предметной области и объединения этих моделей в каноническую модель предметной области всей системы.

ВВЕДЕНИЕ

При проектировании баз данных (БД) сложных автоматизированных информационных систем (АИС) до настоящего времени применяется методология IDEF1X, разработанная по заказу департамента Военно-воздушных сил США в 1981 г., расширенная компанией DACOM и в окончательном виде опубликованная в конце 1985 г. [1].

В реальной предметной области большинства сложных АИС сущности имеют иерархическую структуру, в качестве значений атрибутов могут выступать векторы и повторяющиеся группы, значения некоторых атрибутов могут быть пустыми. Все это создает значительные трудности разработчикам концептуальной схемы БД, использующим методологию IDEF1X. Эти трудности становятся практически непреодолимыми, если нет специалиста, знающего предметную область всей сложной АИС, а есть только ряд специалистов, каждый из которых знает свой фрагмент предметной области АИС.

В докладе представлена разработанная автором методология описания предметной области сложных АИС при проектировании БД, возможности которой во многом превосходят возможности методологии IDEF1X, и при использовании которой отсутствуют перечисленные выше трудности.

1. МЕТОДОЛОГИЯ ОПИСАНИЯ ПРЕДМЕТНОЙ ОБЛАСТИ

Классическая концепция БД в соответствие со стандартом ANSI/SPARC предусматривает три уровня описания данных: внешний, концептуальный и внутренний. На каждом уровне используются соответствующая модель данных и схема БД. Для сложных предметных областей практически невозможно сразу построить концептуальную и внешние схемы БД, поэтому в современных методологиях и CASE-средствах проектирования БД концептуальный уровень описания данных разбит на два: информационно –

логический (инфологический) и даталогический (см. рис. 1).

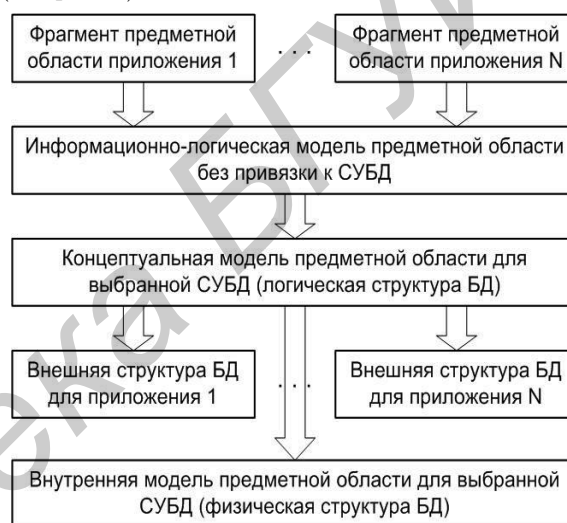


Рис. 1 – Традиционная методология описания предметной области и проектирования БД

Инфологическая модель данных, применяемая для описания предметной области сложной АИС, должна удовлетворять следующим требованиям: средствами модели должны описываться все объекты, предметы, явления предметной области и все отношения между ними, используемые при решении прикладных задач (ПЗ) в АИС; средствами модели должны описываться все ограничения целостности данных, имеющиеся в предметной области; средствами модели должны задаваться все объемные характеристики данных, необходимые для проектирования логической и физической структуры БД; базис понятий модели должен быть максимально приближенным к базису понятий постановок и алгоритмов ПЗ, а также к лексике специалистов по предметной области; должны существовать методы объединения описаний фрагментов предметной области, соответствующих отдельным ПЗ, в глобальное описание всей предметной области; средствами модели должен описываться процесс решения каждой ПЗ с указанием, какие операции над какими данными и с какой

частотой будут выполняться; должны существовать методы и методики отображения глобального описания предметной области в логические, внешние и физические модели данных применяемых СУБД.

В предлагаемой методологии для построения инфологической модели предметной области сложной АИС используется две модели данных [2]. Для построения инфологической модели фрагментов предметной области для ПЗ (одной или нескольких) используется инфологическая модель данных, удовлетворяющих первым четырем требованиям, а для построения канонической модели фрагментов предметной области и глобальной канонической модели предметной области всей АИС используется каноническая модель данных, удовлетворяющей последним трем требованиям (см. рис. 2).



Рис. 2 – Предлагаемая методология описания предметной области и проектирования БД

II. ИНФОЛОГИЧЕСКИЕ И КАНОНИЧЕСКИЕ МОДЕЛИ ПРЕДМЕТНОЙ ОБЛАСТИ

Множество допустимых структурных компонентов инфологической модели данных имеет вид $\{A, F, NR, Q, NQ\}$, где A – множество атрибутов, которые могут быть простыми или составными; F – множество функциональных зависимостей (ФЗ) между атрибутами, существующих в предметной области; NR – множество нефункциональных отношений (НО) между атрибутами, существующих в предметной области; Q – множество типов объектов, для каждого типа объектов задаются имена атрибутов основного ключа, имена обязательных и возможных атрибутов, причем каждый тип объектов может иметь элементарную, простую или сложную иерархическую структуру; NQ – множество отношений между типами объектов, которые определяются на атрибутах их основных ключей, ли-

бо сводятся к НО между атрибутами типов объектов. Для каждого простого и составного атрибута в модели указывается среднее и максимальное число повторений элемента или агрегата данных в значении атрибута. Для каждой ФЗ между атрибутами задается среднее и максимальное число различных значений левой части. Для каждого НО задается среднее и максимальное число кортежей в нем. Для каждого типа объектов задается среднее и максимальное число экземпляров в БД. Для объектов иерархической структуры задается среднее и максимальное число значений каждого сцепленного ключа в одном экземпляре типа объектов. Для каждого возможного атрибута типа объектов задается средняя и максимальная частота появления его значений в БД. Для каждого отношения между объектами задается среднее и максимальное число кортежей в нем.

Множество допустимых структурных компонентов канонической модели данных имеет вид $\{D, A, R, RE\}$, где D – множество доменов; A – множество атрибутов; R – множество отношений в четвертой нормальной форме (4НФ); RE – множество ФЗ между отношениями. Каждая ФЗ между отношениями классифицируется как полная, частичная сильная, частичная слабая или частичная слабая с обратной связью. Алгоритм отображения инфологической модели фрагмента предметной области в каноническую модель этого фрагмента предметной области состоит из следующих шагов: выделение отношений между атрибутами и построение системы образующих структуры ФЗ между атрибутами; построение минимального элементарного базиса структуры ФЗ между атрибутами; синтез отношений в третьей нормальной форме [3]; приведение отношений к 4НФ и выделение возможных атрибутов; выделение и классификация ФЗ между отношениями; удаление транзитивных ФЗ между отношениями. В докладе также представлен разработанный автором алгоритм объединения канонических моделей фрагментов предметной области в каноническую модель предметной области всей системы.

1. Integrated Information Support System. Information Modeling Manual IDEF1-Extended (IDEF1X) / D. Appleton Company Inc. – Manhattan Beach, California, December 31, 1985. – 108 p.
2. Карпук, А. А. Информационное моделирование предметной области автоматизированных информационных систем / А. А. Карпук // Технологии информатизации и управления: Сб. науч. ст. Вып. 2 / Редкол.: А. М. Кадан (отв. ред.) [и др.]. – Минск: БГУ, 2011. – С. 24–30.
3. Карпук, А. А. Об алгоритмах нормализации отношений в реляционных базах данных / А. А. Карпук // Международный конгресс по информатике: информационные системы и технологии: Материалы междунар. науч. конгресса. Ч. 2 / Редкол.: С. В. Абламейко (отв. ред.) [и др.]. – Минск: БГУ, 2011. – С. 283–288.