

УДК 681.3.016

Г.Н.Востров, канд. техн. наук, доц.,
 Е.В.Малахов, канд. техн. наук, доц.,
 В.Н.Кулешов, специалист

ПРОБЛЕМЫ ОПИСАНИЯ СТРУКТУРЫ ПРЕДМЕТНЫХ ОБЛАСТЕЙ

Г.М.Востров, Е.В.Малахов, В.М.Кулешов.

Проблемы опису структур предметних областей. Розглядаються сучасні проблеми і варіанти опису предметних областей та їх взаємодії з середовищем та системою управління. Наведені можливості динамічного розвитку баз даних як моделей предметних областей та місце метаданих в цих моделях.

G.N.Vostrov, E.V.Malakhov, V.N.Kuleshov

The subject domains' structure description problems. The modern problems and variants of a subject domains' description are considered as well as their interaction with environment and control system. The databases' dynamic development opportunities as the subject domain models are shown. Position of metadata in these models is pointed out as well.

Во многих областях деятельности, связанной с управлением, в частности, организационным или административным, существует необходимость принимать решения, направленные на достижение конкретной цели. Для повышения эффективности такого рода деятельности и ускорения решения задачи разработаны и разрабатываются специализированные *системы поддержки принятия решений* (СППР). Их можно разделить на два класса — системы, предназначенные для обеспечения персонала максимально возможным комплексом информации, и системы, вырабатывающие варианты решений, из которых человек или эксперт может выбрать наиболее подходящий по какому-либо ряду критериев.

В любом случае таким системам необходим источник или источники информации о предметной области (ПО), которые можно разделить на две категории:
 — источники динамической информации о состоянии ПО и ее элементов;
 — источники статической информации о прецедентах и процессах, когда-либо имевших место в данной ПО или смежных областях.

В электронном варианте источников *вторая* категория — *хранилища данных*, в которых накапливается информация о характеристиках ПО, возможно, различного *уровня детализации и степени стабильности*.

Для *первой* категории источников возможны два варианта:

— источником данных выступает *сама предметная область* с различными способами кодирования и ввода информации либо непосредственно в систему, либо посредством динамической базы данных;
 — информация *генерируется в системе динамических баз данных* путем прогноза развития состояний предметной области на основе содержимого этих баз, т.е. описания предыдущих состояний ПО, и в соответствии с физическими законами либо установленными правилами поведения.

Подобное взаимодействие можно представить в виде схемы (рис. 1).

Если источник информации о состоянии элементах ПО — сама ПО, то под *различными способами кодирования* понимается чаще всего представление информации на некотором *формализованном языке* L_F . С другой стороны, предметная область взаимодействует, обменивается информацией со средой, в которую она погружена. Такой обмен осуществляется на *естественном языке* посредством множества высказываний L_N . Соответственно, для представления информации в аналитической части системы необходимо выполнить преобразование $\Phi(L_N \cup L_F)$ семантики описания ПО к некоторой единой формализованной грамматике L_C , предложения которой могут быть представлены в информационной части системы.

В информационной части системы, т.е. в хранилищах и базах данных, должно содержаться корректное, адекватное отражение ПО и, в первую очередь, ее структуры.

Для определения структуры ПО необходимо предварительно определить, что мы будем понимать под термином *предметная область*.

Предметную область можно определить как сферу деятельности или исследования человека, описанную множеством высказываний $\{P_1, P_2, \dots, P_n\}$ естественного языка L_N и множеством физических законов, математических функций $\{F_1, F_2, \dots, F_m\}$ или установленных правил.

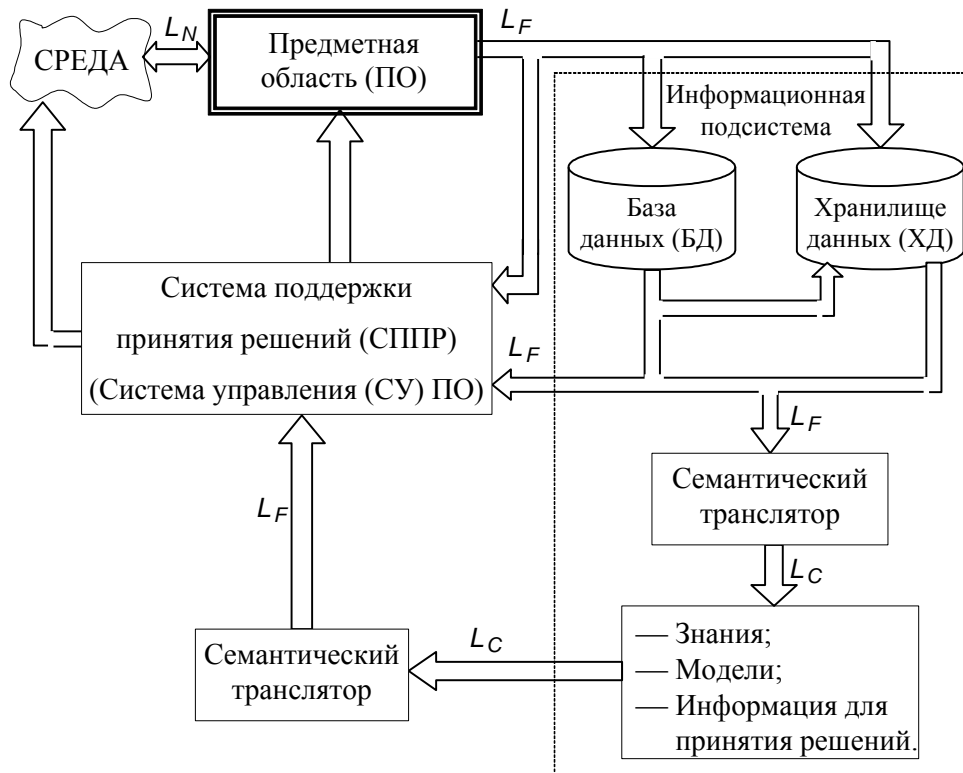


Рис. 1. Схема взаимодействия системы управления с предметной областью

Первое позволяет определить объекты предметной области и связи между ними, что в какой-то мере, согласуется с рекомендацией относительно выбора объектов, их атрибутов и связей между ними на основе грамматических элементов описания на естественном языке L_N [4]. Добавление нового высказывания либо вводит новые объекты в предметную область, либо расширяет или уточняет информацию об уже выделенных объектах

Второе определяет поведение, взаимодействие и "жизненные циклы" объектов, их развитие и развитие самой предметной области.

В теории баз данных каждую сущность предметной области принято описывать набором свойств или атрибутов.

Если использовать на архитектуру баз данных [3], то концептуальная схема базы данных, отображающая информацию обо всей предметной области, должна приобрести бесконечную размерность. Действительно, возьмем, например, сущность человек. Можно описать эту сущность с различной степенью детализации с использованием различных классификаций:

- социальной — фамилия, имя, отчество, социальный статус, место жительства и тому подобное;
- профессиональной — фамилия, имя, отчество, место работы или учебы, должность, образование и так далее;
- медицинской — фамилия, имя, отчество, пол, физическое состояние, перенесенные или имеющиеся заболевания, физиологические параметры и прочее;

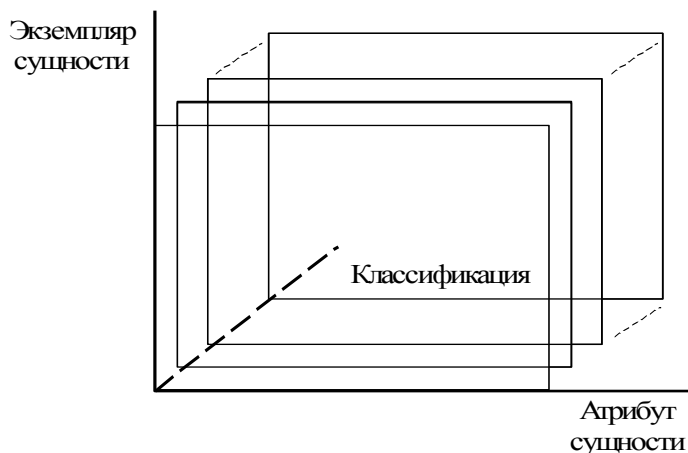


Рис. 2. Структура описания одной сущности ПО

- биологической, биохимической или биофизической — структура ДНК, формула крови, кардио- и энцефалографические данные и тому подобное;
- физико-химической — структура молекул и химические элементы, составляющие организм человека;
- множество иных взглядов и классификаций с позиций микро- и макромира, неизвестных нам.

Получим следующую многоуровневую структуру, описывающую данную сущность (рис. 2).

Очевидно для того, чтобы

обеспечить полноту описания ПО, необходимо определить ее конечные границы.

Например, структурой, аналогичной приведенной, можно описать все сущности, связанные тем или иным образом с данной. Причем, такая связь может существовать как на уровне всех, так и в пределах некоторых или одной приведенной классификации. Соответственно, описание каждой сущности с учетом всех связей будет представлять собой куб размерности $N+1$, где N — максимальное число классификаций конкретной сущности (рис. 3).

Таким образом, предметная область будет представлять собой множество связанных кубов различной размерности или гиперкуб, описанный языком L_F , множества сущностей, описанных в терминах одного языка L_N (см. рисунок 3).

Если к атрибутам сущности добавить описание его поведения, то получим, говоря языком объектно-ориентированного анализа, объект некоторой предметной области.

Таким образом, в информационной части системы описание ПО сводится к определению множества объектов ПО $A = \{A_1, \dots, A_i, \dots, A_n\}$. Причем, каждый объект описывается, множеством свойств $X_i = \{x_{i1}, \dots, x_{ij}, \dots, x_{ik}\}$, которое при представлении в реляционной базе данных соответствует схеме отношения, отражающей этот объект и множеством зависимостей от других объектов или зависимостей, определяющих другие объекты.

Один из критериев сложности систем обусловлен невозможностью полного описания ПО [1, 2, 4]. Соответственно, невозможно точно и полностью определить структуру базы данных, отображающей данную ПО.

Как следствие, возникает или может возникнуть необходимость развития созданной БД без потери хранимой информации о ПО.

Развитие БД может осуществляться следующими способами:

- получение новых и уточнение сведений о предметной области, и соответствующие изменения структуры БД;
- получение новых сведений о ПО путем выполнения операций над имеющимися данными.

В реляционной модели данных в последнем случае возможно как изменение мощности БД — изменение количества информации в существующей структуре, так и изменение степени любого или каждого отношения БД — изменение самой структуры. Если определить методы и алгоритмы таких операций, то подобные изменения можно осуществлять автоматически, т.е. построить саморазвивающуюся базу данных.

Более наглядно изменение мощности базы (рис. 4), при котором определение зависимости между значениями, в реляционной базе данных (РБД) — кортежами, и применение тех или иных методов позволяет либо определить значения, которые до того не существовали (блок В), либо восстановить пропущенные значения, которые до этого не использовались (блок А).

Изменение степени отношений БД представляет собой изменение схемы отношений БД. Такое изменение до недавнего времени осуществлялось путем полной перекомпоновки БД ее разработчиками. Любое манипулирование данными на этот период прекращалось. Если в БД ввести дополнительно надстроечную информацию, описывающую ее структуру, то модификацию схемы отношений можно осуществлять динамически по мере появления или изменения информации о ПО. Такой надстроечной информацией являются метаданные, используемые в современных СУБД.

При традиционном подходе к построению информационной системы в нее вводится информация о сущностях предметной области во время компиляции, например, имена полей, таблиц,

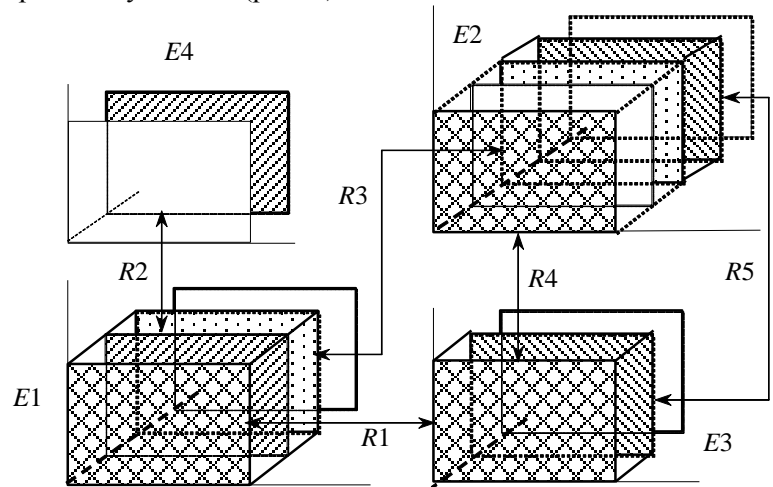


Рис. 3. Гиперкуб структуры предметной области

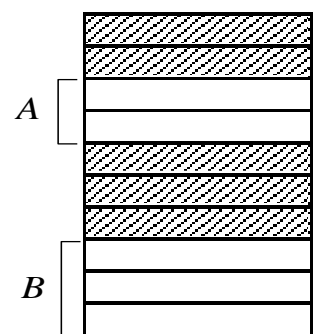


Рис. 4. Развитие мощности БД

индексов и так далее (рис. 5а). То есть прикладная программа непосредственно обращается в информационную часть БД.

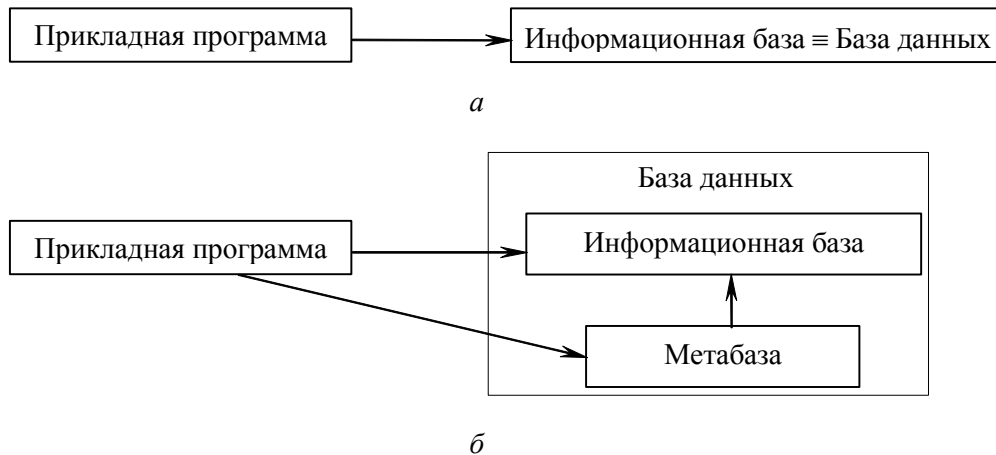


Рис. 5. Схемы доступа к базе данных: а — традиционная схема доступа; б — схема доступа с использованием метаданных

При использовании метаданных схема доступа к информации меняется — программа обращается в метабазу для получения информации о ПО и лишь после анализа хранимых в ней метаданных начинает обработку "полезных" данных (рис. 5б). В схеме отсутствует программа-конфигуратор, основная задача которой — реструктуризация информационной базы при изменении модели ПО.

Таким образом, в *метабаза* хранится *метамодель* системы, описывающая ПО.

Такое построение структуры БД позволяет решать задачи различного уровня.

Задача реализации эффективного *кэширования* данных из базы. Рассмотрим две схемы связи объектов:

Фирма → *Город* → *Страна*

Строка документа → *Документ* → *Фирма*

Если программный код, который кэширует данные, знает только, что сущности находятся в отношении 1:N, он будет считать схемы равноправными. Если же у алгоритма кэширования есть информация, что отношение между *строкой документа* и *документом* есть отношение агрегации, то при создании документа эффективней выбрать все строки документа, т.к. весьма вероятно возникновение задачи подсчета значений на множестве строк документа.

Одна из задачи автоматизации библиотеки.

Рассмотрим объект *библиографическое описание*. Существуют национальные и международные стандарты на машиночитаемые библиографические описания, один из которых UNIMARC, помимо общеизвестных полей — авторов, заглавий, мест издания и т.д., содержит до 2000 полей данных [5]. После нормализации отношений БД пришлось бы работать со 150 таблицами одновременно. Кроме того, этот формат постоянно изменяется. Для проведения изменений таблиц, индексов и программы обработки можно построить метамодель, которая содержит всего 2 элемента: свойства и группы свойств. При добавлении, редактировании или удалении свойств или групп свойств автоматически происходит реструктуризация БД. Программа обработки данных получает информацию из метамодели и на ее основании взаимодействует как с информационной базой, так и с пользователем.

Как видно из примеров, *метамодели* представляют собой эффективный инструмент построения информационных систем, позволяющий развивать базу данных во времени. Следовательно, при проектировании БД необходимо обратить внимание не только на *модели* предметных областей, которые часто уникальны, а, возможно, даже в большей степени, на *метамодели* и их разработку.

Таким образом, предложенный в статье вариант *представления структуры* предметной области и *ее развития* позволяет не только описать ее в базе данных и информационных хранилищах, но и формализовать ее взаимодействие со средой, а также обеспечить систему управления информацией, необходимой для воздействия на предметную область и, возможно опосредованно, на среду. Формализация понятия среды выходит за рамки данной статьи. Однако предполагается, что *среда* представляет собой множество иных предметных областей, описанных подобной структурой, и связанной или взаимодействующей с данной через другие плоскости кубов,

описывающих входящие в нее сущности (незаштрихованные плоскости на рис. 2).

Литература

1. Буч Г. Объектно-ориентированный анализ и проектирование с примерами на C++ — К.: Диалектика, 1997.
 2. Дейт К. Дж. Введение в системы баз данных: Перевод с английского — 6-е издание — К.: Диалектика, 1998.
 3. Цикритзис Д., Лоховски Ф. Модели данных: Перевод с английского — М.: Финансы и статистика, 1985.
 4. Шлеер С., Меллор С. Объектно-ориентированный анализ: моделирование мира в состояниях. - К.: Диалектика, 1993.
 5. Руководство по UNIMARC: Руководство по применению международного коммуникативного формата UNIMARC/Пер. с англ. авт. коллектива под руководством А.И.Земскова, Я.Л.Шрайберга. — М.: ГПНТБ России, 1992.
-