

УДК 004.62:004.942

Г.Н. Востров, канд. техн. наук, доц.,  
Е.В. Малахов, канд. техн. наук, доц.,  
Одес. нац. политехн. ун-т

## БАЗИС ПРОСТРАНСТВА ОБЪЕКТОВ ПРЕДМЕТНЫХ ОБЛАСТЕЙ ПРИ ПРОЕКТИРОВАНИИ БАЗ ДАННЫХ

*Г.М. Востров, Є.В. Малахов. Базис простору об'єктів предметних областей при проектуванні баз даних.* Розглянуто проблеми формування об'єктних ядер предметних областей на підставі запропонованого поняття елементарного об'єкта.

*G.N. Vostrov, E.V. Malakhov. The object space basis of subject domains for databases design.* Problems of forming of object cores of subject domains on the base of the offered concept of elementary object are considered.

Построение любых классов баз данных всегда связано с выделением предметной области (ПО), для которой БД является информационной, структурной и алгоритмической моделью. Сформировалась точка зрения, что понятие “предметная область” не может быть формализована [1]. В то же время, отсутствие достаточно точного формального описания ПО создает значительные проблемы как на этапе проектирования базы данных, так и при манипулировании ею.

Важным элементом любой предметной области является понятие “объект”. Считается, что это понятие, как и понятие “предметная область”, является первичным [1], и поэтому его формализация принципиально невозможна. В то же время в прикладной математике используется значительное количество первичных понятий, которые либо строго определяются, либо для них формулируются конструктивные описания, позволяющее строить их математическую теорию. К первым можно отнести понятия “точка”, “прямая”, “случайная величина”, а ко вторым — понятие “множество”.

Отсутствие конструктивного описания понятий “предметная область” и “объект” обуславливает возникновение проблемы разработки эффективных методов определения степени адекватности баз данных предметным областям и их объектам, создает проблемы построения эффективных алгоритмов информационной поддержки процедур принятия решений, при этом усложняются методы устранения противоречий в базах данных, возрастает вычислительная сложность интеллектуализации баз данных, становится весьма проблематичным раскрытие априорных неопределенностей полноты знаний о предметной области.

Совокупность всех объектов ПО образует объектное ядро. Объекты ядра динамически изменяются и взаимодействуют между собой, что, в общем случае, носит стохастический характер. Известны определения объектного ядра и его особенностей:

“Объектное ядро произвольной предметной области потенциально содержит бесконечно много объектов, которые находятся в потенциально бесконечном множестве ситуаций. При этом одни объекты могут исчезать, другие появляться, меняются взаимосвязи между объектами, возникают новые отношения с новыми свойствами. Однако, несмотря на этот вихрь изменений, “что-то” позволяет говорить о состояниях одной и той же ПО. Эти состояния нужно суметь отобразить в памяти вычислительной системы для того, чтобы с ее помощью эффективно осуществлять информационный поиск и получать новые непротиворечивые знания” [1].

Из сказанного не ясно, каким образом многообразие изменений в предметной области не приводит к выходу за пределы данной предметной области. Следовательно, требуется разработка математических методов конструктивного описания и представления объектных ядер и их отдельных объектов.

Объекты в соответствии с законами выделенной ПО могут образовываться из более простых. Вводится понятие “элементарный объект” — объект, который в заданной системе

ограничений, вытекающей из описания предметной области, не может быть разложен на более простые. Предполагается, что элементарный объект характеризуется следующими свойствами:

- не может быть разложен на более простые объекты в рамках выделенной ПО и интересов всех пользователей;
- обладает некоторой внутренней структурой, недоступной проектировщику базы данных и не представляющей интереса для пользователя;
- характеризуется набором вида:

$$O = \{ \mathbf{X}, \mathbf{Y}, S, F, P_c, P_{dyn} \},$$

где  $\mathbf{X}$  и  $\mathbf{Y}$  — множества (в общем случае) входных  $(x_1, \dots, x_i, \dots, x_r)$  и выходных  $(y_1, \dots, y_l, \dots, y_m)$  переменных (атрибутов), имеющих четкий содержательный смысл, и для которых существует приемлемая формализация, при этом  $X \cap Y = 0$ ; при создании БД приходится ограничиваться конечными множествами входных и выходных переменных, включающих наиболее существенные свойства;

$S$  — множество состояний элементарного объекта с предельным законом перехода объекта из одного состояния в другое; будем предполагать, что для разных объектов одного типа множество  $S$  определяется однозначно; в общем случае это достаточно жесткое ограничение, вводимое с целью упрощения модели объекта;

$F$  — передаточная функция, определяемая классическим способом; с целью упрощения задачи полагается, что она одна и та же для всех однотипных объектов;

$P_c$  — предикат, определяющий семантические ограничения, накладываемые на состояния объекта; как отмечалось, с определенной степенью приближения можно утверждать, что все объекты одного типа имеют одно и то же множество состояний; следовательно, в общем случае предикат постоянен на этом множестве, хотя в каждом конкретном случае он может существенно отклоняться от некоторого своего стандарта;

$P_{dyn}$  — предикат, отражающий семантические ограничения целостности траектории объектов, в общем случае может изменяться для различных объектов.

Необходимо учитывать, что каждая из входных  $\mathbf{X}$  и выходных  $\mathbf{Y}$  переменных изменяется в одной из общепринятых шкал:

- интервальной;
- отношений;
- бальной;
- порядковой;
- номинальной.

При проектировании БД предполагается, что независимо от выбранной шкалы измерения большинство переменных являются случайными величинами либо относятся к классу случайных процессов. Подразумевается, что существует общий подход к построению метода преобразования из одной шкалы измерения в любую другую — оно должно быть однозначным и обратимым, что в настоящее время достаточно сложно. Функция преобразования из одной шкалы в другую БД построена на конечной выборочной совокупности, может быть продолжена за пределы экспериментальной (выборочной) области определения со значительной вероятностью сохранения закономерностей динамики изменения соответствующих переменных.

В общем случае, в математическую модель элементарного объекта целесообразно включать возмущающие воздействия внешней среды. Любой элементарный объект взаимодействует со средой, которая имеет стохастический характер [2, 3]. Поэтому в модель целесообразно ввести стохастическое возмущение в виде вектора  $\mathbf{Z} = (z_1, \dots, z_j, \dots, z_n)$ , каждая компонента которого соответствует случайному процессу возмущающего воздействия внешней среды. Возмущения могут носить характер чистых помех, например, в виде Гауссовского шума либо в виде случайных процессов с достаточно точно описываемыми свойствами.

Возмущающие воздействия можно не учитывать, если они носят стационарный характер и не оказывают существенного влияния на динамические свойства элементарного объекта.

Соответственно, в общем случае элементарный объект (см. рисунок) описывается набором

$$O = \{X, Y, Z, S, F, P_{\tilde{n}}, P_{dyn}\}.$$

Входы  $x_{k1}, \dots, x_{ki}, \dots, x_{kr}$  соответствуют информации, поступающей на вход  $k$ -го от других объектов — элементарных или имеющих сложную структуру.

Совокупность всех элементарных объектов и определенное множество операций над объектами, с помощью которых можно из объектов низшего уровня строить более сложные, будет называться *базисом объектного ядра*. Построение объекта любой степени сложности основано на базисе объектного ядра. Каждый такой объект можно представить в виде графа  $G(O, A)$ , где  $O = \{O_k\}$  — подмножество элементарных объектов базиса,  $A = X \cup Y$  — множество связей между ними в границах объекта ПО [4].

Элементарные объекты в общем случае имеют внутреннюю структуру, которая для заданного уровня анализа и моделирования предметной области не существенна.

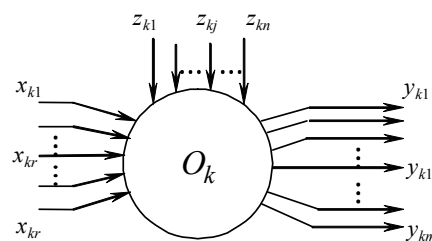
Рассматривая приведенное представление элементарного объекта, необходимо отметить, что априорная информация о нем включает многообразие неопределенностей различной глубины и степени. Остановимся на некоторых наиболее проблематичных неопределенностях.

При разработке БД отсутствует достаточно полная информация о множестве входных и выходных переменных (атрибутов), о степени их важности в описании динамики изменения объектов, зависимости между ними и о закономерностях изменения каждого из выделенных атрибутов как функции времени или переменных другой природы. В некоторых случаях неопределенность может быть настолько глубокой, что невозможно получение данных о структуре кластеров объектов, что не позволяет выделить ее отдельные компоненты, отображающие общие закономерности изменения объектного ядра, при которых не выходим за пределы предметной области.

Три класса неопределенностей связаны с векторами переменных  $(z_1, \dots, z_j, \dots, z_n)$ ,  $(x_1, \dots, x_i, \dots, x_r)$ ,  $(y_1, \dots, y_l, \dots, y_m)$ . Наибольшие неопределенности связаны с возмущающими воздействиями внешней среды. Ряд возмущений носит настолько размытый характер, что существует проблема построения функций, описывающих структуру распределения размытых возмущающих воздействий. Измерения возмущающих воздействий принято считать лежащими за пределами допустимости получения их статистических оценок. Если какое-либо возмущающее воздействие  $z_j$  принадлежит к классу случайных процессов  $\xi(t)$ , то существуют значительные проблемы построения его математической модели на основе данных динамических наблюдений.

Входные и выходные переменные могут измеряться в различных шкалах. Если они относятся к качественному или номинальному типу, то построение мер метрик в пространстве атрибутов требует приведения всех входных и выходных переменных к количественной шкале. Одна из проблем, связанная с входными и выходными переменными, заключается в априорной неопределенности зависимости между столь различными шкалами измерения. Она может быть известна только на основе информации, накопленной в базе данных. С другой стороны, процесс накопления данных является эффективным только при условии, что база данных спроектирована корректно.

Отсутствие априорной информации о шкалах измерения переменных не позволяет строить адекватные меры подобия и метрики, что обуславливает существенную неопределенность априорного анализа структуры множества элементарных объектов. Для решения задачи уменьшения неопределенности предлагается использовать методы последовательного анализа.



Структурное представление элементарного объекта

Например, возможно создание итерационных алгоритмов построения метрик и подобия по мере накопления информации в базах данных.

Необходимо заметить, что один из классов неопределенностей связан с отсутствием достаточно точной информации о степени важности входных и выходных переменных, а кроме того, о степени полноты предоставления элементарных объектов выделенными множествами входных и выходных переменных.

Отмеченные неопределенности могут быть устранены или раскрыты при создании методов проектирования и корректировки баз данных по мере их развития. Для этого предлагается использовать аппарат итерационного последовательного анализа и систем моделирования семантики предметных областей в формируемых базах данных на основе управления процессами их построения.

### **Литература**

1. Цаленко М.Ш. Моделирование семантики в базах данных. — М.: Наука, 1989. — 288 с.
2. Раскина А.А. Семантические основания объектно-признаковых языков / Раскина А.А., Сидоров И.С., Шредер Ю.А. // Науч.-техн. информация. Сер. 2. — 1976. — № 5. — С. 18 — 25.
3. Миркин Б.Г. Анализ качественных признаков и структур. — М.: Статистика, 1980. — 318 с.
4. Малахов Е.В. Оценка степени адекватности баз данных как информационных моделей предметных областей // Тр. Одес. политехн. ун-та. — Одесса, 2004. — Вып. 1(21). — С. 82 — 86.

Поступила в редакцию 7 февраля 2005 г.

---