

## Свойство однозначности построения структур данных

77-30569/243060

№ 10, октябрь 2011

автор: Виноградова М. В.

УДК 004.6

МГТУ им. Н.Э. Баумана

[vinogradova.m@gmail.com](mailto:vinogradova.m@gmail.com)

### Введение

Рассмотрим информационные системы, которые состоят из реляционной базы данных и пользовательского интерфейса к ней. Это широкий класс систем, включающий автоматизированные системы управления и информационно-справочные системы организаций и их подразделений, например, кафедры вуза [2], деканата или методического управления, а также системы организационного управления компаний и агентств [1,8].

Существует несколько причин увеличения затрат на разработку и модификацию систем указанного класса. Рассмотрим их подробнее.

В процессе эксплуатации подобных систем периодически появляется необходимость в наращивании их функциональных возможностей. При этом добавляются прикладные задачи, использующие как старые, так и новые данные. Расширение структуры единой базы данных может потребовать модификации существующих задач, что влечет дополнительные затраты [4].

Для сокращения затрат [9] на разработку и последующую модификацию проводится декомпозиция системы на небольшие и несложные подсистемы, каждая из которых разрабатывается независимо [10]. Поскольку подсистемы обращаются к общим данным, то необходимо согласовывать используемые в них структуры данных. В случае сильной связанности подсистем по данным процесс согласования потребует существенных затрат [6].

По мере вовлечения в автоматизацию различных сфер деятельности организаций и их подразделений происходит слияние или совместная работа различных систем [3].

Происходит объединение баз данных, которое ведет к необходимости согласования и корректировок структур данных, используемых прикладными задачами, что увеличивает затраты.

Поскольку дополнительные затраты на разработку и модификацию указанных выше систем вызываются необходимостью согласования структур данных различных задач, систем и подсистем, то необходимо разработать способ снижения затрат на согласование структур данных.

### **Методы снижения затрат на согласование структур данных**

Сформулируем требования к построению структуры базы данных. Во-первых, описание структуры данных всей системы должно объединять описания структур данных всех задач системы. Во-вторых, расширение структуры данных системы при добавлении новой задачи не должно приводить к изменению существующей структуры данных. В-третьих, параллельная разработка пересекающихся фрагментов структуры базы данных не должна требовать последующего согласования их общих частей.

Известны следующие варианты решения поставленных задач:

- использование представлений, в том числе обновляемых,
- использование хранимых процедур и функций,
- использование хранилищ данных или виртуальных баз данных,
- использование интерфейсов между базами данных.

Применение представлений, хранимых процедур и функций [3] позволяет избежать модификации существующих приложений при добавлении задач или изменении структуры базы данных, но не влияет на проблему согласования структур при объединении баз данных. А также требует модификации кода представлений, процедур и функций при изменении структур базы данных. Затраты на согласование структур данных снижаются только частично.

Применение хранилищ данных и виртуальных баз данных [11] позволяет использовать унаследованные системы, а также иметь единую базу данных для реализации новых прикладных задач. Однако, их создание тоже требует затрат. При этом проблемы расширения самих унаследованных систем никак не решаются, затраты на согласование не уменьшаются.

Создание интерфейсов между базами данных унаследованных систем [11] требует как единовременных затрат при создании интерфейсов, так и постоянных затрат для их согласования при расширении функционала систем. Процесс согласования структур данных централизуется и затраты на него снижаются, но только частично.

В качестве альтернативного решения предлагается формализовать процесс построения структуры базы данных таким образом, чтобы на основе одних и тех же сведений о предметной области, при независимом проектировании одинаковых задач были получены идентичные описания структур данных. Таким образом процесс согласования структур данных становится практически ненужным и затраты на него существенно снижаются. Для получения идентичных описаний при независимой разработке процесс построения структуры данных задачи должен быть однозначным.

### Определение свойства однозначности построения структур данных

Дадим формальное определение свойства однозначности построения структур данных и сформулируем требования его достижения.

Структура данных задачи содержит сведения об объектах, их реквизитах и связях, которые необходимы для выполнения задачи. Для реализации базы данных на реляционных СУБД структура данных задачи должна быть приводима к реляционной модели представления данных. В реляционной модели структура данных задается с помощью множества схем отношений [3]. Каждая схема отношения задается набором атрибутов и множеством функциональных зависимостей. Исходными данными для построения реляционной модели являются множество реквизитов предметной области и множество функциональных зависимостей. Реквизиты – это поименованные характеристики объектов предметной области. Функциональные зависимости – это зависимости между реквизитами, такие, что при совпадении определяющих реквизитов зависимостей зависимые реквизиты будут также совпадать. Рассмотрим процесс построения структуры данных для двух задач.

Пусть разрабатываются структуры данных для некоторых задач  $Z_i$  и  $Z_j$ . Исходными данными для разработки структур данных являются множества реквизитов  $R^i$  и  $R^j$  и множества функциональных зависимостей  $F^i$  и  $F^j$ , определенные на соответствующих множествах реквизитов. На основе множеств реквизитов и функциональных зависимостей для каждой из задач составляются описания структур данных. Описания структур данных задаются множествами описаний переменных:  $D^i$  для задачи  $Z_i$  и  $D^j$  для задачи  $Z_j$ . Для составления описаний переменных используются преобразования

$$D^i = p^{RF}(R^i, F^i) \text{ и } D^j = p^{RF}(R^j, F^j)$$

, где  $p^{RF}$  - преобразование множеств реквизитов и функциональных зависимостей к множеству структур данных.

Процесс разработки информационной системы обладает свойством однозначности построения структур данных, если при независимой разработке задач информационной системы, реализующих одни и те же объекты предметной области, будут получены одинаковые описания структур данных для этих задач:

$$D^{ij} \equiv p^{RF} (R^{ij}, F^{ij})$$

$$\text{где } D^{ij} = D^i \cap D^j, R^{ij} = R^i \cap R^j, F^{ij} = F^i \cap F^j.$$

Однозначность построения структур данных означает, что эквивалентные подмножества реквизитов и функциональных зависимостей должны быть отображены на эквивалентные подмножества структур данных. То есть, если любые две задачи  $Z_i$  и  $Z_j$ , имеют множество общих реквизитов  $R^{ij} = R^i \cap R^j$  и множество общих функциональных зависимостей  $F^{ij} = F^i \cap F^j$ , то структуры данных этих компонентов  $D^i = p^{RF} (R^i, F^i)$  и  $D^j = p^{RF} (R^j, F^j)$  должны иметь общие элементы  $D^{ij} = D^i \cap D^j$ , причем,  $D^{ij} \equiv p^{RF} (R^{ij}, F^{ij})$ .

Основными проблемами, не позволяющими достичь однозначности при построении структур данных, являются: неоднозначность выделения реквизитов, неполнота выявления функциональных зависимостей и нестрогая формализация процесса построения структуры данных.

Неоднозначность выделения реквизитов появляется, если в различных задачах используются одни и те же характеристики объектов предметной области. При этом, в одной из задач на основе характеристик выделяют множество простых реквизитов, а в другой задаче – один составной реквизит. Для устранения проблемы неоднозначности выделения реквизитов следует выделять атомарные реквизиты. Атомарный реквизит – это поименованная характеристика объекта предметной области, которая не может быть представлена как композиция каких-либо других характеристик.

Неполнота выявления функциональных зависимостей появляется из-за того, что в различных задачах являются значимыми различные зависимости между одними и теми же объектами предметной области. Следствием неполноты выявления функциональных зависимостей является различие в структурах данных, описывающих одни и те же

объекты в разных задачах. Для устранения этой проблемы следует при построении структуры данных учитывать все зависимости, которые существуют между объектами, рассматриваемыми в задаче.

Нестрогая формализация процесса построения структуры данных приводит к тому, что на основе одних и тех же исходных данных может быть построено несколько различных структур данных. Например, при построении реляционной модели базы данных задача приведения схемы отношений к третьей нормальной форме в общем случае имеет более одного решения [7]. Для формализации процесса следует использовать однозначное преобразование множеств реквизитов и функциональных зависимостей к множеству структур данных. Кроме того, структура данных, полученная для подмножества атрибутов, должна являться подмножеством структуры данных, полученной для всех атрибутов:

$$\text{если } R^i \subset R^j \text{ и } F^i \subset F^j, \text{ то } p^{RF}(R^i, F^i) \subset p^{RF}(R^j, F^j)$$

Назовем это требование условием сохранения отношения включения.

Исходными данными для преобразования должны быть множество атомарных реквизитов и множество функциональных зависимостей, определенное на множестве атомарных реквизитов. Других исходных данных для преобразования не должно быть.

#### **Лемма об однозначности построения структур данных**

Докажем лемму об однозначности построения структур данных.

Для достижения свойства однозначности построения структур данных необходимо и достаточно провести декомпозицию на атомарные реквизиты, указать полное множество функциональных зависимостей, определенных на множестве атомарных реквизитов, и применить однозначное преобразование множеств реквизитов и функциональных зависимостей к множеству структур данных, при котором сохраняется отношение включения:

$$\{ \forall R_i \in R \exists R^j \subseteq R : R_i = f(R^j), \forall R^i \subseteq R \forall F_j(R^i) : F_j \in F,$$

$$\forall R^i \subseteq R, \forall F^i \subseteq F \exists_1 D_i : D_i = p^{RF}(R^i, F^i(R^i)),$$

$$p^{RF} : p^{RF}(R^k, F^k) \subset p^{RF}(R^j, F^j), \text{ если } R^k \subset R^j, F^k \subset F^j \} \Rightarrow$$

$$\forall R^i, R^j \subseteq R, \forall F^i, F^j \subseteq F \quad p^{RF}(R^i \cap R^j, F^i \cap F^j) \equiv p^{RF}(R^i, F^i) \cap p^{RF}(R^j, F^j)$$

где  $R$  - множество атомарных реквизитов;

$F$  - полное множество функциональных зависимостей;

$F^j(R^j)$  - множество функциональных зависимостей, определенное на множестве реквизитов  $R^j$ ;

$D_i = p^{RF}(R^i, F^i)$  - структура данных, построенная на основе множества реквизитов  $R^i$  и множества функциональных зависимостей  $F^i$ ;

$p^{RF}$  - однозначное преобразование множеств реквизитов и функциональных зависимостей к множеству структур данных.

Доказательство: Пусть необходимо построить структуры данных для двух задач, использующих одни и те же реквизиты предметной области. Для задач будут определены множества атомарных реквизитов  $R_i$  и  $R_j$  и множества функциональных зависимостей

$F_i$  и  $F_j$ . Из атомарности реквизитов следует, что для обеих задач множество общих

реквизитов будет определяться как  $R^{ij} = R_i \cap R_j$ . Множество функциональных

зависимостей  $F^{ij}$ , определенное на множестве общих реквизитов  $R^{ij}$ , будет

определяться как  $F^{ij}(R^i \cap R^j) \equiv F^i(R^i) \cap F^j(R^j)$ . Это следует из полноты

функциональных зависимостей и однозначности выявления реквизитов. Для задач

построены множества структур данных  $D^i = p^{RF}(R^i, F^i)$  и  $D^j = p^{RF}(R^j, F^j)$ .

С помощью преобразования  $p^{RF}$  на основе общих атомарных реквизитов  $R^{ij}$  и общих функциональных зависимостей  $F^{ij}$  может быть построено множество

$D^{ij} = p^{RF}(R^{ij}, F^{ij})$ . Поскольку  $R^{ij} \subseteq R^i$  и  $F^{ij} \subseteq F^i$ , то  $D^{ij} \subseteq D^i$ , - это следует

из однозначности преобразования и выполнения условия включения. Аналогично,

$D^{ij} \subseteq D^j$ . Предположение, что существует структура данных  $D_k \in D^{ij}$  и, при этом,

$D_k \notin D^i$  или  $D_k \notin D^j$ , будет истинным только при неоднозначности преобразования

или неполноте функциональных зависимостей. Следовательно, оно неверно. Если

предположить, что существует структура данных  $D_k \in (D^i \cap D^j)$  и при этом

$D_k \notin D^{ij}$ , то получим утверждение, что на основании различных множеств реквизитов и функциональных зависимостей будут получены идентичные структуры данных. Это утверждение неверно, поскольку структура данных содержит поименованные описания реквизитов. Следовательно,  $D^{ij} \equiv D^i \cap D^j$ . Доказательство леммы завершено.

### Заключение

Для снижения затрат на разработку и модификацию информационных систем, использующих реляционную базу данных, процесс построения структуры базы данных должен обладать свойством однозначности построения структур данных. Дано определение указанного свойства. Определены требования, необходимые для достижения свойства. Сформулирована и доказана лемма о том, что для достижения свойства однозначности построения структур данных необходимо и достаточно провести декомпозицию на атомарные реквизиты, указать полное множество функциональных зависимостей на множестве атомарных реквизитов и применить однозначное преобразование множеств реквизитов и функциональных зависимостей к множеству структур данных, при котором выполняется условие включения.

## Литература

1. Вендров А.М. Проектирование программного обеспечения экономических информационных систем. – М.: Финансы и статистика, 2000. – 352 с.
2. Виноградова М.В. Методика проектирования информационных систем с возможностью наращивания и быстрой модернизации на примере ИС кафедры вуза // Вестник информационных технологий в образовании: Сборник учебно-методических и научных работ (М.). - 2005. - Вып.1.– С.100-118.
3. Дейт К. Введение в системы баз данных. - М.: Издательский дом Вильямс, 2002. – 1072 с.
4. Котов С.Л. Нормирование жизненного цикла программной продукции. – М.: ЮНИТИ-ДАНА, 2002. – 143 с.
5. Лавров С.С. Программирование. Математические основы, средства, теория. – СПб.: БХВ-Петербург, 2001. – 320 с.
6. Липаев В.В. Обеспечение качества программных средств. Методы и стандарты. – М.: СИНТЕГ, 2001. – 380 с.
7. Мамиконов А.Г., Кульба В.В., Косяченко С.А. Оптимизация структур распределенных баз данных в АСУ. – М.: Наука. Главная редакция физ.-мат. литературы, 1990. – 240 с.
8. Системы автоматизации документооборота – от выбора до внедрения: Методические рекомендации. – М.: Компания Интер Траст, 2001. – 23 с.
9. Технологии разработки программного обеспечения: Учебник. / Орлов С.А. – СПб.: Питер, 2002. – 464с.
10. Якобсон А., Буч Г., Рамбо Дж. Унифицированный процесс разработки программного обеспечения. – СПб.: Питер, 2002. – 496 с.
11. Гарсиа-Молина Г., Ульман Д.Д., Уидом Д. Системы баз данных. Полный курс. – М.: Издательский дом «Вильямс», 2004. – 1088 с.

electronic scientific and technical periodical  
**SCIENCE and EDUCATION**

EL № FS 77 - 30569. №0421100025. ISSN 1994-0408

---