

## ОПИСАНИЕ И ПРЕДСТАВЛЕНИЕ ОБЪЕКТОВ В ХРАНИЛИЩАХ ДАННЫХ ВО МНОЖЕСТВЕ ПРЕДМЕТНЫХ ОБЛАСТЕЙ

Малахов В.Е., к.т.н., доцент

Одесский национальный политехнический университет, [mev@utecon.com](mailto:mev@utecon.com)

*Summary. The technique of creation of information model of a subject domain is offered which allows to decide administrative tasks and to accept the reasonable decisions on the basis of the analysis of real objects of a researched subject domain and interrelations between them.*

**Постановка проблемы.** Для развития представлений о выделяемых предметных областях реального мира и формирования отношений между множествами областей в условиях многообразия объектов решающее значение приобретают методы отображения свойств и связей в хранилищах данных. Эффективное прогнозирование изменения состояний или управление развитием некоторой предметной области (ПрО) должно основываться на достаточно полной информации о свойствах сущностей или объектов, составляющих множество, о динамических закономерностях, которые присущи конкретной предметной области. При этом априорные представления о предметных областях часто настолько ограничены, что необходимо вначале накопить достаточный объем информации, а затем путем анализа и ее обработки увеличить степень полноты, непротиворечивости и определенности.

**Анализ публикаций.** Спецификация Common Warehouse Metamodel (Общая метамодель Хранилища данных, далее CWM) - это стандарт, который описывает обмен метаданными при использовании технологий хранилищ данных, Business Intelligence, Knowledge Management (Управление знаний). Опираясь на базовую метамодель, стандарт добавляет метамодели для реляционных, многомерных данных и данных таблиц, а также для преобразования, функций OLAP, data mining и Хранилища данных, включая процессы и операции [1]. Метамодель (модель модели данных) представляет метаданные из области бизнеса и техники, которые в большинстве случаев присутствуют в технологии Хранилищ данных и бизнес аналитики. Она используется как основа для обмена экземплярами метаданных между гетерогенным программным обеспечением, поставляемым различными производителями. Системы, которые идентифицируют метамодель CWM, обмениваются данными в форматах, которые согласуются с этой моделью.

Можно сказать, что CWM расширяет язык UML в том смысле, что каждый метакласс (metaclass) CWM наследуется напрямую, либо непрямою из метаклассов UML. Например, метакласс "Реляционная Таблица (Relational Table) CWM" является непосредственным наследником Класса UML (UML Class), а "Реляционный Столбец (Relational Column)" - прямой потомок Атрибута UML (UML Attribute). Таким образом, CWM можно характеризовать как язык определенной области применения, предназначенный для определения моделей Хранилищ данных.

Другой стандарт OMG – Meta Object Facility (Средство метаобъекта, далее MOF) – определяет общие интерфейсы и семантику для взаимодействующих метамodelей. MOF – это пример модели метамодели (подмножество UML). Он также определяет набор IDL-преобразований (Interface Definition Language, язык описания интерфейса,

который устанавливает спецификацию интерфейса для обнаружения и управления моделями с помощью программных APIs).

Третий стандарт, который непосредственно задействован в обмене метамоделями, – XMI (XML Metadata Interchange, Обмен метаданными XML) – это стандарт OMG, который устанавливает правила преобразования метамodelей MOF в XML. XMI определяет, как использовать XML-теги для представления сериализованных моделей, совместимых с MOF. Метамодели MOF транслируются в XML DTD, а модели – в XML-документы, которые согласуются со своими DTD [1].

Каждая метамодель CWM представляется как XML DTD (в соответствии с правилами XMI), так и определение IDL. В первом случае модели CWM преобразуются в поток (serialize), после чего ими обмениваются, как документами XMI. При экспорте метаданные посредством XMI-документа, необходимо выполнить XMI-преобразование (MXI-rendering) в форме, которая легальна по отношению к DTD. При импорте данных с помощью XMI-документа, следует проверять модель на допустимость по этим DTD.

CWM фактически состоит из ряда составных метамodelей (или суб-метамodelей), которые организованы в виде 4 структур: базовой (Foundation), источников данных (Resources), подсистем анализа (Analysis) и управления Хранилищем (Management).

Уникальность хранилища характеризуется структурой источников данных, которая предоставляет возможность моделировать существующие и новые источники данных предметной области (PrO), в том числе реляционные базы данных, ориентированные на запись базы данных (record oriented databases), а также XML- и основанные на объектах (object-based) источники данных.

**Цель исследований.** Развитие методов моделирования источников данных для обеспечения полноты описания PrO в структурах источников данных хранилищ на основе определения конечных границ PrO.

**Представление объектов PrO.** Каждый объект можно представить многомерной структурой, описывающей его свойства, касающиеся различных аспектов его деятельности или характеризующие его в рамках той или иной классификации. Подобной структурой 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 рассматриваемой классификации. Соответственно, описание каждой сущности с учетом всех связей будет представлять собой куб размерности  $N+1$ , где  $N$  — максимальное число классификаций конкретной сущности (рис. 1).

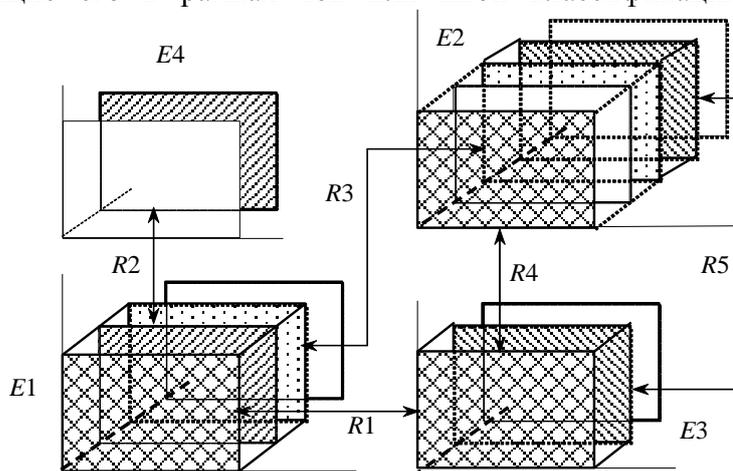


Рис. 1. Гиперкуб комплексной структуры предметных областей

Таким образом, предметная область будет представлять собой множество связанных кубов различной размерности или гиперкуб, описанный некоторым

формальным языком  $L_F$ , множества сущностей, описанных в терминах одного естественного языка  $L_N$  (см. рис. 1).

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

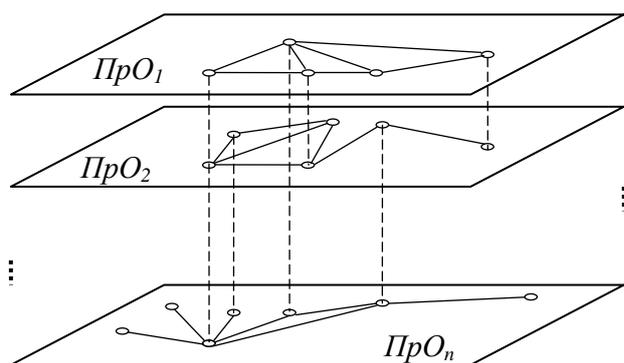


Рис. 2. Граф комплексной модели предметных областей

В графовом представлении такая модель будет выглядеть так, как показано на рис. 2. При таком представлении видно, что каждая плоскость (таблица), описывающая соответствующий объект реального мира (рис. 1), представляет собой проекцию этого объекта на соответствующую предметную область  $PrO_i$  (рис. 2).

Соответствующие математические операции над графами, описывающими предметные области (или подобласти некой предметной области) позволяют:

- получить подобную многомерную структуру (например, операцией *объединения*), что, в свою очередь, составит более целостную картину реального мира и даст возможность „увидеть” объекты, проекции которых, возможно, были „потеряны” в других плоскостях;
- выделить (например, операцией *пересечения*) те объекты, информация о которых может быть вынесена в информационное хранилище того или иного уровня;
- синтезировать новую (виртуальную) предметную область, „населённую” только теми объектами, которые интересуют исследователя.

**Выводы.** Предложена методика создания информационной модели предметной области, которая позволяет решать управленческие задачи и принимать обоснованные решения на основании анализа реальных объектов исследуемой предметной области и взаимосвязей между ними.

### Литература

1. Спецификация "Общая метамодель Хранилища данных" - Common Warehouse Metamodel (CWM). – <http://xmlhack.ru/texts/common.warehouse.model/cwm.html>.
2. 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.

*Рецензент: д.т.н., профессор Шапанов В.М.*