

Инфокоммуникационные ТЕХНОЛОГИИ

УДК 681.3

Об одном из методов разработки концептуальной объектно-ориентированной многомерной метамоделю информационно-аналитической СИСТЕМЫ

А. В. Висков

*Кафедра информационных технологий
Российский университет дружбы народов
ул. Миклухо-Маклая, 6, Москва, Россия, 117198*

Объектно-ориентированный подход предоставляет существенные преимущества в многомерном моделировании. В статье разбираются вопросы разработки многомерной метамоделю информационно-аналитической системы. При этом предлагается широко использовать методы, основанные на традиционных технологиях в области многомерного анализа данных: Meta Object Facility (MOF), Common Warehouse Metamodel (CWM), Unified Modeling Language (UML).

Ключевые слова: информационно-аналитические системы, многомерные модели данных, объектно-ориентированные модели, OLAP, MOF, CWM, UML.

1. Введение

Выбор многомерной модели данных для описания предметной области является одним из наиболее удобных. Такое представление данных не накладывает серьёзных ограничений на область применения. В мире существует богатый опыт удачного многомерного моделирования множества предметных областей: банковская деятельность, страхование, производственная деятельность, торговля, электронная коммерция и др. [1, 2]. Многомерное (многофакторное) представление данных является наиболее распространённым и широко используемым в области анализа данных, т.е. является стандартом де-факто.

Понятия «выразительность» или «семантическая мощь» характеризуют степень, в которой модель может выразить или представить концепции реального мира [3]. Они определяют возможности структур модели представлять концептуальные понятия и быть правильно интерпретированными. Чем более выразительной является модель, тем лучше она представляет реальный мир и тем больше информации предоставляет пользователю. Объектно-ориентированные (ОО) модели семантически более богаты, чем другие (например, E/R или реляционные). Конечно, всегда можно найти возможность обогатить некоторую модель объектно-ориентированными особенностями, но зачем делать это, если можно просто использовать ОО модель.

Возможно, одним из наиболее важных преимуществ моделирования предметной области с помощью ОО моделей является то, что результаты работы оказываются наиболее близким к концепциям и пониманию конечного пользователя, то есть они в действительности отражают способ мышления людей. Каждый моделируемый объект или класс будет иметь соответствие с некоторой реальной

Статья поступила в редакцию 26 мая 2008 г.

Работа выполнена в рамках тематики подпрограммы «Управление инфокоммуникациями» Инновационной образовательной программы, выполняемой РУДН в составе национального проекта «Образование».

сущностью, что облегчает понимание модели. Кроме того, ОО моделирование данных облегчает решение некоторых специфичных задач, таких как проектирование распределённых объектных систем.

ОО моделирование не требует обязательной нормализации сущностей. Можно проектировать объекты, содержащие не атомарные значения. В некоторых случаях это может оказаться действительно ценным по причинам, связанным с вопросами производительности, или потому, что концептуально нет необходимости создавать сущности, не имеющие реального соответствия, только для того, чтобы нормализовать схему.

Цель этой статьи состоит в том, чтобы показать возможность использования преимуществ ОО моделирования для построения многомерных моделей. Преимущества ОО модели данных заключаются в интеграции различных многомерных представлений и в сохранении семантики данных на концептуальном уровне. В случае, если пользователь хочет использовать другую модель, ОО модель всегда может быть преобразована в необходимую. Ниже будет приведено описание разработанной ОО модели концептуального уровня, описывающей основные понятия многомерного моделирования.

Задачей работы является разработка концептуальной модели, которая могла бы стать основой для построения аналитических информационных систем (ИАС) и хранилищ данных (ХД) [4, 5]. Создаваемая модель должна стать языком для описания метаданных в аналитических системах и ХД, следовательно, сама должна являться формальной математической моделью.

Разрабатываемая модель должна иметь строго определённый синтаксис, семантику и, возможно, правила анализа, вывода или доказательства его конструкций. Синтаксис может быть графическим или текстовым. Семантика может быть определена более или менее формально в терминах понятий, существующих в предметной области, или путём транслирования конструкций языка более высокого уровня в другие конструкции, которые имеют строго определённое значение.

2. Стандарты и спецификации, использованные при разработке модели

Разрабатываемая метамодель основана на спецификации CWM [6, 7] и является расширением метамодели CWM на языке UML [8]. Метамодель CWM, в свою очередь, определена на языке UML (рис. 1).

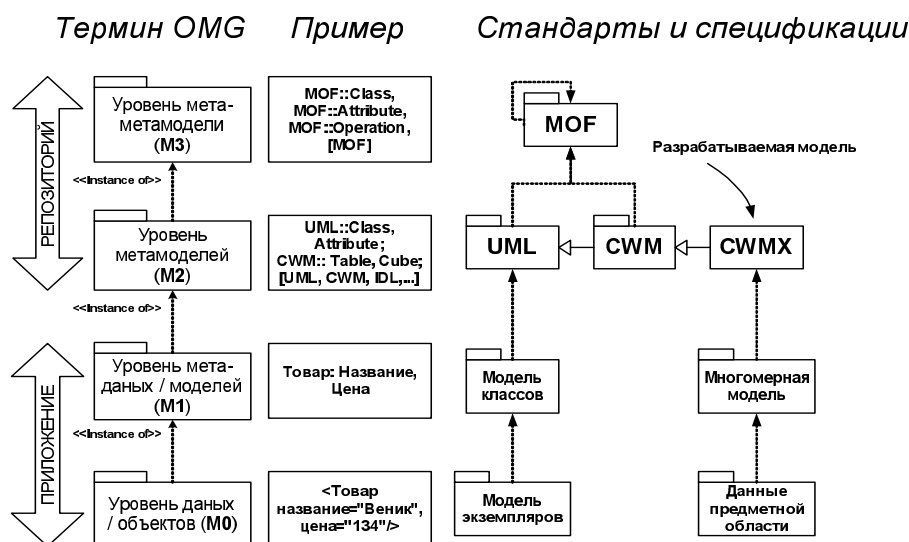


Рис. 1. Используемые стандарты и спецификации

Создавая метамодель как расширение CWM, мы получаем возможность использовать спецификацию MOF [9], что позволит с минимальными трудозатратами получить формальную математическую модель (язык) для описания и обмена метаданными в среде аналитических систем и хранилищ данных. При этом спецификация MOF даёт средства для описания семантики разрабатываемой метамодели (как язык более высокого уровня), UML даёт графическую нотацию и набор базовых конструкций для объектно-ориентированного моделирования.

Далее будут подробно рассмотрены менее распространённые и практически не описанные на русском языке спецификации MOF, которые активно используются при создании модели метаданных информационно аналитических систем. Именно благодаря использованию этих спецификаций разрабатываемая метамодель может считаться формальным языком для описания метаданных.

3. Спецификация Meta Object Facility (MOF)

В качестве языка описания моделей метаданных консорциум OMG предлагает спецификацию Meta Object Facility (MOF) [6]. Спецификация MOF предусматривает описание метаданных с помощью объектно-ориентированных методик. Именно такие метаданные рассматриваются далее в этой работе. Такие метаданные могут описывать любой аспект системы или содержащейся в системе информации на любом уровне детализации и точности в зависимости от решаемой задачи. Выбор объектно-ориентированного подхода не накладывает никаких явных ограничений на множество описываемых систем и их метаданных. MOF — это формальная спецификация, она задаёт формальный язык для описания моделей метаданных. В качестве графической нотации в MOF используется язык объектно-ориентированного проектирования UML [10, 11].

MOF представляет собой основанную на моделях распределённую объектную среду для специфицирования, создания, управления, обмена и интеграции метаданных в программных системах. Целью этой среды является поддержка всех видов метаданных и обеспечение возможности добавления новых видов метаданных при необходимости. Для этого используется четырёхуровневая архитектура метаданных, называемая иногда архитектурой метаданных OMG (табл. 1). В этой архитектуре нижний уровень (M0) является уровнем данных. Метаданные (M1) рассматриваются как данные для (M2). При этом производится формальное моделирование каждого типа метаданных. Полученные формальные модели называются метамоделями (M2) и выражаются средствами единой мета-метамодели (M3), называемой Моделью MOF (с большой буквы).

Таблица 1

Архитектура метаданных OMG [12, 13]

| Мета-уровень | Название согласно MOF | Примеры |
|--------------|--------------------------------|---|
| M3 | Мета-метамодель | Модель MOF |
| M2 | Метамодель; Мета-метаданные | MOF метамодели, в частности: разрабатываемая метамодель; метамодель CWM; метамодель UML. |
| M1 | Модель | UML модели |
| M0 | Объект; Данные | Моделируемые системы; Данные в ХД |

На рис. 2 приведён пример архитектуры метаданных, поясняющий отношения между Моделью MOF (мета-метамодель), метамоделью UML (один из примеров метамодели), пользовательской моделью (пример метаданных) и пользовательскими объектами (пример данных).

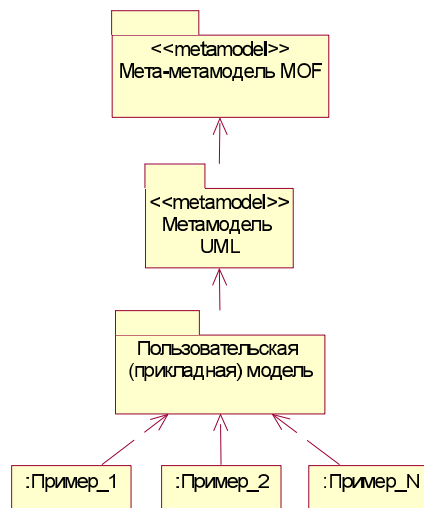


Рис. 2. Пример архитектуры метаданных OMG

Таким образом, благодаря MOF, достаточно четырёх уровней абстракции (мета-уровней) для моделирования любых данных и метаданных. Модель MOF находится на верхнем уровне и описывает саму себя. Любая модель метаданных (метамодель), в том числе и разрабатываемая здесь метамодель, относится к третьему уровню (M2) и может быть описана средствами MOF.

Несмотря на то, что MOF использует графическую нотацию, эта модель является формальной спецификацией. Все элементы Модели MOF имеют строго определённую семантику. Формальность модели MOF подробно показана в спецификации MOF [9], где даны формальные определения всех элементов Модели MOF. Имея такой формальный язык для создания моделей метаданных, можно приступить к описанию интересующей нас модели метаданных области многомерного проектирования. На рис. 3 показана связь между рассмотренными спецификациями и разрабатываемой метамоделью.

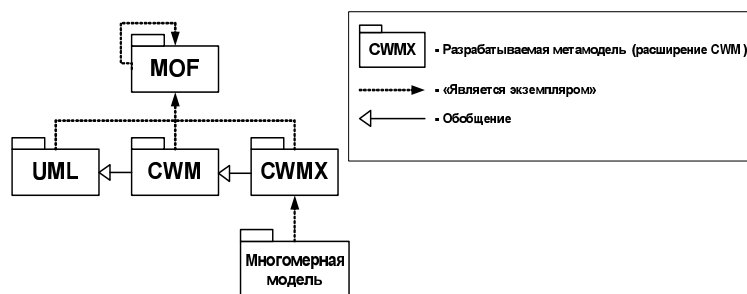


Рис. 3. Роль MOF при создании модели метаданных

Следует обратить внимание, что Модель MOF «является экземпляром» самой себя, т.е. может быть описана на языке MOF. Также может показаться странным использование языка UML — здесь он используется дважды в разных качествах на разных уровнях абстракции. Во-первых, нотация UML (именно нотация, а не сам язык, т.к. семантика у MOF своя) используется в качестве синтаксиса MOF. Во-вторых, язык UML является MOF метамоделью. Подробное исследование этих фактов выходит за рамки данной работы.

4. Модель метаданных как расширение спецификации CWM

Модель метаданных не нужно строить «с нуля», целесообразно использовать метамодель CWM в качестве основы и выполнить создаваемую метамодель в виде расширения CWM. Разрабатываемая метамодель будет оформлена в графической нотации UML с использованием языка OCL, применение которых предусмотрено в спецификации MOF. Несмотря на то, что используется графическая нотация, она является формальным представлением абстрактного синтаксиса MOF и будет однозначно определять выраженную в ней модель метаданных.

Далее необходимо чётко определить, какие аспекты спецификации CWM будут использованы для решения поставленной задачи разработки многомерной метамодели, что требуется изменить или доработать. Несмотря на то, что спецификация CWM предназначена для описания метаданных для хранилищ данных, она не решает всех проблем, возникающих при построении многомерных моделей. Спецификация CWM не предусматривает единого описания предметной области, вместо этого она позволяет создавать множество описаний концептуального уровня и описывать преобразования между ними. Метамодель CWM также не является достаточно подробной для описания конкретных многомерных моделей различных предметных областей.

Предлагаемая модель метаданных описания многомерных объектов будет менее общей, но более богатой (насыщенной деталями) по сравнению с CWM, что позволит на основе этой модели решать практические задачи. В частности, разработанная модель была использована в качестве базовой для разработки метамодели Репозитория Информационно Аналитической Системы (РИАС), которая, в свою очередь, была реализована в программных продуктах компании ЛАНИТ «Технический Каталог Показателей» и «Корпоративный каталог показателей».

Суть модели метаданных РИАС заключается в разбиении всех метаданных на уровни (концептуальный, логический и физический) и формировании единого описания на концептуальном уровне, связанного с множеством описаний логического уровня. Основой для концептуального уровня как раз и является разработанная многомерная метамодель, позволяющая описывать предметную область, для которой строится аналитическая система. К понятиям этого уровня относятся измерения, показатели, кубы. В CWM многомерная модель данных описывается в пакете OLAP и Multidimensional. В предлагаемой архитектуре многомерная модель «поднята» до концептуального уровня метаданных (и затем расширена).

Стандарт CWM разработан очень гибким, что означает его изменчивость по мере развития индустрии, и его расширяемость по требованиям конкретных реализаций. Механизм изменения стандарта обеспечивается и поддерживается OMG. В спецификацию CWM вошли те технологии, которые достаточно широко распространены и для которых существует общее понимание большинства понятий. Остальные технологии в области хранилищ данных являются кандидатами на описание в качестве расширений спецификации CWM. Например, ER-модель удовлетворяет критерию на широкое распространение, но не удовлетворяет критерию по общепринятости используемых понятий. Существует множество реализаций ER-моделей в различных инструментах, но многие из них несовместимы между собой по используемым понятиям. Однако из-за высокой значимости в области программной инженерии ER-модель была описана в виде расширения стандарта CWM.

Разрабатываемая модель метаданных будет содержать дополнительные понятия для более полного описания многомерных объектов. С точки зрения стандарта CWM — это стандартная ситуация, в которой необходимо оформить новые понятия и концепции в виде расширения CWM Extension (CWMX). При разработке многомерной метамодели использовался механизм расширения CWM с помощью наследования классов.

На рис. 4 показана многоуровневая структура метамодели CWM и её пакеты, которые использовались при разработке модели метаданных ИАС (показаны серым цветом). Одним из основных принципов построения CWM является

то, что пакеты, находящиеся на определённом уровне, зависят только от пакетов предыдущего уровня. Такая структура метамодели позволяет при построении конкретных реализаций использовать только те пакеты, которые нужны для решения поставленной задачи. Результатом является отсутствие пакетной зависимости между метамоделями на одном уровне или зависимости пакета нижнего уровня от пакетов верхнего уровня. Это означает, что определённая CWM метамодель (пакет) зависит только от метамоделей (пакетов), находящихся на более низких уровнях на этой диаграмме (но не обязательно от всех таких пакетов).

| | | | | | |
|------------------|----------------------|------------|---------------|---------------------------|-----------------------|
| Управление | Warehouse Process | | | Warehouse Operation | |
| | Transformation | OLAP | Data Mining | Information Visualisation | Business Nomenclature |
| Анализ | Object | Relational | Record | Multi-Dimensional | XML |
| Данные | Business Information | Data Types | Expressions | Keys and Indexes | Software Deployment |
| Базовые принципы | Core | Behavioral | Relationships | Instance | Type Mapping |
| Объектная модель | | | | | |

Рис. 4. Используемые в метамодели ИАС пакеты CWM

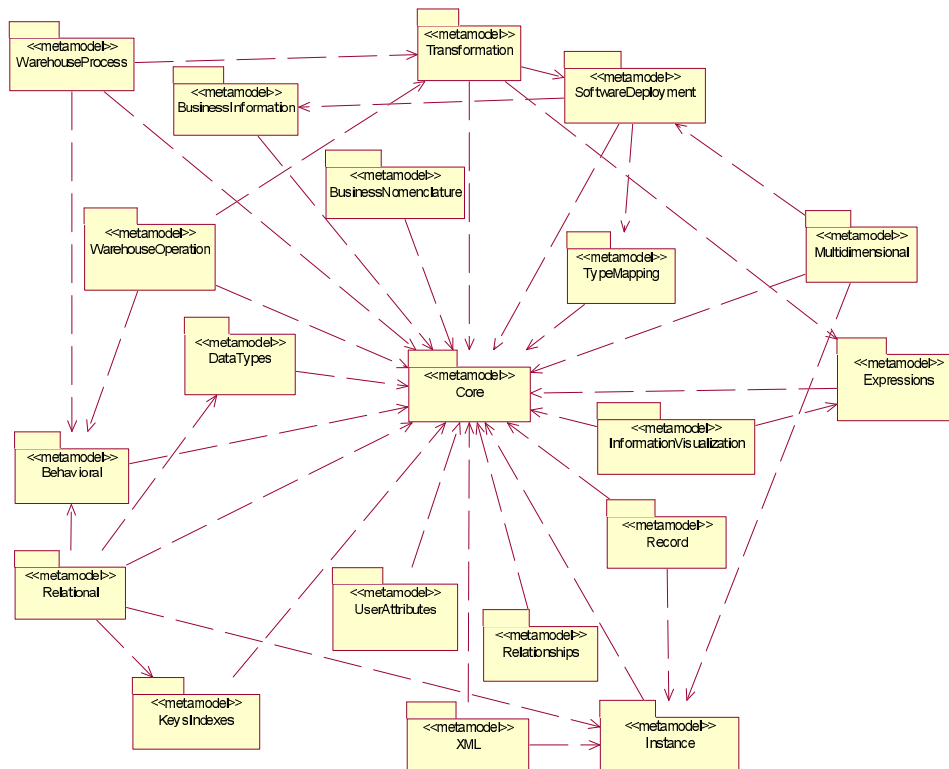


Рис. 5. Зависимость пакетов метамодели ИАС

Структура модели метаданных ИАС повторяет структуру пакетов метамодели CWM. Используемые в многомерной метамодели пакеты выделены серым цветом (рис. 4).

Модель метаданных ИАС отличается от метамодели CWM в основном добавлением новых классов и/или атрибутов классов. Зависимость между пакетами метамодели ИАС показана на рис. 5.

5. Заключение

Разработанная модель метаданных основана на спецификации OMG CWM. Базовые классы многомерной метамодели в основном состоят из классов уровней объектной модели (Object Model) и базовых принципов (Foundation) метамодели CWM. Расширения реализованы в виде новых классов, являющихся производными от базовых и дополнительных атрибутов классов. Предложенная в данной работе многомерная модель метаданных является основой метамодели Репозитория Информационно Аналитической Системы, применённой компанией ЛАНИТ при построении корпоративных информационно-аналитических систем (КИАС) на основе конфедеративного управления метаданными [14, 15].

Литература

1. *Kimball R.* The Data Warehouse Toolkit. — John Wiley & Sons, 1996.
2. *Kimball R., Reeves L., Ross M.* The Data Warehouse Lifecycle Toolkit. — John Wiley & Sons, 1998.
3. *Saltor F., Castellanos M., GarciaSolaco M.* Suitability of Data Models as Canonical Models for Federated DBs. — SIGMOD Record, 1991. — Vol. 20 (4), Pp. 44–48.
4. *Giovinazzo W. A.* Object-Oriented Data Warehouse Design. — Prentice Hall, 2000.
5. *Inmon W. H.* Building the Data Warehouse. — New York: John Wiley & Sons Inc., 1996.
6. Object Management Group: Common Warehouse Metamodel (CWM) Specification, version 1.0. — 2001.
7. Common Warehouse Metamodel: An Introduction to the Standard for Data Warehouse Integration / J. Poole, D. Chang, D. Tolbert, D. Mellor. — New York: John Wiley & Sons, Inc., 2002.
8. Object Management Group: Unified Modeling Language (UML) Specification, version 1.4. — 2001.
9. Object Management Group: Meta Object Facility (MOF) Specification, version 1.4. — 2002.
10. *Буч Г., Рамбо Д., Джекобсон А.* UML. Руководство пользователя. — ДМК, 2000.
11. *Буч Г.* Объектно-ориентированный анализ и проектирование с примерами приложений на C++, 2-е изд. / Пер. с англ. — М.: Издательство Бином, Невский диалект, 1999. — С. 560.
12. Object Management Group: Model Driven Architecture (MDA) – A Technical Perspective. — 2001.
13. Object Management Group: Model Driven Architecture (MDA) Guide / Ed. by J. Miller, J. Mukerji. — 2003.
14. *Шовкун А. В.* Использование корпоративного каталога показателей для управления метаданными в аналитической системе // Семинар «Хранилища данных и аналитические системы — от концепции к решению». — 10.12.2003. — <http://www.events.lanit.ru/seminar121003/materials.htm>.
15. *Шовкун А. В.* Построение корпоративной информационно-аналитической системы в условиях постоянно изменяющегося бизнеса // Научно-техническая информация. Серия 1. — № 9. — 2004. — С. 1–6.

UDC 681.3

On Some Aspects of Information Analytical System Conceptual Object-Oriented Multidimensional Metamodel Development

A. V. Viskov

*Department of Information Technology
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., Moscow, 117198, Russia*

Object-oriented approach provides essential advantages in multidimensional modeling. The question of information analytical system multidimensional metamodel development is considered in this paper very closely. It is offered to use widely the methods based on standards in the field of the multidimensional analysis: Meta Object Facility (MOF), Common Warehouse Metamodel (CWM), Unified Modeling Language (UML).