

ПРИМЕНЕНИЕ ОНТОЛОГИЧЕСКОГО МОДЕЛИРОВАНИЯ ПРИ РАЗРАБОТКЕ ХРАНИЛИЩА ДАННЫХ

Черников Б.В.

Национальный исследовательский университет «Высшая школа экономики», г. Москва
bor-cher@yandex.ru

Алексеев А.О.

ООО «Газпром ВНИИГАЗ», г. Москва
a_alekseev@vniigaz.gazprom.ru

Ключевые слова: информационная система, онтология, моделирование, хранилище данных.

Введение

Масштабность применения информационных технологий существенно изменила не только структуру процессов, но и правила ведения бизнеса. Основные бизнес-процессы компаний реализуются в поддержке корпоративных информационных систем. Чем выше динамика и сложность бизнеса, тем более развитыми должны быть и информационные системы. При этом функциональность самих систем является не единственным важным аспектом. При создании корпоративных информационных систем чрезвычайно важно соблюдение требований к данным: их актуальности, непротиворечивости, полноты и т.п. Отдельные информационные системы имеют свои собственные базы данных, что часто затрудняет и повышает стоимость получения агрегированных аналитик и отчетов. Если говорить об обработке данных в оперативном режиме, то при автономности баз данных отдельных информационных систем агрегирование данных может усложниться до уровня невозможного. Появление технологических решений таких как «Большие данные», «Облачные вычисления» решают только часть задач, связанных непосредственно с обработкой и хранением данных.

Основным инструментом работы с данными в компании должны быть хранилища данных, которые призваны обеспечить получение необходимой информации для всех корпоративных информационных систем с требуемым качеством.

1. Хранилище данных

Наиболее типовой проблемой, решаемой с помощью хранилищ данных, является интеграция разнородных информационных систем в единый комплекс. При этом элементы информации в системах могут систематизироваться по различным основаниям, обозначаться разными терминами (и наоборот – один термин в двух системах может определять схожие, но, тем не менее, отличающиеся понятия), обладать разным составом атрибутов (которые также могли именоваться несовместимыми терминами) и т.д. Аналогичную ситуацию «утери общего смысла» (отсутствие общих словарей, несогласованность терминов, неполнота и противоречивость описаний) можно наблюдать и в сфере нормативно-методологического обеспечения, и в сфере управления корпоративными знаниями, и в организации взаимодействия нескольких участников в рамках крупномасштабных проектов [1].

Хранилище данных в современном понимании давно не воспринимается как отдельная база данных, поскольку представляет собой средство не только хранения, но также сбора, анализа, преобразования, обработки и предоставления данных, как конечным пользователям, так и сопряженному комплексу корпоративных систем или внешних потребителей информации (рис. 1).

Построение структуры хранилища данных – концептуальной модели данных (КМД), – является крайне сложной задачей, и существующих методов визуального проектирования уже недостаточно для решения комплексных задач. Такие методы, как, например, моделирование «сущность-связь», позволяют правильно построить модель из информационных блоков, но не решают проблем их правильного выделения, полноты и непротиворечивости. Необходимо создание и применение более фундаментальных, методологически строгих подходов к созданию КМД и, в более широком смысле, – к созданию общего языка информационного взаимодействия между системами, людьми и организациями.

Наибольшая динамика в последние годы в развитии подобных методов наблюдается в сфере «осмысления» информации, получаемой в Интернете, и базируются они на использовании онтологического моделирования.



Рис. 1 Структура хранилища данных

Онтологическое моделирование

Основная цель метода онтологического моделирования данных состоит в том, чтобы обеспечить структурную методичку, поддерживаемую автоматизированными инструментальными средствами, которые специалист по проблемной области может эффективно разрабатывать и, тем самым, поддерживать полезные и точные онтологии [2].

Онтологическое пространство неоднородно. Выделяются три основных уровня (слоя) онтологий:

- прикладная онтология (ПО) – содержит концептуальные структуры и их связи, с помощью которых выполняется конкретное описание поставленной практической задачи;
- онтология предметных областей (ОПО) – обобщает понятия, используемые в разных прикладных онтологиях и касающиеся отдельных областей знаний;
- базовая онтология (БО) – описывает наиболее общие концептуальные структуры, к которым можно привести все понятия предметных и прикладных онтологий.

Проработка этих слоев выполняется двунаправлено (рис. 2).

Сначала на основании имеющихся описаний проблемной области и требований к создаваемым информационным системам составляется предварительный словарь ПО. Чтобы привести эти объекты к общему знаменателю, в качестве такого «знаменателя» необходимо определить так называемую базовую онтологию и выразить в ее терминах понятия ОПО. Гармонизированные таким образом ОПО будут служить основанием для конструирования терминов ПО. В контексте разработки информационной системы эти термины будут служить элементами КМД, которая по разработанным правилам будет преобразована в логическую схему БД.

Задача трансформации концептуальной модели в логическую модель – намного более простая, чем этап концептуального моделирования. В принципе существует множество инструментов проектирования, обеспечивающих автоматическое преобразование концептуальной модели в логическую модель данных.

Тем не менее, в процессе трансформации требуется произвести достаточно объемный комплекс преобразований, часть из которых имеют альтернативные варианты и поэтому требуют принятия некоторых решений. В таких ситуациях требуется иметь под рукой набор правил, в соответствии с которыми можно будет разрешать спорные ситуации [3].

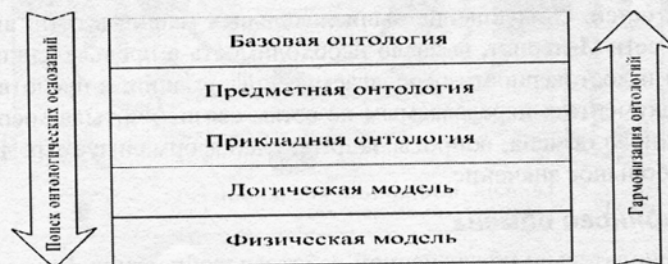


Рис. 2. Структура процесса формирования логической и физической моделей данных

Заключение

Отдельные этапы построения онтологий и их трансформации в модели, применимые для построения хранилищ данных, формализованы и имеют частную практику использования. При этом отсутствие законченных инструментов разработки показывает, что существуют «белые пятна» в онтологии, которые требуют решения. Учитывая темпы роста объема данных, критичность принимаемых на их основе решений, а также потребность в подобных формальных методиках будет только возрастать.

Помимо разработки хранилищ данных метод онтологического моделирования может применяться для решения ряда других проблем:

- несогласованность корпоративной документации;
- невозможность организации единой базы знаний;
- сложность в информационном обмене со сторонними организациями;
- рассогласованность описания корпоративных бизнес-процессов;
- невозможность создания единой системы нормативно-справочной информации;
- сложность интеграции корпоративных систем между собой и с внешними системами.

Литература

1. Асадулаев С. Хранилища данных: Тройная стратегия на практике. [Электронный ресурс] IBM, 2009. Режим доступа к журн.. <http://www.ibm.com/developerworks/ru/library/sabir/strategy/#h6>.
2. Matthew West Ontology Meets Business – Applying Ontology to the Development of Business Information Systems [Electronic resource] / Complex Systems in Knowledge-based Environments: Theory, Models and Applications Studies in Computational Intelligence Volume. – Berlin, 2009. – 229 p. – Mode of access: http://link.springer.com/chapter/10.1007%2F978-3-540-88075-2_9#page-1.
3. IDEF5 Method Report from 1994 / Perakath C. Benjamin, Christopher P. Menzel, Richard J. Mayer. Knowledge Based Systems, Inc., 1994. – 175 p.

ЗАЩИТА СЛАБОФОРМАЛИЗУЕМЫХ ДОКУМЕНТОВ ПРИ МЕЖВЕДОМСТВЕННОМ ОБМЕНЕ

Черников Б.В.

Национальный исследовательский университет «Высшая школа экономики», г. Москва

bor-cher@yandex.ru

Ключевые слова: лексикологический синтез, слабоформализуемые документы, опорные слова, защита информации, модель межведомственного обмена.

Введение

Лексикологический синтез представляет собой прогрессивную технологию подготовки слабоформализуемых документов, сопутствующих основному производству предприятий и организации самого различного профиля деятельности [1]. Переход от прямого ввода текста с клавиатуры к автоматизированному формированию документов на основе применения лексикологического синтеза [2] позволяет обеспечить снижение трудозатрат на формирование документов, повысить их качество, снизить объемы, необходимые для хранения документированной информации, повысить защищенность при передаче документов по каналам связи.

Актуальность проблемы защиты информации подчеркивается тем обстоятельством, что автоматизированное рабочее место является частью систем обработки информации, систем коллективного пользования, вычислительных сетей. Объединение вычислительных машин в сети, впоследствии реализованное в виде всемирной сети Интернет, вызвало необходимость в предъявлении достаточно жестких требований к надежности и достоверности передаваемой информации, к предотвращению несанкционированного доступа к документам, передаваемым по сетям связи. Учитывая особенности межведомственного документационного обмена, вопросы защиты слабоформализуемых документов при передаче информации имеют большое значение.

1. Модель информационного обмена

Для обеспечения законодательно обеспеченной работоспособности информационной системы документационного обеспечения организации или предприятия, использующей автоматизированный