

УДК 004.94

ОНТОЛОГИЯ КАК СРЕДСТВО ИНТЕГРАЦИИ МОДЕЛЕЙ ПРЕДМЕТНЫХ ОБЛАСТЕЙ

В. С. Кавицкая, к. т. н. В. В. Любченко

Одесский национальный политехнический университет

Украина, г. Одесса

kavickaaya@mail.ru

Рассматриваются проблемы уровней интеграции данных. Проводится анализ онтологии как способа интеграции моделей предметных областей. Анализируются основные проблемы интеграции онтологий.

Ключевые слова: интеграция моделей, онтология, способы интеграции, проблемы интеграции.

Проблема интеграции данных очень многообразна. Сложность и характер используемых способов ее решения существенным образом зависят от уровня интеграции, который необходимо обеспечить, от свойств отдельных источников данных и всего множества источников в целом, а также от требуемых способов интеграции. Системы интеграции данных могут обеспечивать интеграцию данных на следующих трех уровнях [1].

1. Интеграция на физическом уровне сводится к конверсии данных из различных источников в требуемый единый формат их физического представления. Проблема интеграции на этом уровне заключается в том, что в источниках данных могут использоваться различные форматы представления данных.

2. Интеграция на логическом уровне предусматривает возможность доступа к данным, содержащимся в различных источниках, в терминах единой глобальной схемы, которая описывает их совместное представление с учетом структурных и, возможно, поведенческих (при использовании объектных моделей) свойств данных. Семантические свойства данных при этом не учитываются. Проблема интеграции на этом уровне заключается в том, что может иметь место неоднородность используемых моделей данных для различных источников или различаются схемы данных, хотя используется одна и та же модель данных. Например, одни источники могут быть веб-сайтами, а другие — объектными базами данных.

3. Интеграция на семантическом уровне обеспечивает поддержку единого представления данных с учетом их семантических свойств в контексте единой онтологии модели предметной области. Проблема интеграции на этом уровне заключается в том, что различным источникам данных могут соответствовать различные онтологии.

Анализ ситуации в решении проблемы интеграции данных на семантическом уровне позволяет на основе работ [2, 3] сделать вывод о недостаточной изученности интеграции онтологий в целом. В связи с этим, целью данной работы является анализ принципов и методов интеграции онтологий моделей предметных областей.

В настоящее время проблема интеграции онтологий обсуждается довольно широко. Уже сформулировано множество условий, определений и методов, представлены различные уровни интеграции. Но даже сейчас нет четкого соглашения о том, что включает в себя интеграция онтологий. В общем, интеграцию онтологий принято определять как процесс нахождения сходства двух онтологий А и В и, как результат, создание новой онтологии С, объединяющей и согласующей семантические представления исходных онтологий. В результате две системы, основанные на онтологиях А и В, получают возможность взаимодействовать между собой, используя онтологию С.

Можно выделить два метода интеграции онтологий:

— с замещением новой онтологией исходных (вновь созданная онтология используется вместо интегрируемых);

— с совместным использованием интегрированной и исходных онтологий.

Методы второго типа обладают большей гибкостью, поскольку позволяют в большей степени сохранить и в дальнейшем использовать структуру уже имеющихся онтологий. В то же время, в случае использования одной монолитной онтологии необходимо включать в нее все термины исходных онтологий, что влечет за собой трудности, связанные с перестройкой связей с уже имеющимися терминами и разрешением семантических конфликтов. По этой причине методы первого типа оказываются применимы лишь в том случае, когда набор интегрируемых онтологий известен заранее и его расширение не предполагается [4].

Рассмотрим три уровня интеграции онтологий, в зависимости от числа изменений, которые необходимо сделать, чтобы получить общую онтологию из частных.

1. *Соответствие* (alignment) — отображение понятий и отношений одной онтологии на другую. Соответствие может быть определено не полностью. Так, может существовать несколько понятий в одной онтологии, не имеющих своих эквивалентов в другой. Иногда для приведения онтологий в соответствие в них добавляют новые подклассы и надклассы понятий. Никаких других изменений аксиом, определений, доказательств или вычислений не производится.

2. *Частичная совместимость* (partial compatibility) — соответствие онтологий, которое поддерживает также эквивалентные выводы и вычисления для всех эквивалентных понятий и отношений. Если две онтологии являются частично совместимыми, то любой вывод или вычисление, которые могут быть выражены в одной онтологии с использованием только соответствующих понятий и отношений, могут быть транслированы в эквивалентный вывод или вычисление в другой онтологии.

3. *Унификация* (unification) — взаимнооднозначное соответствие всех понятий и отношений в двух онтологиях, которое позволяет любой процесс вывода или вычислений, выраженных в одной онтологии, отображать в эквивалентный процесс вывода или вычислений в другой. Обычным способом унификации двух онтологий является усовершенствование каждой из них в более детальные онтологии, чьи категории взаимнооднозначно эквивалентны.

Самой слабой формой интеграции является соответствие, так как она требует минимальных изменений исходных онтологий, но может поддерживать глубокие выводы и вычисления. Частичная совместимость требует больших изменений, при этом она обеспечивает более широкую способность к взаимодействию. Унификация или полная совместимость требуют, как правило, значительных изменений или, в некоторых случаях, полной перестройки исходных онтологий, но ее результатом является наиболее полная способность к взаимодействию, то есть все, что может быть сделано в одной онтологии, может быть сделано полностью эквивалентным способом в другой. Несмотря на то, что интеграция на уровне соответствия является наиболее слабой, для системы семантического поиска информации ее вполне достаточно. В этом случае отсутствует необходимость обеспечения эквивалентности выводов для идентичных понятий, а также взаимнооднозначного соответствия, что должно быть сделано при интеграции на уровне частичной совместимости или унификации.

Важной проблемой при интеграции онтологий является проблема семантической неоднородности информации. Для решения этой проблемы необходимо [4]:

- описать предметную область;
- привязать к предметной области спецификации структуры, поведения и экстенсионалов объектов;
- проверить и согласовать понимание предметных областей взаимодействующих ресурсов;
- построить взаимодействие ресурсов, основываясь на согласованной семантике предметной области.

Соглашение об интерпретации онтологических понятий позволяет избегать неоднородности на уровне семантики при работе в рамках одной онтологии. Создание общедоступных онтологий предметных областей в определенной мере решает проблему неоднородности онтологических спецификаций. Однако в условиях открытого информационного пространства при решении задач предъявляются различные требования к предметным областям, глубине и формальности их описания, используются неоднородные онтологические описания предметной области, представленные в неоднородных онтологических моделях. Неоднородность онтологических спецификаций появляется на уровнях модельной и понятийной семантики. Соответственно, возникают задачи согласования онтологических моделей и онтологических контекстов.

На уровне моделей факторами, создающими неоднородность, становятся различия:

- в синтаксисе языков, определяющих онтологические модели;
- в выразительной способности моделей;
- в семантике примитивов, используемых в моделях.

На онтологическом уровне неоднородность порождает различия:

- в именах понятий и отношений;
- в подходах к определению понятий;
- в разбиении предметной области на понятия;
- в покрытии предметной области;
- в точках зрения на предметную область.

Можно сделать вывод, что разрабатываемые принципы и методы интеграции онтологий в основном неформальны и имеют множество открытых вопросов. Проблему интеграции онтологий предпочитают избегать, несмотря на то, что постановка данной проблемы является актуальной с самого начала использования онтологий при создании информационных систем. В связи с этим необходимо признать целесообразность выполнения исследований, которые позволят формализовать процесс интеграции онтологий, а также выполнять интеграцию онтологий в условиях семантической неоднородности информации.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

1. Коголовский М.Р. Методы интеграции данных в информационных системах // Сб. трудов Третьей Всероссийской конференции «Стандарты в проектах современных информационных систем». — Москва, 2003. — С. 1—8.
2. Гаврилова Т.А. Онтологический инжиниринг // Сб. докладов Восьмой науч.-практич. конфер. «Реинжиниринг бизнес-процессов на основе современных информационных технологий. Системы управления знаниями». Москва. — 2005. — С. 79—82.
3. Ломов П. А., Шишаев М. Г. Интеграция семантически связанных информационных ресурсов на основе онтологий для эффективного информационного обеспечения рационального природопользования // Глубокая переработка минеральных ресурсов: Сб. материалов IV школы молодых ученых и специалистов «Сбалансированное природопользование» (6-8 ноября 2007 г.). — Апатиты: Изд-во КНЦ РАН, 2008. — С. 243—247.
4. Тузовский А.Ф. Работа с онтологической моделью организации на основе дескриптивной логики // Известия Томского политехнического университета. — 2006. — Т. 309. — № 7. — С. 134—137.

V. S. Kavitskaya, V. V. Liubchenko

Ontology as a means of domain models integration.

The problems of data integration levels are considered. The authors analyze ontology as a means of domain models integration. The main problems of ontologies integration are discussed.

Keywords: *integration model, ontology, methods of integration, integration challenges.*