

Вопросы согласования неоднородных онтологических моделей и онтологических контекстов ¹

Скворцов Н. А.

Институт проблем информатики РАН
Россия, г. Москва, 117333, ул. Вавилова, 44/2
nsvk@ipi.ac.ru

Аннотация. Представленный обзор исследований и проектов, посвящённых работе с неоднородными онтологическими спецификациями, показывает сегодняшнее состояние проблемы. В нем затрагиваются три аспекта: неоднородность онтологических моделей, неоднородность спецификаций на уровне понятийной семантики и преобразование неоднородных данных между контекстами, определяемыми онтологиями.

Ключевые слова: неоднородные онтологические модели, неоднородные онтологические контексты, отображение онтологий.

1 Введение

Для решения проблемы семантической неоднородности информации необходимо описывать предметную область, привязывать к ней спецификации структуры, поведения и экстенционалов объектов, проверять и согласовывать понимание предметных областей взаимодействующих ресурсов и строить их взаимодействие, основываясь на согласованной семантике предметной области. При спецификации любых информационных ресурсов явно или неявно производится концептуализация предметной области. При описании предметной области явным образом используется онтология.

Онтология предметной области включает спецификации понятий и отношений между ними. В различных онтологических моделях эти спецификации достаточно формализованы, чтобы определить допустимые интерпретации аннотируемых онтологическими понятиями объектов, которые могут существовать в данной предметной области. Тем самым,

¹ Работа выполнена при финансовой поддержке РФФИ (гранты 06-07-189188-а, 08-07-00157-а) и программы ОИТВС РАН «Фундаментальные основы информационных технологий и систем» (проект 1-10).

онтологические описания задают весьма точную подразумеваемую семантику прикладной области для объектов и определяют их онтологический контекст.

Соглашение об интерпретации онтологических понятий позволяет избегать неоднородности на уровне понятийной семантики при работе в рамках одной онтологии. Создание общедоступных онтологий предметных областей в определённой мере решает проблему неоднородности онтологических спецификаций для определённых групп агентов. 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;
- выразительной способности моделей;
- семантике примитивов, используемых в моделях.

На онтологическом уровне неоднородность рождают различия в:

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

Для обеспечения семантически корректной интероперабельности неоднородных информационных ресурсов в контексте предметной области задачи необходимо выяснить общность и различия онтологий, лежащих в их основе, согласовать неоднородные онтологические спецификации и на базе соответствий онтологических контекстов, осуществлять преобразование информации. Как следствие, обеспечивается совместная работа неоднородных информационных ресурсов в контексте предметной области задачи на семантически значимом уровне.

Анализ состояния исследований [1] соответствующих методов показывает, что эти темы исследованы до сих пор недостаточно глубоко, преимущественно, для частных случаев. Разрабатываемые методы, в основном, неформальны и имеют множество открытых вопросов. Принципы и методы интеграции онтологий остаются предметом дискуссий, при создании систем такую интеграцию предпочитают избегать, несмотря на то, что постановка данной проблемы является актуальной с самого начала использования онтологий при создании

информационных систем. Наименее исследованы методы согласования онтологических контекстов, разработанных в неоднородных онтологических моделях. Согласование контекстов предметной области для организации совместной работы неоднородных информационных ресурсов и идентификация семантически релевантных элементов информационных ресурсов производится без использования онтологических спецификаций предметной области, вместо этого эксперт принимает определённые решения, либо онтологиями называют концептуальные схемы баз данных.

2 Неоднородность онтологических моделей

2.1 Разновидности онтологических моделей

В качестве онтологий используются спецификации, как в простых неформальных, так и в выразительных логических моделях. Сложность модели выбирается, исходя из необходимой ее выразительной мощности с одной стороны и возможности решения задач логического вывода с другой.

В неформальном представлении онтология может специфицировать коллекцию понятий в форме определений, изложенных на естественном языке. Мы будем называть такие онтологии вербальными. Формальная онтология может быть специфицирована посредством аксиом и определений, выраженных в логике или на языке, который может автоматически транслироваться в логику.

Неформальные лингвистические модели часто используются для спецификации онтологий. Это вербальные онтологические модели. J. Sowa также называет такие модели терминологическими. В таких моделях онтологические понятия описываются вербальным определением, как в толковом словаре. Может устанавливаться ограниченное количество видов семантических связей между понятиями. К таким онтологиям применяются методы информационного поиска, как для установления отношений между вербально определёнными понятиями, так и для поиска понятий, релевантных запросу. Вербальная онтология определяется кортежем $\langle C, S \rangle$, где C – множество онтологических понятий, $S: C \times C$ – множество семантических связей понятий. Множество S является объединением множеств: $S = \cup S_k$, где S_k – множества семантических связей одного вида.

В частности, в онтологии Sowa [2] акцент ставится на различии между классами и единичными сущностями. Имеются двуместные примитивные отношения, такие как *иметь*, *быть экземпляром*, *быть подклассом*, *быть пространственной частью*.

Формальные отношения в онтологии N. Guarino [3] рассматриваются как отношения, которые могут возникать между сущностями во всех материальных сферах. Его онтология наиболее приближена к философским изысканиям в области онтологии. Примеры формальных отношений включают *конкретизацию, принадлежность, частичность, связь, местоположение, расширение, зависимость*. Формальные свойства включают *конкретность, абстрактность, экстенциональность, единство, множественность* и другие. Guarino уделяет большое внимание различным свойствам онтологических отношений. Как будет показано далее, они оказываются весьма полезными при проверке корректности, как внутренней структуры онтологии, так и связей понятий разных онтологий.

Онтологии, основанные на логиках, определяются формально и имеют возможность формального вывода. Формальные онтологии включают спецификации аксиом и определений. Одна из первых широко используемых логических онтологических моделей была реализована в языке Ontolingua [4]. Он подразумевался как язык для однородного представления онтологических знаний, созданных в различных системах представления знаний, и для создания общих онтологий для групп агентов. Онтологическая модель построена на объектах, представляемых фреймами. Принимается деление на индивиды и множества. Фрейм, представляющий множество объектов, называется классом. Индивидом является объект, не являющийся множеством. Слоты фреймов определяют бинарные отношения. Функции и отношения вводятся как множества конечных списков. Таким образом, онтология в модели Ontolingua определяется кортежем: $\langle C, R, I, A \rangle$, где C – множество классов как унарных отношений над объектами, $R: C \times \dots \times C$ – множество отношений, введённых как множества кортежей объектов, и функций как особой разновидности отношений, I – множество индивидов, A – утверждения, определяющие классы и отношения. Определения включают вербальные описания объектов и формальные описания на языке KIF (Knowledge Interchange Format) [5]. В связи с выразительностью языка KIF, основанного на логике первого порядка, спецификации на языке Ontolingua не поддаются автоматическому логическому выводу для многих задач. Разработка онтологий в этой модели подчиняется специальной технике дополнения непротиворечивых спецификаций.

Некоторые эксперты в качестве основы выражения онтологий выбирают языки RDF или RDF-Schema [6]. RDF – простой декларативный язык, служащий для представления метаданных в форме утверждений над свойствами и отношениями ресурсов в вебе с использованием XML, которые могут быть практически чем угодно, в частности, онтологическими понятиями. Этот язык используется для онтологий благодаря своей простоте, однако сам по себе не предоставляет желаемого

уровня формальной семантики определений. RDF стал базисом, на котором построены некоторые современные языки описания онтологий.

Целый класс современных формальных онтологических моделей основан на всевозможных видах дескриптивных логик. Это семейство логических формализмов, прекрасно подходящих для представления терминологических знаний и онтологий, а также вывода в них. Они характеризуются набором конструкций, которые позволяют создавать сложные понятия и отношения из простых и при этом оставляют возможность автоматического логического вывода для решения определённых задач. Онтологии в моделях, основанных на дескриптивной логике, составляют кортеж $\langle C, R, H_C, H_R, I, A \rangle$, где C – множество классов, $R: C \times C$ – множество свойств классов, $H_C: C \times C$ – иерархия классов, $H_R: R \times R$ – иерархия отношений, I – множество индивидов, A – множество аксиом над классами и ограничений отношений. В некоторых моделях могут отсутствовать H_R и/или I , а A составляют только конструкции определённых видов, предусмотренные в данной модели. Этот набор определяет сложность вывода в задачах непротиворечивости и поглощения спецификаций.

В процессе развития теории дескриптивных логик, следует отметить такие известные разработки как KL-ONE и CLASSIC. Позднее появился язык OIL – язык представления онтологий, который должным образом основан на стандартах W3C. Это основанный на веб-технологиях язык, специально предназначенный для онтологий с формальной семантикой и средствами вывода, представляемыми дескриптивной логикой.

Наиболее заметную разработку логические онтологические модели получили в рамках развития семантического веба. DAML+OIL [7] – известная модель, использующая дескриптивную логику SHIQ [8], обладающую достаточной выразительной мощностью. Она является результатом слияния языков DAML и OIL. Она также построена на предшествующих стандартах W3C, таких как RDF и RDF Schema и расширяет их богатыми элементами онтологического моделирования. Эта модель планировалась как основа для языка, который ложится в основу технологии семантического веба.

Основанием языка OWL (Web Ontology Language) [9] служат возможности DAML+OIL, также язык предоставляет встроенную поддержку отображения между онтологиями. У языка OWL есть три диалекта: OWL Lite, OWL DL и OWL Full. Эти стили представления онтологий отличает друг от друга сложность моделей, тот или иной диалект может использоваться приложениями в зависимости от требований к простоте вывода и формальности описаний. Диалект OWL DL составлен на основе дескриптивной логики SHOIN. Сегодня имеются попытки использования OWL для спецификации схем баз данных и сервисов.

2.2 Унификация онтологических моделей

Унификация информационных моделей является необходимым шагом при работе с неоднородными онтологическими спецификациями. Модели приводятся к некоему каноническому однородному представлению, в котором будут производиться все дальнейшие манипуляции спецификациями. Модели онтологий называются исходными, а каноническая модель – целевой. Задача унификации множества исходных информационных моделей становится актуальной при необходимости масштабирования по количеству неоднородных моделей. При унификации моделей должно быть построено отображение исходных моделей в целевую. При отображении моделей производится поиск близких конструкций моделей и их выражение друг через друга.

Известные работы по отображению информационных моделей в основном связаны с конкретными моделями или их классами. Работ по методикам отображения произвольных информационных моделей немного. В основном, они основаны на подходах конструирования функций отображения элементов моделей (например, [10]). Следует отметить исследования [11], посвящённые коммутативному отображению моделей на основе отношения уточнения спецификаций.

Решение задачи отображения онтологий для конкретного проекта, сводится, в частности, к построению интерфейса к системе, использующего наиболее общие средства спецификации онтологий, допускающий расширение семантики отображаемых моделей. Так онтологическая модель ОКВС (Open Knowledge Base Connectivity) [12] используется в качестве базиса фреймовую модель. Она рассчитана на расширение для различных систем представления знаний. Программный интерфейс, разработанный для этой модели, рассматривается как ключевое средство для архитектуры распределённого репозитория онтологий. Модель ОКВС служит языком обмена для онтологий, использующих ОКВС как средство коммуникации. ОКВС является частью сервера Ontolingua. Состав конструкций ограничивается работой с классами, фреймами, слотами, фасетами и некоторыми свойствами отношений.

Специфика онтологических моделей не находит должного отражения в методиках их отображения. Поэтому в данной области приходится обращаться к опыту исследований отображения информационных моделей, не связанных с онтологией. Тем временем, особенности, связанные с разновидностями сущностей и связей, используемых во многих онтологических моделях, а также связанные с основанием обширного класса моделей на представлении знаний, вполне могли бы быть использованы для разработки специфических подходов и помощи экспертам в корректном отображении онтологических моделей.

Многообразие примитивов, используемых в онтологических моделях, может быть сведено к ограниченному набору элементарных составляющих. Примитивы конкретных моделей можно выразить онтологически при помощи композиции понятий и свойств некоторой метаонтологии верхнего уровня (например, DOLCE [13]), включающей элементарные составляющие, из которых могут формироваться примитивы, входящие в онтологические модели. Таким образом, в спецификации онтологической модели U и V (рис. 1), описывающей набор её конструкций и связей между ними, элементы могут быть аннотированы описаниями A_U и A_V в терминах метаонтологии верхнего уровня. Такой подход позволяет использовать методы семантической интеграции при решении задачи отображения моделей. Найдя в A_V подпонятия описаний A_U , имеем основание предполагать, что примитивы, связанные в этих аннотациях, однотипны, и для них можно построить отображение. Основой такого подхода послужила работа [14], в которой для трансляции схем и данных сложные элементы моделей описываются композициями элементарных составляющих (сама работа не связана с онтологиями).

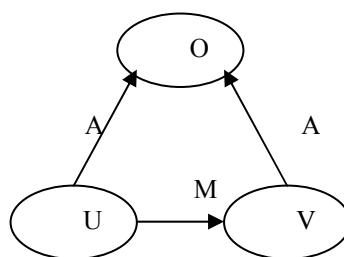


Рис. 1 Аннотация примитивов моделей для отображения

3 Согласование онтологических контекстов

3.1 Онтологические контексты и интеграция онтологий

Неоднородные онтологии могут являться различными точками зрения на одну и ту же предметную область или, по меньшей мере, на пересекающиеся предметные области. Эти точки зрения задают онтологические контексты, описанные конкретными онтологиями. Только в этом случае есть смысл говорить о задаче интеграции онтологий для согласования онтологических контекстов.

Задача интеграции онтологий появилась практически с началом использования онтологий в информационных технологиях, актуальна до

сих пор и будет актуальна ещё долгое время, пока сохраняется баланс между созданием унифицированных онтологий и признанием невозможности использовать их в специфических и изменчивых областях знаний. Задача интеграции нетривиальна и в основном полуавтоматическая, потому что на текущем уровне развития онтологических определений только эксперт может окончательно подтверждать корректность семантики установленных отношений между онтологическими понятиями.

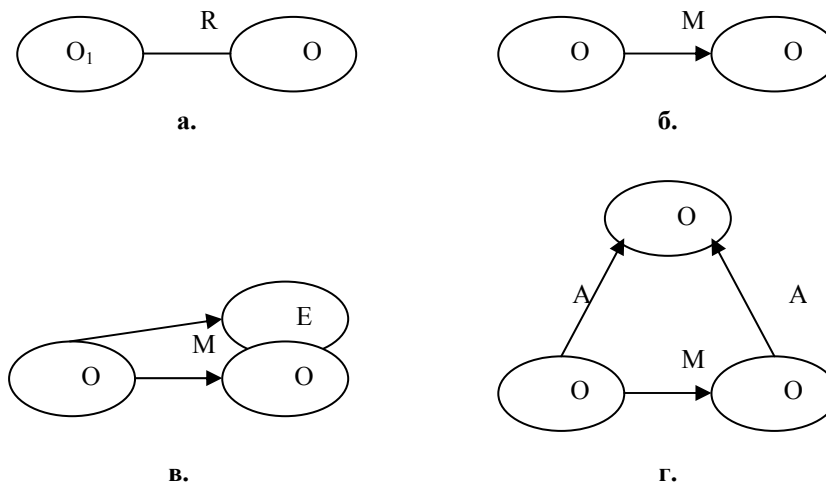


Рис. 2 Согласование онтологических контекстов: а. интеграция, б. отображение, в. расширение, г. многоуровневая интеграция.

Терминология в области интеграции онтологий используется разная. Под интеграцией мы понимаем процесс поиска связей между онтологиями для различных целей (рис. 2а). Происходит накопление информации, которая облегчает достижение интероперабельности между взаимодействующими информационными ресурсами, основанными на разных исходных онтологиях. Унификация онтологий – процесс приведения онтологических спецификаций к однородной онтологии. При этом производится поиск общих частей онтологий, разрешаются конфликты и строятся правила преобразования понятий одной онтологии в другую.

Функция преобразования одной онтологии $O_1 = \langle C_1, A_1 \rangle$ к другой $O_2 = \langle C_2, A_2 \rangle$ называется отображением онтологий $M_{12}: C_1 \rightarrow C_2$ (рис. 2б). Отображение может быть частичным в том смысле, что не для каждого понятия в исходной онтологии имеются эквивалент в другой онтологии. Полное отображение предполагает, что интерпретации, верные для онтологии O_2 , верны и для O_1 , то есть $A_2 \models M_{12}(A_1)$. Частичное

отображение подразумевает, что можно выделить подонтологию $O_1' = \langle C_1', A_1' \rangle$, $C_1' \subseteq C_1$, $A_1' \subseteq A_1$, для которой существовало бы полное отображение O_1' в O_2 .

Прежде, чем две онтологии могут быть отображены одна в другую, может появиться необходимость ввести новые подтипы или супертипы понятий или отношений, чтобы обеспечить подходящих кандидатов для точного отображения. Эти понятия и отношения составляют расширение E_1 онтологии (рис. 2в) в задаче унификации.

Глубина интеграции онтологий может быть различной. Слабая интеграция может использоваться для классификации и поиска информации, но не поддерживает логические выводы и вычисления. Часто интеграция требует большего количества модификаций для поддержания большей интероперабельности, даже если могут оставаться некоторые понятия или отношения в одной из взаимодействующих систем, неидентифицируемые в другой. Взаимооднозначная унификация онтологий может потребовать серьезных модификаций в онтологиях.

Подходы, используемые для согласования онтологических контекстов, в основном, неформальны, рекомендательны и требуют прямых действий эксперта. Спектр подходов очень широк и включает следующие разновидности:

- вербальные – обрабатывают вербальные определения и имена понятий и связей, выделяют основы слов, разбирают словосочетания, исключают стоп-слова, учитывают расстояния между словами, синонимы из тезаурусов и так далее, с целью оценить степень связи онтологических понятий лексически; используются в той или иной мере большинством проектов;
- структурные – множество различных подходов, в основном, эвристических, учитывающих таксономию, множественность, вывод в логике предикатов или в дескриптивной логике;
- с использованием внешних источников информации, в частности, WordNet [15], экстенционалов понятий, соотношения схем и онтологий;

Многие методы работают интерактивно. Они включают инициализирующее соответствие некоторых понятий, эвристическое предложение альтернатив предполагаемых соответствий и предоставление их эксперту, дальнейшие шаги вычислений зависят от выбора эксперта, например, отклик на происшедшее изменение по уже известным связям или по набору правил.

Методы, используемые в основе вычислений, также различны:

- эвристические формулы и правила;
- анализ графов, включая правила на графах, подходы к их сравнению и алгоритмы на взвешенных графах;
- обучающиеся подходы: использование статистики, значений и экземпляров;

- вероятностные подходы, вероятностное сопоставление результатов работы эвристических методов;
- формализация вывода и доказательства.

3.2 Проекты с объединением и отображением онтологий

Задача объединения онтологий, как и отображение онтологических понятий, включает поиск общностей и различий в спецификациях. Поэтому полезно рассмотреть проекты, направленные на задачи обоих видов. Некоторые проекты, использующие термин «онтология», фактически являются средствами интеграции схем.

Система ONION [16] направлена на определение правил разрешения неоднородностей между структурированными онтологиями. Правила учитывают только релевантные части исходных онтологий. Подход использует лексические методы и методы на основе графов. Метод нахождения лексического подобия между понятиями использует словари и семантическую индексацию, основанную на местонахождении группы слов в тексте. Учитываются простые отношения: *подкласс*, *часть/целое*, *атрибут*, *экземпляр*, *значение*. Первый шаг состоит в поиске связей понятий по их классификации, затем связываются части, определяющие онтологию верхнего уровня, и после этого согласовываются отдельно понятия в остальных частях онтологий. Решение по применению тех или иных правил ложится на эксперта, помощь которому заключается в предложении правил разрешения неоднородностей в соответствии с их эвристической оценкой. В системе присутствует элемент обучения, учитывающий ответы экспертов для улучшения предлагаемых решений для близких онтологий.

Система PROMPT [17] реализована как дополнительный модуль к распространённому редактору онтологий Protege и служит для объединения онтологий. Система в интерактивном режиме предлагает список возможных действий по объединению онтологий. Эксперт выбирает одну из предлагаемых операций или определяет производимую операцию самостоятельно. Система выполняет выбранную операцию, учитывает дополнительные конфликты, вызванные этой операцией, и модифицирует список предлагаемых операций. Это повторяется до тех пор, пока существуют конфликты и эксперт продолжает работу. Anchor-PROMPT [18] выбирает пары релевантных терминов из исходных онтологий и проводит пути между такими парами. Затем сравнивает термины в полученных путях и генерирует новые пары семантически близких терминов.

Интерактивный инструмент для объединения онтологий Chimaera [19] является частью сервера Ontolingua. Он предоставляет средства для анализа логической верности онтологии и диагностики типичных ошибок

при проектировании онтологий. Пользователь запрашивает анализ, и инструмент указывает проблемные места в объединении онтологий, однако не делает никаких действий самостоятельно. В своих предложениях Chimaega главным образом полагается на имена понятий и их таксономию.

OntoMerge [20] начинает объединение онтологий с приведения онтологий к общему синтаксическому представлению. Затем определяются аксиомы соединения между понятиями двух онтологий. OntoMerge предоставляет инструменты для трансляции данных, являющихся экземплярами понятий, в объединенную онтологию по спецификациям исходных онтологий и аксиомам соединения.

OntoMorph [21] предоставляет язык правил для спецификации отображений, производит объединение онтологий и генерацию трансляторов баз знаний. Эксперт использует начальный список пар из исходных онтологий для определения набора операторов, которые должны применяться к исходным онтологиям для устранения различий между ними. Однако, эксперт не получает никакого руководства, за исключением начального списка пар. Подход комбинирует синтаксическое и семантическое переписывание. Синтаксическое переписывание производится с помощью правил, основанных на образцах. Семантическое переписывание основано на семантических моделях и логическом выводе.

Система OBSERVER [22] обеспечивает доступ к неоднородным, распределённым, независимо разработанным репозиториям данных, используя несколько онтологий и информацию об отображении между ними. Вначале пользователи определяют набор межонтологических отношений *синонимии* и *выше/ниже*. Система помогает справиться с этой задачей, производя поиск синонимов в исходных онтологиях. Определив отображения, пользователи могут формулировать запросы в терминах собственной онтологии. Применяется технику расширения запросов, пользовательские запросы формулируются в терминах одной онтологии, а система старается расширить запрос, учитывая связанную онтологию, и объединяет возвращённую релевантную информацию, возвращённой по запросу.

Подход Similarity Flooding (SF) [23] работает с помеченными графами, основным принципом поиска связанных понятий является предположение, что элементы двух онтологий подобны, если подобны их смежные элементы. Предположение о близости двух элементов далее распространяется по их соседям. Инициализирующие предположения о близости вершин находятся простым сравнением имён. Для улучшения результатов инициализирующих данных используется внешний источник типа WordNet, без его привлечения качество работы метода сильно страдает.

В проекте Cupid [24] используется развитый алгоритм оценки коэффициента близости по именам с привлечением внешнего тезауруса, а затем алгоритм сопоставления структуры, основанный на сравнении листьев графов.

SMatch [25] позиционируется как система сопоставления схем, однако также упоминается постоянно, когда речь идёт об интеграции онтологий. Проект транслирует понятия в логические формулы и сопоставляет понятия онтологий первым делом с использованием WordNet. Решатель SAT [26] используется для проверки отношений *эквивалентности, включения, пересечения*.

В проекте OLA [27] утверждается, что методы оценки расстояния строк имеют большую производительность и эффективность в сравнении методами оценки близости определений, основанными на использовании внешних источников, в частности WordNet, за счёт времени обращения к внешним источникам. OLA основан на оценке терминологических и структурных расстояний между понятиями онтологий численно в интервале от 0 до 1. Близость представляется как множество формул, каждая переменная которой представляет подобие сущностей. Определения формул соответствуют определению функции близости и определениям онтологических сущностей. Проект работает с онтологиями в модели OWL Light.

SODA [28] – инструмент для выравнивания структурных онтологий в модели OWL DL с использованием методов оценки близости. Обе онтологии преобразуются в графы DL-GRAPH, затем вычисляется локальное подобие лингвистическими и структурными методами, а затем оценивается семантическая близость.

В подходе QOM [29] замечено, что на время работы алгоритма поиска отображения непосредственно влияет количество вероятных пар. Здесь применяется эвристический метод оценки структуры онтологий, позволяющий уменьшить количество кандидатов на отображение. На этапе оценки близости QOM избегает полной попарной оценки деревьев онтологии и ограничивает число дорогостоящих сравнений. Там, где используются итерации, ограничивается их количество, утверждая на тестах, что дальнейшие итерации не сильно влияют на результат. Констатируется, что оптимизация операций уменьшает качество отображения, а использование комбинации подходов его увеличивает. В целом, QOM показывает неплохие результаты при разнице во времени работы на порядок относительно других методов.

FCA-Merge [30] – метод для сравнения онтологий, которые имеют набор общих экземпляров или набор общих документов, аннотируемых с помощью концептов исходных онтологий. Основываясь на этой информации, система производит решетку понятий, связывающую концепты исходных онтологий. Алгоритм предлагает отношения *эквивалентности и подкласса*. Затем эксперт анализирует результат и

использует его как руководство для создания объединенной онтологии. Однако предположение, что две объединяемые онтологии используют общий набор экземпляров или имеют набор документов, представительных для данной проблемной области, и каждый документ должен аннотироваться терминами обоих источников слишком жесткое и на практике такая ситуация происходит редко. В качестве альтернативы, авторы предлагают использовать методы обработки естественного языка для аннотации набора документов понятиями этих двух онтологий.

Проект GLUE [31] представляет оригинальный подход к связыванию онтологий, использующий обучающиеся машины для предположения близости элементов онтологий по данным экземпляров понятий. Онтологии определяются как таксономии понятий с атрибутами. Во время фазы обучения находятся шаблоны и правила сопоставления элементов онтологий. Точность предположения зависит от качества обучения. Используется несколько машин обучения и метаобучатель, взвешивающий их результаты в соответствии с тем, какие результаты выдавал конкретный подход во время обучения. Для многих онтологий не существует данных, состоящих из экземпляров понятий, в этом случае метод не применим.

RiMOM [32] – средство для выравнивания онтологий, комбинирующее различные стратегии с целью достичь наилучших результатов. Процесс выравнивания состоит из следующих шагов. Производится оценка подобия структуры и текстового подобия. По результатам оценок выбирается стратегия между распространением по структуре и текстовыми методами. Далее выполняются задачи в соответствии с выбранными стратегиями, комбинируются их результаты. Если у этих двух онтологий высокая оценка структурного подобия, RiMOM использует распространение связей по структуре. В конце производится оценка выравнивания и применяются несколько эвристических правил для удаления маловероятных связей.

Falcon-AO [33] объединяет два подхода, как и большинство других методов: сравнение на основе лингвистического соответствия (LMO) и на основе графов (GMO). Графовые методы используют лингвистические как дополнительные.

SOMA [34] – инструмент для сравнения схем, поддерживающий расширяемую библиотеку алгоритмов сравнения, гибкие средства для их комбинации и платформу для оценки эффективности различных алгоритмов. Также инструмент способен выполнить итерации процесса сравнения.

В течение нескольких лет выполняются работы по оценке алгоритмов выравнивания онтологий, где разработан общий интерфейс выравнивания онтологий и средства оценки алгоритмов, реализующих интерфейсы. Результаты этой работы публикуются в [35].

3.3 Необходимость экспертной работы

«Хороший» путь интеграции онтологий сегодня является предметом интенсивных исследований. Применяя существующие методики, невозможно автоматически интегрировать онтологии, созданные разными рабочими группами.

Некоторые авторы, разрабатывающие методы интеграции онтологий утверждают, что это невозможно до тех пор, пока не развиты технологии семантической обработки естественного языка. Однако и понимание текстов на естественном языке может быть также недостаточным. По аналогии общения между людьми, для понимания агента, действующего в другом онтологическом контексте, мало разобрать его утверждения на естественном языке, но необходимо задавать наводящие вопросы, получать на них ответы и оценивать адекватность ответов с точки зрения своего контекста. Однако и это часто не приводит к взаимопониманию.

Семантические различия похожих понятий могут выясняться зачастую только в процессе дискуссий, на основе применения онтологий к одним и тем же примерам моделей и объектов реального мира и проверки, какие классы объектов реального мира могут или не могут соответствовать данному понятию. Рабочие группы по гармонизации онтологий обнаруживают множество различий в понимании понятий, которые с первого взгляда казались схожими. Усилия экспертов и их групп для создания «хороших» онтологий занимают порой годы. При интеграции пересекающихся областей сталкиваются продуманные решения, и находить консенсус между ними непросто.

Интенсивно занимаясь оценкой алгоритмов выравнивания онтологий, исследователи, тем не менее, понимают ограничения этих подходов и отмечают необходимость и перспективность разработок, позволяющих оптимально поддерживать интеллектуальную работу экспертов при интеграции онтологических контекстов [36].

3.4 Корректность отображения онтологий

Один из путей поддержки экспертов при поиске отображения онтологических контекстов заключается в проверке корректности установленных соответствий понятий и построенных отображений. Проект OntoClean [3] проверяет онтологии с использованием семантических свойств понятий. Эта методология может использоваться и для проверки связей, установленных между интегрируемыми онтологиями.

Онтология в проекте представляет собой набор отношений, которые являются утверждениями об объекте. С отношениями разных видов в качестве метаданных связаны их различные свойства, такие как:

- *Идентификация* – является ли утверждение об объекте идентифицирующим его свойством;
- *Неизменность* – может ли свойство меняться во времени;
- *Зависимость* – может ли объект существовать без других;
- *Постоянность* – как долго сущность остаётся таковой, и множество других свойств.

Некоторые из этих свойств совместимы друг с другом, другие исключают друг друга. В частности, при построении иерархии, некоторые свойства более специфических отношений должны наследоваться, и обнаружение конфликтов свойств между отношениями в иерархии будет означать некорректность построения онтологии.

Однако тот же принцип может быть использован для проверки результатов отображения онтологий. Если свойства связанных отношений из двух онтологий противоречат друг другу, это означает, что отображение было составлено некорректно, отношения имеют разную семантику.

4 Преобразование данных между онтологическими контекстами

В основном, для интеграции информации используются однородные описания в терминах одной онтологии. Интеграцию онтологических контекстов для преобразования информации между контекстами применяют немногие. В частности, проект OBSERVER использует расширения запросов к данным по связям между онтологиями и объединяет возвращённые данные из разных онтологических контекстов.

Проект ECOIN [37] представляет подход к интеграции информации из разных онтологических контекстов. Здесь присутствует общая онтология и контексты, определённые как уточнение спецификаций общей онтологии путём присвоения значений, введения подпонятий. Также есть логические правила, которые определяют преобразование данных при переходе из контекста в контекст. Данные имеют метаатрибуты в терминах общей онтологии, которые могут принимать различные значения в разных контекстах в соответствии с их определением. Метаатрибуты определяют семантику, свойства и структуру данных. Переходя из контекста в контекст, такие данные подвергаются контекстуализации, вычисляются их новые значения, соответствующие текущему контексту.

Принцип метаинформации используется и в проекте SemPro [38], где с объектами связываются дополнительные описания, взятые из общей онтологии. Множество таких описаний составляет контекст объекта, моделирующий характеристики в предметной области. SemPro обеспечивает механизм для сравнения и манипуляции различными

контекстами, такими как контекст запроса и контекст определения объекта. Этот механизм основан на отношении специализации между двумя контекстами, в соответствии с которым контексты выстраиваются в решётку, и возможно определить наибольший общий контекст, являющийся специализацией двух разных контекстов.

5 Заключение

Проведённый анализ позволяет сделать вывод о том, что для работы с неоднородными онтологическими спецификациями часто используются методы, не учитывающие специфику онтологической информации.

Методы построения отображений неоднородных онтологических моделей для общего случая практически не отражены в публикациях, вместо этого исследования, в основном, посвящены преобразованию между конкретными онтологическими моделями и созданию интерфейсов над наиболее общими моделями. Специфические свойства представления онтологических спецификаций при отображении моделей никак не используются.

Похожая ситуация видна и в области интеграции онтологических контекстов. Преобладающее число подходов для интеграции используют методы, применяемые для интеграции объектных или реляционных схем. В основном, это лексический анализ имён и определений на естественном языке и анализ структурных спецификаций. Согласование понятийной семантики онтологических спецификаций отдаётся на откуп экспертам, так как именование и структура описаний может существенно отличаться, не влияя на подразумеваемую семантику. Даже проекты, работающие в моделях, имеющих формальные методы вывода, используют в основном оценочные подходы. Неформальность преобладающего большинства методов интеграции онтологий и широкий спектр различных подходов без выраженного лидирующего подхода говорят о том, что «хорошая» методика интеграции онтологий до сих пор не найдена.

Отображение онтологических контекстов сегодня слабо используется для семантического преобразования данных между контекстами. Чаще для интеграции данных и сервисов используются спецификации в терминах одной онтологии.

Литература

1. L. Kalinichenko, M. Missikoff, F. Schiappelli, N. Skvortsov. Ontological Modeling. RCDL'2003. St.-Petersburg, 2003

2. J. F. Sowa. Ontology, metadata, and semiotics. In B. Ganter & G. W. Mineau, eds., *Conceptual Structures: Logical, Linguistic, and Computational Issues*, LNAI 1867, Springer-Verlag, Berlin, 2000
3. N. Guarino, C. Welty. Evaluating ontological decisions with OntoClean. *Communications of the ACM*, 45 (2), pp. 61-65, 2002
4. T. Gruber. *Ontolingua: a Mechanism to Support Portable Ontologies*. Knowledge Systems Laboratory of Stanford University, Jun 1992
5. M. Genesereth, R. Fikes. Knowledge interchange format, version 3.0, reference manual. Technical report, Logic-92-1, Computer Science Dept., Stanford University, 199
6. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation. <http://www.w3.org/TR/rdf-schema/>
7. D. McGuinness, R. Fikes, J. Hendler, L. Stein. IEEE Intelligent Systems: DAML+OIL: An Ontology Language for the Semantic Web. *IEEE Distributed Systems Online* 3(11), 2002
8. I. Horrocks and U. Sattler and S. Tobies. Practical Reasoning for Very Expressive Description Logics. *Logic Journal of the IGPL*, Volume 8, Issue 3, May 2000
9. OWL Web Ontology Language Guide. W3C. <http://www.w3.org/TR/owl-guide/>
10. G. Caplat, J. L. Sourrouille. Model Mapping in MDA. *Workshop in Software Model Engineering*, 2002
11. Kalinichenko L.A. Methods and tools for equivalent data model mapping construction. *Advances in Database Technology: Proc. of the International Conference on Extending Database Technology EDBT'90*. LNCS 416. -- Berlin-Heidelberg: Springer-Verlag, 1990. -- P. 92-119.
12. V. Chaudhri, A. Farquhar, R. Fikes, P. Karp, J. Rice. OKBC: A Programmatic Foundation for Knowledge Base Interoperability. *AAAI'98*, 1998
13. C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, L. Schneider. *The WonderWeb Library of Foundational Ontologies Preliminary Report*. ISTC-CNR, Italy, 2003
14. P. Atzeni, P. Cappellari, and P. A. Bernstein. Model independent schema and data translation. *EDBT*, Munich, 2006
15. C. Fellbaum. *WordNet – an Electronic Lexical Database*. MIT Press, Cambridge, 1998
16. P. Mitra and G. Wiederhold. Resolving terminological heterogeneity in ontologies. In *Proceedings of the ECAI'02 workshop on Ontologies and Semantic Interoperability*, Lyon, 2002
17. N. Noy, M. Musen. *The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping*. Stanford Medical Informatics, Stanford University, 2003
18. N. Noy, M. Musen. Anchor-PROMPT: Using NonLocal Context for Semantic Matching. In *Proceedings of the Workshop on Ontologies and Information Sharing at the International Joint Conference on Artificial Intelligence (IJCAI)*, 2001
19. McGuinness D., Fikes R., Rice J., Wilder S. An environment for merging and testing large ontologies. *KR2000*, San Francisco, 2000
20. D. Dou, D. McDermott, P. Qi. Ontology translation by ontology merging and automated reasoning. *EKAW'02 workshop on Ontologies for Multi-Agent Systems*, Siguenza, 2002
21. H. Chalupsky. *OntoMorph: A translation system for symbolic knowledge*. *KR2000*, San Francisco, 2000

22. E. Mena, A. Illarramendi, V. Kashyap, A. Sheth. OBSERVER: An approach for query processing in global information systems based on interoperation across preexisting ontologies. *Distributed and Parallel Databases, An International Journal*, Vol.8, No.2, 2000
23. S. Melnik, H. Garcia-Molina, E. Rahm. Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching. In *Proc. 18th ICDE*, San Jose CA, 2002
24. J. Madhavan, P. A. Bernstein, E. Rahm. Generic Schema Matching with Cupid. In *Proc. of the 27th Conference on Very Large Databases*, 2001
25. F. Giunchiglia, P. Shvaiko, and M. Yatskevich. Semantic Schema Matching. In *Proc. of CoopIS'05*, volume 3760 of LNCS, pages 347 – 360, 2005
26. F. Giunchiglia, P. Shvaiko. Semantic Matching. *CEUR-WS*, vol: 71, 2003
27. J. Euzenat, D. Loup, M. Touzani, P. Valtchev. Ontology Alignment with OLA. In *Proceedings of the 3rd EON Workshop, 3rd Intl. Semantic Web Conference*, Hiroshima, 2004
28. S. Zghal, S. Ben Yahia, E. Mephu Nguifo, Y. Slimani. SODA: an OWL-DL based ontology matching system In *Proceedings of the first French Conference on Ontology (JFO 2007)*, Sousse, 2007
29. Ehrig, Marc and Staab, Steffen. QOM – Quick Ontology Mapping. in S.A. McIlraith et al. (Eds.): *ISWC 2004*, LNCS 3298, pp. 683–697, 2004
30. G. Stumme, A. Medche. FCA-Merge: Bottom-up merging of ontologies. *IJCAI'01*, Seattle, WA, 2000
31. Doan, A.H., J. Madhavan, P. Domingos, A. Halevy: Learning to Map between Ontologies on the Semantic Web. *WWW 2002*
32. J. Tang, J. Li, B. Liang, X. Huang, Y. Li, and K. Wang. Using Bayesian Decision for Ontology Mapping. *Journal of Web Semantics*, Vol(4) 4:243-262, 2006
33. N. Jian, W. Hu, G. Cheng and Y. Qu: Falcon-AO: Aligning Ontologies with Falcon. *Proceedings of the K-CAP Workshop on Integrating Ontologies 2005*
34. H. Do and E. Rahm. Coma: A System for Flexible Combination of Schema Matching Approaches. In *Proceedings of 28th International Conference on Very Large Data Bases*. Hong Kong, China. 2002
35. *International Workshop on Ontology Matching*. <http://www.ontologymatching.org/>
36. S. M. Falconer, N. F. Noy, M.-A. Storey. Towards Understanding the Needs of Cognitive Support for Ontology Mapping. *International Workshop on Ontology Matching*, Athens, 2006
37. F. Madnick, E. Stuart, and B. Grosz. Contextual alignment of ontologies in the eCOIN semantic interoperability framework. *Information Technology and Management*, Volume 8, Issue 1, Pages: 47 – 63, 2007
38. V. Kashyap, A. Sheth. Schematic and semantic similarities between database objects: A context-based approach. *VLDB Journal*, vol. 5, no. 4, October 1996