

Связывание объектных спецификаций по семантике онтологического уровня*

© Н. А. Скворцов

Институт проблем информатики РАН
nskv@synth.ipi.ac.ru

Аннотация

Онтологический подход к решению задач над неоднородными информационными источниками предполагает использование согласованных онтологических контекстов предметной области задачи и предметных областей используемых источников. На основе связи согласованных онтологий решается задача семантического связывания спецификаций задачи и схем источников. В данной статье обосновывается модель связи онтологических спецификаций с объектными схемами информационных источников и рассматривается вопрос о том, какие связи на уровне схем можно почерпнуть из связей онтологической информации. Учитываются разные виды связей онтологического уровня: синонимия, суперпонятие/подпонятие, атрибуты и ассоциации понятий, часть/целое, функции преобразования значений.

1 Введение

Одной из проблем использования неоднородных информационных источников для решения задач является согласование прикладных контекстов источников и решаемой задачи. Эта проблема решается на основе онтологического подхода. Онтология составляет описание в некотором формализме предметной области, соответствующей информационному источнику или задаче, в виде словаря понятий, отношений между ними и универсальных ограничений, свойственных данным понятиям и отношениям. В общем случае спецификации информационных источников и решаемой задачи связываются с контекстами разных онтологий. При использовании онтологий для согласования прикладных контекстов приходится решать следующие задачи:

- связывание спецификаций информационных источников со спецификациями предметной

области для отражения их подразумеваемой семантики;

- приведение формализмов различных онтологий к одному для возможности сравнения представлений о предметной области;
- отображение онтологических контекстов в одном формализме с целью их согласования;
- семантическое связывание элементов объектных схем информационных источников и решаемой над ними задачи на основе связи онтологических понятий.

В предыдущих работах были освещены методы и средства решения вышеуказанных задач. В [6] на основе языка спецификаций СИНТЕЗ определена онтологическая модель, выразительная мощность которой достаточна для представления различных видов онтологий, от вербальных до формальных. В [4] описана методика расширения онтологической модели при отображении онтологических формализмов в каноническую модель для представления в ней произвольных онтологий. В [5] и [6] представлены подходы отображения онтологических контекстов на основе эвристических вербальных методов и точного формального вывода. Простой принцип аннотирования объектных схем онтологическими понятиями и связывания релевантных элементов объектных схем по информации о связи онтологических понятий, полученных в результате отображения онтологических контекстов, показан в [5].

Однако возникает необходимость дополнительно обосновать и усовершенствовать указанный подход к связыванию объектных схем информационных источников с онтологическими спецификациями предметной области. Данная статья является очередным шагом исследований, проведённых в указанных работах, и фокусируется на исследовании аннотаций элементов схем онтологическими понятиями. Ниже в разделе 2 коротко описываются онтологическая модель, используемая в исследованиях, подходы, применяемые для отображения описаний понятий разных онтологий в данной модели, а также способ связывания элементов спецификаций схем с онтологиями. Раздел 3 посвящён обоснованию указанного способа связывания схем и онтологий. В разделе 4 рассматриваются различные виды онтологических

связей и их использование для семантического связывания спецификаций объектных схем источников и решаемой задачи.

2 Онтологическая модель и принципы работы в ней

Для понятийного описания предметных областей и выражения семантики объектов реального мира используются онтологические спецификации. Они содержат словарь понятий предметной области и отношения между ними, ограничивающие подразумеваемую семантику понятий. Учитываются два подхода к работе с онтологическими спецификациями. Вербальный подход использует лингвистические спецификации онтологических понятий и эвристические методы работы с ними. Формальный подход интерпретирует онтологические понятия как абстрактные типы данных и для манипуляций понятиями использует формальные методы вывода. В средствах спецификации онтологий информационных источников и задач отражены оба подхода.

Спецификации онтологических понятий определяются в онтологических модулях. В разделе описания классов определяются метаклассы, соответствующие понятиям. Экстенсионалы этих метаклассов содержат произвольные объекты, которые семантически соответствуют онтологическому понятию. Метаклассы могут быть связаны отношением класс/подкласс.

Типами экземпляров данных метаклассов являются типы онтологических понятий, они определяются внутри спецификаций метаклассов. Эти абстрактные типы данных составляют формальную спецификацию онтологического понятия. Понятия могут состоять в иерархии обобщения/специализации. Отношения между понятиями описываются атрибутами типа, которые могут определяться типом атрибута и/или метаклассом ассоциаций. Метаклассы ассоциаций, ограничивающие интерпретацию отношений между типами онтологических понятий определяются в разделе классов онтологического модуля. Для представления интенциональных свойств понятия используются инварианты. В разделе функций могут определяться универсальные функции преобразования значений при переходе между онтологическими контекстами.

В следующем примере на языке спецификаций СИНТЕЗ определен пример класса онтологического понятия `human` (человек) и его подклассов `man` (мужчина) и `woman` (женщина). Они вводятся в модели как метаклассы. В качестве типов экземпляров этих классов определяются соответствующие типы онтологических понятий `Human`, `Man`, `Woman`. Экстенсионалы классов `man` и `woman` не пересекаются, что определяется в инварианте `disjoint` одного из типов их экземпляров.

```
{ human;
  in: metaclass;
  instance_section:
  { Human;
    in: type, concept;
  }
}
{ man;
  in: metaclass;
  superclass: human, male;
  instance_section:
  { Man;
    in: type, concept;
    supertype: Human, Male;
  }
}
{ woman;
  in: metaclass;
  superclass: human, female;
  instance_section:
  { Woman;
    in: type, concept;
    supertype: Human, Female;
    disjoint: {in: predicate, invariant;
      { predicative:
        { intersect(man, woman) = {} }
      }
    }
  }
}
```

Лингвистическая составляющая спецификаций онтологического понятия определяется в метафрейме типа понятия. Понятие может быть определено вербально. Вербальное определение служит основой для эвристических методов поиска близких понятий. В слотах определяются семантические отношения разных видов: позитивные, синонимические, ниже, выше, часть, целое, и родственные с другими понятиями. Для этих отношений может быть установлена сила связи в интервале от 0 до 1.

Как говорилось выше, экземплярами метаклассов, соответствующих онтологическим понятиям, являются объекты, семантически соответствующие данным понятиям. Такими объектами могут становиться, в том числе, элементы спецификаций объектных схем, такие как типы, классы, метаклассы ассоциаций, атрибуты, функции, формальные параметры. Элементы спецификаций аннотируются онтологическими понятиями с помощью метаопределений, описываемых внутри элементов и определяющих, в какие онтологические метаклассы входят данные элементы. Например, следующее определение типа `Person` аннотируется определённым выше понятием `human`. По семантике языка СИНТЕЗ это выражается в том, что тип `Person` как объект является экземпляром метакласса `human`.

```
{ Person;
  in: type, human;
  ...
}
```

Аннотацией о принадлежности понятиям онтологии снабжаются как элементы локальных спецификаций информационных источников, так и спецификации требований задачи, решаемой над данными информационными источниками.

В зависимости от предметных областей и используемых онтологий могут иметь место три варианта принадлежности им элементов спецификаций источников и спецификаций задачи:

- источники и задача используют одну общую онтологию предметной области;
- ряд информационных источников использует общую онтологию, независимую от онтологического контекста предметной области задачи;
- спецификации информационного источника содержат определение специфического онтологического контекста, независимого от онтологического контекста предметной области задачи.

Семантически релевантные спецификациям задачи компоненты информационных источников находятся по принадлежности одному и тому же или связанным онтологическим понятиям. В первом из перечисленных случаев тривиально предполагается, что элементы спецификаций, принадлежащие одним и тем же онтологическим понятиям, онтологически релевантны. В остальных случаях приходится сталкиваться с проблемой согласования онтологических контекстов. Понятия общих онтологий и локальных онтологий информационных источников необходимо отобразить на понятия онтологии предметной области задачи. Связь между понятиями онтологий обеспечит, в свою очередь, возможность установления связей между элементами спецификаций схем, являющимися экземплярами связанных понятий.

Отношения между онтологическими спецификациями устанавливаются с использованием различных подходов в зависимости от разновидности онтологий и сложности используемых онтологических моделей. Связи между понятиями разных онтологических контекстов в случае их вербального определения устанавливаются эвристическими методами, основанными на трактовке определения понятия как вектора дескрипторов и оценки соотношения этих векторов по определённым критериям. При формальном определении онтологий предварительное соотношение понятий устанавливается так же, как и в случае вербальных онтологий, а затем на базе отношения уточнения понятий как типов. Состав используемых методов отображения онтологических понятий может расширяться произвольно.

Все перечисленные в этом разделе конструкции модели онтологических понятий и пути установления отношений между понятиями различных онтологий приведены здесь для того, чтобы далее рассуждать о возможности использования информации из онтологических спецификаций при семантическом связывании элементов объектных схем неоднородных информационных источников и спецификаций схемы решаемой задачи.

3 Использование онтологических описаний для спецификации информационных источников

Онтологический подход призван отделить абстрактное описание подразумеваемой семантики объектов реального мира, могущих существовать для агентов в данной предметной области, от описания, обслуживающего технические детали представления этих объектов в базах данных и информационных системах. Таким образом, различают онтологический уровень спецификаций и уровень представления. Последнее время появилось достаточное количество работ, посвящённых различению онтологий и концептуальных схем [3]. Хотя признаётся, что чёткую грань между этими понятиями провести сложно, однако указываются признаки различия между ними. Согласно N. Guarino [2], онтология может использоваться при разработке информационных систем, а также информационной системой в процессе её работы, в то время как концептуальные модели являются строительными блоками при анализе и проектировании информационной системы. Это путь наиболее частого использования, но не признак онтологии или концептуальной схемы. Цель определения онтологического уровня – выражение ограничений интерпретации понятий предметной области вне зависимости от реализации задачи или информационных ресурсов. Цель определения концептуальной схемы – описание понятий, необходимых в решении текущей задачи.

В результате нечёткого разграничения между онтологиями и концептуальными схемами сегодня представление знаний часто путается с онтологией, и онтологией называется практически любое понятийное описание. В частности, пользуясь возможностями автоматического вывода в языке OWL [7], изначально разработанном для представления онтологий, OWL начинают применять просто как язык представления знаний. При этом название онтология никуда не девается и используется некорректно. Конкретным примером является практика использования языка OWL-S [8]. Он предназначен для функциональной и процессной спецификации Вэб-сервисов с использованием онтологий. Это значит, что в основе описания OWL-S должна использоваться онтология, утверждения в терминах которой специфицировали бы Веб-сервис. На деле, на некоторых примерах использования OWL-S видно, что основой для описания OWL-S выбирается интерфейс сервиса и классы, необходимые для логического выражения спецификаций сервиса. Ничего общего с онтологией такое описание не имеет, однако в контексте языка OWL, который является средством описания, вся спецификация преподносится как онтология.

Онтология может использоваться корректным образом для построения концептуальной схемы, а далее схемы базы данных или информационной системы и для использования полученных

спецификаций на уровне представления данных. Однако, превращаясь в схему, она перестаёт быть онтологией. В схеме могут оставаться ссылки на лежащую в основе онтологию, для того чтобы можно было использовать онтологический уровень для нужд обеспечения семантической интероперабельности или интеграции. Спецификации всякого информационного ресурса могут изначально разрабатываться на базе онтологии либо связываться с онтологическими определениями в последствии. Так или иначе, существует разделение онтологических спецификаций и спецификаций схем.

Онтологическая модель и способ аннотации объектных спецификаций схем онтологическими определениями, описанные в предыдущем разделе, придерживаются принципа разделения онтологических спецификаций и спецификаций уровня представления. Элементы объектных спецификаций становятся экземплярами классов онтологических понятий, и обратно, онтологические описания фактически становятся метаинформацией, аннотирующей спецификации элементов схем.

Некоторые информационные источники разрабатываются с использованием общих онтологий, с изначальным соглашением об общей интерпретации определений предметных областей. Это упрощает поддержку семантической интероперабельности и корректного взаимодействия ресурсов. Достижение же семантической интероперабельности неоднородных информационных источников будет включать связывание онтологических контекстов между собой и связывание схем источников с использованием онтологии. В следующем разделе рассмотрены пути выявления семантических связей объектных спецификаций информационных источников на уровне представления по известным отношениям между онтологическими понятиями, возможно, разных онтологических контекстов.

4 Выделение семантики уровня представления по онтологическим отношениям

Результатом отображения онтологических контекстов предметных областей информационных источников на онтологический контекст задачи являются обнаруженные между данными онтологиями отношения. В зависимости от сложности изначальных онтологических моделей и методов отображения эти отношения могут быть различных видов. Эти отношения, а также отношения понятий внутри одной онтологии, используемой для нескольких информационных источников, могут быть использованы для предположения онтологической релевантности элементов объектных схем источников. Рассмотрим некоторые виды отношений онтологических понятий внутри или между онтологическими контекстами и предположим по этим отношениям

релевантность аннотируемых онтологическими понятиями элементов объектных схем.

Элементы схем совпадающих видов (тип с типом, атрибут с атрибутом и так далее), находящиеся в одном и том же онтологическом классе, то есть принадлежащие одному и тому же понятию, предполагаются эквивалентными или сильно онтологически релевантными. То же касается элементов, принадлежащих понятиям-синонимам с силой связи 1 (или неуказанной силой связи). Элементы совпадающих видов, принадлежащие понятиям, связанным синонимией или положительным отношением с нечёткой силой связи больше 0, но меньше 1, предполагаются слабо онтологически релевантными.

Элементы схемы, состоящие в уточняемых классах понятий между онтологиями или в классах понятий, состоящих в отношении обобщения/специализации в одной онтологии, сами предполагаются связанными отношением подтипа или подкласса.

Например, тип `Person` в объектных спецификациях требований решаемой задачи является экземпляром класса понятия `human`, как определено выше. А тип `Individual` схемы информационного источника, который предполагается использовать при решении задачи, является экземпляром класса понятия `character` другой онтологии, определяющей контекст источника. В результате отображения онтологии источника на онтологию предметной области задачи выясняется, что понятие `character` является подпонятием понятия `human`. Соответственно, экземпляры класса `character` входят также и в полный экстенционал класса `human`. Элементы схем из класса `character` семантически соответствуют подмножеству класса `human`. В частности, тип `Individual` предположительно является подтипом или эквивалентен типу `Person`.

Одни понятия онтологии могут отражать идентификацию объектов реального мира, образовывать тип. Другие – создавать категории объектов. Третьи – относиться к ролям объектов реального мира. Так, понятие `human` (человек) может образовывать тип и подразумевать идентификацию объекта как человека, тогда как понятие `man` (мужчина) создавать один из двух классов людей, а понятие `father` (отец) относиться к роли человека.

Релевные понятия могут быть подпонятиями отражающих идентификацию. Категорийные понятия могут быть суперпонятиями ролевых или отражающих идентификацию объектов. Релевные друг другу элементы объектных схем, соответствующие понятиям, отражающим идентификацию объектов, следует искать скорее среди типов и атрибутов. Соответствующие категорийным понятиям – среди классов и ассоциаций. К ролевым понятиям относятся классы объектов, множества значений свойств.

По связи часть/целое между понятиями могут иметь место следующие варианты рассуждений.

Совокупность всех частей составляет целое, это может быть использовано на уровне схемы для предположения связи элементов. Если часть является экстенсией, целое может быть идентифицировано по части. Если целое является инвариантным, то части можно идентифицировать по целому. Связь часть/целое является асимметричной и транзитивной. Соответственно, могут использоваться триангулярные правила распространения связей. Варианты выражения частей и целого на уровне представления могут быть различны, вероятнее всего, связью по атрибутам, имеющим тип части.

В качестве примера часть/целое приведём понятие `marriedCouple` (супружеская чета). Его составляют две части: `husband` (муж) и `wife` (жена). Понятие `marriedCouple` является инвариантным целым, так как при отсутствии одной из двух частей не существует и целого. Обратное, понятия `husband` и `wife` являются экстенсией частями, так как каждая из этих частей подразумевает наличие целого.

Здесь были использованы особенности понятий, определяющие, является ли понятие категорическим, ролевым, или отражающим идентификацию объектов, является ли целое инвариантным, а часть экстенсией. Эти и другие подобные метасвойства онтологических понятий более подробно описаны в [3]. В модели СИНТЕЗ для каждого вида метасвойств онтологических понятий определяется соответствующий ему метакласс. Онтологические понятия, обладающие данным свойством, становятся экземплярами соответствующего метакласса.

В описании модели упоминались универсальные функции преобразования значений при переходе между онтологическими контекстами. Такие функции могут быть определены внутри онтологии [1], а также генерироваться в процессе согласования онтологий. К примеру, функция вычисления возраста человека по дате его рождения. Функции преобразования могут использоваться на уровне представления для преобразования значений связанных элементов информационных источников.

5 Заключение

В данной статье обоснована модель аннотации объектных схем неоднородных информационных источников онтологическими определениями. Рассмотрены различные виды отношений между онтологическими понятиями с целью выявления связи между элементами объектных схем неоднородных информационных источников, относящимися к данным понятиям. Выявленные отношения могут быть использованы в процессе интеграции для обеспечения семантической интероперабельности информационных источников.

Литература

- [1] D. Briukhov, L. Kalinichenko, N. Skvortsov, S. Stupnikov. Value Reconciliation in Mediators of Heterogeneous Information Collections Applying Well-Structured Context Specifications In Proceedings of the Fifth International Baltic Conference on Databases and Information Systems BalticDB&IS'2002, Tallinn, Estonia, June 3-6, 2002
- [2] N. Guarino. Formal Ontology and Information Systems. In N. Guarino (ed.), Formal Ontology in Information Systems. Proc. of the 1st International Conference, Trento, Italy, 6-8 June 1998. IOS Press
- [3] N. Guarino. Ontology-Driven Conceptual Modeling.
- [4] L. Kalinichenko, N. Skvortsov. Extensible ontological modeling framework for subject mediation In Proceedings of the 4-th Russian Scientific Conference "Digital Libraries: Advanced Methods and Technologies, Digital Collections", Dubna, Russia, Oct 15-17, 2002
- [5] L. Kalinichenko, N. Skvortsov. Ontology reconciliation in terms of type refinement In Proceedings of the 6th Russian Conference on Digital Libraries RCDL2004, Pushchino, Russia, Sep 2004
- [6] N. Skvortsov, L. Kalinichenko. An Approach to Ontological Modeling and Establishing Intercontext Correlation in the Semistructured Environment 2-nd Russian Scientific Conference "Digital Libraries: Advanced Methods and Technologies, Digital Collections", Protvino, Russia, Sep 26-28, 2000
- [7] OWL Web Ontology Language Guide. W3C. <http://www.w3.org/TR/owl-guide/>
- [8] OWL-S: Semantic Markup for Web Services. <http://www.daml.org/services/owl-s/>

Object specification relevance detection by semantics of ontological level

N. A. Skvortsov

Ontological approach to solving tasks with the use of heterogeneous information sources implies reconcilability of ontological contexts of information source's and task's subject domains. On basis of relationships between reconciled ontologies, the problem of semantic relevance of task and source object schema specifications may be solved. The paper gives ground for the model of relating of information source object schemas to ontological specifications and considers which kinds of relationships may be derived from relationships in ontological information. There are different kinds of relations of ontological level taken to consideration: synonym, hypernym/hyponym, attributes, associations, part/whole, value contextualization functions.

* Исследование поддержано грантами РФФИ 06-07-89188-а, 05-07-90413-в, 06-07-08072-офи.