

Отображение модели данных RDF в каноническую модель предметных посредников

Скворцов Н. А.
Институт проблем информатики РАН
nskv@ipi.ac.ru

RCDL-2013, Ярославль

План

- Контекст работы
- Подходы к отображению RDF
- Семантический подход к отображению RDF
- Модель RDF и целевая модель языка СИНТЕЗ
- Отображение модели RDF
- Отображение модели RDF Schema
- Отображение языка запросов SPARQL
- Производные модели
- Заключение

Предметные посредники

- Промежуточный слой предметных посредников
 - Каноническая информационная модель
 - Однородное представление спецификаций и информации
 - Спецификации предметной области
 - Онтологии
 - Концептуальные схемы для определения структуры и поведения информационных объектов
 - Набор неоднородных информационных ресурсов, интегрированных в спецификации предметной области
 - Средства переписывания и планирования запросов
 - Запросы в терминах концептуальной схемы предметной области
 - Решение задач в терминах спецификаций предметной области

Создание канонической модели предметного посредника

- Каноническая модель формируется из
 - Ядра модели
 - В качестве ядра используется язык спецификаций СИНТЕЗ: язык фреймов, объектная модель, язык формул первого порядка
 - Набора расширений ядра для отображения неоднородных моделей данных
- Унификация неоднородных моделей данных
 - Отображение моделей данных в ядро канонической модели
 - Посредством использования существующих расширений или создания новых

Информационные ресурсы и отображение моделей

- Неоднородность информационных ресурсов
 - На уровне моделей данных
 - На уровне схем
 - Семантическая неоднородность
- Интеграция информационных ресурсов в посредник
 - Модель данных информационных ресурсов отображается в каноническую модель посредника
 - Отображение концептуальных схем ресурсов в спецификации предметной области
- До сих пор рассмотрены отображения для моделей
 - Традиционные модели данных
 - Реляционные, объектные
 - Перспективные модели
 - Онтологические модели (Ontolingua, OWL, OWL 2)
 - Модели многомерных массивов данных
 - Модели баз данных NoSQL (ключ-значение, колоночные, документные)
 - Графовые модели

Модель данных RDF и необходимость её отображения

- Модель данных RDF
 - Триплеты «субъект – предикат – объект»
 - Идентификация и описание ресурсов в открытой среде
 - Базовая модель семантического веба
- RDF – разновидность графовой модели
 - Субъекты и объекты – узлы
 - Предикаты – направленные именованные рёбра
- Специфика RDF-графов
 - Глобальная идентификация узлов посредством URI
 - Определённая семантика рёбер, определяемая словарями
 - Определённые правила вывода неявных связей
 - Отображение модели RDF заслуживает отдельного рассмотрения
- Обогащение модели RDF связанными моделями
 - RDF Schema – определение схем для RDF-описаний (определение данных)
 - SPARQL – язык запросов к RDF-триплетам (манипуляция данными)
 - OWL – в контексте семантического веба

Представление данных RDF и доступ к ним

- Документы
 - В формате одного из синтаксисов RDF
 - Документы, распределённые в открытой информационной среде
 - RDF-данные о ресурсах, вкраплённые в XML и HTML
 - Собранные поисковыми механизмами на основе SPARQL
- Специализированные RDF-хранилища
 - На основе реляционных отношений
 - С колоночной организацией
- Интерфейсы API к генерируемым данным в формате RDF
 - Точки доступа SPARQL
- Использование правил вывода
 - Без вывода
 - С выбранным или полным набором RDF-правил
 - С расширенным набором правил (например, для данных в OWL)

Подходы к отображению RDF

- Большинство RDF-данных хранится и обрабатывается в других моделях
 - Обмен данными в RDF-формате
 - Публикация данных в открытое пространство RDF-триплетов
- Инструменты работы с RDF имеют в интерфейсах триплеты
 - Отображение в RDF отдаётся на откуп прикладных программ
- Различные представления RDF в реляционных хранилищах
 - Единственное отношение с атрибутами для хранения субъектов, предикатов и объектов
 - Разбиение отношений
 - По классам субъектов
 - По предикатам
 - Хранение в кортеже набора предикатов и объектов, связанных с одним субъектом
- Отображение, учитывающее семантику данных
 - При трансляции данных из объектной модели в RDF
- Отображение языков запросов
 - Из SPARQL в SQL
 - Зависит от способа отображения RDF в реляционную модель

Модель RDF

- RDF
 - Триплеты: ресурсы и предикаты
 - Пространства имён для предикатов и ресурсов
 - Идентификация ресурсов и предикатов (URI)
 - Безымянные узлы
 - Примитивные типы данных в объектах
 - Контейнеры bag, seq, alt; коллекции list
- RDF Schema
 - Классы, предикат type
 - Свойства, области определения и значений
 - Отношения обобщения/специализации на классах и свойствах
 - Список правил вывода
 - транзитивность и рефлексивность отношений обобщения/специализации
 - отнесение ресурсов к классу в соответствии с областью определения или значения свойства
 - правила о веществе
- SPARQL
 - Конъюнкции триплетов с переменными
 - Объединение
 - Фильтрация по условию
 - Необязательные элементы
 - Представление результатов в виде списка кортежей или триплетов

Целевая модель СИНТЕЗ

- Язык фреймов
 - Идентификатор фрейма
 - Слоты, значения слотов
 - Метафреймы, метаслоты, метазначения
- Объектная модель
 - Объекты
 - Фреймы как типизированные значения объектов
 - Абстрактные типы данных
 - Определение структуры и поведения объектов
 - Решётка типов
 - Типовые выражения
 - Классы
 - Множества объектов определённого типа
 - Иерархия класс/подкласс, иерархия классификации
 - Метаклассы ассоциаций
- Язык формул
 - В предикативных спецификациях ограничений целостности типов
 - В правилах логических программ
 - В запросах над спецификациями

Выбор способа отображения

- Возможное тривиальное отображение:
абстрактный тип с атрибутами для субъектов,
предикатов и объектов
 - Не удовлетворяет нуждам предметных посредников
- В среде предметных посредников важна
семантическая интеграция информационных
ресурсов
 - Выявление семантики предметной области в данных
 - Выявление структуры данных
 - Создание спецификаций информационного ресурса в
канонической модели для дальнейшего связывания с
предметной областью посредника в соответствии с
выявленными семантикой и структурой

Отображение модели RDF-данных

- RDF-данные отображаются средствами фреймовой модели языка СИНТЕЗ (в базу фреймов)
 - Обе модели служат для описания ресурсов произвольной природы

- Пример данных:

```
prefix vCard: <http://www.w3.org/2001/vcard-rdf/3.0#> .  
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
prefix : <#> .  
<http://somewhere/MattJones/>  
  vCard:FN "Matt Jones" ;  
  vCard:N [ vCard:Family "Jones";  
            vCard:Given "Matthew" ] .
```

Отображение модели RDF-данных

- Пространства имён (словари) отображаются в миры фреймов

```
{ vCard;  
  in: world; ...  
}
```
- Триплеты с общим субъектом отображаются в один фрейм
 - Предикаты – в слоты
 - Объекты с одним предикатом – в список значений слотов
 - Неименованные узлы – во встроенные фреймы без идентификатора

```
{ MattJones;  
  in: frame;  
  FN: Matt Jones;  
  N: { in: frame;  
      Family: Jones;},  
  { in: frame;  
    Given: Matthew;};  
}
```

Отображение модели RDF Schema

- Описания в модели RDF Schema отображаются в объектную модель языка СИНТЕЗ
 - Обе модели определяют схемы данных посредством задания структуры, множеств и ограничений

prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>

prefix vcard: <http://www.w3.org/2001/vcard-rdf/3.0#>

vCard:Individual rdf:type rdf:Class .

vCard:FN rdf:type rdf:Property;

 rdfs:domain vCard:Individual .

vCard:NPROPERTIES rdf:type rdfs:Class .

vCard:N rdf:type rdf:Property;

 rdfs:domain vCard:Individual;

 rdfs:range vCard:NPROPERTIES .

vCard:Family rdf:type rdf:Class;

 rdfs:subClassOf vCard: NPROPERTIES .

vCard:Given rdf:type rdf:Class;

 rdfs:subClassOf vCard: NPROPERTIES .

Отображение модели RDF Schema

- Ресурс RDF Schema – во фрейм языка СИНТЕЗ
 - или в объект -экземпляр класса
- Класс RDF Schema – в класс языка СИНТЕЗ с типом экземпляров
 - subclass – в отношении подкласса между классами
 - bag, seq, list – в коллекции языка СИНТЕЗ
- Свойства RDF Schema – в метакласс ассоциаций
 - domain и range – в определения домена и диапазона значений в метаклассе ассоциаций
 - subproperty – в отношении подкласса между метаклассами ассоциаций
- Правила вывода игнорируются, если применены на уровне инструмента RDF

```
{ n;  
  in: association, metaclass;  
  instance_section:  
  { association_type: {0, inf},{0, inf};  
    domain: individual;  
    range: nproperties;  
  }  
},
```

```
{ individual;  
  in: class;  
  instance_section:  
  {  
    FN: {set_of: Frame};  
    N: {set_of: nproperties};  
    metaslot in: n; end  
  };  
},  
{ nproperties;  
  in: class;  
  instance_section: Frame  
},  
{ family;  
  in: class;  
  superclass: nproperties  
  instance_section: Frame  
},  
{ given;  
  in: class;  
  superclass: nproperties  
  instance_section: Frame  
};
```

Конфликт парадигм

RDF

- парадигма открытого мира
- Принадлежность ресурсов нескольким классам
- Свойства – самостоятельные сущности, используемые в разных классах и ресурсах
- Структура экземпляра не ограничивается спецификацией класса

Объектные модели

- парадигма закрытого мира
- Принадлежность объекта строго одному тип
- Атрибуты инкапсулированы в типах, наследуются с типом
- Объект жёстко соответствует структуре, заданной типом

Решение проблемы конфликта парадигм средствами языка СИНТЕЗ

- Разделение на классы как множества и абстрактные типы данных как интенциональные описания
 - Объекты одного типа могут принадлежать разным классам
- Метаклассы ассоциаций
 - Множества значений атрибутов как самостоятельные сущности
- Возможность использовать типовые выражения для комбинации типов
- Язык фреймов для описания объектов

Отображение SPARQL

- Необходимо обратное отображение: из языка СИНТЕЗ в запрос на языке SPARQL
 - Для переписывания запросов из терминов предметной области посредника в запросы к информационным ресурсам
 - Из формул языка СИНТЕЗ в конъюнктивной форме

```
prefix vcard: <http://www.w3.org/2001/vcard-rdf/3.0#>
```

```
SELECT ?x ?fname ?gname
```

```
{
```

```
  ?x vcard:N ?vc .
```

```
  ?vc vcard:Family ?fname .
```

```
  ?vc vcard:Given ?gname .
```

```
}
```

- отображение формул языка СИНТЕЗ в конъюнктивной форме – в триплеты правила
 - Переменные результата – в переменные в голове правила
 - Безымянным узлам соответствует утверждение с квантором существования
 - Ограничения классов в запросе не должны противоречить определениям domain и range в RDF Schema
- В случае отображения RDF во фреймы канонической модели в качестве запросов применяются формулы канонической модели над базой фреймов

$q([x, \text{fname}, \text{gname}]) :-$
 individual(x),
 ex y (family(y) & is_in(y, x.N) & fname=y),
 ex z (given(z) & is_in(z, x.N) & gname=z)

Выводы

- Определены принципы семантического отображения модели RDF в целевую модель в случаях, как для случая возможного выявления схемы данных, так и произвольных данных без схемы
- Учтены особенности модели данных, определяемой совместно моделями RDF и RDF Schema и языком запросов SPARQL
- Когда RDF-данные ограничены спецификациями в производной модели с собственной семантикой, это существенно влияет на семантическое отображение RDF-данных в целевую модель
 - Отображение для OWL и OWL 2 было определено в предшествующих работах