

Интеллектуальный поиск медиаконтента с применением онтологического подхода

Соколенко Вл.В.; Соколенко В.Вл.

Кафедра САПР, факультет компьютерных технологий и информатики (ФКТИ)
Санкт-Петербургский государственный электротехнический университет им. В.И. Ульянова (Ленина) «ЛЭТИ»

Санкт-Петербург, Россия

e-mail: sokolenko88@gmail.com

Аннотация — Предложен способ, позволяющий оперативно осуществлять поиск локального учебного контента при наличии одного, или нескольких условий для поиска.

Ключевые слова: Ресурсы спутникового телерадиовещания, контент, онтология, семантическая сеть

В СПбГЭТУ, в рамках проведения работ по реализации конкурсных проектов по межвузовской комплексной работе «Инновационные технологии образования» (МКР ИТО), в период 2008-2010гг. был построен спутниковый приемный телепорт. Межкафедральной рабочей группой были разработаны и созданы оригинальные программно-технические средства и предложена перспективная технология обучения иностранным языкам на базе мультимедийного комплекса с использованием аутентичных учебных ресурсов спутникового телерадиовещания.

Основная цель представляемого проекта состоит в создании условий для поддержки системного внедрения и активного использования информационно-спутниковой образовательной технологии (ИСТ) в учебный процесс СПбГЭТУ. В ходе реализации проекта предполагается разработка комплексных учебно-методических материалов, включающих цифровые (электронные) компоненты, по дисциплинам кафедры иностранных языков, а также образовательные модули по отдельным темам, соответствующим учебным планам. Разрабатываемые учебные материалы будут в своей основе содержать единую, «корпоративную» общедоступную коллекцию цифровых образовательных ресурсов (ЦОР). Коллекция ЦОР будет представлять собой систематизированное собрание цифровых ресурсов, включающее предметные и тематические коллекции, объединенные единой системой описания ресурсов и единой поисковой системой. Создаваемые в ходе реализации проекта учебные материалы будут объединяться в учебно-методические комплексы (УМК), содержащие не только образовательные ресурсы, но и методику организации учебного процесса.

За год функционирования мультимедийного спутникового комплекса в СПбГЭТУ был накоплен значительный объем актуальных ЦОР, общий объем которого превышал сотни гигабайт. Необходимо было решить вопрос об оперативном доступе к нужным ресурсам, например, оперативный поиск аутентичных видео-сюжетов по радиоэлектронике (робототехнике) на немецком языке, или оперативный поиск аудио-сюжетов по истории Великобритании на английском языке.

В результате, было принято решение о введении онтологий, что позволяло ввести метаданные, описывающие семантические свойства ресурсов и концептуальные схемы, описывающие объекты ресурсов и отношения между ними.

В качестве предметной области для онтологии рассматривался контент, хранящийся локально в мультимедийной лаборатории.

Онтология проектировалась в среде Protégé.

Во-первых, было произведено толкование важных понятий в рамках вербальной модели и создан глоссарий. Рассмотрены следующие понятия: мультимедиа, файл, расширение, тема, протокол, доступность, язык.

Далее были определены категории онтологии (классы):

- multimedia – мультимедийные ресурсы
- type – тип файла
- availability – открытость файла (бесплатность)
- extension – расширение файла
- theme – тематика файла
- protocol – протокол доступа к файлу

Во-вторых, были определены предикаты, константы и функции. Был введен следующий предикат: $name(x, y)$ – предикат, показывающий наличие у объекта x свойства «имя» со значением y .

Затем были определены отношения между классами:

- $select_by(x,y)$ - предикат, задающий отношение «выбран» между мультимедийным файлом x и типом мультимедийного файла y .
- $scan_by(x,y)$ - предикат, задающий отношение «просканирован» между мультимедийным файлом x и доступностью мультимедийного файла y .
- $record_in(x,y)$ - предикат, задающий отношение «записан» между мультимедийным файлом x и расширением мультимедийного файла y .
- $control_by(x,y)$ - предикат, задающий отношение «контроль» между мультимедийным файлом x и тематикой мультимедийного файла y .
- $locate_to(x,y)$ - предикат, задающий отношение «располагается» между мультимедийным файлом x и расположением мультимедийного файла y .

В результате, была построена следующая семантическая сеть (Рис. 1.):

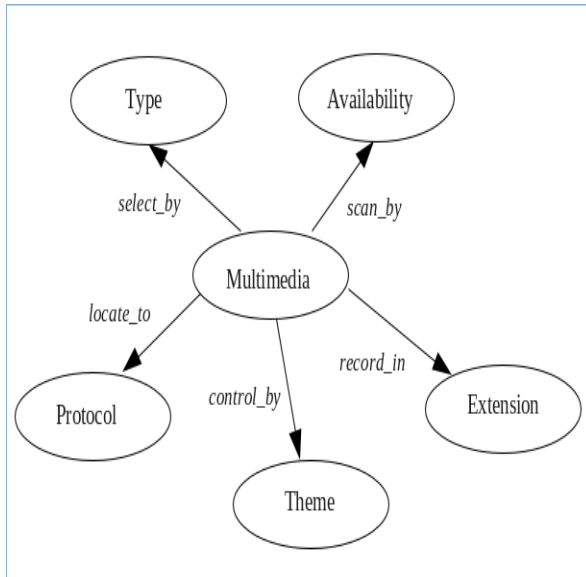


Рис. 1. Семантическая сеть

Следующим шагом было составление конкретного экземпляра задачи:

```

isMemberOf(audio, Type)
name(audio, audio)
isMemberOf(video, Type)
name(video, video)
isMemberOf(graphic, Type)
name(graphic, graphic)
isMemberOf(text, Type)
name(text, text)
isMemberOf(free, Availability)
name(free, free)
isMemberOf(nonfree, Availability)
name(nonfree, nonfree)
isMemberOf(avi, Extension)
name(avi, avi)
isMemberOf(mkv, Extension)
name(mkv, mkv)
isMemberOf(mpeg, Extension)
name(mpeg, mpeg)
isMemberOf(robotics, Theme)
name(robotics, robotics)
isMemberOf(programming, Theme)
name(programming, programming)
isMemberOf(lan, Protocol)
name(lan, lan)
isMemberOf(ssh, Protocol)
name(ssh, ssh)
select_by(Multimedia, Type)
scan_by(Multimedia, Availability)
record_in(Multimedia, Extension)
control_by(Multimedia, Theme)
locate_to(Multimedia, Protocol)
  
```

Для извлечения явных знаний из онтологии использовался SPARQL. Например: какие существующие файлы имеют тематику программирования?

```
PREFIX M: <http://www.owl-ontologies.com/Multimedia.owl#>
```

```
SELECT ?Multimedia WHERE {
  ?Multimedia M:control_by M:programming
}
```

Еще пример: какие существующие файлы имеют тематику робототехники, имеют расширение .mkv и записаны с “открытых спутников”?

```
PREFIX M: <http://www.owl-ontologies.com/Multimedia.owl#>
```

```
SELECT ?Multimedia WHERE {
  ?Multimedia M:control_by M:robotics;
  M:record_in M:mkv;
  M:scan_by M:free
}
```

В рамках работы в аспирантуре, одной из краткосрочных задач будет разработка пользовательского приложения для работы с онтологией, которое будет позволять с помощью SPARQL-запросов извлекать явные и неявные знания.

Описанный метод интеллектуального поиска информации с применением онтологии позволит студентам и преподавателям мультимедийной лаборатории существенно упростить поиск и улучшить его эффективность, благодаря возможности задания нескольких условий для поиска.

- [1] The 8th International Conference «Television: Image Transmission and Processing» Spb. May 30-31, 2011, pp. 152-155
- [2] I.V.Gerasimov, A.I.Majga, L.N.Lozaovoj Ontological Engineering. Means and Specifications of Ontological Modelling. - SPb.: SPbGETU "LETI", 2009.
- [3] SPARQL - Query Language for RDF (<http://www.w3.org/TR/rdf-sparql-query/>).
- [4] SPARQL - Query Language for RDF (<http://www.semanticweb.narod.ru/3.html>).
- [5] Ogoltsova, N.N.Multimedia Projects of Improvement of Teachers and Teachers of High Schools Professional Skills / N.N.Ogoltsova, V.A.Starodubtsev//News of Tomsk Polytechnical University. – 2007. - T. 310, № 2. - 254-257.