

Средства автоматической интеграции разнородных ОНТОЛОГИЙ

Панасенко Алексей, 525 группа
Научный руководитель Большакова Е.И.

Понятие онтологии

Онтология – конструкция для представления знаний в Интернете, представляющая собой множество объектов, классифицированных в соответствии с некоторыми критериями, плюс описание свойств этих объектов. Онтология кодирует объекты и свойства в понятном для компьютера формате.

Применение онтологий

- Информационный поиск (Information retrieval)
 - Обеспечение концептуального индексирования
- Семантический Веб (Semantic Web)
 - Автоматизация «интеллектуальных» задач обработки значения ресурсов сети

Методы создания

- 3 подхода к созданию онтологий
 - Автоматическое извлечение знаний
 - Использование экспертов
 - Использование сообщества
- У каждого – свои недостатки

Для получения наилучшего результата необходимо комбинировать подходы

Цель дипломной работы

Цель – разработка средств интеграции двух разнородных онтологий – русскоязычной Wikipedia и базы знаний, составленной сотрудниками НИВЦ МГУ.

Задачи

- Анализ существующих проектов
- Выбор формата результирующей онтологии
- Проектирование и кодирование средств автоматической интеграции
- Выборочное тестирование результирующей онтологии

Обзор аналогов

- Наиболее близкий аналог – YAGO
 - Авторы – Fabian M. Suchanek, Gerhard Weikum
 - Автоматическое получение знаний из WordNet и Wikipedia
 - 1.7 миллиона сущностей
 - 15 миллионов фактов
 - Язык описания онтологии – RDF/RDFS

Язык описания результата

RDF/RDFS – Resource Description Framework

- Разработан World Wide Web Consortium
- Создан специально для обмена онтологиями через Internet
- Синтаксис на основе XML

Средства интеграции

- Язык разработки – Python 2.6.2
- Сначала онтология НИВЦ конвертируется в RDF, затем пополняется из XML-dump Wikipedia
- Для пополнения используются категории и панели информации Wikipedia

Результат

- Проведен обзор аналогов
- Выбран стандарт RDF/XML в качестве языка описания результирующей онтологии
- На языке Python созданы средства интеграции онтологий

Направления дальнейшей работы

- Улучшение эвристик для определения категорий сущностей
- Тестирование новых эвристик
- Оптимизация программы по скорости выполнения

Спасибо за внимание