

Подход к созданию персональной семантической электронной библиотеки

Атаева О.М. (ВЦ РАН)
Серебряков В.А. (ВЦ РАН)

Предпосылки

- Semantic Web
- Web of Data
 - LOD
- Linked Library Data (W3C)

Электронная библиотека

- Стандарты
 - Delos, ...
- Проблемы
 - Концентрация на библиографии
 - «Плоское» представление контекстной информации

Семантическая электронная библиотека

- Семантические технологии
 - Расширение функциональности
 - Интеграция ресурсов разных типов
 - Интероперабельность
- Онтологии
 - Библиографические
 - Структурные
 - Пользовательские

Персональная семантическая электронная библиотека

информационная система

- тематика наполнения определяется пользователем
- источники данных (из LOD)
при этом пользователь может не быть осведомлен о структуре данных источника
- не требует от пользователя специальных знаний

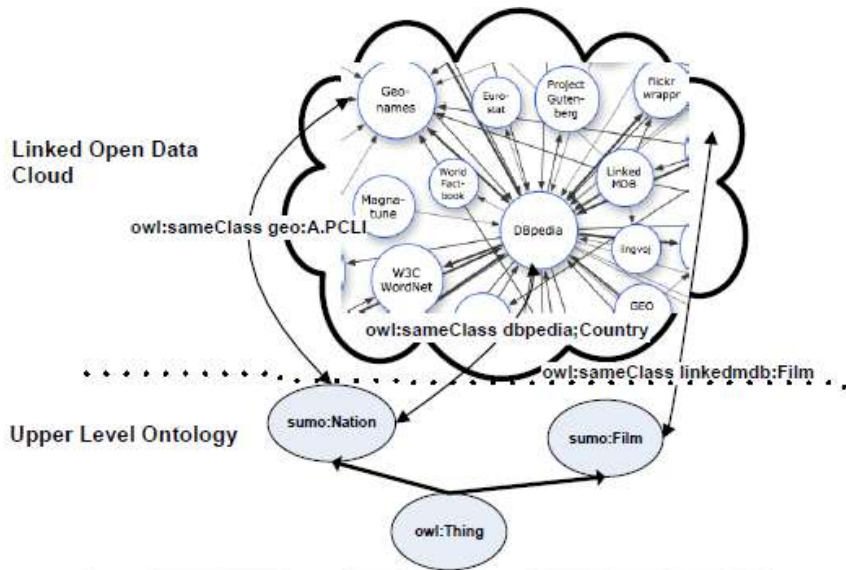
Linked Open Data (LOD)

- Задачи
 - наполнение данными в стандартных форматах Semantic Web
 - устанавливание связей
- Основные принципы
 - использовать URI для идентификации
 - использовать HTTP URIs для доступа к информации о ресурсах.
 - предоставлять информацию о ресурсе в RDF формате.
 - предоставлять RDF-ссылки на другие ресурсы.

ПОИСК В LOD

- Falcons (IWS, China)
- Sig.ma (DERI, Ireland)
- Swoogle (UMBC, USA)
- VisiNav (DERI, Ireland)
- Watson (Open University, UK)

Существующие подходы LOQUS

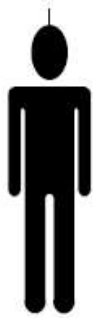


1. Identify movies, countries where they were shot and the current population of these countries.

2. Identify concepts from upper level ontology to serialize into the query.

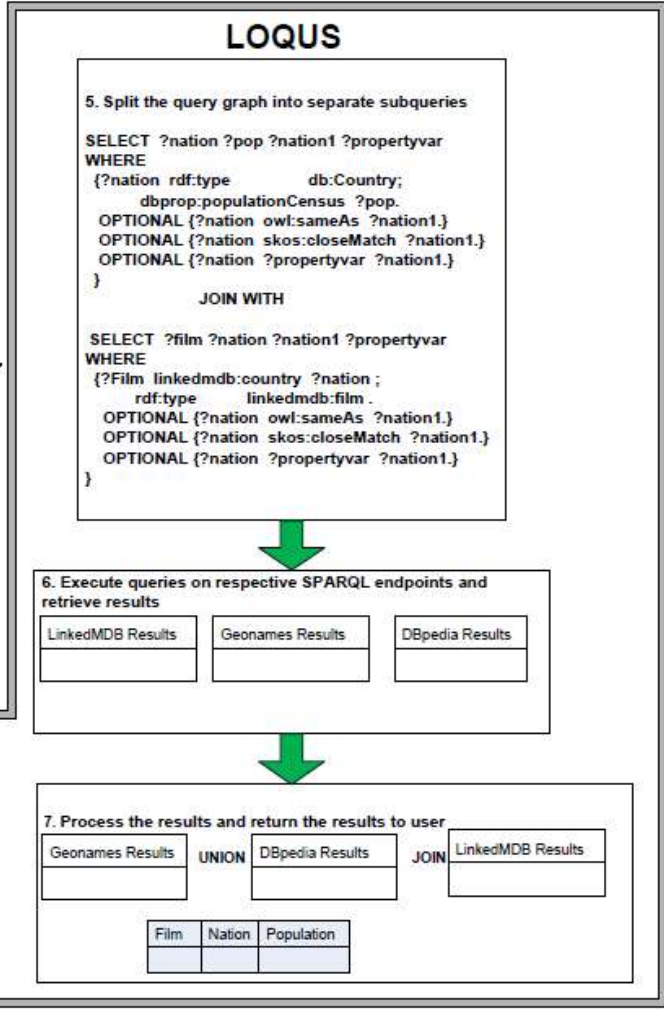
3. Submit query to LOQUS

```
Select ?film ?nation ?pop
WHERE
{ ?film sumo:location ?nation;
  rdf:type silo:Film.
  ?nation rdf:type silo:Nation.
  sumo:population
  ?pop.}
```



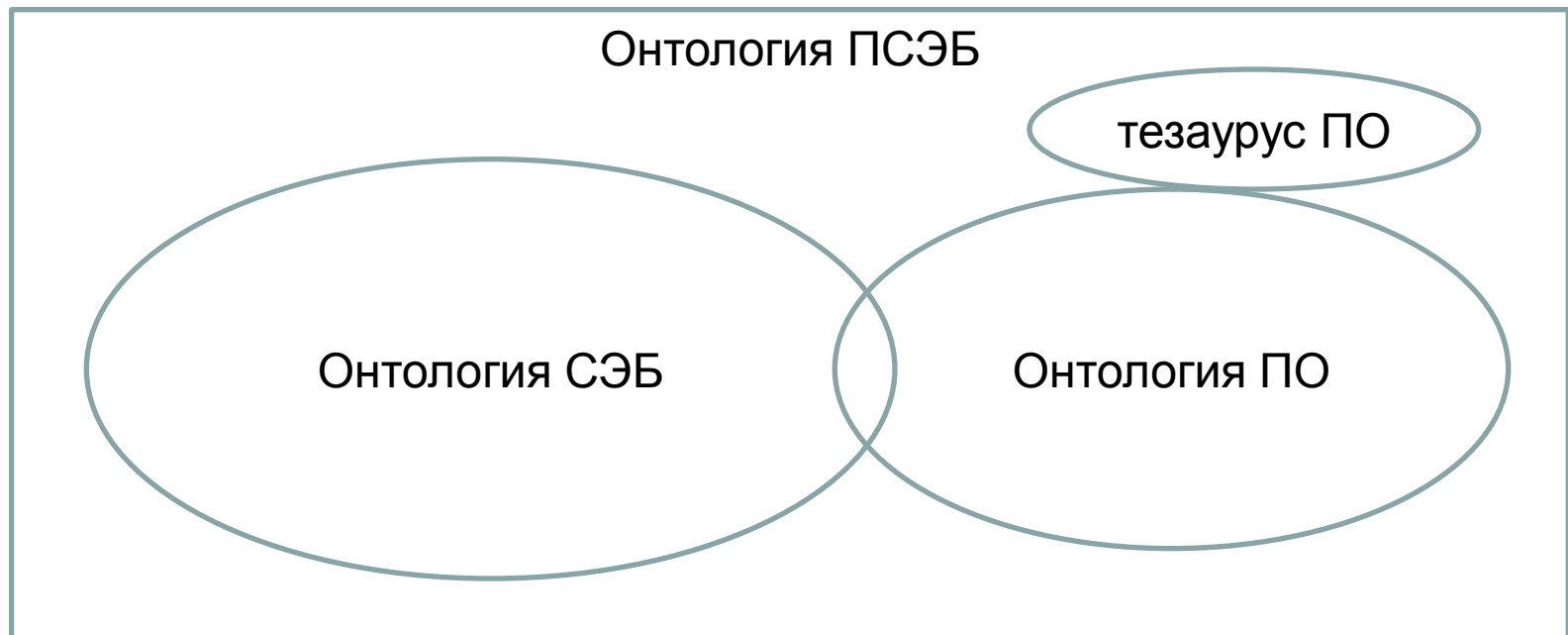
4. Create LOD specific query using the mappings

```
SELECT ?film ?nation ?pop
WHERE
{ ?film linkedmdb:country ?nation ;
  rdf:type linkedmdb:film .
  ?nation rdf:type dbpedia:Country ;
  geo:featureCode geo:A.PCLI ;
  dbprop:populationCensus ?pop ;
  geo:population ?pop .
}
```

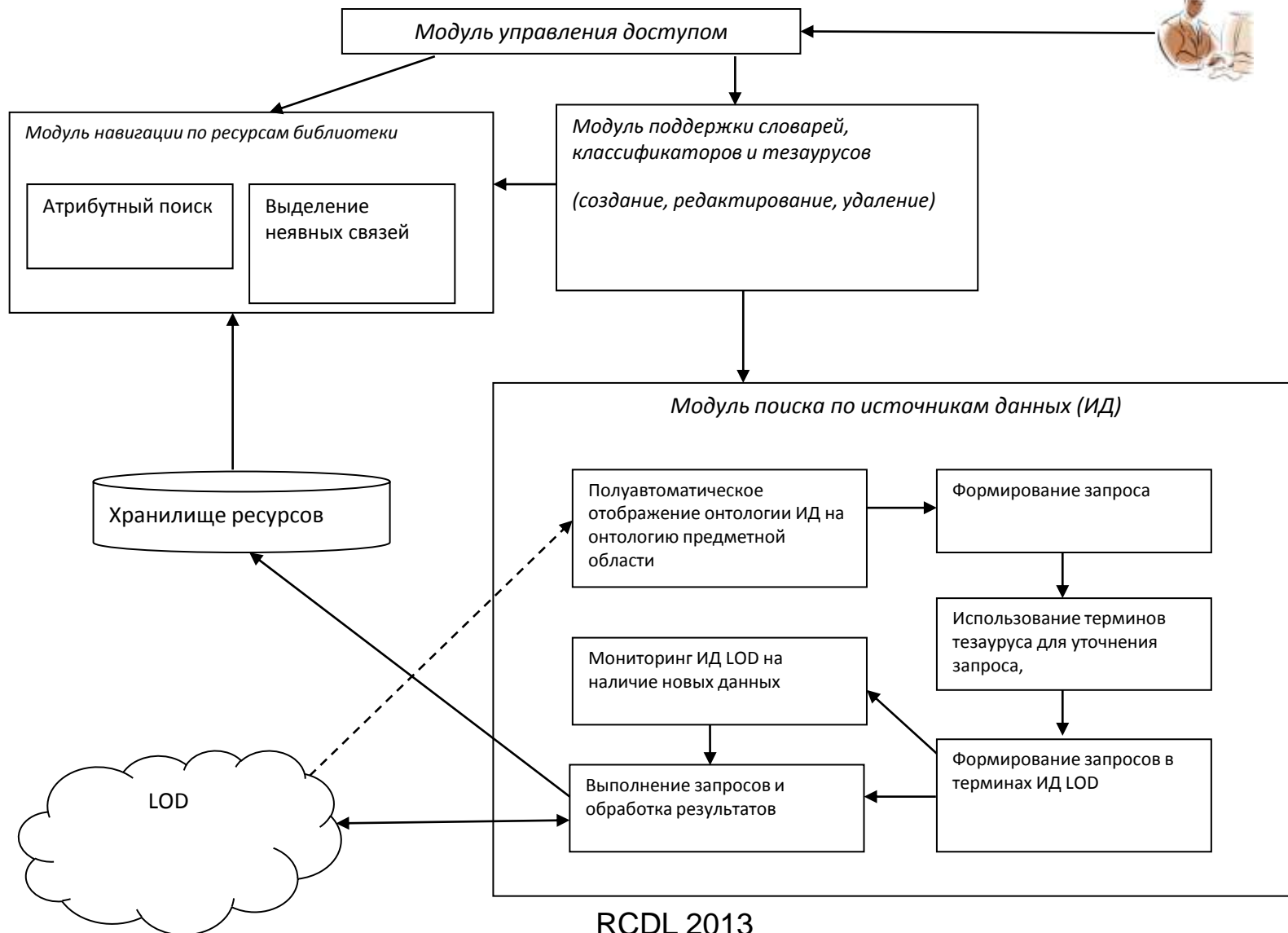


Онтология ПЭСБ

- Основана на онтологии описания информационных систем WWW



Архитектура



Модуль управления доступом

- Роли – Пользователи
- Доступ к различным функциям в зависимости от роли пользователя
 - Поиск
 - Атрибутный
 - Поиск новых связей между ресурсами

Модуль навигации по ресурсам библиотеки

- *Функции атрибутивного поиска*
- *Функция выделения неявных связей между ресурсами по их описаниям*
- *Функции работы с коллекциями*
 - *Создание/Просмотр/Редактирование/Объединение/Вложенные коллекции*

Модуль поиска по источникам данных (ИД)

- Функция отображения онтологии ИД*
- Функция детализации обеспечивает преобразование в подзапросы, соответствующих различным ИД*
- Функция для выполнения запросов и обработки результатов и предоставления окончательного результата пользователю*
- Функция автоматического мониторинга ИД на наличие новых/измененных данных*

Представление результатов поиска

- Проблемы
 - Идентификация
 - Устранение избыточности
 - Достоверность
 - Преобразование

Модуль поддержки словарей, классификаторов, тезаурусов

- *Создание словарей, классификаторов, тезаурусов*
- Редактирование элементов
- Поддержка («гибкой») классификации ресурсов
- Поддержка настройки уровней доступа к различным ветвям тезауруса

Виды функций ПСЭБ

- Функции работы с «внутренними ресурсами»
 - Собственно «обычная» библиотека
- Функции работы с «внешними ресурсами»
 - Все что связано с онтологией предметной области и поиском в LOD

Источники данных в LOD

- Проблемы неоднородности
 - Структурной
 - Семантической
- Преодоление неоднородности
 - Предварительный анализ
 - на сходство символических имен терминов, структурное положение понятий, степень сходства множеств необходимых и достаточных атрибутов.

Источники данных в LOD

- Частичное соответствие
 - Соответствие есть отображение понятий и отношений одной онтологии на другую.
 - Соответствие может быть определено не полностью и является *частичным*, если, может существовать несколько понятий в одной онтологии, не имеющих своих эквивалентов в другой.

Источники данных в LOD

- Проблемы связанности
 - На уровне схем источников
- Преодоление проблем связанности
 - Анализ связей sameAs, seeAlso на уровне данных

Дальнейшее развитие

Методы text mining как вспомогательные методы

- для выявления связей между ресурсами
- для уточнения тематической направленности...

Спасибо за внимание!