РАЗРАБОТКА ЯДРА НАСТРАИВАЕМОГО ПОРТАЛА НАУЧНЫХ ЗНАНИЙ СРЕДСТВАМИ ТЕХНОЛОГИИ SEMANTIC WEB

Шибков Алексей Викторович

Научный руководитель:

к.т.н., зав.лаб. ИСИ СО РАН,

Загорулько Юрий Алексеевич

ВВЕДЕНИЕ

- в сети интернет огромный объем информационных ресурсов по различным областям научных знаний;
- доступ к информации ограничен из-за того, что она не достаточно структурирована и систематизирована;
- решением данной проблемы может стать реализация концепции Semantic Web;
- одной из реализаций концепции Semantic Web является воплощение в жизнь порталов научных знаний.

ПРЕДМЕТНАЯ ОБЛАСТЬ

Semantic Web – это направление развития всемирной паутины, целью которого является предоставление информации в виде, пригодном для машинной обработки.

Машинная обработка возможна благодаря двум ее важнейшим характеристикам:

- Использование унифицированных идентификаторов ресурсов(URI)
- Использование семантических сетей и онтологий

Семантическая сеть – это информационная модель предметной области, имеющая вид ориентированного графа, вершины которого соответствуют объектам предметной области, а дуги задают отношения между ними.

Онтология – это спецификация некоторой области знаний с помощью концептуальной схемы.

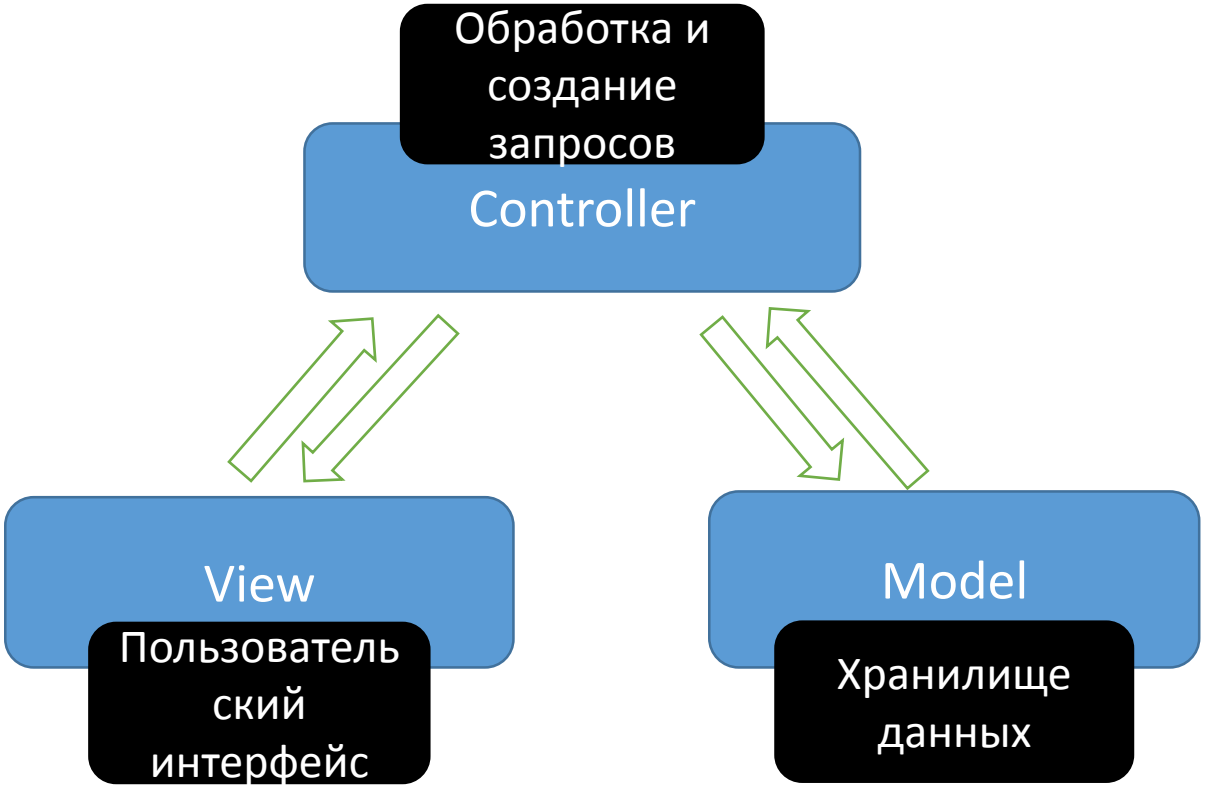
ЦЕЛЬ

Разработка ядра настраиваемого портала научных знаний, система знаний, контент и функциональность которого базируются на онтологии, выступающей в качестве информационной модели портала.

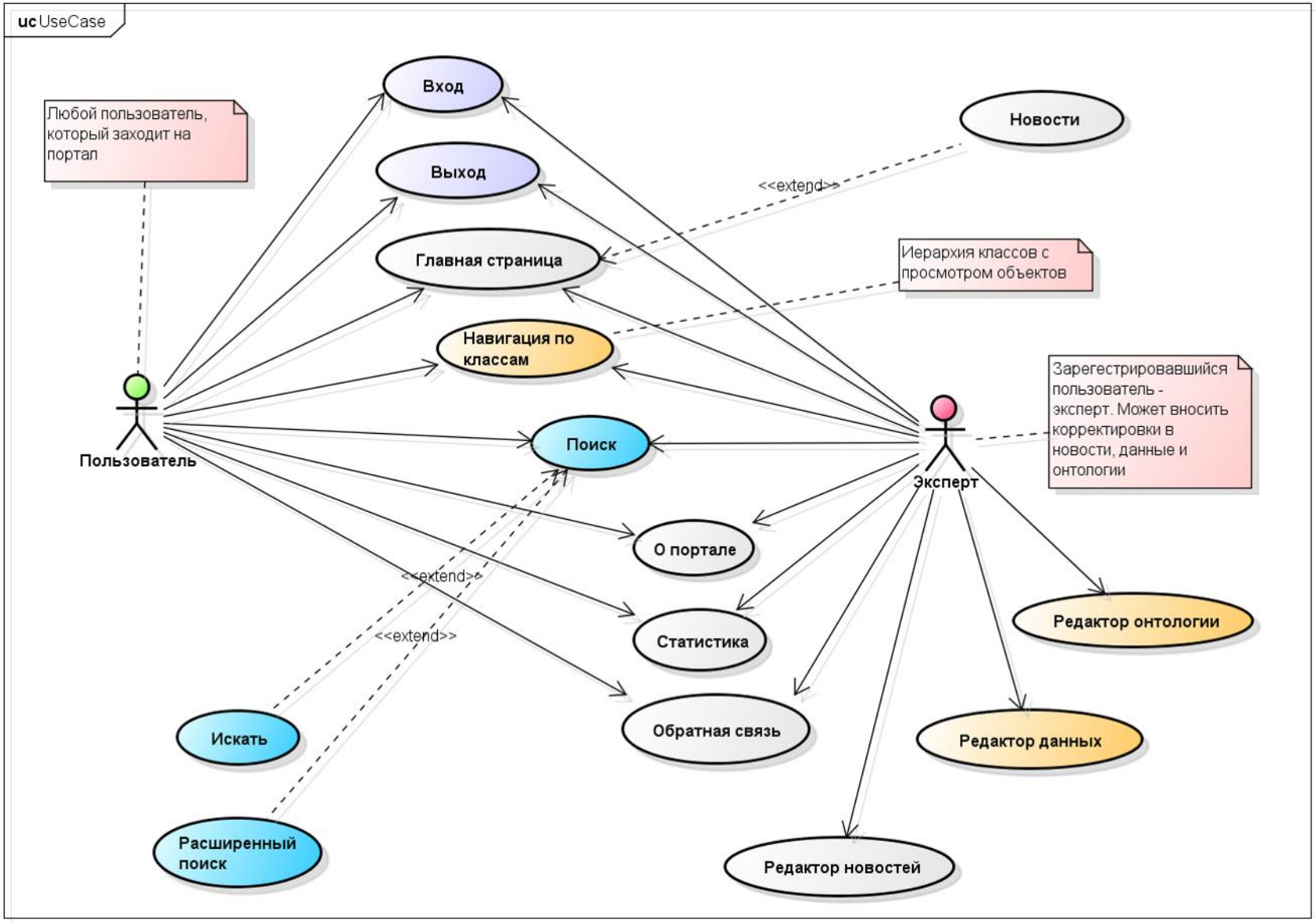
Задачи:

- Исследовать и выбрать средства технологии Semantic Web, подходящие для разработки портала научных знаний.
- Разработать архитектуру портала научных знаний.
- Спроектировать хранилище данных портала.
- Спроектировать и реализовать функциональность подсистем навигации и поиска портала, а также функциональность подсистем редактирования данных и онтологии.

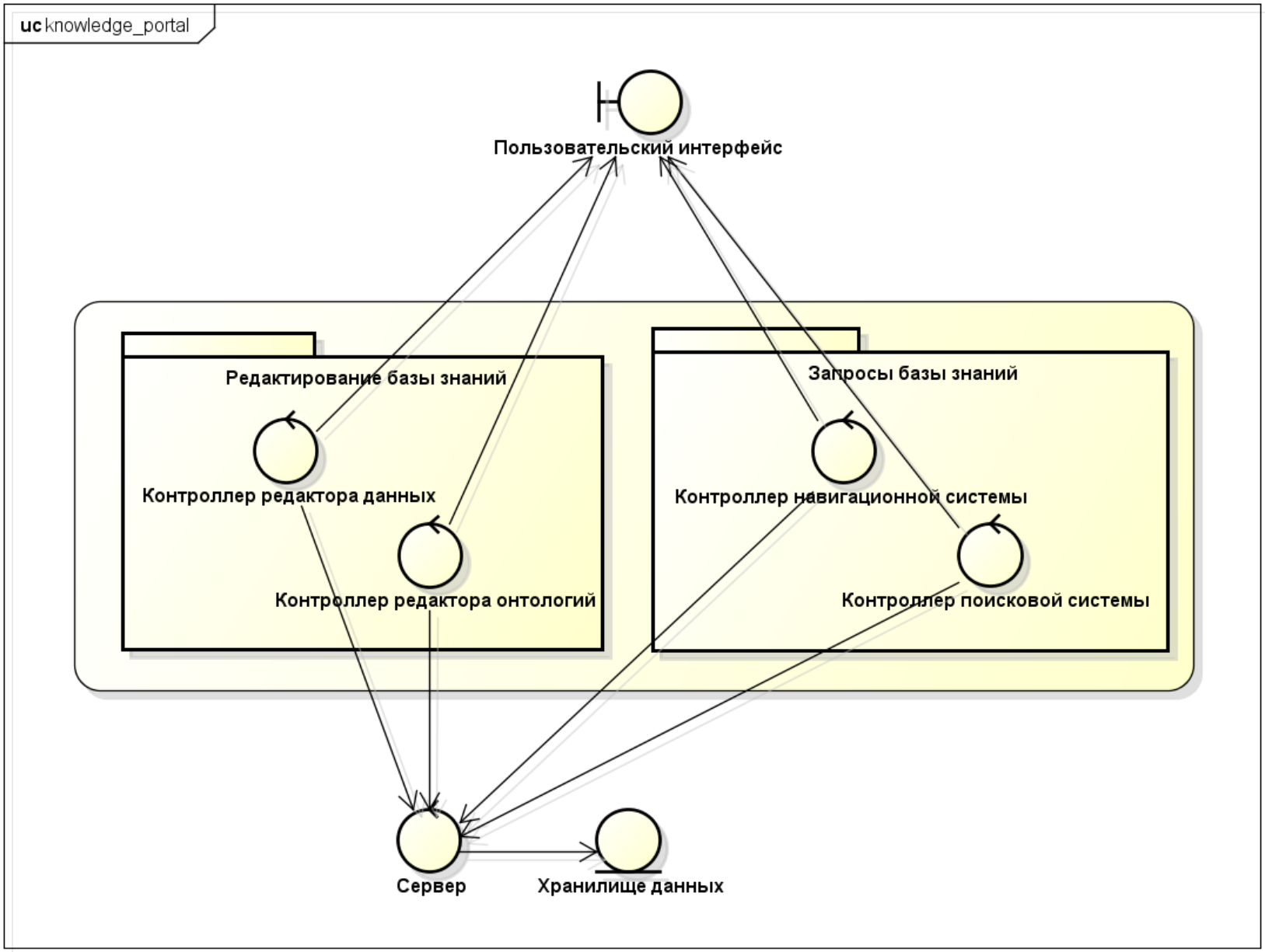
АРХИТЕКТУРА MVC



USE CASE ДИАГРАММА



АРХИТЕКТУРА ПОРТАЛА ЗНАНИЙ



ХРАНИЛИЩЕ ДАННЫХ JENA FUSEKI ВЕБ-СЕРВЕР

ТРЕБОВАНИЯ

- Клиент SPARQL HTTP
- Язык запросов: SPARQL Select
- Поддержка логического вывода
- Поддержка изменения данных: Files import, SPARQL Update

АНАЛОГИ

- BigOWLIM
- BigData
- Virtuoso

ПОДСИСТЕМЫ ВЗАИМОДЕЙСТВИЯ С ДАННЫМИ

SPARQL Select

- Подсистема навигации
- Подсистема поиска

SPARQL Update

- Подсистема редактора данных
- Подсистема редактора онтологии

ПРИМЕРЫ ШАБЛОНОВ НАВИГАЦИИ И ПОИСКА

ИЕРАРХИЯ КЛАССОВ ОДНОГО УРОВНЯ

```
PREFIX xsd:     <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:    <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl:     <http://www.w3.org/2002/07/owl#>
SELECT ?subClass
WHERE {
      ?subClass rdfs:subClassOf SOME_CLASS.
      FILTER NOT EXISTS {
            ?subClass rdfs:subClassOf ?X .
            ?X rdfs:subClassOf SOME_CLASS .
            FILTER (?X != ?subClass && ?X != SOME_CLASS)
      }
}
```

ОБЫЧНЫЙ ПОИСК

```
SELECT ?class ?subject ?dataProp ?value
      WHERE {
            ?subject
                  ?dataProp ?value FILTER REGEX(?value ,
                                    “STRING_SEARCH_VALUE”, "i" );
                  rdf:type ?class . }
```

ПРИМЕР РАСШИРЕННОГО ПОИСКА

```
SELECT *
      WHERE {
            ?subject
                  rdf:type kp:Публикации;
                  kp:Публикация_имеет_название ?name
                                    FILTER REGEX(?name, "информатика", "i")
                  kp:Автор_Публикации [ kp:Проживает_в kp:Новосибирск ];
                  kp:Научное_напрвление
                        kp:Моделирование_языка_и_языковой_деятельности . }
```

ПРИМЕРЫ ШАБЛОНОВ РЕДАКТОРА ДАННЫХ И ОНТОЛОГИИ

СОЗДАНИЕ ОБЪЕКТА

```
INSERT DATA {
      INDIVIDUAL_URI
            rdf:type SOME_CLASS;
       1    rdf:type owl:NamedIndividual;
            DATA_PROPERTY_1 VALUE_1;
       2    …
            DATA_PROPERTY_N VALUE_N;
            OBJECT_PROPERTY_1 OBJECT_URI_1;
       3    …
            OBJECT_PROPERTY_N OBJECT_URI_N;
}
```

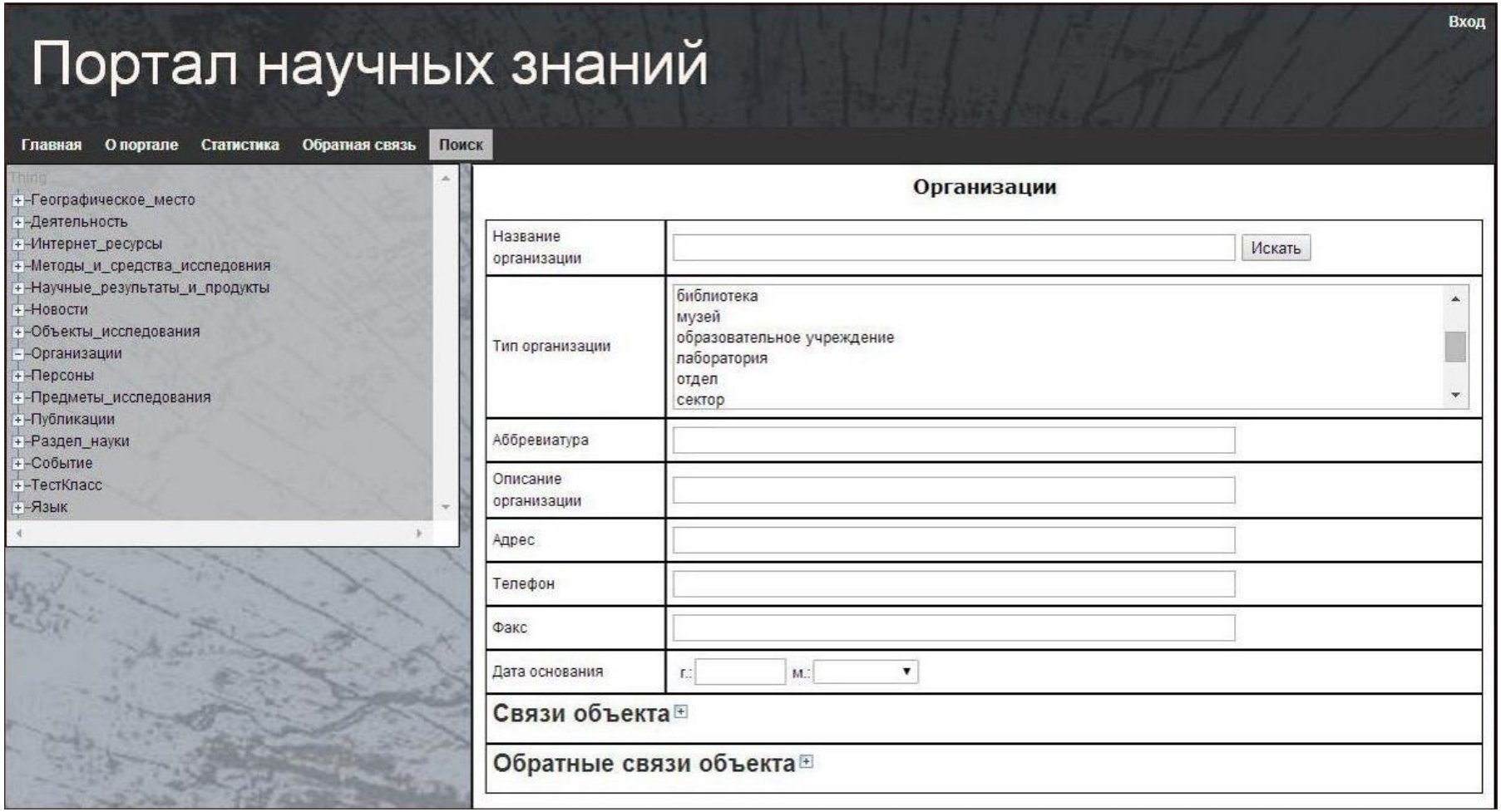
СОЗДАНИЕ КЛАССА

```
INSERT DATA {
      SOME_CLASS rdf:type rdfs:Class;
            rdfs:subClassOf owl:Thing;
            rdfs:subClassOf PARENT_CLASS_1;
            …
            rdfs:subClassOf PARENT_CLASS_N;
}
```

ТЕСТИРОВАНИЕ ПОРТАЛА

- Для проверки работоспособности портала знаний все подсистемы были протестированы на данных портала знаний по компьютерной лингвистике.
- Данные портала знаний по компьютерной лингвистике были представлены в формате xml, что не позволяло их непосредственно использовать серверу данных, в связи с этим был разработан программный модуль, который переводит базу знаний из xml-ориентированного формата в owl.

ПРИМЕР РАБОТЫ ПОРТАЛА



ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ

Исследованы средства технологии Semantic Web и возможность их применения к разработке порталов научных знаний.

При выполнении работы было разработано ядро портала научных знаний, а именно:

- Разработана архитектура портала научных знаний, базирующаяся на технологиях Semantic Web и шаблоне проектирования информационных систем MVC.
- Выбрано и настроено хранилище данных портала.

Спроектирована и реализована функциональность следующих подсистем портала:

- Подсистема навигации;
- Подсистема поиска;
- Подсистема редактора данных;
- Подсистема редактора онтологии.

Работоспособность портала была протестирована на данных портала знаний по компьютерной лингвистике. Для получения данных портала знаний по компьютерной лингвистике был разработан программный модуль, переводящий данные из xml-ориентированного формата в owl.

ЗАКЛЮЧЕНИЕ

Достоинства средств технологии Semantic Web:

- Удобство представления данных и знаний в едином формате, следование стандартам делает возможным использовать уже готовые средства вывода и обработки данных.
- Использование языка SPARQL для описания запросов дает возможность строить запросы проще, по сравнению с SQL.
- Возможность создания больших распределенных хранилищ.

Недостатки:

- Более бедные возможности для описания структур данных, чем в прототипном языке описания порталов знаний.
- Более низкая скорость обработки запросов по сравнению с реляционными СУБД.