МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего профессионального образования

**"Южно-Уральский государственный университет"**

**(национальный исследовательский университет)**

**Факультет Вычислительной математики и информатики**

**Кафедра системного программирования**

**Разработка веб-приложения для мониторинга позиций сайтов по запросам в поисковой системе Яндекс.**

КУРСОВАЯ РАБОТА

по дисциплине «Программная инженерия»

ЮУрГУ – 010300.62.2014.11-027-1909.КР

|  |  |
| --- | --- |
| Нормоконтролер,кандидат физико-математических наук, доцент каф. СП  \_\_\_\_\_\_\_\_\_\_\_\_\_\_Г.И. Радченко  “\_\_\_”\_\_\_\_\_\_\_\_\_\_\_ 2014 г. | Научный руководитель:  кандидат физико-математических наук, доцент каф. СП  \_\_\_\_\_\_\_\_\_\_Г.И. Радченко  Автор работы:  студент группы ВМИ-311  \_\_\_\_\_\_\_\_\_\_Е.А. Неповинных  Работа защищена  с оценкой: \_\_\_\_\_\_\_\_\_\_\_  “\_\_\_”\_\_\_\_\_\_\_\_\_\_\_ 2014 г. |

Челябинск 2014

Оглавление

[Оглавление 2](#_Toc387271468)

[Словарь 2](#_Toc387271469)

[Введение 4](#_Toc387271470)

[1. Обзор литературы 6](#_Toc387271471)

[2. Требования к системе мониторинга позиций сайтов 9](#_Toc387271472)

[*2.1.* *Система мониторинга позиций сайтов* 9](#_Toc387271473)

[*2.2.* *Варианты использования системы мониторинга позиций сайтов* 9](#_Toc387271474)

[*2.3.* *Нефункциональные требования к системе мониторинга позиций сайтов* 12](#_Toc387271475)

[*2.4.* *Вывод* 13](#_Toc387271476)

[3. Архитектура системы мониторинга позиций сайтов 14](#_Toc387271477)

[*3.1.* *Архитектура MVC* 14](#_Toc387271478)

[*3.2.* *Интерфейс веб-приложения* 18](#_Toc387271479)

# Словарь

*Пользователь* – конкретное лицо или представитель организации, использующий систему для того, чтобы выполнять конкретные задачи или функции.

*Организация* — группа людей, деятельность которых сознательно координируется для достижения общих целей.

*Ресурс* — веб-сайт, для которого осуществляется мониторинг позиций по заданному набору поисковых запросов.

*Аккаунт* — учётная запись, содержащая сведения об организации в системе.

*Мониторинг* —анализ позиции ресурса в результатах поиска.

*Позиция ресурса* в результатах поиска (далее: позиция) — порядковый номер ссылки на страницу определенного ресурса в результатах поиска по определенному поисковому запросу.

*Поисковый запрос (далее: запрос)* – строка для осуществления поиска с помощью поисковой системы.

*Регион мониторинга* – это параметр поискового запроса, осуществляющий территориальную специализацию результатов поиска.

*Результаты поиска* — упорядоченный список ссылок на различные ресурсы, которые пользователь видит в результате запроса к поисковой системе Яндекс.

# Введение

Продвижение интернет-сайта в поисковых системах позволяет значительно увеличить количество просмотров и качество активных пользователей сайта. На сегодняшний день продвижение сайтов в русско-язычном сегменте сети Интернет осуществляется в 5-ти наиболее популярных поисковых системах [4]: Яндекс, Google, поиск Mail.ru, Rambler, Bing. Существует целый ряд различных способов продвижения интернет-сайта в поисковых системах: генерация и размещение на сайте уникального и релевантного контента, продвижение сайта в каталогах и социальных сетях, использование определенных ключевых слов, покупка внешних ссылок на сайт и др. Для скорейшего достижения высокой позиции сайта и привлечения большего числа клиентов, необходимо постоянно отслеживать положение продвигаемого веб-ресурса в поисковых системах. Для решения этой задачи существуют системы мониторинга [5]. Система мониторинга отслеживает и позволяет сохранять для дальнейшего анализа позиции веб-сайта в поисковых системах за определённый период времени. В настоящий момент наиболее развиты системы мониторинга в виде веб-приложений. Их популярность объясняется, главным образом, тем, что клиенты не зависят от конкретной операционной системы пользователя, поэтому веб-приложения являются межплатформенными сервисами.

Актуальность данной темы обусловлена необходимостью постоянного анализа позиций коммерческих сайтов в поисковой выдаче для увеличения их посещаемости в процессе поисковой оптимизации (Search Engine Optimization, SEO) сайта. Оперативное определение неэффективных запросов, по которым позиции сайта долгое время остаются низкими, позволяет скорректировать стратегию и оптимизировать расходы продвижения сайта.

*Целью данной работы* является разработка веб-приложения для мониторинга позиций сайтов по запросам в поисковой системе Яндекс.

Для достижения указанной цели необходимо решить следующие задачи:

1. изучить особенности существующих систем мониторинга позиций сайтов в поисковых системах;
2. изучить особенности работы с платформами Django и Twitter Bootstrap;
3. изучить существующие подходы разработки систем мониторинга позиций сайтов в поисковых системах;
4. спроектировать архитектуру системы мониторинга позиций сайтов в поисковых системах;
5. реализовать и протестировать веб-приложение для мониторинга позиций сайтов по запросам в поисковой системе Яндекс.

*Первый раздел* посвящён теории. Мы рассмотрим основные плюсы и минусы ведения «электронного бизнеса», постараемся выделить основные современные технологии, позволяющие решить данную задачу, проведём их сравнительный анализ и выберем наиболее, на наш взгляд, подходящие.

Во *втором разделе* мы рассмотрим основные составляющие интернет-магазина с программной точки зрения. Сюда входят описания: Схемы базы данных, взаимодействия отдельных частей магазина, особенности реализации предложенных моделей, методов и алгоритмов.

*Третий раздел* содержит экспериментальную часть. Здесь мы попытаемся оценить адекватность и эффективность предложенных и реализованных ранее алгоритмов и методов. Проведём некоторое сравнение с уже существующими подобными системами.

В *заключении* мы сделаем вывод о проделанной работе и поставим основные цели развития.

# Обзор литературы

Для успешного продвижения и мониторинга позиций сайта необходимо изучить и понять алгоритмы работы поисковых систем, в которых планируется продвигать сайт. Поисковая система получает и обрабатывает запрос, сортирует все имеющиеся в своей базе данные о сайтах, к которым может иметь отношение данный запрос, то есть производит ранжирование, и выдает результат. На странице результатов поиска в определенном порядке располагаются [ссылки](http://www.webeffector.ru/wiki/Ссылка) на веб-ресурсы, которые, по мнению поисковой системы, наиболее соответствуют указанному запросу [1, 6].

Основные отличия моделей ранжирования поисковых систем Яндекс и Google заключаются в алгоритмах расчета характеристик веб-ресурса, определяющих рейтинг сайта на странице результатов поиска. Такие алгоритмы являются запатентованными разработками и держатся в строгом секрете. Анализ выдачи позволяет сделать некоторые предположения о том, на основе каких признаков производится ранжирование результатов поисковой выдачи. Благодаря независимости признаков ранжирования от смыслового наполнения сайта, в рамках одной поисковой системы можно использовать одну и ту же модель ранжирования для запросов из совсем разных областей знаний [2, 3].

Результаты исследований показывают, что более 70% пользователей сети Интернет находят интересующие их ресурсы, используя поисковые системы []. SEO оптимизация сайта направлена на повышение ранга сайта в поисковой выдаче, что позволяет увеличить посещаемость сайта. Мероприятия по продвижению сайта нужно проводить параллельно с мониторингом показателей – это дает возможность вовремя принимать важные решения и корректировать [стратегию продвижения](http://searchenginez.ru/strategii-prodvizheniya-sajtov/)[6].

Мониторинг позиций сайта дает возможность осуществлять контроль за рейтингом сайтов в поисковых системах по целому или частичному запросу и решает следующие задачи:

* контроль действий конкурентов;
* своевременная корректировка позиции сайта в соответствии с целевыми запросами;
* изменение текстового блока, находящегося на страницах сайта для улучшения позиции в поисковых машинах.

В настоящее время существует множество различных сервисов, которые позволяют проводить мониторинг позиций сайта в поисковых системах. Наиболее простым в использовании является бесплатный сервис «MegaIndex» ([http://www.megaindex.ru](http://www.megaindex.ru/)). Система «MegaIndex» позволяет создать проекты под каждый из своих Интернет-ресурсов и добавить поисковые запросы, по которым будет производиться мониторинг. На странице запросов можно увидеть, на каких позициях в поисковых системах России, Украины и Белоруссии находится продвигаемый ресурс, количество запросов в месяц. Недостатком данного сервиса является то, что его основная функция – платное размещение ссылок на сторонних ресурсах в целях увеличения показателей продвигаемых ресурсов либо повышения их посещаемости, поэтому отслеживать позиции сайтов там неудобно.

Множество функциональных возможностей предоставляет сервис «AllPositions» ([http://allpositions.ru](http://allpositions.ru/)), включая:

* контроль позиций сайтов в выдаче поисковых систем;
* анализ информации о текущих позициях конкурирующих ресурсов;
* получение отчетов о результатах продвижения в удобном формате, которые максимально подробно отобразят все запрашиваемые данные.

Однако данный сервис платный, и при первом использовании довольно трудно разобраться в настройках [7].

После изучения существующих решений было принято решение реализовать систему мониторинга позиций сайтов в поисковой системе Яндекс в виде отдельного django-приложения с использованием средств Twitter Bootstrap. Django [8] — это высокоуровневая веб-платформа для языка программирования Python, поощряющий быструю разработку и чистый, прагматичный дизайн. Twitter Bootstrap – это HTML, CSS и Javascript платформа для Web-разработки [9].

# Требования к системе мониторинга позиций сайтов

## *Система мониторинга позиций сайтов*

Система мониторинга позиций сайтов – это система, позволяющая отслеживать рейтинг исследуемых сайтов в поисковых системах по выбранным поисковым запросам. Под рейтингом понимается позиция ресурса на странице результатов поиска.

Разрабатываемая система мониторинга позиций сайтов не использует специальных программных интерфейсов для доступа к результатам поиска. Она имитирует действия посетителей поисковых ресурсов и работает только с той информацией, которая представлена на html-страницах.

Приложение работает с наиболее популярной в русскоязычном сегменте сети Интернет поисковой системой Яндекс.

## *Варианты использования системы мониторинга позиций сайтов*

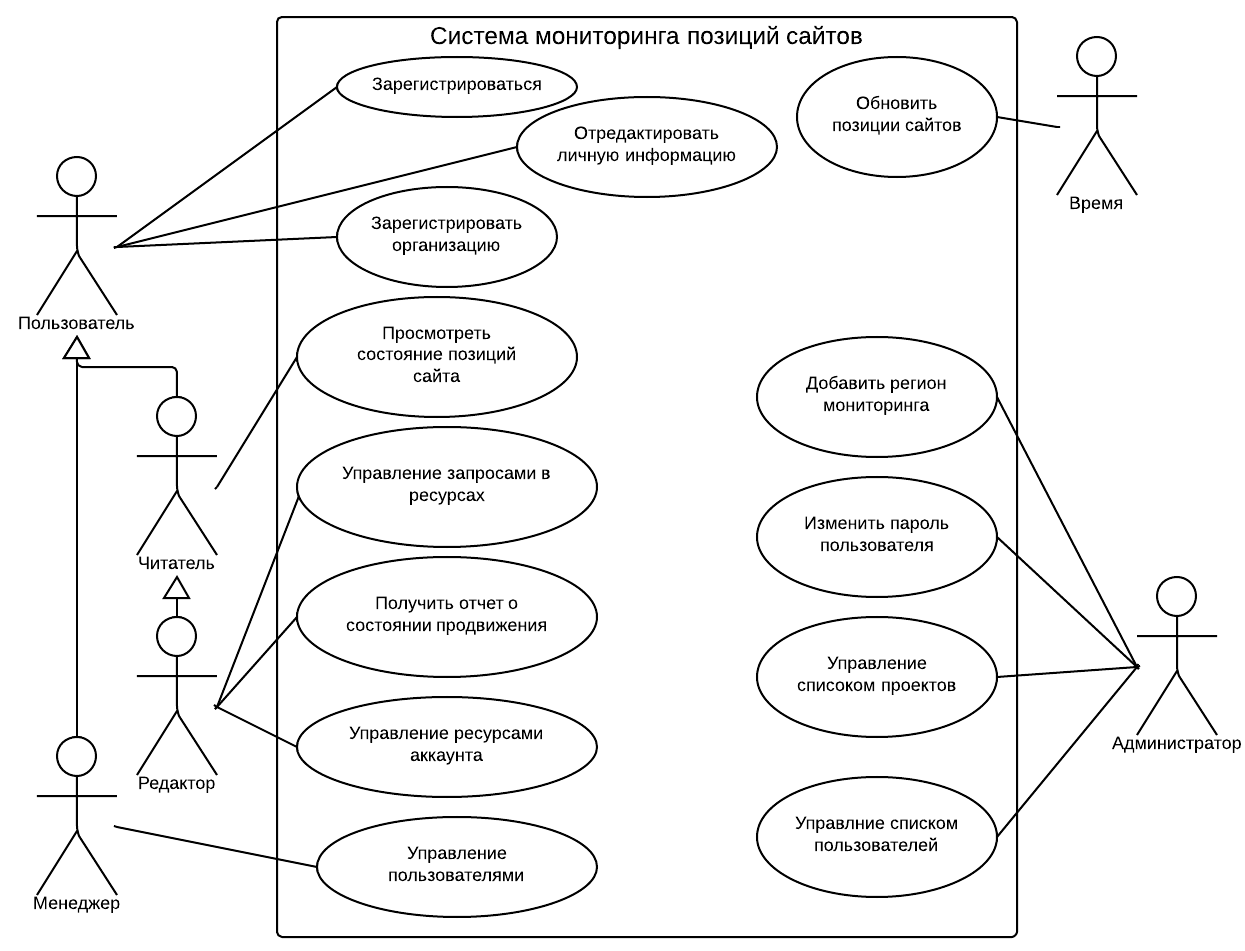
Можно выделить следующих актеров, взаимодействующих с системой мониторинга позиций сайтов (см. рис. 1):

* *администратор* – актер, отвечающий за поддержку работоспособности системы;
* *пользователь* – актер, использующий систему;
* *читатель* – пользователь, просматривающий результаты мониторинга;
* *редактор* – читатель, управляющий параметрами мониторинга;
* *менеджер* – пользователь, управляющий доступом других пользователей к ресурсам;
* *время* – актер, обновляющий состояние системы.

Определим основные варианты использования данными актерами нашей системы.

Администратор инициирует исполнение следующих прецедентов:

1. *управление списком пользователей*: зайдя на страницу редактирования пользователей, администратор может просмотреть список пользователей, добавить нового пользователя или удалить существующего;
2. *управление списком ресурсов*: зайдя на страницу редактирования ресурсов, администратор может просмотреть список ресурсов, добавить новый ресурс или удалить существующий;
3. *изменение пароля пользователя*: зайдя на страницу редактирования пользователя, администратор может задать новый пароль пользователя;
4. *добавление региона мониторинга*: зайдя на страницу редактирования регионов, администратор может добавить новый регион мониторинга.



**Рис. 1.** Диаграмма вариантов использования системы мониторинга позиций сайтов

Пользователь инициирует исполнение следующих прецедентов:

1. *отредактировать личную информацию*: пользователь заходит на форму редактирования личной информации, где он может указать следующую информацию о себе:
   * Ф.И.О.
   * E-mail;
2. *зарегистрироваться:* пользователь выбирает соответствующий пункт при входе на сайт, после чего заполняет форму регистрации для физического лица;
3. *зарегистрировать организацию:* пользователь выбирает соответствующий пункт при входе на сайт, после чего заполняет форму регистрации для организации. После регистрации для организации создается *аккаунт* организации в системе мониторинга.

Менеджер инициирует исполнение следующих прецедентов:

1. *управление пользователями организации*: менеджер может предоставить пользователю следующие права для всех или определенного ресурса организации:
   * просмотреть состояние позиций ресурса;
   * редактирование ресурса;
   * управление правами пользователей на доступ к ресурсу.

Читатель инициирует исполнение следующих прецедентов:

1. *просмотреть состояние позиций ресурса*: пользователь заходит на страницу ресурса и просматривает таблицу позиций сайта по заданным запросам в выбранном регионе и диапазоне дат.

Редактор инициирует исполнение следующих прецедентов:

1. *управление запросами в ресурсах*: нажав на кнопку «Редактировать» на странице ресурса, пользователь может добавить, удалить или изменить один или несколько запросов;
2. *получение отчета о состоянии продвижения*: нажав на кнопку «Экспорт» на странице ресурса, пользователь получает себе на компьютер отчет, в котором содержится таблица позиций сайта по выбранным запросам за заданный период времени в заданном регионе;
3. *управление ресурсами аккаунта:* на странице аккаунта организации пользователь может просмотреть список ресурсов и добавить новый ресурс.

Время инициирует исполнение следующих прецедентов:

1. *обновить позиции сайтов:* каждый день в заданное время запускается сборщик позиций сайтов; собранные позиции записываются в базу данных.

## *Нефункциональные требования к системе мониторинга позиций сайтов*

В результате интервью с заказчиком были сформулированы следующие нефункциональные требования.

Поскольку приложение ориентировано на клиентов, которые могут находится в географически удалённых точках , интерфейс системы мониторинга позиций сайтов в поисковой системе Яндекс должен быть доступен из сети Интернет. При этом все страницы веб-интерфейса должны корректно отображаться в следующих браузерах: Google Chrome 33 и выше, включая мобильную версию; Mozilla Firefox 29.0 и выше.

Для повышения удобства использования был выполнен анализ сценариев работы пользователей. Необходимо учесть результаты анализа для максимально удобного расположения основных элементов интерфейса.

Веб-интерфейс к системе мониторинга позиций сайтов в поисковой системе Яндекс должен быть выполнен с использованием платформы Django, поскольку с ее использованием реализовано большинство проектов заказчика.

## *Вывод*

Таким образом, система мониторинга позиций сайтов в поисковой системе Яндекс должна быть ориентирована на удобство использования клиентами и легко доступна через сеть интернет на различных устройствах. Также система должна интегрироваться в имеющуюся инфраструктуру заказчика.

# Архитектура системы мониторинга позиций сайтов

## *Архитектура MVC*

Веб-приложение — представляет собой вариацию модели MVC (Model-View-Controller). Шаблон проектирования MVC предполагает разделение данных приложения, пользовательского интерфейса и управляющей логики на три отдельных компонента: Модель, Представление и Контроллер – таким образом, что модификация каждого компонента может осуществляться независимо.

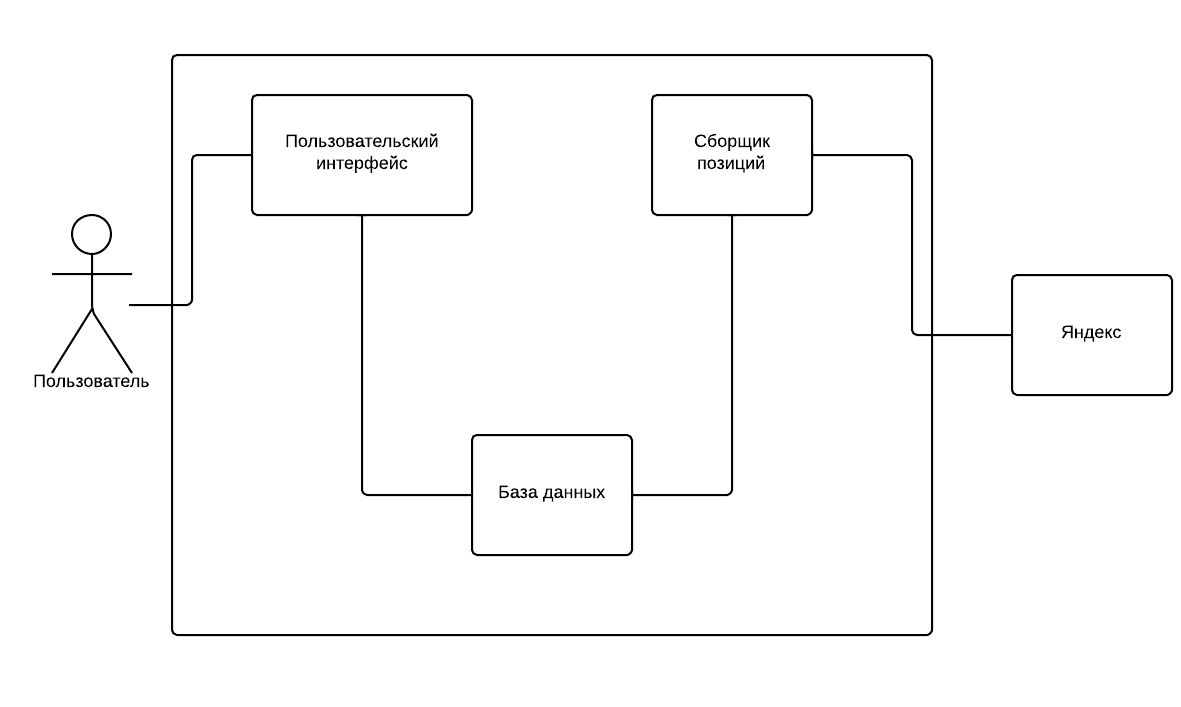
* Модель. Модель предоставляет знания: данные и методы работы с этими данными, реагирует на запросы, изменяя своё состояние. Не содержит информации, как эти знания можно визуализировать.
* Представление, вид. Отвечает за отображение информации (визуализацию). Часто в качестве представления выступает форма с графическими элементами.
* Контроллер. Обеспечивает связь между пользователем и системой: контролирует ввод данных пользователем и использует модель и представление для реализации необходимой реакции.

Представление и контроллер зависят от модели. Но модель не зависит ни от представления, ни от контроллера. Таким образом достигается назначение такого разделения: оно позволяет строить модель независимо от визуального представления, а также создавать несколько различных представлений для одной модели.

В Django модели представляют собой классы, отвечающие за ORM и взаимодействие с базой данных, что представляет собой бизнесс-логику системы.

Представления в django —это функции, которые принимают объект класса Request, представляющего собой упрощенное представление http-запроса, и возвращающие объект класса Response, представляющего собой возвращаемый клиенту ответ сервера. Представления могут использовать шаблоны при получении ответа.

Контроллером является веб-платформа django и её механизм распределения путей по представлениям (urls).



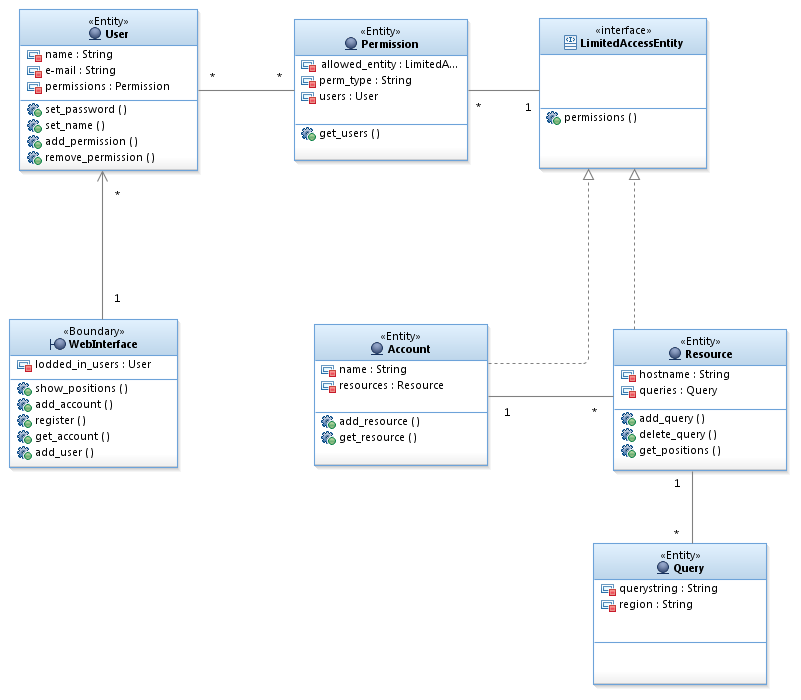
**Рис. 2.** Диаграмма архитектуры системы мониторинга позиций сайтов в поисковой системе Яндекс.

В результате анализа требований к системе было решено выделить следующие классы-сущности (рис. 2):

* User – класс, соответствующий сущности пользователя;
* Account – класс, соответствующий сущности организации;
* Resource – класс, соответствующий сущности ресурса;
* Permission – класс объектов, которые представляют из себя право пользователя на совершение определенных действий с другой конкретной сущностью.

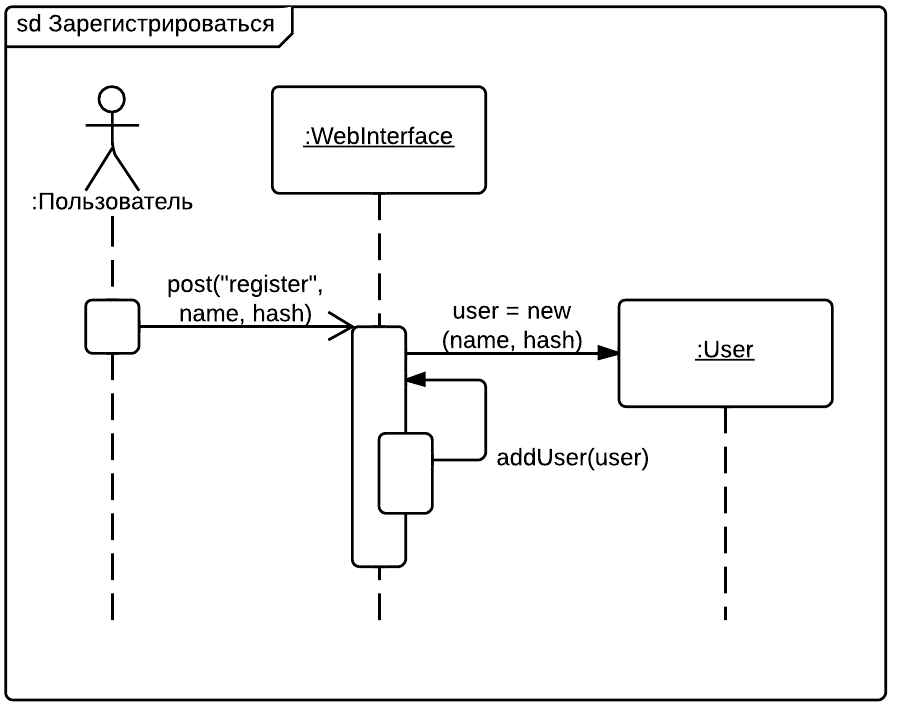
Для обеспечения взаимодействия системы с внешней средой были выделены следующие классы-интерфейсы:

* LimitedAccessEntity – общий интерфейс, представляющий собой сущность, права доступа к которой можно ограничивать с помощью Permission;
* WebInterface – класс, обеспечивающий взаимодействие системы с пользователем.

****

**Рис. 3** Диаграмма классов анализа

На диаграмме последовательностей (рис. 4) изображены экземпляры объектов и сообщения, которыми они обмениваются в рамках прецедента «Зарегистрироваться». Пользователь отправляет сообщение о регистрации в веб-интерфейс, после чего создаётся новый объект «пользователь» и добавляется в хранилище, становясь таким образом доступным через веб-интерфейс.

****

**Рис.4** Регистрация

На рис. 5 изображена диаграмма последовательности прецедента «Просмотреть позиции». Пользователь, имеющий право на чтение ресурса, делает запрос на просмотр позиций этого ресурса в поисковой системе Яндекс с помощью веб-интерфейса, указав интересующий промежуток времени. После этого система возвращает требуемый объект типа «Ресурс», и таблицу позиций за указанный промежуток времени. Веб-интерфейс отображает пользователю желаемые данные.

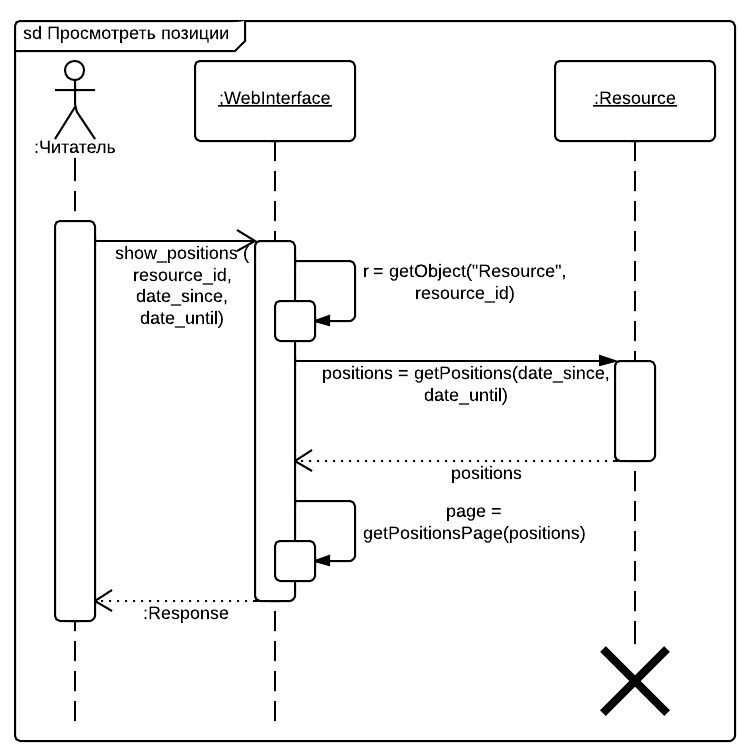


Рис. 5 Просмотр позиций

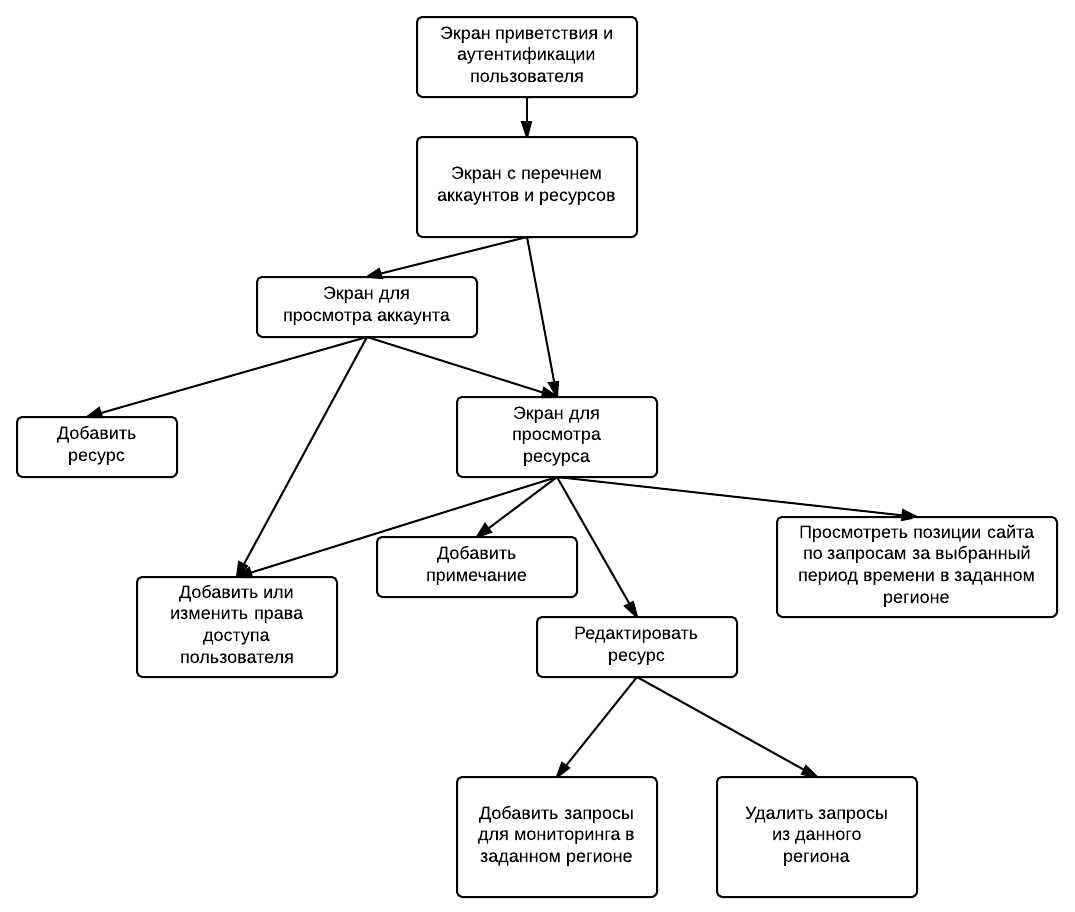
## *Интерфейс веб-приложения*

Интерфейс веб-приложения предоставляет отдельные методы работы с системой в зависимости от того, какими правами обладает пользователь, вошедший в систему:

* интерфейс пользователя;
  + интерфейс читателя
    - интерфейс редактора
  + интерфейс менеджера
* интерфейс администратора;

Один пользователь может получить доступ сразу к нескольким интерфейсам, если соответствующие роли прописаны в его пользовательском профиле.

Диаграмма доступных пользовательских интерфейсов представлена на Рис. 6.



**Рис. 6.** Диаграмма доступных пользовательских интерфейсов системы мониторинга позиций сайтов

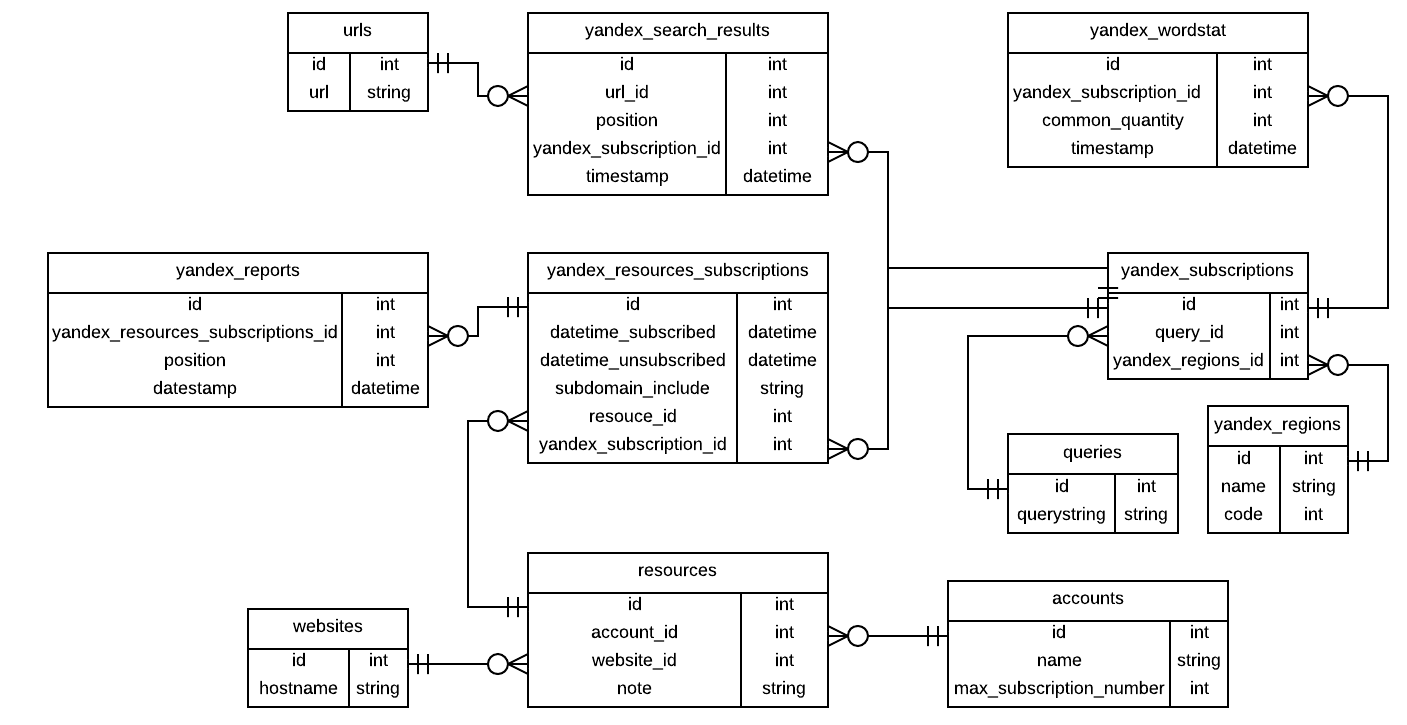
# Реализация системы мониторинга позиций сайтов

## *Схема базы данных*

Для реализации веб-приложения для работы с системой мониторинга позиций сайтов была использована база данных, состоящая из следующих таблиц:

* таблица аккаунтов, состоящая из данных обо всех организациях, использующих систему;
* таблица ресурсов, содержащая информацию обо всех ресурсах, используемых системой;
* таблица для связи информации о запросе с конкретным регионом;
* таблица, содержащая данные для формирования отчета о состоянии продвижения;
* таблица, содержащая имена хостов веб-сайтов, которые использует система для мониторинга позиций;
* таблица, содержащая строки запросов, по которым производится мониторинг;
* таблица, содержащая коды регионов, по которым производится мониторинг;
* таблица, связывающая поисковую строку и код региона, по которым производится мониторинг;
* таблица, в которую записывается информации о частотности запросов по регионам;
* таблица, содержащая информацию о позиции запроса для каждого исследуемого региона в поисковой системе Яндекс;
* таблица, содержащая пути доступа к информации о результатах мониторинга.

Схема используемой базы данных изображена на Рис. 7.



**Рис.7.** Схема базы данных

## *Организация безопасности приложения*

Безопасность приложения осуществляется на основе создания пользователей и ролей. Ролям соответствуют логические роли внутри приложения. Логические роли описывают правила, согласно которым разрешается доступ к различным частям приложения. Каждому пользователю назначается одна или несколько логических ролей.

Подключение к приложению происходит в два этапа:

1. Аутентификация. Система безопасности проверяет права доступа пользователя на основании его имени и пароля.

2. Авторизация. После прохождения этапа авторизации система безопасности передает приложению информацию о пользователе в виде списка предоставленных ему ролей.

В зависимости от имени подключившегося пользователя и его прав можно скрыть или сделать недоступными определенные части приложения.

Для того, чтобы иметь доступ к системе, каждый пользователь должен быть зарегистрирован администратором с помощью интерфейса администрирования. После того, как созданы пользователи, нужно создать роли и распределить по ним пользователей (Рис. 12). В веб-приложении для организации работы системы мониторинга позиций сайтов создано 3 роли:

1. reader – роль читателя приложения;
2. editor – роль редактора приложения;
3. manager – роль управляющего доступом пользователей к приложению;

Для каждой сущности роли пользователям присваиваются отдельно. Так, если пользователю дано право читать аккаунт, то ему автоматически присваивается право чтения для всех ресурсов, принадлежащих аккаунту. Интерфейс управления доступом пользователей представлен на рис.8.

**Рис. 8.** Интерфейс управления доступом пользователей

Модель безопасности приложения разделяет все страницы на несколько подмножеств в зависимости от того, какие группы пользователей могут иметь к ним доступ.

1. *Не подключившиеся пользователи.* Этим пользователям доступна только начальная страница.
2. *Читатели.* Этим пользователям доступны начальная страница в состоянии после входа в систему и страницы просмотра ресурсов аккаунта в случае предоставления права всему аккаунту. Иначе – только список доступных ему ресурсов. На странице ресурса пользователи могут просматривать позиции в любом регионе за любой выбранный период времени, в котором производился мониторинг.
3. *Редакторы.* Этим пользователям доступен весь интерфейс, к которому имеют доступ читатели. При предоставлении права редактирования аккаунта, пользователю, обладающему ролью редактора, предоставляется интерфейс добавления ресурсов для мониторинга у этого аккаунта. При предоставлении права редактирования ресурса пользователю, обладающему ролью редактора, предоставляется интерфейс добавления, удаления запросов из данного ресурса.
4. *Менеджеры.* Этим пользователям доступны начальная страница в состоянии после входа в систему и страницы аккаунта и ресурса, на которых ему отображены только таблицы управления доступом пользователей к данному объекту (рис. 8).
5. *Администраторы.* Этим пользователям доступны все страницы приложения, включая интерфейс администрирования.

Задача ограничения доступа к страницам задача решается средствами авторизации, предоставляемыми платформой Django.

**Литература**

1. Bar-Ilan J., Mat-Hassan M., Levene M. Methods for comparing rankings of search engine results // Comput. Networks. 2006. Vol. 50, № 10. P. 1448–1463.

2. В.В. Зосимов, В.С. Степашко А.С.Б. Построение и сравнительный анализ моделей ранжирования результатов работы поисковых систем google и яндекс. 2011. P. 69–77.

3. Сегалович И., Маслов М. Некоторые аспекты полнотекстового поиска и ранжирования в Яндекс. 2004.

4. Статистика поисковых систем . URL: http://www.liveinternet.ru/stat/ru/searches.html (28.01.2014).

5. Определение позиций сайта в поисковиках Яндекс и Гугл . URL: http://www.semonitor.ru/positioning.html (28.01.2014).

6. Как работают поисковые системы - SearchEngines.ru . URL: http://www.searchengines.ru/articles/004556.html (28.01.2014).

7. SEO анализ . URL: http://www.seobuilding.ru/ (28.01.2014).

8. Django . URL: https://www.djangoproject.com/ (28.01.2014).

9. Bootstrap . URL: http://getbootstrap.com/ (28.01.2014).