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**

|  |  |
| --- | --- |
| РАБОТА ПРОВЕРЕНА  Рецензент  генеральный директор  ООО «АппБит Софтвейр»  \_\_\_\_\_\_\_\_\_\_ Д.А. Диков  “\_\_\_”\_\_\_\_\_\_\_\_\_\_\_ 2015 г. | ДОПУСТИТЬ К ЗАЩИТЕ  Заведующий кафедрой,  д.ф.-м.н., профессор  \_\_\_\_\_\_\_\_\_\_ Л.Б. Соколинский  “\_\_\_”\_\_\_\_\_\_\_\_\_\_\_ 2015 г. |

**РАЗРАБОТКА ИнтеллектуальнОЙ системЫ монИторинга и продвижения веб-сайтов в поисковых системах**

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

ЮУрГУ – 010400.62.2015.11-027-1909.ВКР

|  |  |
| --- | --- |
|  | Научный руководитель  к.ф.-м.н., доцент  \_\_\_\_\_\_\_\_\_\_Г.И. Радченко  Автор работы, студент группы ВМИ-411  \_\_\_\_\_\_\_\_\_\_Е.А. Неповинных  Ученый секретарь  (нормоконтролер)  \_\_\_\_\_\_\_\_\_\_\_\_\_ О.Н. Иванова  “\_\_\_”\_\_\_\_\_\_\_\_\_\_\_ 2015 г. |

Челябинск-2015

**Оглавление**

[Глоссарий 5](#_Toc421288889)

[Введение 6](#_Toc421288890)

[1. Обзор литературы 9](#_Toc421288892)

[1.1. Обзор алгоритмов 10](#_Toc421288893)

[1.2. Анализ существующих решений 12](#_Toc421288894)

[1.3. Вывод 13](#_Toc421288895)

[2. Требования к системе мониторинга и продвижения веб-сайтов 14](#_Toc421288896)

[2.1. Интеллектуальная система мониторинга и продвижения веб-сайтов 14](#_Toc421288897)

[2.2. Варианты использования системы 15](#_Toc421288898)

[2.3. Нефункциональные требования к системе 17](#_Toc421288899)

[2.4. Вывод 17](#_Toc421288900)

[3. Архитектура системы мониторинга и продвижения веб-сайтов 18](#_Toc421288901)

[3.1. Компоненты системы мониторинга 18](#_Toc421288902)

[3.2. Схема базы данных 20](#_Toc421288903)

[3.3. Применение алгоритма кластеризации для запросов 21](#_Toc421288904)

[3.4. Вывод 25](#_Toc421288905)

[4. Реализация системы 27](#_Toc421288906)

[4.1. База данных 27](#_Toc421288907)

[4.2. Реализация компонентов системы 29](#_Toc421288908)

[4.3. Организация взаимодействия компонентов системы 31](#_Toc421288909)

[4.4. Веб-интерфейс 31](#_Toc421288910)

[4.5. Вывод 36](#_Toc421288911)

[5. Тестирование 37](#_Toc421288912)

[Заключение 42](#_Toc421288913)

[Литература 43](#_Toc421288914)

[Приложения 45](#_Toc421288915)

[Приложение 1 45](#_Toc421288916)

[Приложение 2 48](#_Toc421288917)

# Глоссарий

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

*Сбор позиций* – проверка позиций сайтов на странице результатов поиска в поисковой системе по заранее заданным параметрам.

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

*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.

*Продвижение интернет-сайта* – это комплекс мер по обеспечению [посещаемости сайта](https://ru.wikipedia.org/w/index.php?title=%D0%9F%D0%BE%D1%81%D0%B5%D1%89%D0%B0%D0%B5%D0%BC%D0%BE%D1%81%D1%82%D1%8C_%D1%81%D0%B0%D0%B9%D1%82%D0%B0&action=edit&redlink=1) потенциальными потребителями, которые заинтересованы в приобретении товаров или услуг, представленных на продвигаемом сайте.

*Анализ конкурентов* –метод определения компаний, работающих на рынке и предлагающих сходную продукцию, задачей которого является выявление сильных и слабых мест компании для последующей разработки правильной стратегии поведения компании на рынке и диверсификации рисков.

*Семантическое ядро сайта* – это набор поисковых слов, их морфологических форм и словосочетаний, которые наиболее точно характеризуют вид деятельности, товар или услугу, предлагаемые сайтом.

# 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 пяти наиболее популярных поисковых системах [14]: 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. [7]. 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 [16]. 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, а данные о состоянии конкурентов помогают адекватно оценить перспективы продвигаемого проекта.

*Целью данной работы* является разработка интеллектуальной системы мониторинга и продвижения сайтов по запросам в поисковых системах.

Для достижения указанной цели необходимо решить следующие задачи:

1. изучение и анализ особенностей существующих систем мониторинга позиций сайтов в поисковых системах;
2. изучение особенностей работы с платформами Django и Twitter Bootstrap;
3. изучение существующих алгоритмов интеллектуального анализа данных;
4. разработка архитектуры интеллектуальной системы мониторинга и продвижения веб-сайтов в поисковых системах;
5. разработка пользовательского интерфейса системы мониторинга и продвижения позиций сайтов по запросам в поисковых системах;
6. реализация интеллектуальной системы мониторинга и продвижения позиций сайтов по запросам в поисковых системах;
7. тестирование интеллектуальной системы мониторинга и продвижения веб-сайтов в поисковых системах на реальных данных.

## Структура и объем работы

Работа состоит из введения, пяти разделов, заключения, библиографии и приложения. Объем работы составляет 47 страниц, объем библиографии – 19 источников.

В первой главе дается обзор технологий работы поисковых систем и рассматриваются основные алгоритмы интеллектуального анализа данных. В этом разделе рассмотрены алгоритмы ранжирования наиболее популярных поисковых систем и выявлены наиболее подходящие алгоритмы кластеризации данных для решения задачи интеллектуального анализа данных.

Вторая глава содержит описание и анализ требований к интеллектуальной системе мониторинга и продвижения сайтов по запросам в поисковых системах.

В третьей главе представлена архитектура интеллектуальной системы мониторинга и продвижения сайтов по запросам в поисковых системах. В этом разделе подробно рассмотрена общая архитектура системы, интерфейсы компонентов и детали их реализации.

Четвертая глава посвящена реализации интеллектуальной системы мониторинга и продвижения сайтов по запросам в поисковых системах с использованием алгоритма четкой кластеризации k-means и технологий django, AJAX и MVC.

В пятой главе приводятся результаты тестирования интеллектуальной системы мониторинга и продвижения сайтов по запросам в поисковых системах.

В заключении сделаны выводы о проделанной работе.

Приложение 1 содержит детальную реализацию основных прецедентов.

Приложение 2 содержит перечень проектов, на которых была протестирована система.

# Обзор литературы

Для успешного продвижения сайта и определения его конкурентов необходимо 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, т.е. 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, 17].

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 [7, 18].

Анализ конкурентов способствует достижению целого спектра маркетинговых и производственных целей организации. На его основе можно спрогнозировать исход конкурентной борьбы или выбрать оптимальную модель продвижения и позиционирования продукции компании.

Качественный анализ ближайших конкурентов позволяет смоделировать все многообразие вариантов развития событий и своевременно вносить необходимые изменения в деятельность компании [9].

Сравнительный анализ конкурентов по временным и другим параметрам позволяет выявить динамику изменений маркетинговой стратегии, провести оценку наиболее эффективных ходов лидера рынка, обеспечивающих наилучшие результаты. Внедрение этих знаний в практику работы компании позволит минимизировать ошибки в планировании и реализации собственной маркетинговой стратегии, а также добиться увеличения доходов и доли рынка [6, 9].

Наиболее эффективным методом анализа текущего состояния конкурентов компании является интеллектуальный анализ данных. Интеллектуальный анализ данных представляет собой процесс обнаружения пригодных к использованию сведений в крупных наборах данных [4]. В интеллектуальном анализе данных применяется математический анализ для выявления закономерностей и тенденций, существующих в данных. Обычно такие закономерности нельзя обнаружить при традиционном просмотре данных, поскольку связи слишком сложны или из-за чрезмерного объема данных. Эти закономерности и тренды можно собрать вместе и определить как модель интеллектуального анализа данных [2]. Модели интеллектуального анализа данных могут применяться к конкретным сценариям, а именно:

* прогноз: оценка продаж, прогнозирование нагрузки сервера или времени простоя сервера;
* риски и вероятности: выбор наиболее подходящих заказчиков для целевой рассылки, определение точки равновесия для рискованных сценариев, назначение вероятностей диагнозам или другим результатам;
* рекомендации: определение продуктов, которые с высокой долей вероятности могут быть проданы вместе, создание рекомендаций;
* определение последовательностей: анализ выбора заказчиков во время совершения покупок, прогнозирование следующего возможного события;
* группирование: разделение заказчиков или событий на кластеры связанных элементов, анализ и прогнозирование общих черт [9, 19].

## Обзор алгоритмов

Алгоритм интеллектуального анализа данных − это набор эвристики и вычислений, который создает модель интеллектуального анализа данных из данных. Чтобы создать модель, алгоритм сначала анализирует предоставленные данные, осуществляя поиск определенных закономерностей и тенденций. Алгоритм использует результаты этого анализа для выбора оптимальных параметров создания модели интеллектуального анализа данных. Затем эти параметры применяются ко всему набору данных, чтобы выявить пригодные к использованию закономерности и получить подробную статистику [6, 7].

Кластеризация представляет собой объединение объектов в группы (кластеры) на основе схожести признаков для объектов одной группы и отличий между группами. Большинство алгоритмов кластеризации не опираются на традиционные для статистических методов допущения; они могут использоваться в условиях почти полного отсутствия информации о законах распределения данных. Кластеризацию проводят для объектов с количественными (числовыми), качественными или смешанными признаками. В данной работе рассматривается кластеризация только для объектов с количественными признаками. Исходной информацией для кластеризации является матрица (Х), каждая строчка которой представляет собой значения n признаков одного из Х объектов кластеризации.



Задача кластеризации состоит в разбиении объектов из X на несколько подмножеств (кластеров), в которых объекты более схожи между собой, чем с объектами из других кластеров [5, 13].

Методы кластеризации можно классифицировать на четкие и нечеткие. Четкие методы кластеризации разбивают исходное множество объектов X на несколько непересекающихся подмножеств. При этом любой объект из X принадлежит только одному кластеру. Нечеткие методы кластеризации позволяют одному и тому же объекту принадлежать одновременно нескольким (или даже всем) кластерам, но с различной степенью. Нечеткая кластеризация во многих ситуациях более «естественна», чем четкая, например, для объектов, расположенных на границе кластеров [3, 5, 15].

Методы кластеризации также классифицируются по тому, определено ли количество кластеров заранее или нет. В последнем случае количество кластеров определяется в ходе выполнения алгоритма на основе распределения исходных данных [3, 5].

В данной работе для решения задачи разбиения запросов используется алгоритм четкой кластеризации k-средних [3, 6].

## Анализ существующих решений

В настоящее время существует множество различных сервисов, которые занимаются интеллектуальным анализом больших объемов данных. Наиболее схожим с разрабатываемой системой является платный сервис SpyWords.

Концепция SpyWords – это функциональная платформа по анализу запросов конкурентов в контексте и поиске. Сервис позволяет специалистам по интернет-маркетингу и seo отслеживать полезную и уникальную информацию о сайте конкурентов: ключевые слова в контекстной рекламе и органическом поиске, тексты объявлений, позиции в контексте и органике, дневной бюджет и так далее. Функционал сервиса полезен как на стадии начального формирования семантического ядра для контекста/продвижения, так и на этапе оптимизации действующей компании (поиска новых эффективных слов/анализа активности ключевых конкурентов). Недостатком данного сервиса является то, что в нем отсутствует разделение запросов на регионы и пользователь получает отчеты в неудобном для себя виде [12].

Сервис Quicksprout проводит анализ веб-сайтов на предмет того, как они продвигаются в популярных социальных сетях и насколько активно посетители делятся их контентом. На основе этих данных можно узнать, какие страницы конкурентов пользуются наибольшей популярностью в социальных сетях, какая социальная сеть наиболее популярна в каждой тематике, и делать акцент на продвижение именно в ней. На сервисе можно анализировать свой сайт в сравнении с тремя сайтами конкурентов, причем как главные страницы, так и внутренние. Недостатком сервиса является то, что он англоязычный и, как следствие, не поддерживает работу с популярными в России социальными сетями, такими как «Вконтакте» и «Одноклассники» [11].

Также полезным при продвижении сайтов является сервис Ahrefs. С помощью него можно узнать, какие позиции и по каким поисковым запросам занимает та или иная внутренняя страница любого сайта, причем будет отображаться изменение позиций и дата последнего изменения. Ahrefs имеет свою собственную базу данных о сайтах и внешних ссылках , которая ежедневно пополняется. Список появившихся и удаленных внешних ссылок, данные можно просматривать по дням, а также получать ежедневную статистику в удобной для пользователя форме. К недостаткам можно отнести отсутствие русскоязычного интерфейса [8].

## Вывод

Проведенный обзор существующих систем и технологий мониторинга и продвижения веб-сайтов в поисковых системах показал, что разработка системы, осуществляющей мониторинг и продвижение веб-сайтов в поисковых системах, является актуальной задачей. Проведенный обзор алгоритмов кластеризации данных позволил выявить наиболее подходящий для решения поставленной задачи.

# Требования к системе мониторинга и продвижения веб-сайтов

## Интеллектуальная система мониторинга и продвижения веб-сайтов

Интеллектуальная система мониторинга и продвижения веб-сайтов в поисковых системах является расширением системы мониторинга позиций сайтов в поисковых системах и направлена на ускорение и оптимизацию процесса продвижения веб-ресурса, а также анализа его состояния на фоне конкурентов.

Для корректной работы системы необходимы данные о сборе позиций интересующего веб-ресурса за любой период времени (от одного дня) и список запросов, по которым эта информация собиралась. Запросы могут быть добавлены в систему как самим пользователем, так и специалистом по продвижению сайтов.

Сборщик позиций ежедневно обрабатывает данные наиболее популярных поисковых систем, таких как Яндекс и Google, до глубины 50, 200 или 500 позиций, после чего система анализирует полученную информацию и предоставляет ее в удобном для пользователя виде.

Благодаря системе анализа данных, на вкладке «Конкуренты» пользователю будет доступна информация о других ресурсах, которые занимали наиболее высокие позиции на страницах результатов поиска по интересующим его запросам в течение длительного времени. Также пользователь может увидеть список запросов, по которым каждый конкурент встречается наиболее часто (рис. 1).



**Рис. 1.** Выявление конкурентов

Система мониторинга и продвижения веб-сайтов позволит специалистам по продвижению веб-ресурсов сформировать новые запросы для продвижения и семантическое ядро сайта, на основе которого осуществляется ранжирование большинства поисковых систем.

Для подбора семантического ядра достаточно ввести ключевое слово, и сервис подберет информацию о том, как продвигаются конкуренты. Система обеспечивает анализ уже введенных пользователем запросов, отбирает наиболее эффективные и качественные из них, а также предлагает однородные запросы, по которым конкуренты осуществляют продвижение наиболее успешно.

Все табличные данные можно легко сортировать по возрастанию / убыванию. Содержимое любых отчетов легко экспортируется в популярный формат .odt для дальнейшего анализа.

## Варианты использования системы

Можно выделить следующих основных актеров, взаимодействующих с системой мониторинга и продвижения веб-сайтов:

1. *клиент* – это актер, использующий систему для мониторинга состояния интересующего его веб-ресурса;
2. *система мониторинга* – это актер, который обновляет данные, которые обрабатывает система с течением времени.

Можно определить следующие основные варианты использования системы мониторинга и продвижения веб-сайтов (рис. 2).

Клиент может *выявить конкурентов своего веб-сайта,* то есть просмотреть веб-сайты, которые наиболее часто встречаются на странице результатов поиска по интересующим клиента запросам.

Клиент может *просмотреть группы интересующих его запросов,* т.е. подмножество из заданных им запросов, по которым каждый из конкурентов занимает высокие позиции на странице результатов поиска в течение длительного времени.

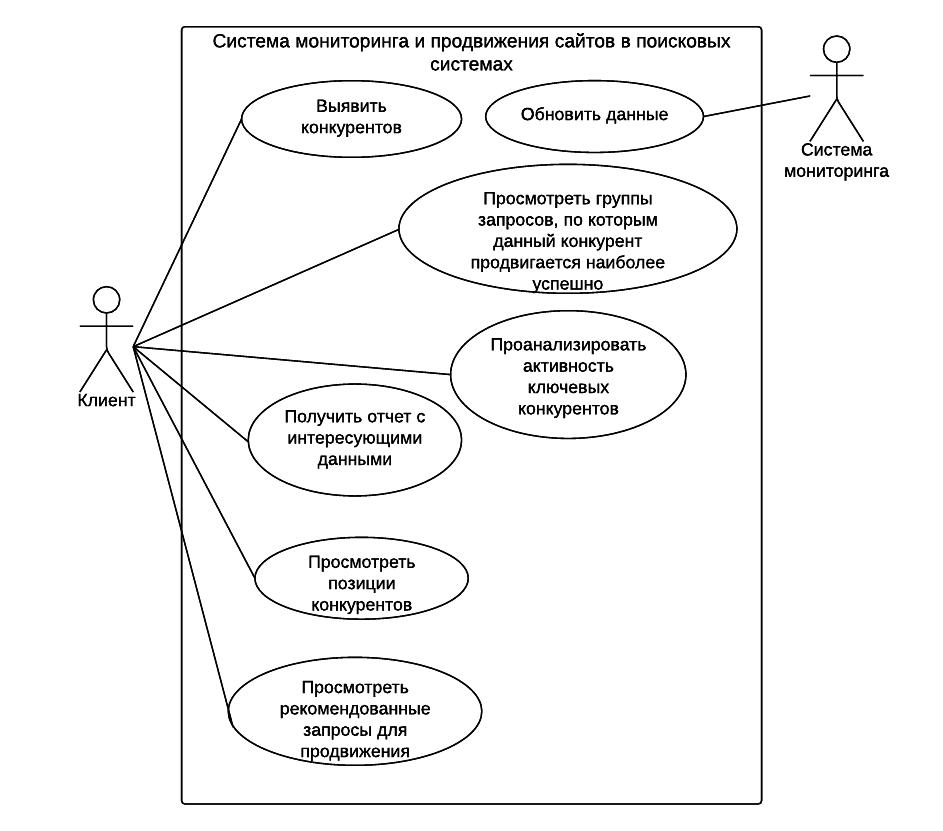


Рис. 2. Варианты использования системы мониторинга и продвижения

веб-сайтов

Клиент может *проанализировать активность каждого из конкурентов* за заданный период времени, т.е. увидеть, как часто и на каких позициях оказывался сайт конкурента в поисковой выдаче по каждому запросу за заданный период времени.

Клиент может *просмотреть позиции конкурентов* за заданный период времени по интересующим его запросам. Клиент может *просмотреть запросы*, которые рекомендованы ему для продвижения. Пользователь видит список запросов, схожих с теми, которые клиент заранее задал в системе.Клиент может *получить отчет о состоянии продвижения его сайта*, в котором также содержится информация о конкурентах. Все отчеты предоставляются клиенту в распространённом формате .odt.

Внешняя система мониторинга каждый день *собирает информацию* в поисковых системах *и записывает ее в базу данных*. В базе данных содержатся все веб-ресурсы, доменные имена которых встретились при сборе позиций по конкретному запросу на глубине 50, 200 или 500 позиций.

Подробное описание вариантов использование системы представлено в приложении 1.

## Нефункциональные требования к системе

В результате интервью с заказчиком были сформулированы следующие нефункциональные требования.

Поскольку приложение ориентировано на клиентов, которые могут находится в географически удаленных 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 интеллектуальной системе мониторинга и продвижения веб-сайтов в поисковых системах должен быть выполнен с использованием платформы Django, поскольку с ее использованием реализовано большинство проектов заказчика [14, 15].

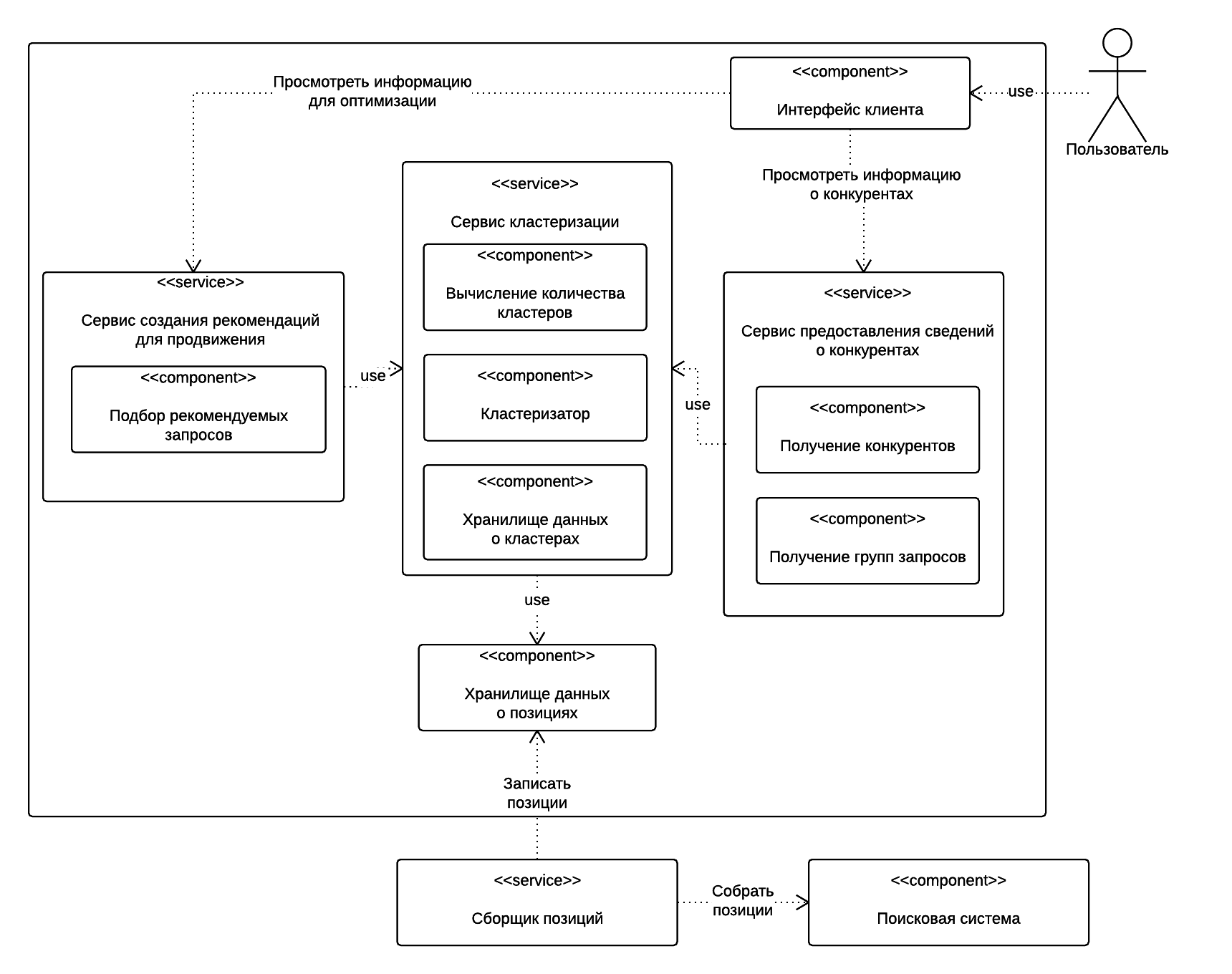
## Вывод

Таким образом, интеллектуальная система мониторинга и продвижения веб-сайтов в поисковых системах должна быть ориентирована на удобство использования клиентами и легко доступна через сеть интернет на различных устройствах. Также система должна интегрироваться в имеющуюся инфраструктуру заказчика.Система рассчитана на предоставление пользователям актуальной информации о текущем состоянии интересующих его веб-ресурсов и оказание помощи в продвижении сайтов SEO-специалистам.

# Архитектура системы мониторинга и продвижения веб-сайтов

## Компоненты системы мониторинга

На рисунке 3 изображена диаграмма компонентов интеллектуальной системы анализа данных мониторинга позиций сайтов в поисковых системах. Система состоит из следующих независимых сервисов: сервис кластеризации, сервис создания рекомендаций для продвижения, сервис предоставления сведений о конкурентах и сервис «хранилище данных о позициях». Сервис кластеризации является внутренним и не предоставляет открытого API для внешних пользователей. Сервис кластеризации использует данные о позициях из хранилища. Сервис создания рекомендаций для продвижения и сервис предоставления сведений о конкурентах используют данные о кластерах, предоставленные сервисом кластеризации*.*



**Рис. 3.** Компоненты системы мониторинга

Разберем более подробно компоненты, составляющие программную систему.

*Сборщик позиций* – это внешний сервис, обеспечивающий постоянный мониторинг позиций сайтов по запросам в поисковых системах и запись информации по позициям сайтов в хранилище.

*Сервис кластеризации* – это сервис, обеспечивающий разбиение запросов на кластеры. Объектом кластеризации является текст запроса к поисковой системе, привязанный к региону поиска. Реализация кластеризации в данной системе подробно рассматривается в п. 3.3 данной работы.

*Хранилище данных о позициях* – это база данных, содержащая информацию о результатах сбора позиций веб-сайтов по всем проектам за весь период времени.

*Компонент вычисления количества кластеров* – система, определяющая оптимальное количество кластеров, на которые следует разбить исходные данные.

*Кластеризатор –* система, разбивающая исходные данные на заранее заданное количество кластеров.

*Хранилище данных о кластерах* – база данных, содержащая информацию о точках кластеризации и кластерах.

*Получение конкурентов* – система анализа данных хранилища кластеров, выявляющая конкурентов сайта, запрошенного пользователем.

*Получение групп запросов* – система анализа данных хранилища кластеров, выявляющая запросы, по которым сайт, заданный пользователем, является конкурентом для исходного сайта.

*Подбор рекомендуемых запросов* – система анализа данных хранилища, определяющая запросы, по которым рекомендуется начать продвижение по сайту, запрошенному пользователем.

*Интерфейс клиента* – пользовательский интерфейс, с помощью которого можно получить сведения о конкурентах.

## Схема базы данных

Для организации работы мониторинга и продвижения веб-сайтов была разработана схема баз данных, содержащих информацию о поисковых запросах, найденных веб-сайтах и степени связи между запросами и веб-сайтами. Схемы разработанных хранилищ данных представлены на рисунках 4 и 5.



**Рис. 4.** Хранилище данных о позициях



**Рис. 5.** Хранилище данных о кластерах

Рассмотрим структуру базы данных о позициях:

* queries – это таблица для связи строки запроса с конкретным регионом;
* urls – это таблица, содержащая URL веб-сайтов, которые использует система для мониторинга позиций;
* regions – это таблица, содержащая коды регионов, по которым производится мониторинг;
* subscriptions – это таблица, реализующая связь «многие-ко-многим» между таблицами queries и regions;
* search\_results – это таблица, содержащая информацию о позиции запроса для каждого исследуемого региона в поисковой системе Яндекс.

Рассмотрим структуру базы данных о кластерах:

* points – это таблица, содержащая информацию о кластеризуемых элементах;
* clusters – это таблица, содержащая информацию о полученных в ходе кластеризации подмножествах;
* clusres\_points – это таблица, содержащая информацию о принадлежности элементов кластерам.

## Применение алгоритма кластеризации для запросов

Сборщик позиций производит постоянный мониторинг позиций ресурсов в результатах поиска по заданному набору запросов в день . Основной характеристикой отслеживания является максимальная глубина поиска , определяющая максимальное количество позиций, анализируемых сборщиком.

Исходной информацией для алгоритма кластеризации в рамках решаемой задачи является матрица , каждая строка которой представляет собой поисковый запрос, а столбец представляет определенный анализируемый веб-сайт.

Для заполнения ячейки матрицы с номером , отвечающей за усредненную позицию -того сайта по -тому запросу за дней, воспользуемся следующей метрикой:

*.*

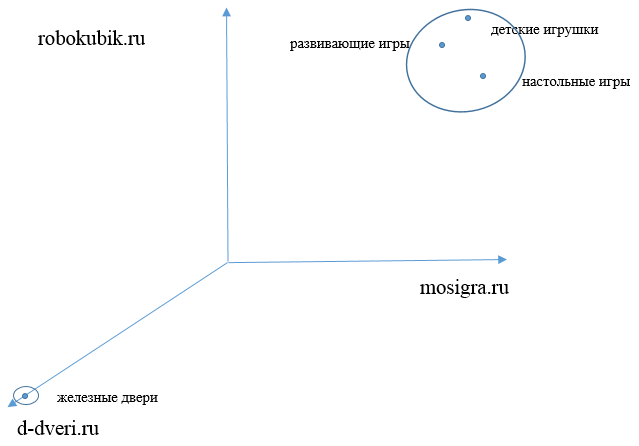
Таким образом, в ячейках матрицы будут располагаться величины, являющиеся разницей между максимальной глубиной поиска и текущим положением сайта по запросу в день . В этом случае, чем выше сайт находится в поисковой выдаче, тем больше будет его вес в матрице по данному запросу.

Рассмотрим пример анализа. Пусть имеются 4 запроса: «детские игрушки», «настольные игры», «развивающие игры», «железные двери» и 3 веб-сайта: «robokubik.ru», «d-dveri.ru», «zapovednikrest.ru». В таблице 1 приведены результаты расчета усредненной метрики для соответствующих сайтов по указанным запросам при глубине поиска . Как видно из таблицы, сайты «robokubik.ru» и «zapovednikrest.ru» присутствуют в TOP-200 результатах по запросам «детские игрушки», «настольные игры», «развивающие игры». Сайт «d-dveri.ru» по данным запросам в 200 первых результатов выдачи не входил, зато он появляется в выдаче по запросу «железные двери».

**Табл. 1.** Результаты сбора позиций

|  |  |  |  |
| --- | --- | --- | --- |
| Веб-сайт  Запрос | robokubik.ru | mosigra.ru | d-dveri.ru |
| детские игрушки | 199 | 190 | 0 |
| развивающие игры | 196 | 170 | 0 |
| настольные игры | 168 | 195 | 0 |
| железные двери | 0 | 0 | 180 |

Можно провести кластеризацию представленных данных, для выявления близких по смыслу запросов. Визуальное представление решения задачи кластеризации представлено на рисунке 6. В качестве осей координат на графике используются веса сайтов. Точками на графике отображаются соответствующие запросы.



**Рис. 6.** Пример работы алгоритма кластеризации

Таким образом, из анализа видно, что веб-сайты robokubik.ru и mosigra.ru конкурируют друг с другом по запросам: «детские игрушки», «настольные игры», «развивающие игры», а запрос «железные двери» не входит в их кластер.

Кроме того, в системе предусмотрен алгоритм выявления конкурентов для каждой группы запросов. Для каждого центра кластера упорядочим метрики кластеризации по значению центра в данной метрике. Поскольку каждой метрике соответствует один веб-сайт, мы сможем получить список сайтов, отсортированный по значимости в данном кластере. Выбираются 5 наиболее значимых веб-сайтов-конкурентов.

В результате работы алгоритма пользователь сможет увидеть конкурентов для каждого проекта, зарегистрированного системе, а также список запросов для каждого конкурента, по которым он соперничает с искомым сайтом.

*Алгоритм четких k-средних применительно к решаемой задаче*

Поставим задачу кластеризации.

1. В качестве объектов кластеризации возьмем множество запросов .
2. В качестве признаков кластеризации по запросу возьмем строку матрицы , содержащую усредненную позицию всех анализируемых запросу за дней.
3. В качестве признака сходства запросов и возьмем евклидову метрику расстояния между координатами запросов в мерном пространстве, определенных соответствующими строками матрицы :

Алгоритм реализуется следующим образом.

1. Случайным образом инициализировать центроиды кластеров запросами так, что количество совпадает со значением оптимального количества кластеров, вычисленным ранее.

2. Для каждого центроида вычислить расстояния до всех точек

3. Сформировать кластеры. Для каждого центроида из множества отобрать подмножество точек с минимальным расстоянием до

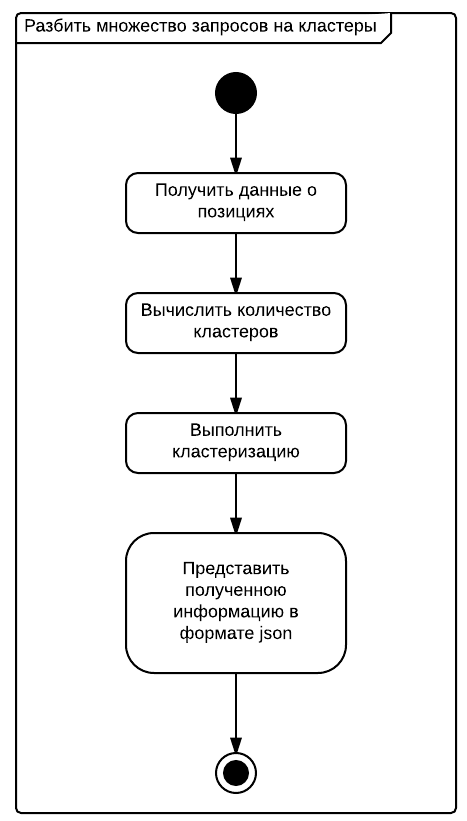
4. Вычислить новые центроиды как среднее всех точек кластера :

где *-* количество точек в кластере *.*

5. Сравнить значения с полученными на предыдущей итерации. В случае нахождения хотя бы одного несовпадения перейти к шагу 2.

6. Конец алгоритма.

Использование алгоритма k-means в разрабатываемой системе представлено на рис. 7.



**Рис. 7.** Диаграмма деятельности алгоритма кластеризации

## Вывод

В соответствии с системными требованиями была спроектирована архитектура системы мониторинга и продвижения сайтов в поисковых системах. Система имеет сервис-ориентированную архитектуру, состоящую из трех сервисов. Кроме того, в соответствии с функциональными требованиями были выделены компоненты, представляющие собой интерфейсы для взаимодействия с внешними сущностями. Для реализации основной функциональности был разработан алгоритм поиска конкурентов веб-сайта на основе алгоритма четкой кластеризации k-средних. Также для хранения результатов работы системы были спроектированы хранилище данных о кластерах и хранилище данных о позициях. В соответствии с нефункциональными требованиями была подробно разработана реализация наиболее важнейших прецедентов. В соответствии с анализом сценариев работы пользователя были разработаны доступные пользовательские интерфейсы системы мониторинга и продвижения сайтов в поисковых системах.

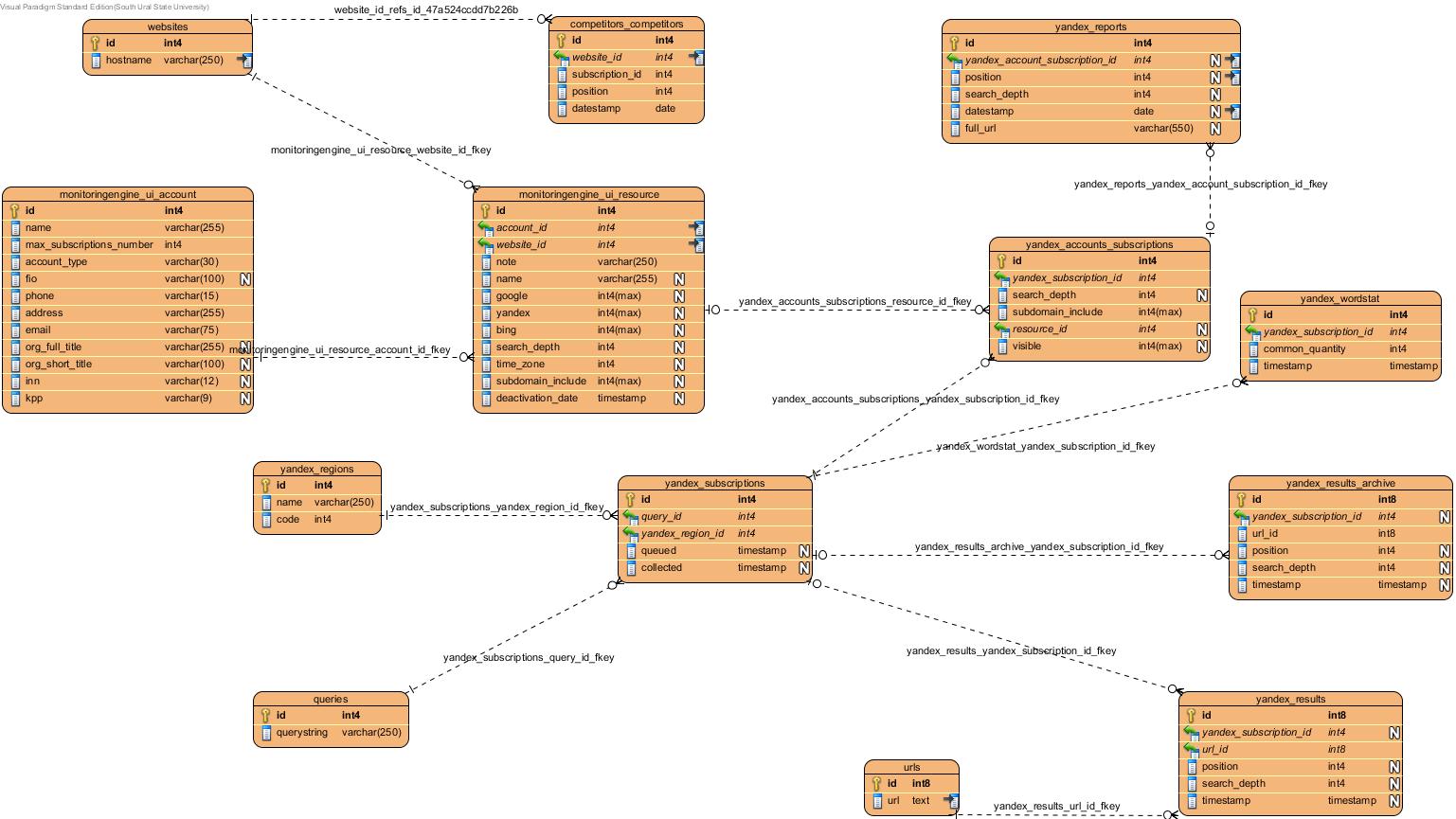
# Реализация системы

## База данных

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

* *monitoringengine\_ui\_account* – таблица аккаунтов, состоящая из данных обо всех организациях, использующих систему;
* *monitoringengine\_ui\_resource –* таблица ресурсов, содержащая информацию обо всех ресурсах, используемых системой;
* *yandex\_subscriptions* – таблица для связи информации о запросе с конкретным регионом в поисковой системе Яндекс;
* *yandex\_reports* – таблица, содержащая данные для формирования отчета о состоянии продвижения;
* *websites* – таблица, содержащая имена хостов веб-сайтов, которые использует система для мониторинга позиций;
* *queries* – таблица, содержащая строки запросов, по которым производится мониторинг;
* *regions* – таблица, содержащая коды регионов, по которым производится мониторинг;
* *yandex\_account\_subscriptions* – таблица, связывающая поисковую строку и код проекта, по которым производится мониторинг;
* *yandex\_wordstat* – таблица, в которую записывается информации о частотности запросов по регионам;
* *yandex\_results* – таблица, содержащая информацию о позиции запроса для каждого исследуемого региона в поисковой системе Яндекс;
* *competitors* – таблица, содержащая информацию о конкурентах каждого веб-сайта.

Схема используемой базы данных изображена на рис. 8.



**Рис. 8.** Схема базы данных системы мониторинга и продвижения сайтов

## Реализация компонентов системы

*Сервис кластеризации*

Сервис обеспечивает работу со следующими ресурсами.

*Кластеры:* конечная точка /clusters. Допустимые действия:

* GET /clusters (получить список кластеров) возвращает полученные данные с информацией о принадлежности каждого элемента к определенному кластеру ф формате массива, каждый элемент которого словарь, где идентификатору кластера соответствует степень принадлежности каждого элемента множества этому кластера, например [{id: 1, points: {«1»: 0.1, «2»: 0.1, «3»: 0.9 …}}, {id: 2, points {…}}].
* GET /clusters/<id> (получить информацию о кластере с id в формате словаря, состоящего из идентификатора кластера и степеней принадлежности каждого элемента исходного множества к этому кластеру). Например, GET /cluster/1 -> {id: 1, points: {«1»: 0.1, «2»: 0.1, «3»: 0.9 …}}

*Запросы*, конечная точка /query. Допустимые действия: GET /query – получить список запросов. GET /query/1 – получить информацию о запросе с id 1. GET /query/1 ->{id: 1, positions: {«2015-02-10»: 5, «2015-02-11»: 4, «2015-02-12»: 4 …}}

*Сервис сведений о конкурентах*

*Конкуренты*. Конечная точка /resource/1/competitors. Допустимые действия: GET /resource/1/get\_competitors — получить список конкурентов для ресурса с id 1.

*Запросы*. Конечная точка /queries. Допустимые действия: GET /queries?competitor1=5&competitor2=10 – получить список запросов для ресурса с id 1 и конкурентов с id=1 или 2.

Для определения оптимального количества кластеров используется CS индекс (рис. 9). Для его расчета вычисляется отношение максимального расстояния между элементами кластера к минимальному расстоянию между кластерами. Оптимальным считается то количество кластеров, при котором значение CS минимально.

def quality(clusters):

max\_clusters\_points = 0

min\_clusters\_dist = 0

for cluster\_center, points in clusters:

for point in points:

max\_point = 0

for point1 in points:

dif = abs\_clust(diff(point1, point))

if dif > max\_point:

max\_point = dif

max\_clusters\_points += max\_point / len(points)

min\_clust = 100000000000000000

for cluster\_center1, points1 in clusters:

dif\_clust = abs\_clust(diff(cluster\_center1, cluster\_center))

if not compare(cluster\_center, cluster\_center1) and (dif\_clust < min\_clust):

min\_clust = dif\_clust

min\_clusters\_dist += min\_clust

return max\_clusters\_points / min\_clusters\_dist

**Рис. 9.** Листингфункции для определения количества кластеров

Для выполнения кластеризации используется конструктор KMeans с параметром n\_clusters=n и метод fit, который разбивает искомый объект на кластеры. Для получения конкурентов каждого кластера метрики сортируются по убыванию относительно значения центра кластера в данной метрике. После этого результат приводится к виду, пригодному для обработки (рис. 10).

def clusterize(data, items, n):

kmeans = KMeans(n\_clusters=n)

kmeans.fit(data)

clusters = []

result = {}

clusters\_for\_quality = []

centers = kmeans.cluster\_centers\_

for c in centers:

srt = range(len(c))

srt.sort(key=lambda i: -c[i])

clusters.append(srt)

clusters\_for\_quality.append((c, []))

for i in range(0, len(kmeans.labels\_)):

if not kmeans.labels\_[i] in result:

result[kmeans.labels\_[i]] = set()

result[kmeans.labels\_[i]].add(items[i])

clusters\_for\_quality[kmeans.labels\_[i]][1].append(data[i])

value = quality(clusters\_for\_quality)

return clusters, result, value

**Рис. 10.** Листингфункции кластеризации

*Сервис создания рекомендаций для продвижения*

Ресурсы /resources:

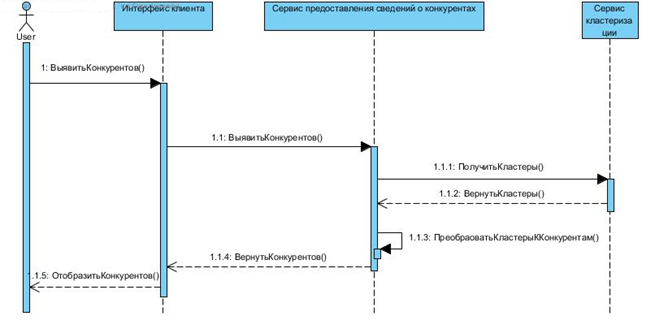
* GET /resources – выдает каталог доступных пользователю ресурсов
* GET /resources/<id> - выдает данные о ресурсе с id 1

Рекомендуемые запросы:

Конечная точка /resources/1/recommended\_queries. Допустимые действия: GET /resource/1/recommended\_queries — получить список рекомендуемых запросов для ресурса с id 1.

## Организация взаимодействия компонентов системы

Получение информации о конкурентах (рис. 11).



**Рис. 11.** Диаграмма последовательностей для получения конкурентов

## Веб-интерфейс

Веб-интерфейс системы поиска конкурентов по группам запросов было решено реализовать с помощью двумерного графика. Для реализации данной функциональности была использована библиотека d3.js [16]. Код функции инициализации графика приведен на рис. 12. График расположен внутри элемента svg и представляет собой два подграфика. Первый подграфик представляет собой множество точек на плоскости, расположенных в соответствии с заданными координатами и радиусом. Радиус каждой точки вычисляется отдельно. Другой подграфик представляет собой диагональную прямую линию, с помощью которой моно облегчить восприятие информации.

var initPlot=function(element\_id, options){

var xAxis = options.x, yAxis = "";

var svg = d3.select(element\_id)

.append("svg")

.attr("width", 800)

.attr("height", 460);

var div = d3.select("body").append("div")

.attr("class", "tooltip")

.style("opacity", 0);

var chart = svg.append("g")

.classed('chart', true)

.attr('transform', 'translate(80, -220)');

$("#plot").click(function(){

div.transition()

.duration(200)

.style("opacity", 0);

});

chart.append('line')

.attr('class', 'diagonal').attr({'x1': xScale(1), 'y1': yScale(1), 'x2': xScale(200), 'y2': yScale(200)});

// Axis labels

chart.append('text')

.attr({'id': 'xLabel', 'x': 250, 'y': 670, 'text-anchor': 'middle'})

.text(xAxis);

chart.append('text')

.attr('transform', 'translate(-60, 420)rotate(-90)')

.attr({'id': 'yLabel', 'text-anchor': 'middle'})

.text(yAxis);

chart.append('g')

.attr('transform', 'translate(0, 630)')

.attr('id', 'xAxis')

.call(makeXAxis);

chart.append('g')

.attr('transform', 'translate(-10, 0)')

.attr('id', 'yAxis')

.call(makeYAxis);

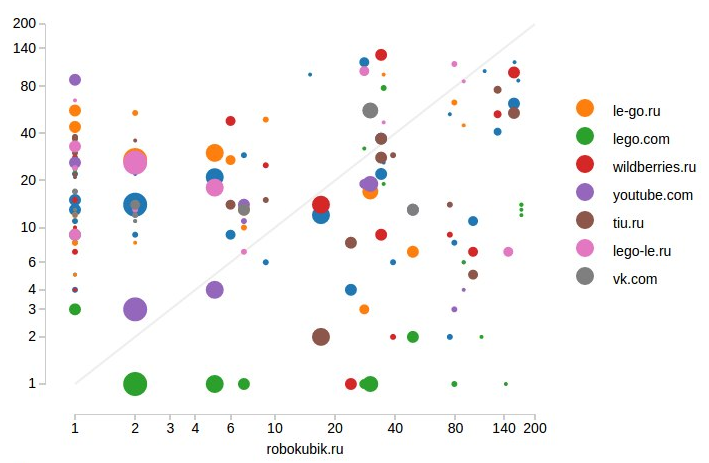
return svg;

}

**Рис. 12.** Листингфункции отображения графика

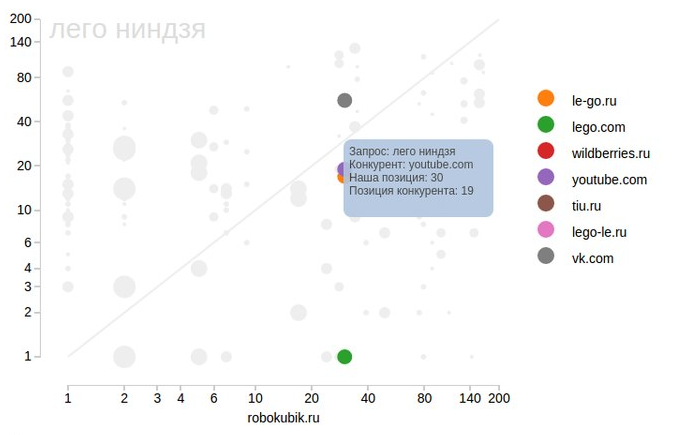
На рисунке 13 приведен пример работы системы для веб-сайта robokubik.ru. На осях графика отмечены позиции сравниваемых веб-сайтов с использованием логарифмической шкалы. Горизонтальная ось представляет собой позиции искомого сайта, вертикальная – сравниваемых веб-сайтов конкурентов. Точки на графике представляют собой поисковые запросы. Их координаты определяются позицией искомого сайта в поисковой системе по горизонтальной оси и координатой сайта конкурента по вертикальной оси. Цвет точки определяется цветом сайта конкурента. Соответствие цветов веб-сайтам описано в легенде. Каждая точка на графике имеет диаметр, который определяется частотностью запроса в поисковой системе: чем выше частотность, тем больше диаметр точки.

Кроме того, для удобства восприятия, на график добавлена диагональная линия. Точки, оказавшиеся выше диагональной линии, представляют запросы, по которым позиция искомого сайта выше позиции конкурента, причем, чем дальше точка от диагональной линии, тем позиция искомого сайта выше позиции конкурента. То же касается точек, оказавшихся ниже диагональной линии. Они представляют собой запросы, по которым позиция конкурентов превышает позицию искомого сайта. Чем ближе запросы к диагональной линии, тем ближе находятся их позиции в поисковой выдаче.



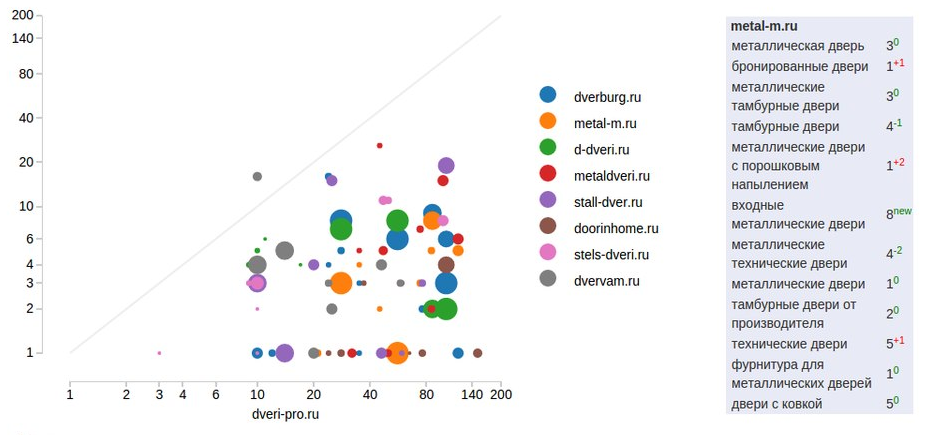
**Рис. 13.** Интерфейс поиска конкурентов

Для сравнения состояния позиций искомого сайта с позициями конкурентов по конкретному запросу необходимо нажать на одну из точек на графике, представляющую этот запрос. При этом на графике подсветятся точки, соответствующие выбранному запросу, но для других веб-сайтов. Таким образом пользователь сможет увидеть все позиции конкурентов по этому запросу. Кроме того, пользователь увидит всплывающее окно с текстом запроса, информацией о конкуренте и его позиции по этому запросу в поисковой системе. Пример отображения информации для сравнения конкурентов по конкретному запросу представлен на рис. 14.



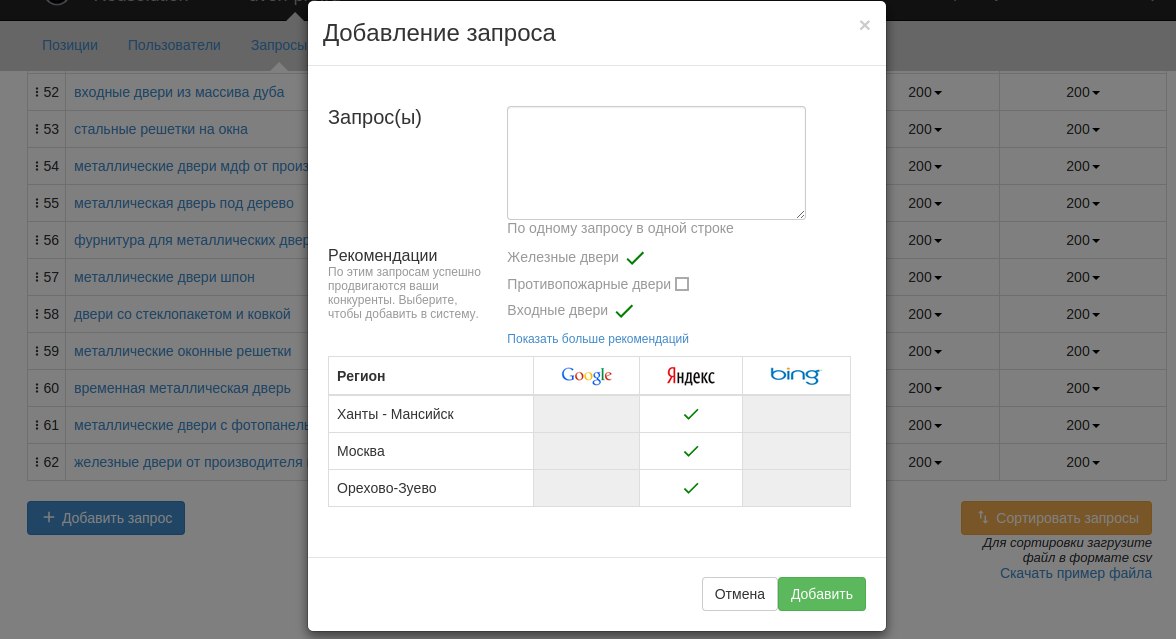
**Рис. 14.** Интерфейс конкурентов для запроса

Для просмотра групп запросов по конкурентам, необходимо нажать на имя интересующего веб-сайта в легенде. После этого справа от графика отображается перечень запросов, по которым данный веб-сайт конкурирует с искомым. Справа от каждого запроса отображается позиция веб-сайта конкурента, занимаемая им по данному запросу, и разница с его позицией в прошлом месяце. Если в прошлом месяце данный веб-сайт не конкурировал с искомым по данному запросу, то возле позиции по данному запросу отобразится значок «new» (рис. 15).



**Рис. 15.** Интерфейс отображения групп запросов для конкурента

Интерфейс создания рекомендаций для продвижения встроен в интерфейс добавления запросов в систему мониторинга позиций сайтов. Пользователю доступно окно ввода, куда он может самостоятельно вписать интересующие запросы. Ниже представлены запросы, которые система определила как рекомендуемые на основе уже добавленных. Пользователь может выбрать понравившиеся из видимых или раскрыть список и просмотреть все рекомендованные запросы (рис. 16).



**Рис. 16.** Интерфейс предоставления рекомендаций для продвижения

## Вывод

На основе разработанной архитектуры была реализована система, предоставляющая пользователю информацию о позициях интересующего его веб-сайта по заданным запросам в поисковых системах, выявлять основных конкурентов веб-сайта, получать рекомендации по продвижению веб-сайта. Информация для отображения берется из базы данных, созданной с помощью СУБД PostreSQL. Выполнена реализация сервиса кластеризации, сервиса поиска конкурентов, сервиса создания рекомендаций для продвижения и пользовательского интерфейса с использованием веб-фреймворка Django на языке программирования Python. Взаимодействие компонентов системы было реализовано через API посредством отправки HTTP-запросов.

# Тестирование

Для тестирования веб-приложения использовались следующие методы тестирования:

* функциональное тестирование;
* проверка эргономичности;
* тестирование ссылок;
* тестирование интерфейса пользователя.

Тестирование проводилось на реальных данных пользователей, которые являются сотрудниками организации.

**Функциональное тестирование** – это тестирование программного обеспечения в целях проверки реализуемости функциональных требований, то есть способности программного обеспечения в определенных условиях решать задачи, нужные пользователям. Функциональные требования определяют, что именно делает программное обеспечение, какие задачи оно решает. Набор тестов на функциональность представлен в таблице 2.

**Проверка эргономичности** – метод оценки удобства продукта в использовании, основанный на привлечении пользователей в качестве тестировщиков, испытателей и суммировании полученных от них выводов.

При прохождении данного теста были привлечены специалисты по продвижению сайтов компании. Тестировщикам было предложено решить следующие задачи:

* зарегистрироваться;
* просмотреть конкурентов для веб-сайтов, продвижением которых они занимаются;
* выделить запросы, по которым каждый конкурент продвигается наиболее успешно;
* просмотреть позиции конкурентов по выбранным запросам;
* проанализировать активность конкурентов;
* добавить запросы в созданные проекты, основываясь на рекомендациях для продвижения;
* просмотреть собранные позиции для добавленных проектов и проектов-конкурентов за несколько дней;
* получить отчет о состоянии продвижения;
* удалить запросы из проекта.

Все задачи были решены всеми тестировщиками без каких-либо затруднений. Тест пройден успешно.

Далее проводилось тестирование ссылок на работоспособность. Тестирование прошло успешно.

Тестирование интерфейса пользователя. Интерфейс был протестирован на браузерах Internet Explorer, Mozilla Firefox, Opera, Google Chrome. Не возникло никаких проблем с версткой. Интерфейс во всех браузерах был идентичным. Тест пройден успешно.

**Табл. 2.** Тестирование системы

| **№** | **Название теста** | **Шаги** | **Ожидаемый результат** | **Тест пройден?** |
| --- | --- | --- | --- | --- |
|  | Регистрация | 1. На странице аутентификации выбрать кнопку «Зарегистрироваться». 2. Заполнить форму регистрации. 3. Нажать на кнопку «Готово». | В базе данных появляется информация о новом пользователе. | Да |
|  | Вход в систему | 1. На странице аутентификации ввести зарегистрированный адрес электронной почты и верный пароль. 2. Нажать на кнопку «Войти». | Пользователю становится доступна главная страница сайта. | Да |
|  | Выявление конкурентов | 1. На странице одного из проектов выбрать вкладку «Конкуренты». | На странице отображается график запросов и список конкурентов | Да |
|  | Просмотр групп запросов по конкурентам | 1. На странице одного из проектов выбрать вкладку «Конкуренты». 2. Выбрать интересующего конкурента | Рядом с выбранным конкурентом отображается список запросов, по которым он продвигается наиболее успешно. Справа от каждого запроса отображается позиция, занимаемая конкурентом по данному запросу. | Да |

Продолжение таблицы 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Название теста** | **Шаги** | **Ожидаемый результат** | **Тест пройден?** |
|  | Анализ активности конкурентов | 1. На странице одного из проектов выбрать вкладку «Конкуренты». 2. Нажать на кнопку «Сравнить с результатом прошлого месяца». | На странице отображается график сравнения позиций конкурентов за прошлый месяц и в настоящее время | Да |
|  | Просмотр позиций конкурентов | 1. На графике отображения запросов по конкурентам выбрать одну из точек 2. Нажать на точку запроса | Появляется окно с информацией о позиции конкурента по данному запросу | Да |
|  | Просмотр позиций | 1. На странице ресурса выбрать с помощью слайдера промежуток времени. 2. Выбрать регион с помощью таблеток. | Пользователю отображается таблица позиций для данного ресурса за выбранный промежуток времени в выбранном регионе. | Да |
|  | Ошибка при просмотре позиций | 1. На странице ресурса выбрать с помощью слайдера промежуток времени. 2. Выбрать регион с помощью таблеток. | Пользователю отображается сообщение «Нет позиций» | Да |
|  | Добавление запросов, просмотр рекомендаций | 1. На странице редактирования проекта вписать строки запросов в текстовое поле. 2. Выбрать запросы из рекомендуемых. 3. Нажать на кнопку «Добавить». 4. С помощью кликов левой кнопки мыши по пустым ячейкам добавить регионы мониторинга. | В таблице запросов появляется новое поле с только что добавленным запросом. Ему присвоены параметры по умолчанию. | Да |
|  | Получение отчета | 1. На странице просмотра позиций выбрать интересующие регионы. 2. Выбрать интересующий промежуток времени. 3. Нажать на кнопку «Получить отчет». | Данные со страницы сохраняются на компьютер пользователя в формате ods. | Да |
|  | Удаление запроса | 1. На странице редактирования проекта нажать на крестик в непустой ячейке. 2. Подтвердить действие. | Выбранная ячейка оказывается пуста.  Запрос удаляется из базы данных. На странице позиций запрос не отображается. | Да |

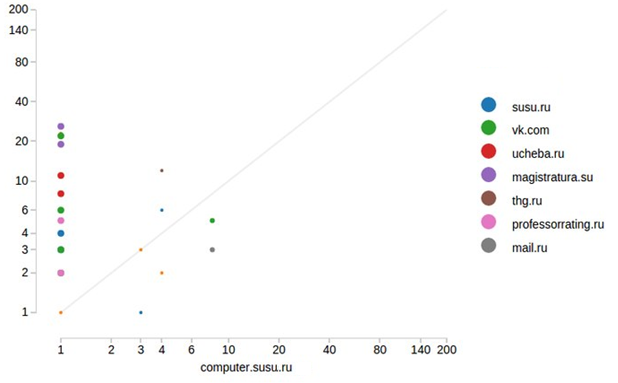
Окончание таблицы 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Название теста** | **Шаги** | **Ожидаемый результат** | **Тест пройден?** |
|  | Добавление прав | 1. На странице аккаунта или ресурса выбрать право для пользователя из выпадающего списка. 2. Подтвердить действие. | Пользователь получил добавленное право. | Да |
|  | Удаление прав | 1. На странице аккаунта или ресурса нажать на крестик, расположенный на таблетке права для обозначения права пользователя. 2. Подтвердить действие. | Пользователь лишился удаленного право. | Да |

Работа системы была протестирована на 50 проектах (приложение 2), имеющихся в системе. В каждом проекте содержится около 100 запросов. На всех запросах были получены корректные результаты при тестировании сервиса поиска конкурентов и сервиса создания рекомендаций для продвижения.

На рисунке 17 представлен результат поиска конкурентов для веб-сайта computer.susu.ru. Тестирование было проведено в поисковой системе Яндекс по региону Челябинск и включало следующие запросы:

* факультет вми юургу;
* факультет вычислительной математики и информатики;
* суперкомпьютер юургу;
* вми;
* программирование юургу;
* куда пойти учиться на программиста;
* фиит;
* прикладная математика и информатика;
* суперкомпьютер скиф аврора;
* фундаментальная информатика и информационные технологии
* программирование на суперкомпьютере



**Рис. 17.** Тестирование сервиса поиска конкурентов для сайта computer.susu.ru

В ходе тестирования оказалось, что в результатах, полученных с помощью системы сбора позиций, не учитывается разница в доменах второго уровня и ниже. В результате тестирования нами была произведена корректировка работы хранилища сборщика, чтобы не учитывались результаты домена более высокого уровня.

# Заключение

В ходе проделанной работы были решены следующие задачи.

1. Изучены особенности существующих систем мониторинга позиций сайтов в поисковых системах.
2. Изучены особенности работы с платформами Django и Twitter Bootstrap.
3. Изучены существующие алгоритмы интеллектуального анализа данных.
4. Разработана архитектура интеллектуальной системы мониторинга и продвижения веб-сайтов в поисковых системах.
5. Разработан пользовательский интерфейс системы мониторинга и продвижения позиций сайтов по запросам в поисковых системах.
6. Реализована интеллектуальная система мониторинга и продвижения веб-сайтов в поисковых системах.
7. Система протестирована на реальных данных.
8. Разработанная система внедрена в эксплуатацию активно используется сотрудниками и клиентами IT-компании ООО «Редсолюшн».

## Литература

* + - 1. Bar-Ilan J., Mat-Hassan M., Levene M. Methods for comparing rankings of search engine results // Comput. Networks, 2006. − Vol. 50. − No. 10. P. 1448-1463.
      2. Celebi M.E. A comparative study of efficient initialization methods for the k-means clustering algorithm // Expert Systems with Applications January 2013. – Vol. 40. – Issue 1. – P. 200-210.
      3. Ghosh S., Dubey S.K. Comparative Analysis of K-Means and Fuzzy CMeans Algorithms // International Journal of Advanced Computer Science and Applications, 2013. – Vol. 4. – No. 4. – P. 45-47.
      4. Kiselev M.V., Ananyan S.M., Arseniev S.B. Principles of Data Mining and Knowledge Discovery Lecture Notes//Computer Science, 1999. – Vol. 1704. – P.366-371.
      5. Segaran T. Programming Collective Intelligence by Toby Segaran. – USA: O`Reilly Media, 2007. – P. 29-38.
      6. Wagstaf K., Cardie C. Constrained K-means Clustering with Background Knowledge // Proceedings of the Eighteenth International Conference on Machine Learning, 2001. − Р. 577-584.
      7. Севостьянов И. Поисковая оптимизация. Практическое руководство по продвижению сайта в Интернете. 2-е изд. – СПб.: Питер, 2012. – 272 с.
      8. Ahrefs. Site Explorer. [Электронный ресурс] URL: https://ahrefs.com/ (дата обращения: 12.05.2015).
      9. BCGroup. Маркетинговые исследования и аналитика. [Электронный ресурс] URL: http://www.bcgroup.su/analiz-konkurentov (дата обращения: 01.11.2014).
      10. Microsoft Developer Network. [Электронный ресурс] URL: http://msdn.microsoft.com/ru-ru/library/ms175595.aspx (дата обращения: 01.11.2014).
      11. QuickSpout, сайт системы. [Электронный ресурс] URL: http://www.quicksprout.com/ (дата обращения: 01.11.2014).
      12. SpyWords, сайт системы. [Электронный ресурс] URL: http://spywords.ru/ (дата обращения: 01.11.2014).
      13. Алгоритм нечеткой кластеризации. [Электронный ресурс] URL: http://home.deib.polimi.it/matteucc/Clustering/tutorial\_html/cmeans.html (дата обращения: 19.05.2015).
      14. Документация Django. [Электронный ресурс] URL:https://www.djangoproject.com/ (дата обращения: 28.05.2015).
      15. Документация Twitter Bootstrap. [Электронный ресурс] URL: http://getbootstrap.com/ (дата обращения: 28.05.2015).
      16. Документация по работе с библиотекой d3. [Электронный ресурс] URL: D3js.org (дата обращения: 21.05.2015)
      17. Как работают поисковые системы - SearchEngines.ru . [Электронный ресурс] URL: http://www.searchengines.ru/articles/004556.html (дата обращения: 30.05.2015).
      18. Определение позиций сайта в поисковиках [Электронный ресурс] URL: http://www.semonitor.ru/positioning.html (дата обращения: 29.05.2015).
      19. Статистика. [Электронный ресурс] URL: http://www.liveinternet.ru/stat/ru/searches.html (дата обращения: 28.05.2015).

# Приложения

# Приложение 1

Описание вариантов использования системы мониторинга и продвижения веб-сайтов в поисковых системах

***UseCase:* Выявление конкурентов**

*ID:* 1

*Аннотация:* Выявление конкурентов веб-сайта

*Главный актер:* Клиент

*Предусловия:* Клиентом в Системе мониторинга:

1. указан веб-сайт для анализа;
2. созданы запросы для анализа;
3. собраны данные результатах поиска по запросам в течение интервала времени

*Описание:* Клиент просматривает позиции для конкретного ресурса, система определяет часто встречающиеся веб-сайты в данных о сборе для каждого запроса, система выделяет основных конкурентов данного ресурса и группы запросов для каждого конкурента

*Постусловия:* Данные отображены Клиенту

***UseCase:* Просмотр групп запросов**

*ID:* 2

*Аннотация:* Просмотр групп запросов, по которым конкуренты наиболее успешны

*Главный актер:* Клиент

*Предусловия:* Выявлены конкуренты

*Описание:* Клиент запрашивает позиции по ресурсу, система на основе проанализированных данных сбора отображает пользователю запросы, по которым каждый конкретный конкурент занимал высокую позицию на странице результатов поиска длительный период времени

*Постусловия:* Клиент просматривает группы запросов

***UseCase:* Просмотр позиций конкурентов**

*ID:*4

*Аннотация:* Просмотр позиций конкурента за заданный период времени по заданным запросам

*Главный актер:* Клиент

*Предусловия:* Выявлены конкуренты, выбран конкретный веб-ресурс для просмотра

*Описание:* Клиент запрашивает данные о конкретном конкуренте, система запрашивает у базы данных информацию о сборе для конкретного веб-ресурса по заданным запросам и периоду времени, система отображает данные пользователю в табличном виде

*Постусловия:* Данные отображены пользователю

***UseCase:* Просмотр запросов, рекомендованных для продвижения**

*ID:*5

*Аннотация:* Просмотр группы запросов, по которым рационально продвигать сайт клиента.

*Главный актер:* Клиент

*Предусловия:* Заданы интересующие запросы, есть данные о сборе не менее чем за месяц

*Описание:* Клиент просматривает позиции, система предлагает пользователю список из запросов, однородных уже заданным, по которым рекомендовано осуществлять продвижение

*Постусловия:* Данные отображены пользователю

***UseCase:* Получение отчета**

*ID:*7

*Аннотация:*Экспорт содержимое любых отчетов в формат .odt для дальнейшего анализа

*Главный актер:* Клиент

*Предусловия:* Выбраны данные для отчета

*Описание:* Клиент запрашивает отчет, система конвертирует данные предоставленные пользователю в документ формата .odt, затем создает документ

*Постусловия:* Отчет получен пользователем

***UseCase:* Сбор данных**

*ID:*8

*Аннотация:* Ежедневный сбор данных о позициях веб-сайтов в поисковых системах и запись в БД

*Главный актер:* Система мониторинга

*Предусловия:* Заданы веб-сайты и запросы, по которым необходимо собирать данные в поисковых системах

*Описание:* Cистема сбора позиций в заданное время суток осуществляет запись в базу данных информации о позициях каждого веб-ресурса, который оказался в поисковой выдаче поисковой системы по конкретному запросу до заданной глубины поиска

*Аннотация:* Ежедневный сбор данных о позициях веб-сайтов в поисковых системах и запись в БД

# Приложение 2

Перечень веб-сайтов, на которых была протестирована система:

* rodniki-vip.ru
* electrosad.ru
* uralice.ru
* doorsmetall.ru
* rg.ru
* logotek.ru
* legorubik.ru
* erg74.ru
* balver.ru
* paritet-dveri.ru
* churilovo-agro.ru
* inteldoors.ru/
* doors-torex.ru
* dverka.ru
* credit-broker74.ru
* clinica-repromed.ru
* citywoman74.ru
* protivopozharnie-dveri.ru
* garant-srv.ru
* impulse-plus.ru
* rnga.ru
* vizit-doors.ru
* dveri-pro.ru
* redsolution.ru
* ural-pricep.ru
* mirdverey.com
* zlatmedika.ru
* robokubik.ru
* stroistal-invest.ru
* dvery.ru
* 1990.ru
* zapovednikrest.ru
* dh74.ru
* kolesastok.ru
* diagnoz74.ru
* massage-masters.ru
* teploventkom.ru
* all-doors.ru
* stroistal-invest.ru
* impulse-plus.ru
* dvery.ru
* dveri-krit.ru
* 1995.ru
* luck-okna.ru
* kupibuket74.ru
* dveri-pro.ru
* sp.susu.ru
* computer.susu.ac.ru
* dolcshik74.ru