

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
Информатики и радиоэлектроники
Кафедра инженерной психологии и эргономики

На правах рукописи

УДК 004.77

Назаренко
Артур Игоревич

МОНИТОРИНГ ПРОЦЕССА ПРОДВИЖЕНИЯ ВЕБ-САЙТОВ
В ПОИСКОВЫХ СИСТЕМАХ

Автореферат диссертации на соискание академической степени
магистра технических наук

1-23 80 08 Психология труда, инженерная психология, эргономика

Магистрант А.И. Назаренко

Научный руководитель
С.Л. Флерко, кандидат
экономических наук, доцент

Заведующий кафедрой ИПиЭ
К.Д. Яшин, кандидат
технических наук, доцент

Нормоконтролёр
Е.С. Иванова,
ассистент кафедры ИПиЭ

Минск 2016

ВВЕДЕНИЕ

В настоящее время доля интернет-бизнеса в ВВП стран СНГ неуклонно растет, множество коммерческих компаний ориентируются именно на интернет как чуть ли не единственный источник клиентов, с каждым годом число таких компаний растет. Основным источником посетителей почти любого интернет-ресурса являются поисковые системы, Google в масштабе всего мира, Яндекс в масштабе СНГ. Изначально, до появления поисковых систем найти интернет-ресурс можно было только по каталогам или информации вне интернета. С появлением все большего количества сайтов такой способ «навигации» перестал быть удобным, появились более продвинутые системы, позволяющие находить интернет-ресурсы непосредственно по информации, которая в них содержится.

Основные принципы работы поисковых систем строятся на нескольких понятиях: индексация, релевантность и поведенческие факторы, а также вес страницы (PageRank). Процесс индексации представляет собой загрузку страниц различных сайтов и преобразование этой информации в удобный для хранения и поиска вид. На данный момент на серверах поисковой системы Google хранится около 40 миллиардов веб-страниц. Релевантность – это соответствие страницы запросу пользователя, современные же поисковые системы учитывают морфологию языков, исключают незначимые слова (предлоги, междометия) и символы (знаки препинания), дополняют текст наиболее часто используемыми фразами и близки в системе расчета релевантности к языкам программирования четвертого поколения (4GL). Поведенческие факторы — это история фраз пользователя и примерный объем действий на посещенном сайте. Вес страницы – численное выражение авторитетности ресурса.

Владельцы сайтов заинтересованы, чтобы их страницы были как можно выше в результатах поиска. Продвижение сайтов в поисковых системах представляет собой процесс улучшения релевантности страниц сайта и увеличения веса этих страниц. Мониторинг процесса продвижения – получение своевременной информации о позициях различных страниц сайта по различным поисковым запросам, а также информация о числе посетителей, которые перешли на сайт, используя данную поисковую фразу, и совершили какое-либо действие (например, покупку товара или услуги, звонок). В поисковых системах существуют свои средства мониторинга, однако их данные обновляются слишком редко и с запаздыванием.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Объектом исследования в диссертации являются веб-сайты.

Предмет исследования - методы и способы мониторинга процесса продвижения веб-сайтов в поисковых системах.

Задачи исследования были поставлены следующие:

1. Изучить научную литературу по работе поисковых систем, ранжированию веб-сайтов, способам мониторинга позиций веб-сайтов в ПС.
2. Провести теоретическое исследование по проблеме мониторинга процесса продвижения веб-сайтов в поисковых системах.
3. Разработать и испытать программное средство для мониторинга процесса продвижения веб-сайтов.

Целью работы является изучение способов мониторинга позиций веб-сайтов в поисковых системах, разработка для изученных способов программного средства.

Проблема и способы мониторинга процесса продвижения веб-сайтов в поисковых системах были изложены в докладе [1-А.] к 51-ой научной конференции аспирантов, магистрантов и студентов в 2015 году.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе рассматривается литература по теме магистерской диссертации, в частности, по поисковым системам Google и Яндекс, платформе Node.js, языку программирования Javascript.

Разрабатываемое программное обеспечение предназначено для мониторинга процесса продвижения веб-сайтов в поисковых системах путем анализа позиций веб-сайта по различным запросам в поисковых системах Яндекс и Google. Лучшей средой клиентской части является веб-браузер, так как он присутствует на подавляющем числе всевозможных устройств, от персональных компьютеров, до мобильных телефонов и планшетов. Клиентская часть представляет собой одну веб-страницу стандарта HTML5, вся программная часть клиента реализована на языке JavaScript. Платформа Node.js и язык программирования JavaScript, фреймворки Express и Backbone.js являются удобными для разработки программного средства, предназначенного для мониторинга процесса продвижения веб-сайтов в поисковых системах. В качестве базы данных лучшим решением, исходя из языка программирования и задач, является документо-ориентированная СУБД MongoDB. Документо-ориентированные СУБД предназначены для хранения иерархических структур, реализуются с помощью подхода NoSQL. Такой тип организации СУБД был назван так из-за того, что парадигма кардинально отличается от SQL-решений и реляционных баз данных.

Для определения позиций веб-сайта по определенному запросу в определенной поисковой системе для получения самой актуальной информации необходимо отправлять HTTP-запросы на сервер поисковой системы и распознавать позиции сайта. Для этого лучше всего использовать объектную модель документа DOM, т.к. используя ее можно легко получить отдельные элементы структуры HTML-документа без регулярных выражений и поиска подстрок. Для определения числа посетителей, перешедших из поисковых систем по определенным запросам за определенный период необходимо работать с API сервисов статистики. От компании Google это сервис Google Analytics, который предоставляет свой API. У компании Яндекс есть тоже подобный сервис, Яндекс.Метрика, который также позволяет получать данные о веб-сайте через API.

Если каких-либо возможностей языка программирования или платформы недостаточно, всегда можно использовать любые другие языки

программирования для компонентов системы. Node.js позволяет создавать процессы и исполнять Shell-команды.

Во второй главе диссертации описано структурное и функциональное проектирование. Программное средство для мониторинга представляет собой клиент-серверное приложение, следовательно, все модули поделены на модули клиента и сервера.

Клиент-серверная система выбрана для того, чтобы, во-первых, перенести основную нагрузку на сторону сервера, во-вторых, позволять выполнять различные процессы без участия пользователя, в-третьих, дать возможность использовать нескольким пользователям одну инсталляцию приложения и позволить работать с любых устройств (компьютеры, планшеты, мобильные телефоны), поддерживающих технологии клиентской части приложения. Программа поделена на модули для удобной и прозрачной разработки и отладки компонентов системы. Функционал программного средства может быть улучшен путем добавления новых модулей или модернизации существующих. Задача клиентской части приложения – отображение пользовательского интерфейса и связь с сервером приложения. Задача сервера – выполнение различных вычислений, хранение данных. Модуль сервера отвечает в первую очередь за создания сервера, во вторую очередь за пересылку данных клиенту и связь с клиентом остальных модулей серверной части.

Программа написана на языке программирования JavaScript. Среда выполнения серверной части – Node.js, среда выполнения клиентской части – веб-браузер с поддержкой HTML5 и JavaScript. У языка JavaScript несколько другая парадигма, чем у других языков программирования. JavaScript – событийно-ориентированный язык, при этом основные концепции ООП на нем можно реализовать.

В событийно-ориентированном программировании выполнение кода определяется событиями. Эти события включают себя с клиентской стороны действия пользователя: нажатие на кнопки, перемещение указателя мыши, нажатие клавиш, изменение величины экрана браузера. С серверной стороны события включают в себя сообщение сторонних библиотек, таких как библиотека для работы с базой данных, и библиотека создания HTTP-запросов. Код приложения разделяется на две части: выбор определенного события и обработка события. При этом в JavaScript некоторые операции можно делать синхронно, однако основная парадигма языка – ориентация на события.

При проектировании интерфейса необходимо учесть факты, такие как удобство и функциональность. Для удобства пользования программным средством интерфейс поделен на три части. Навигационная панель, отображаемая сверху, содержит ссылки по иерархии интерфейса: главная

страница, проект, запрос. Также содержит ссылку на выход из системы. Рабочее пространство располагается в центре экрана и отображает все рассмотренные ранее представления. Панель состояния расположена снизу и служит для отображения различных служебных сообщений системы, включающих в себя ошибки при совершении каких-либо операций. Все три части пользовательского интерфейса адаптированы под различное разрешение, что позволяет использовать программное средство на устройствах с различным разрешением экрана.

В третьей главе описано тестирование разработанного в рамках диссертации программного средства, описаны методы его работы и пользовательский интерфейс.

В результате ручного тестирования интерфейса было подтверждено, что все вышеописанные функции работают правильно. Помимо тестирования интерфейса и функций программного средства есть и другие виды тестирования. Нагрузочное тестирование можно произвести путем добавления большого множества поисковых фраз и проектов. Программное средство успешно справляется с мониторингом веб-сайта по нескольким сотням поисковых фраз. Увеличение производительности может быть достигнуто путем установки программного средства на более мощном сервере. Успешное юзабилити-тестирование обуславливается единым интерфейсом, выполненном в одном стиле. Интерфейс прост и понятен, успешно работает на устройствах с различным разрешением экрана и манипулятором (в планшетах и мобильных устройствах тачскрин, в компьютерах мышь). Тестирование безопасности произведено только на программном уровне. Пользователь не может управлять чужими проектами, также без авторизации в системе нельзя ничего сделать. За остальные уровни отвечает операционная система и платформа Node.js.

Руководство по использованию программного средства для мониторинга процесса продвижения веб-сайтов в поисковых системах можно поделить на следующие части:

- инструкция по установке сервера программного средства на собственный компьютер или удаленный сервер;
- инструкция использования интерфейса клиентской части программного средства.

Установку сервера программного средства можно поделить на следующие стадии:

- установка платформы Node.js, на которой работает сервер программного средства;
- установка и настройка документо-ориентированной базы данных MongoDB;

- запуск базы данных;
- запуск сервера.

На главной странице программного средства доступна только форма авторизации, первый раз пользователь может ввести любой действующий адрес электронной почты и пароль. Ввиду того, что приложение представляет из себя не простой сайт, а одностраничное JavaScript-приложение, все элементы интерфейса и страницы доступны по одному URL-адресу.

После успешной авторизации пользователь перейдет в систему, где будет доступно добавление проекта (веб-сайта) для мониторинга. Для добавления проекта необходимо указать URL-проекта и имя, имя может быть любым, он отображается пользователю для удобства навигации в проектах. После перехода в проект будет доступна форма добавления поисковых фраз. Также добавить поисковые фразы можно с помощью интерфейса возле колонки фраз каждого проекта. URL проекта ведет непосредственно на сам проект. Для каждой фразы доступны действия по удалению фразы и обновлению какой-либо статистики, интерфейс интуитивно понятен и простой. Для каждой позиции доступна ссылка на график изменения позиции вместе с посещаемости.

Освоить интерфейс среднестатистическому пользователю очень легко. Интерфейс успешно работает на устройствах с различным разрешением экрана.

ЗАКЛЮЧЕНИЕ

В рамках диссертации были изучены способы мониторинга позиций веб-сайтов в поисковых системах, было успешно спроектировано и разработано программное средство для мониторинга процесса продвижения веб-сайтов в поисковых системах. Цель диссертации достигнута, программное средство реализовывает поставленные задачи.

В ходе работы над диссертацией были решены следующие задачи:

- изучены принципы работы поисковых систем, факторы, влияющие на ранжирование веб-сайтов в поисковой системе, способы получения позиций веб-сайта по различным поисковым запросам;

- реализован алгоритм получения позиций веб-сайтов в поисковых системах Google и Яндекс, который в отличие от средств мониторинга, разработанных самими поисковыми системами, получает актуальную информацию;

- реализовано получение статистики посещений веб-сайта с поисковой системы Яндекс с помощью API сервиса Яндекс.Метрика;

- реализовано получение статистики посещений веб-сайта с поисковой системы Google с помощью API сервиса Google Analytics;

- реализовано удобное отображение информации о посещениях и позициях веб-сайта в виде графиков с использованием стандарта HTML5;

- реализована система уведомлений пользователя в случае значимых перепадов статистики веб-сайта.

Программное средство реализовано на удобной кроссплатформенной программной платформе Node.js, клиентская часть программного средства реализовано с использованием HTML и JavaScript, выполняется в среде веб-браузера, что дает возможность использовать программное средство на любых устройствах, поддерживающих соответствующие технологии, таких как персональные компьютеры, современные мобильные телефоны и планшеты. Главное достоинство программного средства по сравнению с аналогами – удобная платформа, которая позволяет как пользоваться одному пользователю одним экземпляром приложения, если на одном компьютере будет расположен и сервер, и клиент, так и нескольким пользователям, когда сервер будет обслуживать различные проекты. В разработанном программном средстве можно расширять функции путем добавления новых поисковых систем, сервисов статистики.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

[1-А.] Назаренко, А.И. Способы мониторинга процесса продвижения веб-сайтов в поисковых системах / А.И. Назаренко // Мониторинг процесса продвижения веб-сайтов в поисковых системах: Тезисы докл. к конф. – Минск, 2015 – С.7-9.

Библиотека БГУИР