

Министерство образования Республики Беларусь

Учреждение образования

Белорусский государственный университет

информатики и радиоэлектроники

УДК 004.775

Соколов

Александр Вячеславович

Специализированная система поиска информации в сети интернет

АВТОРЕФЕРАТ

на соискание степени магистра информатики и вычислительной техники

по специальности 1-40 81 01 «Информатика и технологии разработки программного обеспечения»

Научный руководитель

Новиков Владимир Иванович

кандидат технических наук, доцент

Минск 2015

ВВЕДЕНИЕ

Актуальность темы

Информация – неотъемлемая часть современного мира. В настоящее время информация сравнима с продуктом первой необходимости, т.к. она является двигателем для всех многих сфер деятельности человека, государства и общества в целом. Экспоненциальное увеличение объёма информации поставило общество перед сложной задачей: как информацию получать и как усваивать полученный объём знаний.

Остановимся на первом вопросе. Для решения данной задачи человечество разработало своего рода аккумулятор информации – место, где можно найти абсолютно всё – поисковую систему. Всем известные поисковые системы Google и Yandex уже способны и обрабатывают более 40 млрд. запросов и индексируют порядка 25 млрд. веб-страниц ежемесячно, при этом выполнение обычного поискового запроса занимает всего 0.5 секунд.

В современном мире всё большую популярность начинают приобретать специализированные поисковые системы, направленные на поиск информации в какой-то одной сфере.

Для оптимизации поиска могут использоваться различные инструменты, такие как горизонтальное масштабирование, вертикальное масштабирование, совершенствование архитектуры приложения. Мы рассмотрим один из способов индексации и поиска информации, используя обратный индекс, некоторые алгоритмы и математические модели для ранжирования результатов поиска, а также возможность масштабирования системы.

Краткая история развития объекта исследования

Специализация поисковой системы позволяет добиться больших результатов и сделать поиск ещё более гибким и удобным. Рассмотрим сферу покупки/продажи продуктов и услуг.

Количество ресурсов, предлагающих продукты и услуги, с каждым днём возрастает (Вам известны такие онлайн каталоги как Ebay, Amazon, Тао-Бао), что приводит к невозможности пользователя составить общую картину рынка и найти необходимые услуги и товары в минимальные сроки. Для решения данной проблемы разрабатывается поисковая систем, способная искать продукты и услуги в рамках ресурсов, предоставляющих данную информацию – это могут быть интернет магазины, онлайн каталоги и другие ресурсы,

удовлетворяющие определенным критериям. Предоставление единого онлайн каталога товаров и услуг позволит пользователю составить наиболее точную картину рынка и амплитуду цен, что в свою очередь сократит время поиска и выбор провайдера услуг.

Недостатки объекта исследования

Основными недостатками существующих систем поиска является их множественность и на данный момент невозможность составления общей картины спроса-предложения по тем или иным продуктам. На данный момент на рынке специализированных систем поиска в сфере интернет коммерции можно выделить Yandex Market и Google.

Цель работы

Объект исследования – алгоритмы поиска информации, способы индексирования информации, поисковые системы, специализированные поисковые системы.

Цель работы – изучение принципов реализации поисковых систем, разработка системы поиска информации в интернет для поиска информации, связанной с товарами и услугами.

Методы исследования – сравнительный анализ алгоритмов поиска информации, разработка программного обеспечения.

Цели исследования конкретизируются в следующих задачах:

1. Изучение алгоритмов поиска информации на основе анализа существующих тенденций и алгоритмов поиска.
2. Определение архитектуры системы поиска и построение формализованной модели.
3. Выбор способов проектирования структуры программных систем, информационных систем.
4. Практическая реализация специализированной поисковой системы для агрегирования и поиска продуктов.

СОДЕРЖАНИЕ

В работе рассматриваются существующие алгоритмы и методы поиска, используемые в современных поисковых системах. Произведено рассмотрение архитектуры поисковой системы на примере ПС Google. Выведены требования для реализации собственной ПС, разработана архитектура, а также предоставлен список используемых технологий с описанием их достоинств и недостатков. Дана оценка производительности системы. Определены сферы совершенствования реализованной поисковой системы.

Библиотека БГУИР

ЗАКЛЮЧЕНИЕ

В данной работе изучены алгоритмы и модели поиска, применяющиеся в поисковых системах. Выявлены их достоинства и недостатки. Рассмотрена архитектура поисковой системы Google, её технологические и алгоритмические достоинства.

Произведён сравнительный анализ алгоритмов и моделей поиска информации в интернете. В качестве основных алгоритмов были выбраны векторная модель поиска и использование инвертированных файлов - наиболее зарекомендовавшие себя концепции при построении поисковых систем.

Разработана поисковая система, позволяющая производить агрегирование информации по продуктам, а также быстрый поиск, с возможностью просмотра информации о продукте (описание, рейтинг, магазин и др.).

Произведена оценка производительности, актуальности поисковой системы, а также выработаны идеи для дальнейшего совершенствования:

- обновление архитектуры для поддержки горизонтального масштабирования
- добавление динамических критериев поиска, основываясь на структуре продукта
- возможность кластеризации (разделения продуктов на категории)
- интернационализация поисковых запросов
- возможность поддержки различных типов продуктов в зависимости от требований сайта, на котором будет производиться индексация – различные типы парсеров.
- подключение платёжных систем
- совершенствование производительности системы

СПИСОК ПУБЛИКАЦИЙ

1. Отческий В.И., Соколов А.В. Методическое пособие по работе с Nadoor // Методическое пособие по работе с Nadoor. Задачи. Минск БГУ, 2013, 20с

Библиотека БГУИР