

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра информатики и программирования

**РАЗРАБОТКА И ПРОГРАММНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА
ТАРГЕТИРОВАННОГО ПОИСКА И АНАЛИЗА ДАННЫХ О
ПОЛЬЗОВАТЕЛЯХ СОЦИАЛЬНЫХ СЕТЕЙ
АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

студента 4 курса 441 группы
направления 02.03.03 «Математическое обеспечение и администрирование
информационных систем»
факультета компьютерных наук и информационных технологий
Киселева Вадима Юрьевича

Научный руководитель,
проф. кафедры информатики
и программирования, д. т. н.
должность, уч. степень, уч. звание

подпись, дата

А.С. Фалькович
инициалы, фамилия

Зав.кафедрой информатики
и программирования,
канд. ф.-м. наук
должность, уч. степень, уч. звание

подпись, дата

М.В. Огнева
инициалы, фамилия

Саратов 2018

ВВЕДЕНИЕ

Актуальность темы. С развитием сети Интернет люди получили массу возможностей ведения бизнеса, установления коммуникаций с географически недостижимыми людьми за доли секунды, получения знаний из всевозможных источников и проведения статистических социальных экспериментов и опросов.

Информации становится настолько много, что люди вынуждены изобретать новые математические алгоритмы, новые подходы разработки программного обеспечения, чтобы иметь возможность обрабатывать подобный поток данных, находить релевантную для поискового запроса информацию и визуализировать статистические данные, используемые в социологии.

Помимо обработки больших потоков информации, человечество сталкивается с новыми задачами, которые ведут к новым возможностям, но сильно усложняются в технической реализации. Бизнесмены сталкиваются с проблемами усложнения условий таргетирования (релевантной фильтрации) покупателей, статистики и социологи - с проблемой возможности удобного получения больших объемов данных, отфильтрованных по специально описанным заранее критериям.

Но есть в проблемах поиска, анализа, обмена знаниями и продажах кое-что общее, это люди. И развитие Интернета помогает решить большую часть таких задач, Интернет позволяет найти людей, данные которых в невероятных объемах располагаются в социальных сетях.

Цель бакалаврской работы – разработка абстрактной масштабируемой и легко конфигурируемой поисковой платформы для таргетированного поиска людей в любой социальной сети, которая располагает публичным интерфейсом доступа для манипулирования внутренними данными о сущностях в сети, и механизма удобно визуализированного представления результата поисковых запросов для последующего анализа или проведения социологических и статистических экспериментов.

Поставленная цель определила **следующие задачи**:

1. Исследовать основные характеристики поисковых механизмов, с целью выделения признаков построения гибкого поиска
2. Изучить технологии удовлетворения поисковых потребностей, хранения, индексации и визуализации данных.
3. Изучить особенности языка программирования Java 8, применимые для реализации поискового механизма
4. Изучить особенности и недостатки существующих механизмов поиска, таргетирования и анализа данных в социальных сетях и поисковых системах
5. Разработать поисковую платформу, позволяющую взаимодействовать с абстрактной социальной сетью, с целью предоставить модернизированный функционал поиска и анализа пользователей.
6. Применить программу для решения набора прикладных задач, не осуществимых с помощью существующих технологий

Теоретическая значимость бакалаврской работы заключается в том, что тема исследования находится на стыке нескольких областей знаний: информатики, программирования, социологии, маркетинга и статистики.

Практическая значимость бакалаврской работы. В ходе выполнения практической части бакалаврской работы было разработано web-приложение «Social search». Достоинством разработанного web-приложения является то, что оно предоставляет абстрактный и расширенный функционал поискового механизма, независимый от конкретной социальной сети.

Структура и объём работы. Бакалаврская работа состоит из введения, четырех разделов, заключения, списка использованных источников и одного приложения. Общий объем работы – 63 страницы, из них 57 страниц – основное содержание, включая 23 рисунка, цифровой носитель в качестве приложения, список использованных источников информации – 23 наименования.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «ТЕОРИЯ ПОИСКА» посвящен описанию теоретических основ и характеристик поискового механизма, из которых выделяются наиболее релевантные признаки построение гибкого поиска, а также описанию теории поиска, разделяющего поисковую функциональность на две группы, концепты и контракты.

«Гибкий поиск» – в этом подразделе перечисляются характеристики поиска, которые описывают спектр параметров поиска от ресурсозатратности до степени конфигурирования запроса и результативных данных.

«Концепты» – в этом подразделе перечисляются теоретические критерии, которым обязан соответствовать поисковой механизм, построенный на данной теории.

«Контракты» – в этом подразделе описывается набор динамических критериев, описывающих как типы данных, с которыми будет оперировать поисковой механизм, так и переходы между данными, то есть переходы, описывающие отношения типов данных.

Второй раздел «АНАЛИЗ И СТАТИСТИКА» посвящен описанию необходимых компонент для организации системы анализа и статистики данных, предоставляемых источником данных в виде сервиса или социальной сети.

«Хранение» – в этом подразделе описывается необходимость обеспечения доступности и защищенности данных для осуществления возможности работы аналитической платформы в любой момент, независимо от работы поисковой платформы.

«Индексация» – в этом подразделе описывается необходимость реализации полнотекстового индексирующего движка между этапами хранения

и визуализации для обеспечения высокой скорости доступа и поиска релевантной выборки данных.

«Визуализация» – в этом подразделе описан механизм визуализации, получающий индексируемые данные из полнотекстового движка, как из источника данных.

Третий раздел «ПРОГРАММНАЯ РЕАЛИЗАЦИЯ» посвящен описанию реализации поисковой платформы и сервиса анализа и визуализации данных и практическому применению с примером использования программы.

«Требования к системе» – в этом подразделе перечислены функциональные особенности, которым должна удовлетворять разрабатываемая платформа.

«Методологии и парадигмы» – в этом подразделе перечислены теоретические основы и подходы, которым будет следовать разрабатываемая программа. [1-7]

«Языки программирования» – в этом подразделе перечислены выбранные языки программирования для реализации приложения. [8-10]

«Внешняя архитектура» – в этом подразделе проводится анализ и решение о выборе компонент и инструментов, которые будут наиболее подходящими для решения заявленных задач, без описания внутреннего представления сервисов. [11-14]

«Внутренняя архитектура» – в этом подразделе раскрывается архитектура поисковой платформы без описания моментов взаимодействия с внешними системами. [15-16]

«Способы взаимодействия» – в этом подразделе описаны функциональные способы использования разрабатываемой программы в зависимости от типа решаемых задач.

«Кэширование» – в этом подразделе описаны способы оптимизации работы с поисковым механизмом, а именно с повторяемыми запросами. [17]

«Сериализация» – в этом подразделе перечислены возможности приложения предоставлять результативные данные в разных форматах, в зависимости от пожелания пользователя. [18-20]

«Практическое использование» – в этом подразделе рассмотрено пошаговое использование разработанного приложения от взаимодействия с поисковым механизмом через веб интерфейс, до построения статистик и графиков в инструменте визуализации.

Четвертый раздел «ПРИКЛАДНОЕ ПРИМЕНЕНИЕ» посвящен описанию и сравнению задач, решаемых с помощью существующих механизмов и реализованного в рамках практической части приложения.

«Социальные сети» – в этом подразделе описаны определения социальной сети как в обществе, так и в Интернете. [21]

«Маркетинг» – в этом подразделе рассмотрен пример решения задачи маркетингового таргетирования с помощью существующих механизмов и предложенного в работе.

«Социология» – в этом подразделе рассмотрен пример задачи социологического исследования с помощью сервиса опросов и предложенного в работе. [22]

«Статистика» – в этом подразделе описаны предлагаемые разработанным приложением возможности для упрощения и функционального обогащения для решения статистических задач. [23]

«Знакомства» – в этом подразделе описаны предлагаемые разработанным приложением возможности облегчения решения задач поиска конкретных людей для коммуникации или рекрутинга.

«Прикладные задачи» – в этом подразделе представлены прикладные примеры задач из описанных ранее сфера, в решение которых разработанная программа приносит большой вклад.

ЗАКЛЮЧЕНИЕ

В результате работы была разработана поисковая платформа «Social search», предоставляющая возможность осуществлять гибкий поиск в социальных сетях, на языке программирования Java. Абстракция над источником данных позволяет платформе взаимодействовать в любой социальной сети, модуль общения с которой добавлен прямо в исходный код приложения. Произведя многоуровневый повторяемый и фильтруемый поиск, программа позволяет получить результативные данные в разных форматах, а также загружает выходную выборку в визуализационный инструмент, который обеспечивает возможность гибко строить графики и статистики, основанные на свойствах полученных данных.

Анализ задач, основанных на осуществлении поиска персональной информации в Интернете, показал актуальность разработанной поисковой платформы, которая позволит усложнить и модернизировать механизмы удовлетворения задач маркетинга и социологии.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ И ЛИТЕРАТУРЫ

1. Грэди Буч. Объектно-ориентированный анализ и проектирование с примерами приложений на C++ = Object-Oriented Analysis and Design with Applications / Пер. И.Романовский, Ф.Андреев. — 2-е изд. — М., СПб.: «Бином», «Невский диалект», 1998. — 560 с. — 6000 экз. — ISBN 5-7989-0067-3.
2. Основные принципы ООП [Электронный ресурс]. — URL: http://vschol.ru/TurboPascal/gl10/gl10_1.html (дата обращения: 18.03.2018).
3. The DCI Architecture: A New Vision of Object-Oriented Programming [Электронный ресурс]. — URL: https://www.artima.com/articles/dci_vision.html (дата обращения: 18.03.2018).
4. Model – view – controller [Электронный ресурс]. — URL: <http://rstdn.org/article/patterns/generic-mvc.xml> — 2007. (дата обращения: 18.03.2018).
5. Машнин Тимур Сергеевич. Технология Web-сервисов платформы Java. — БХВ-Петербург, 2012. — С. 115. — 560 с. — ISBN 978-5-9775-0778-3.
6. Distributed Application Architecture [Электронный ресурс]. — URL: <https://web.archive.org/web/20110406121920/http://java.sun.com/developer/Books/jdbc/ch07.pdf> (дата обращения: 01.04.2018).
7. Fluent Interface [Электронный ресурс]. — URL: <https://martinfowler.com/bliki/FluentInterface.html> (дата обращения: 01.04.2018).
8. Что такое технология Java и каково ее применение? [Электронный ресурс]. — URL: https://www.java.com/ru/download/faq/whatis_java.xml (дата обращения: 01.04.2018).
9. Stream API [Электронный ресурс]. — URL: <https://docs.oracle.com/javase/8/docs/api/java/util/stream/package-summary.html>(дата обращения: 01.04.2018).
10. Флэнаган Д. JavaScript. Карманный справочник. Сделайте веб-страницы интерактивными! / Перевод А.Г. Сысонюк. — Москва.: Издательский дом "Вильямс", 2015. — С. 320. — 1000 экз. — ISBN 978-5-8459-1948-9.

11. Embedding Jetty [Электронный ресурс]. – URL: <http://www.eclipse.org/jetty/documentation/9.3.x/embedding-jetty.html>(дата обращения: 28.04.2018).
12. The Best Way to Run MongoDB in Your Data Center. Supported. Secure. Automated. [Электронный ресурс]. – URL: <https://www.mongodb.com/products/mongodb-enterprise-advanced> (дата обращения: 29.04.2018).
13. PostgreSQL [Электронный ресурс]. – URL: <https://www.postgresql.org> (дата обращения: 29.04.2018).
14. Presentation and use of H2 Database Engine [Электронный ресурс]. – URL: <https://baptistewicht.com/posts/2010/08/presentation-usage-h2-database-engine.html> (дата обращения: 29.04.2018).
15. Maven [Электронный ресурс]. – URL: <http://www.apache-maven.ru/> (дата обращения: 05.05.2018).
16. Spring Framework [Электронный ресурс]. – URL: <http://spring-projects.ru/projects/spring-framework/> (дата обращения: 05.05.2018).
17. Spring Cache [Электронный ресурс]. – URL: <https://docs.spring.io/spring-boot/docs/current/reference/html/boot-features-caching.html> (дата обращения: 05.05.2018).
18. Сериализация [Электронный ресурс]. – URL: <https://docs.microsoft.com/ru-ru/dotnet/csharp/programming-guide/concepts/serialization/> (дата обращения: 05.05.2018).
19. JSON [Электронный ресурс]. – URL: <https://www.json.org/json-ru.html> (дата обращения: 06.05.2018).
20. How to extract information from Office files by using Office file formats and schemas [Электронный ресурс]. – URL: <https://support.microsoft.com/kb/840817/> (дата обращения: 06.05.2018).

21. Wasserman, Stanley; Faust, Katherine (1994). "Social Network Analysis in the Social and Behavioral Sciences". *Social Network Analysis: Methods and Applications*. Cambridge University Press. pp. 1–27. ISBN 9780521387071.
22. Опрос социологический // Социология. Словарь / Сост. Т. Е. Зерчанинова. — Екатеринбург: УрАГС, 2006. — 64 с.
23. Малая советская энциклопедия. — М.: Советская энциклопедия, 1960. — Т. 8. — С. 1090.