

Министерство образования и науки РФ  
Федеральное государственное автономное образовательное учреждение  
высшего профессионального образования  
**«Национальный исследовательский ядерный университет «МИФИ»**  
**Кафедра №12 «Компьютерные системы и технологии»**

## **ТЕХНИЧЕСКОЕ ЗАДАНИЕ**

**«Разработка локального поискового движка с ранжированием результатов»**

Составили:

Муров Вадим Б19-503

Харке Денис Б19-503

Согласовано:

Овчаренко Евгений Сергеевич

## Оглавление

|  |          |
|--|----------|
| <b>1. Введение .....</b>   | <b>3</b> |
| <b>1.1 Краткая характеристика области применения системы .....</b>       | <b>3</b> |
| <b>2. Назначение разработки системы .....</b>                            | <b>3</b> |
| <b>2.1 Существующие решения .....</b>                                    | <b>3</b> |
| <b>3. Требования к системе .....</b>                                     | <b>4</b> |
| <b>3.1 Требования к функциям сервиса .....</b>                           | <b>4</b> |
| <b>3.2 Требования к составу и параметрам технических средств .....</b>   | <b>4</b> |
| <b>3.3 Требования к информационной и программной совместимости .....</b> | <b>4</b> |
| <b>4. Техничко-экономические показатели .....</b>                        | <b>4</b> |

## **1. Введение**

- Полное наименование системы: «Локальный поисковый движок с ранжированием результатов»
- Условное обозначение: система
- Плановые сроки начала работы по созданию системы: 01.10.2022
- Плановые сроки окончания работы по созданию системы: 01.12.2022

### **1.1 Краткая характеристика области применения системы**

Система предназначена для поиска информации по какой-либо части всемирной сети (нескольким сайтам). Такого рода поисковые машины обычно используются внутри частных (корпоративных) сетей или в системах для электронной коммерции (интернет-магазинах). Однако область применения этим не ограничивается.

## **2. Назначение разработки системы**

- Назначение системы: Поиск необходимой информации одновременно на одном или несколькими сайтами с ранжированием результатов.
- Цель создания системы: Получение ранжированного списка ссылок по заданному запросу.
- Поисковой движок обеспечивает поиск информации на нескольких сайтах одновременно. Система может быть использована корпорациями, для найма сотрудников с нескольких сайтов. Веб-ресурсы могут использовать поисковой движок для навигации и получения нужных товаров из базы.

### **2.1 Существующие решения**

На сайте Statcounter – сервиса веб-аналитики – есть данные о рынке глобальных поисковых систем в России. Самый популярный поисковик – Google, в ноябре 2021 года им пользуются 55,29% всех пользователей. Яндексом – 42,11%. В то же время, в такого рода поисковиках нет возможности искать информацию только по выбранным заранее сайтам.

Для решения данной задачи существуют локальные поисковые машины: Elasticsearch, Solr или Sphinx. Данные веб-приложения являются поисковыми решениями с открытым исходным кодом, с хорошей поддержкой от собственных сообществ участников. Эти решения отличает высокая производительность, масштабируемость и гибкость, но каждая из них все же имеет свои особенности, которые могут повлиять на проект. Тем не менее, не существует отечественных локальных поисковых машин для работы с русским языком. Также большинство существующих решений больше не доступно для использования на территории Российской Федерации.

### **3. Требования к системе**

#### **3.1 Требования к сервису**

##### **3.1.1 Требование к функциональности сервиса**

- *Поиск ссылок*: Поиск ссылок на страницы и другие документы сайтов;
- *Рекурсивный поиск*: Поисковая машина ищет ссылки со страниц сайтов;
- *Ручной режим*: Пользователи сами добавляют в базу данных поисковой машины ссылки на страницы своих сайтов;
- *Индексация документов сайтов*: Извлечение из документов информации, важной для поиска, преобразование этой информации в формат, удобный для поисковой машины и сохранение этой информации в базу данных поисковой машины;
- *Поиск по запросу*: Поиск по базе данных проиндексированных документов;
- *Релевантность результатов*: Ранжирование документов в соответствии с их релевантностью поисковым запросам;
- *Регистрация и авторизация*: Регистрация, авторизация, а также управление аккаунтом: изменение пользовательских данных (почты, пароля).

##### **3.1.2 Требование к визуальной составляющей сервиса**

- *Сбор статистики*: Возможность сбора статистики проиндексированных сайтов, а именно: затраченное время, количество обнаруженных страниц и лемм;
- *Управление ссылками*: Возможность добавлять/удалять ссылки
- *Поиск лемм*: Страница, позволяющая пользователю искать информацию среди выбранных проиндексированных сайтов.
- *Управление аккаунтом*: Опция, позволяющая пользователю изменять свои пользовательские данные

#### **3.2 Требования к составу и параметрам технических средств**

Для работы системы необходим арендованный сервер для хранения данных.

#### **3.3 Требования к информационной и программной совместимости**

Поисковый движок должен представлять из себя Spring-приложение (веб-ресурс), работающее с базой данных MySQL(PostgreSQL), имеющее простой веб-интерфейс и API, через который им можно управлять и получать результаты поисковой выдачи по запросу.

- Язык программирования бэкенд составляющей: Java 8 (Spring, hibernate)
- Язык программирования фронтенд составляющей: JavaScript
- База данных: MySQL
- Хостинг: МИФИ

### **4. Техничко-экономические показатели**

- Расходы на создание и размещение сайта - при условии предоставления учебной организацией НИЯУ МИФИ серверного оборудования, дополнительные расходы на реализацию проекта не потребуются.
- Предполагаемая месячная потребность продукта исходя из анализа поисковых запросов – 5000 (проанализированы ключевые фразы: «поисковая машина» и «поисковый движок»)

- Зарубежные аналоги: Elasticsearch, Solr или Sphinx
- Отечественных аналогов нет.