ИТОГОВЫЙ ПРОЕКТ

На тему: Реализация системы поиска по корпоративному порталу компании «АйТиБокс»

Москва 2024

Содержание

[Введение 3](#_Toc156684318)

[Этап 1 Создание плана разработки проекта 4](#_Toc156684319)

[1.1 Цели этапов разработки 4](#_Toc156684320)

[1.2 Задачи этапа 5](#_Toc156684321)

[1.3 Календарь 5](#_Toc156684322)

[1.4 Оценка себестоимости проекта 6](#_Toc156684323)

[Этап 2 Изучение принципов работы поискового движка 7](#_Toc156684324)

[2.1 Задачи этапа 7](#_Toc156684325)

[Этап 3 Подготовка ПО 8](#_Toc156684326)

[3.1 Задачи этапа 8](#_Toc156684327)

[Этап 4 Разработка класса для взаимодействия с json-файлами 9](#_Toc156684328)

[4.1 Задачи этапа 9](#_Toc156684329)

[Этап 5 Подключение системы тестирования 10](#_Toc156684330)

[5.1 Задачи этапа 10](#_Toc156684331)

[Этап 6 Создание инвертированного индекса для документов 11](#_Toc156684332)

[6.1 Задачи этапа 11](#_Toc156684333)

[Этап 7 Разработка системы индексации документов 12](#_Toc156684334)

[7.1 Задачи этапа 12](#_Toc156684335)

[Этап 8 Размещение проекта в GitHub 13](#_Toc156684336)

[8.1 Задачи этапа 13](#_Toc156684337)

[Этап 9 Подготовка к сдаче проекта 14](#_Toc156684338)

[9.1 Задачи этапа 14](#_Toc156684339)

[Заключение 15](#_Toc156684340)

[Список используемой литературы 16](#_Toc156684341)

Введение

**Цель итогового проекта:** разработка поискового движка для корпоративного портала компании – аналог поисковой строки Яндекса или Google.

Для выполнения цели итогового проекта необходимо выполнить следующие **задачи**:

1. Создать план разработки проекта;
2. Изучить принципы работы поискового движка;
3. Подготовить необходимое программное обеспечение для программирования поискового движка;
4. Разработать класс для взаимодействия с файлами JSON в описанном формате;
5. Подключить систему тестирования к проекту;
6. Создать инвертированный индекс для документов;
7. Реализовать систему определения релевантности поискового запроса;
8. Разместить исходные данные проекта в удаленном публичном репозитории GitHub;
9. Подготовить проект к сдаче на проверку куратору.
10. Создание плана разработки проекта
    1. Цели этапов разработки

В Таблице Таблица 1 представлены цели каждого из этапов разработки.

Таблица 1 – Цели этапов разработки

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Этап** | **Цель** |
| 1 | Создание плана разработки проекта | Разработка данного документа |
| 2 | Изучение принципов работы поискового движка | Изучить основные принципы работы разрабатываемого и готовых решений поисковых движков |
| 3 | Подготовка ПО | Подготовить необходимое ПО для программирования поискового движка |
| 4 | Разработка класса для взаимодействия с json-файлами | Добавить в созданный проект search\_engine классы ConverterJSON для работы с файлами формата JSON |
| 5 | Подключение системы тестирования | Подключить к проекту search\_engine библиотеку Google Test для проведения автоматических модульных тестов |
| 6 | Создание инвертированного индекса для документов | Реализовать инвертированную индексацию документов |
| 7 | Разработка системы индексации документов | Реализовать систему определения релевантности поискового запроса |
| 8 | Размещение проекта в GitHub | Научиться размещать проекты в публичном доступе для демонстрации при презентации проектов и при общении с потенциальными работодателями |
| 9 | Подготовка к сдаче проекта | Подготовить проект к сдаче на проверку куратору |

* 1. Задачи этапа

Для выполнения цели этапа необходимо выполнить следующие задачи:

* разработать календарный план, согласно которому будет вестись разработка проекта, и обозначить этапы разработки;
* определить цели каждого этапа;
* оценить себестоимость проекта.
  1. Календарь

В Таблице Таблица 2 представлен календарный план по разработке итогового проекта.

Таблица – Календарный план по разработке проекта

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№ п/п** | **Название этапа** | **Начало** | **Конец** | **Дней** |
| 1 | Создание плана разработки проекта | 20.01.2024 | 26.01.2024 | 7 |
| 2 | Изучение принципов работы поискового движка | 27.01.2024 | 28.01.2024 | 2 |
| 3 | Подготовка ПО | 29.01.2024 | 31.01.2024 | 3 |
| 4 | Разработка класса для взаимодействия с json-файлами | 01.02.2024 | 10.02.2024 | 10 |
| 5 | Подключение системы тестирования | 11.02.2024 | 20.02.2024 | 10 |
| 6 | Создание инвертированного индекса для документов | 21.02.2024 | 01.03.2024 | 10 |
| 7 | Разработка системы индексации документов | 02.03.2024 | 11.03.2024 | 10 |
| 8 | Размещение проекта в GitHub | 12.03.2024 | 14.03.2024 | 3 |
| 9 | Подготовка к сдаче проекта | 15.03.2024 | 19.03.2024 | 5 |
| Итого | | | | 60 |

* 1. Оценка себестоимости проекта

1. Изучение принципов работы поискового движка
   1. Задачи этапа

Для выполнения цели этапа необходимо выполнить следующие задачи:

* изучить принципы работы поискового движка, представленные в техническом задании;
* провести анализ принципов работы поисковых движков в глобальной сети Интернет.
  1. Принципы поисковых движков на примере из глобальной сети Интернет

TBD

* 1. Принципы поискового движка проекта

Принципы работы поискового движка разрабатываемого проекта следующие:

1. В конфигурационном файле перед запуском приложения задаются названия файлов, по которым движок будет осуществлять поиск.
2. Поисковый движок должен самостоятельно обходить все файлы и индексировать их так, чтобы потом по любому поисковому запросу находить наиболее релевантные документы.
3. Пользователь задаёт запрос через JSON-файл requests.json. Запрос — это набор слов, по которым нужно найти документы.
4. Запрос трансформируется в список слов.
5. В индексе ищутся те документы, на которых встречаются все эти слова.
6. Результаты поиска ранжируются, сортируются и отдаются пользователю, максимальное количество возможных документов в ответе задаётся в конфигурационном файле.
7. В конце программа формирует файл answers.json, в который записывает результаты поиска.
8. Подготовка ПО
   1. Задачи этапа

Для выполнения цели этапа необходимо выполнить следующие задачи:

* утвердить список ОС, в которых проект должен функционировать и подготовить устройства с данными ОС или виртуальное окружение, позволяющее смоделировать данные ОС;
* скачать и установить компилятор C++ и утвердить его версию;
* скачать и установить интегрированную среду разработки, утвердить его версию, а также перечень плагинов (при необходимости), используемых при разработке;
* скачать и установить систему контроля версий Git и утвердить его версию;
* создать публичный репозиторий на GitHub и слинковать удаленный репозиторий GitHub с локальным репозиторием;
* определить директории и файлы в репозитории, необходимые для корректного процесса контроля версий и разработки ПО;
* добавить стороннее ПО в исходный код проекта.
  1. Характеристики ПО

В Таблице Таблица 3 представлены характеристики разрабатываемого ПО.

Таблица – Характеристики ПО

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Параметр** | **Значение** |
| 1 | Целевые ОС | Windows 10, Windows 11, MacOS TBD (?) |
| 2 | Компилятор | g++ (x86\_64-posix-seh-rev1, Built by MinGW-Builds project) 13.1.0 |
| 3 | IDE | Visual Studio Code версии 1.85.2 |
| 4 | СКВ | Git версии 2.42.0.windows.2 |
| 5 | Репозиторий GitHub | <https://github.com/Sarnavskiy-Alexey/Developer_CPP_Diploma.git> |
| 6 | C/C++ Extension Pack | Идентификатор: ms-vscode.cpptools-extension-pack  Версия: 1.3.0  Ссылка на Visual Studio Marketplace: <https://marketplace.visualstudio.com/items?itemName=ms-vscode.cpptools-extension-pack> |
| 7 | CMake Tools | Идентификатор: ms-vscode.cmake-tools  Версия: 1.16.32  Ссылка на Visual Studio Marketplace: <https://marketplace.visualstudio.com/items?itemName=ms-vscode.cmake-tools> |
| 8 | Библиотека nlohmann\_json | Версия 3.11.2 |
| 9 | Google Test |  |

* 1. Состав репозитория

Дерево репозитория представлено на Рисунке Рисунок 1.

Рисунок – Дерево репозитория

1. Разработка класса для взаимодействия с json-файлами
   1. Задачи этапа

Для выполнения цели этапа необходимо выполнить следующие задачи:

* создать класс для работы с данными в формате JSON. Класс выполняет следующие функции:
  + считывает конфигурационные данные из JSON,
  + преобразовывает запросы в формате JSON,
  + формирует ответы в заданном формате JSON.
  1. Описание файла конфигурации

TBD

* 1. Описание файла запросов

TBD

* 1. Описание файла ответов

TBD

1. Подключение системы тестирования
   1. Задачи этапа

Для выполнения цели этапа необходимо выполнить следующие задачи:

* подключить Google Test к проекту в CMakeLists.txt.
  1. Описание подключения GTest

TBD

1. Создание инвертированного индекса для документов
   1. Задачи этапа

Для выполнения цели этапа необходимо выполнить следующие задачи:

* создать класс InvertedIndex, который будет хранить и индексировать слова.
  1. Описание класса InvertedIndex

TBD

1. Разработка системы индексации документов
   1. Задачи этапа

Для выполнения цели этапа необходимо выполнить следующие задачи:

* создать класс SearchServer для определения наиболее релевантных документов.
  1. Алгоритм выдачи результатов поиска

1. Размещение проекта в GitHub
   1. Задачи этапа

Для выполнения цели этапа необходимо выполнить следующие задачи:

* дополнить README.md описанием проекта и стеком используемых технологий;
* разместить проект в удаленном репозитории GitHub и запустить размещенный проект на другом устройстве.

1. Подготовка к сдаче проекта
   1. Задачи этапа

Для выполнения цели этапа необходимо выполнить следующие задачи:

* проверить проект по чек-листу и убедиться, что все пункты чек-листа выполнены.
  1. Чек-лист для проверки

TBD

Список используемой литературы

1. Техническое задание на разработку итогового проекта. – Москва, ЧУО ДПО «Образовательные технологии «Скиллбокс (Коробка навыков)». – 19 с.