

SkillBox (коробка навыков)

Search Engine

Разработка локального поискового
движка по файлам

Выполнил: Катков Денис

Итоговый проект в курсе
«Профессия Разработчик C++ с нуля» на
образовательной платформе «SkillBox»

Июнь 2022 г.



Содержание презентации

1

О себе

2

Постановка задачи

3

Состав программы

4

Применённые библиотеки

5

Заключение



О себе:

Возраст: 41 год

Образование: магистр радиопизик

Род деятельности: инженер-электроник, разработчик РЭА

Цель обучения: Смена рода деятельности, упорядочить, углубить, расширить имеющиеся знания в области программирования C++

Хобби: фотография, разработка и ремонт РЭА, строительство, садоводство



Постановка задачи

- Поисковый движок должен представлять консольное приложение
- Конфигурирование должно осуществляться посредством JSON-файла config.json
- Приложение должно обходить все файлы, индексировать и на основании индексов и запросов производить поиск наиболее подходящих документов
- Запросы пользователя задаются через JSON-файл requests.json
- Результаты поиска возвращаются в JSON-файл answers.json
- Должны быть разработаны тесты проверяющие правильность функционирования методов программы



Состав программы

Структурно программа построена из следующих классов:

ConverterJSON

Класс работы с JSON-файлами. Обеспечивает чтение конфигурационных данных, преобразует запросы в формате JSON, формирует ответы в заданном формате JSON

InvertedIndex

Класс реализует чтение заданных в конфигурации файлов, реализует инвертированную индексацию документов. Система индексации многопоточная. Каждый файл обрабатывается в отдельном потоке.

SearchServer

Класс обеспечивает поиск документов согласно заданным пользователем запросам. В процессе поиска рассчитывается релевантность каждого документа данному запросу. В классе реализовано три типа поиска:

***Полного
совпадения***

***Частичного
совпадения***

Best Match 25



Применённые инструменты

STL

Использованы контейнеры **map**, **set**, **vector** для хранения и упорядочивания элементов словарей при индексации

nlohmann/JSON

Библиотека для манипуляций с JSON-файлами. Чтение/запись, сериализация/десериализация данных.

bsoshany/thread-pool

Библиотека для организации пула потоков

CMake, GCC

Инструменты компиляции, сборки отладки программного продукта

GoogleTest

Написание и проведение тестов методов и функций.

VS Code

Удобный в работе кроссплатформенный редактор



Заключение

- Разработано консольное приложение поиска документов
- Опробована и внедрена идея конфигурирования с помощью JSON-файлов
- Реализовано несколько методов вычисления релевантности документа
- Углублены знание и понимание структуры CMake-проекта
- Проведено автоматизированное тестирование полученного продукта

Как это использовать

- Предполагаю использовать полученные результаты при разработке программы с графическим интерфейсом на базе Qt для оформления рентгенологических заключений с автоматическим поиском шаблона и вставкой в текущий документ.



Контактная и прочая информация



Катков Денис



fotmast@ya.ru



@DikoStone



8-918-548-54-53



<https://github.com/DikoSt>

