

**Московский авиационный институт
(национальный исследовательский университет)**

**Институт информационные технологии и прикладной
математики**

Кафедра вычислительной математики и программирования

**Лабораторная работа №4 по курсу
«Информационный поиск»**

Студент:	Е.М. Стифеев
Преподаватель:	А.А. Кухтичев
Группа:	М8О-109М-21
Дата:	10.11.21
Оценка:	
Подпись:	

Москва, 2021

Лабораторная работа №4 «Булев поиск»

Нужно реализовать ввод поисковых запросов и их выполнение над индексом, получение поисковой выдачи.

Синтаксис поисковых запросов:

- Пробел или два амперсанда, «&&», соответствуют логической операции «И».
- Две вертикальных «палочки», «||» – логическая операция «ИЛИ»
- Восклицательный знак, «!» – логическая операция «НЕТ»
- Могут использоваться скобки.

Парсер поисковых запросов должен быть устойчив к переменному числу пробелов, максимально толерантен к введённому поисковому запросу.

Примеры запросов:

- [московский авиационный институт];
- [(красный || желтый) автомобиль];
- [руки !ноги].

Для демонстрации работы поисковой системы должен быть реализован веб-сервис, реализующий базовую функциональность поиска из двух страниц:

- Начальная страница с формой ввода поискового запроса.
- Страница поисковой выдачи, содержащая в себе форму ввода поискового запроса, 50 результатов поиска в виде текстов заголовков документов и ссылок на эти документы, а так же ссылку на получение следующих 50 результатов.

Так же должна быть реализована утилита командной строки, загружающая индекс и выполняющая поиск по нему для каждого запроса на отдельной строчке входного файла.

В отчёте должно быть отмечено:

- Скорость выполнения поисковых запросов.
- Примеры сложных поисковых запросов, вызывающих длительную работу.
- Каким образом тестировалась корректность поисковой выдачи.

1. Описание

Корпус и индекс

Напомню, что корпус документов имеет следующую структуру, полученную по результатам ЛР1 (доступен по ссылке <https://cloud.mail.ru/public/ZfkX/gccM7hnDR>):

- Корпус документов
 - films1.txt (94 Мб, 15000 документов)
 - films2.txt (96 Мб, 15000 документов)
 - films3.txt (184 Мб, 15000 документов)
 - films4.txt (219 Мб, 15000 документов)
 - films5.txt (322 Мб, 15000 документов)
 - films6.txt (711 Мб, 15000 документов)
 - films7.txt (823 Мб, 15000 документов)
 - films8.txt (226 Мб, 15000 документов)
 - films9.txt (67 Мб, 15000 документов)
 - films10.txt (75 Мб, 15000 документов)
 - films11.txt (99 Мб, 15000 документов)
 - films12.txt (78 Мб, 15000 документов)
 - films13.txt (41 Мб, 6109 документов)

$$\Sigma_{Gb} = 2,899 \text{ Gb}, \Sigma_{docs} = 186109$$

Также, напомню, что получение одного документа могло включать проход по нескольким html-страницам и обработку динамически подгружаемых страниц, поэтому общее количество обработанных страниц было >800'000.

В каждом файле *.txt документы хранятся следующим образом:

- 1 строка 1 документ {...}
- 2 строка 2 документ {...}
- n строка n документ {...}

Каждый документ снабжён прямой ссылкой на источник, откуда был скачен, и хранит только выделенный из html-кода текст в кодировке UTF-8. Например, 234 строка файла films1.txt выглядит так:

```
{"page_url": "https://www.kinopoisk.ru/media/article/1773537/", "title": "Артур Смольяников: «Я сомневался, что смогу сыграть ангела»", "body": "2 января в российский прокат вышла романтическая комедия Веры Сторожевой „Мой
```

парень — ангел“, главные роли в которой исполнили Артур Смольянинов и Анна Старшенбаум. Мы подготовили небольшой видеосюжет с участием создателей картины...Студентка Саша с большим трудом верит в чудеса. Ангелу Серафиму приходится приложить немало усилий, чтобы доказать ей, что ангелы существуют. Но он не учел одного: если девушка тебе поверит, она, скорее всего, тебя полюбит.\n\n\n\n\n\n\n\nАвтор: Дарико Цулая", "comments": ""}.

Готовый индекс хранится в трёх файлах:

- **docs_id.data** (35 Мб)

Файл служит для отображения индекса документа (doc_id) в его текстовое представление в файлах *.txt. Поддерживается переменная длина пути до файлов с документами.

- **terms.data** (81 Мб)

Файл служит для хранения словаря с терминами и ссылок на файл с словопозициями. Поддерживается переменная длина термина. Термины упорядочены в лексикографическом порядке.

- **postings_list.data** (889 Мб)

Файл служит для хранения словопозиций и частот терминов в документе. Словопозиции упорядочены по возрастанию идентификаторов документов.

Алгоритм поиска

Скажу сначала пару слов о парсинге запроса из линейной строчной формы в форму, подходящую для обработки. Для начала входящая строка трансформируется в инфиксный запрос в виде последовательности операторов и операндов:

(самый лучший фильм) || (интересный сериал || !документалка)

->

(самый & лучший & фильм) | (интересный & сериал | ! документалка)

Далее, преобразуем в постфиксный запрос с помощью стандартного алгоритма Дейкстры:

самый лучший фильм & & интересный сериал & документалка ! | |

Полученное выражение можно вычислять с помощью стека.

Основные операции со словопозициями запрограммированы согласно [1].

Приведу код операции пересечения словопозиций:

```
void intersect_postings(const vector<int> &p1, const vector<int> &p2,
                      vector<int> &answer)
{
    int i=0, j=0;

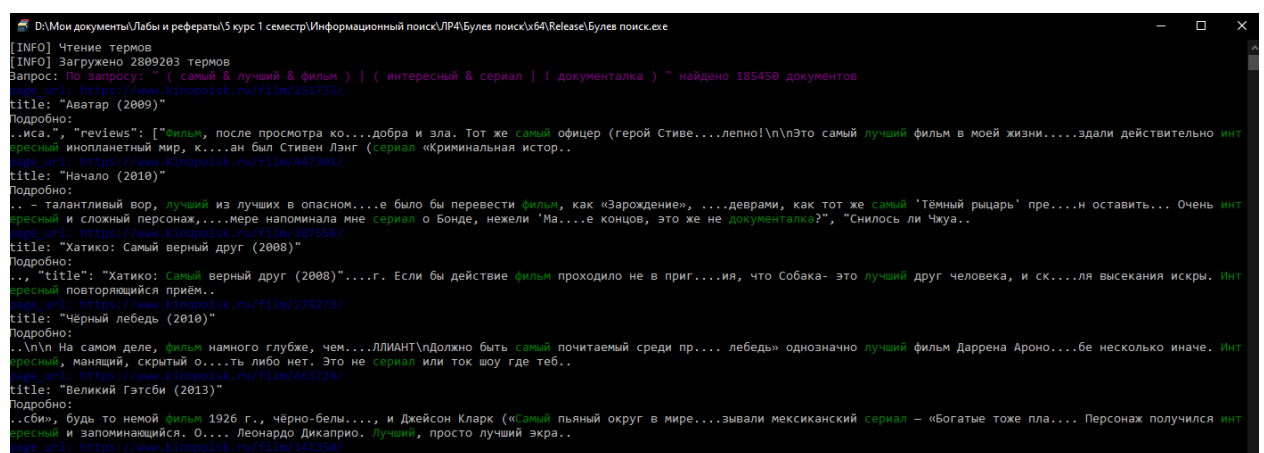
    answer.resize(0);

    while (i < p1.size() && j < p2.size())
    {
        if (p1[i] == p2[j])
        {
            answer.push_back(p1[i]);
            i++;
            j++;
        }
        else if (p1[i] < p2[j])
        {
            i++;
        }
        else
        {
            j++;
        }
    }
}
```

Интерфейс

Программа была реализована в двух интерфейсах: консольном и оконном.

Для последнего использовалась библиотека Windows.h (winapi). Примеры ниже:



The screenshot shows a Windows command prompt window with the following text:

```
D:\Мои документы\Лабы и рефераты\5 курс 1 семестр\Информационный поиск\IP4\Булев поиск\h64\Release\Булев поиск.exe
[INFO] Чтение термов
[INFO] Загружено 2809203 термов
Вопрос: По запросу: " ( самый & лучший & фильм ) | ( интересный & сериал | ! документалка ) " найдено 185450 документов
page_url: https://www.kinopoisk.ru/film/251733/
title: "Аватар (2009)"
Подробнее:
...иса.", "reviews": ["фильм, после просмотра ко...добра и зла. Тот же самый офицер (герой Стиве...ленно!\n\nЭто самый лучший фильм в моей жизни....эдали действительно инт
ресный инопланетный мир, к...ан был Стивен Лэнг (сериал «Криминальная истор..
page_url: https://www.kinopoisk.ru/film/447301/
title: "Начало (2010)"
Подробнее:
... - талантливый вор, лучший из лучших в опасном....е было бы перевести фильм, как «Зарождение», ....деврами, как тот же самый 'Тёмный рыцарь' пре....н оставить... Очень инт
ресный и сложный персонаж,....мере напоминала мне сериал о Бонде, нежели 'Ма....е концов, это же не документалка?", "Снилось ли Чжуа..
page_url: https://www.kinopoisk.ru/film/463756/
title: "Хатико: Самый верный друг (2008)"
Подробнее:
...,"title": "Хатико: Самый верный друг (2008)"....г. Если бы действие фильм проходило не в приг....ия, что Собака- это лучший друг человека, и ск....ля высекая искры. Инт
ресный повторяющийся приём..
page_url: https://www.kinopoisk.ru/film/278273/
title: "Чёрный лебедь (2010)"
Подробнее:
...\n\n На самом деле, фильм намного глубже, чем...ЛЛИАНТ\нДолжно быть самый почитаемый среди пр.... лебедь" однозначно лучший фильм Даррена Ароно....бе несколько иначе. Инт
ресный, манящий, скрытый о....ть либо нет. Это не сериал или ток шоу где теб..
page_url: https://www.kinopoisk.ru/film/463724/
title: "Великий Гэтсби (2013)"
Подробнее:
...сби", будь то немой фильм 1926 г., черно-белы...., и Джейсон Кларк («Самый пьяный округ в мире....звали мексиканский сериал – «Богатые тоже пла.... Персонаж получился инт
ресный и запоминающийся. О.... Леонардо Дикаприо. Лучший, просто лучший экра..
page_url: https://www.kinopoisk.ru/film/141350/
```

Поиск по корпусу документов

Параметры

(самый лучший фильм) || (интересный сериал || !документалка)

Поиск Построить индекс Загрузить индекс

Ссылка	Заголовок	Подробности
https://www.kinopoisk.ru/film/2...	Аватар (2009)	..иса.", "reviews": ["Фильм, после просмотра ко...добра и зла. Тот же САМЫЙ офицер (герой Стиве...лепно!\nЭ
https://www.kinopoisk.ru/film/4...	Начало (2010)	.. – талантливый вор, ЛУЧШИЙ из лучших в опасном...е было бы перевести ФИЛЬМ, как «Зарождение», ...деврам
https://www.kinopoisk.ru/film/3...	Хатико: Самый верный друг (2008)	.., "title": "Хатико: САМЫЙ верный друг (2008)"....г. Если бы действие ФИЛЬМ проходило не в приг...ия, что Собака
https://www.kinopoisk.ru/film/2...	Чёрный лебедь (2010)	..!\n\n На самом деле, ФИЛЬМ намного глубже, чем...ЛИАНА!\nДолжно быть САМЫЙ почитаемый среди при... ле
https://www.kinopoisk.ru/film/4...	Великий Гэтсби (2013)	..сби», будь то немой ФИЛЬМ 1926 г., чёрно-белый, и Джейсон Кларк («САМЫЙ пьяный округ в мире...зывали м
https://www.kinopoisk.ru/film/1...	Росарио — Ножницы (2005)	.. N99.", "reviews": ["ИНТЕРЕСНЫЙ подход к пришельцам...сическом понимании 'ФИЛЬМ о пришельцах' подра...с
https://www.kinopoisk.ru/film/3...	Район №9 (2009)	..льниками.\n\nИ хотя ФИЛЬМ и основан на комикс...ильм понравился, он ИНТЕРЕСНЫЙ, захватывающий и са...
https://www.kinopoisk.ru/film/4...	Мстители (2012)	..е успеваю. Если бы ФИЛЬМ был в жанре 'театр ...ния, будь то книга, СЕРИАЛ или компьютерная иг...аже Ноа Р
https://www.kinopoisk.ru/film/2...	Повелитель стихий (2010)	..ажные вопросы через ИНТЕРЕСНЫЙ сюжет, полный иноск...рый Добби.\n\nЭтот ФИЛЬМ идеально охватыв
https://www.kinopoisk.ru/film/1...	Гарри Поттер и Дары Смерти: Часть I (2010)	.. даже библейский. А ФИЛЬМ — хлопущка, плотно - это одновременно САМЫЙ злой организм и сам...ствова
https://www.kinopoisk.ru/film/3...	Джокер (2019)	..товое кино? Почему ФИЛЬМ 1997 года смотрится...ть его здорово =)], "ЛУЧШИЙ фильм в истории!\nЯ ...твный Дж
https://www.kinopoisk.ru/film/1...	Титаник (1997)	..рош!\nГравитация - ФИЛЬМ Альфонсо Куарона. И...минация на Оскар за ЛУЧШИЙ фильм года. Как все... снял
https://www.kinopoisk.ru/film/4...	Макао (1952)	..аботка погружения в ФИЛЬМ очень дорогого стои...г задел был положен ИНТЕРЕСНЫЙ.\n\nБольше всего из...а
https://www.kinopoisk.ru/film/5...	Прометей (2012)	..елся бы по шикарно. ФИЛЬМ словно пришелец из ...но человек он очень ИНТЕРЕСНЫЙ и за ним стоит прис...р
https://www.kinopoisk.ru/film/8...	Покémon. Апокалипсис (2010)	..ивый, меланхоличный ФИЛЬМ, вмещающий в себе д...енных фильмах, этот САМЫЙ «эффект бабочки», и... что
https://www.kinopoisk.ru/film/3...	Yes: 9012 Live (1985)	..1", "description": "ФИЛЬМ о мужчине, который ...актёрский состав не САМЫЙ малоизвестный. Итог... "Мы може
https://www.kinopoisk.ru/film/2...	Господин Никто (2009)	.. описываемый мир, а ФИЛЬМ в свою очередь уже ...Жаль, правда, что в САМЫЙ последний раз, ибо ...го лица, «
https://www.kinopoisk.ru/film/8...	Загадочная история Бенджамина Баттона (2008)	... \n\nЯ считаю этот ФИЛЬМ современным прочтен...ности, как например СЕРИАЛ ВВС, он на мой взгл... просм
https://www.kinopoisk.ru/film/3...	Гордость и предубеждение (2005)	..ной.", "reviews": ["ФИЛЬМ дает больше чем от ...ников, но далеко не САМЫЙ худший.\n\nСюжет фа...такого фи
https://www.kinopoisk.ru/film/2...	Трон: Наследие (2010)	..щущение, что кино и ФИЛЬМ связаны!\n\nМомент Но перед нами не СЕРИАЛ, в котором легко мо...ментов
https://www.kinopoisk.ru/film/4...	Варкрафт (2016)	.. не сомневался, что ФИЛЬМ стоящий. Тем более Это и создает тот САМЫЙ необходимый шарм, Л...ь, что сн
https://www.kinopoisk.ru/film/1...	Волк с Уолл-стрит (2013)	..го Макс уверен, что ЛУЧШИЙ способ выжить - ски...и пересматриваю этот ФИЛЬМ, и у меня к нему мн... вопро
https://www.kinopoisk.ru/film/4...	Безумный Макс: Дорога ярости (2015)	..нять заключительный ФИЛЬМ какой-то франшизы. ...!\nВ 2000 году вышел САМЫЙ первый фильм 'Люди ... мен
https://www.kinopoisk.ru/series/...	Логан (2017)	..аист, пай-мальчик, САМЫЙ 'крутой' парень в ш...но писали, что этот СЕРИАЛ по атмосфере скорее...ками кон
https://www.kinopoisk.ru/film/4...	1+1 (2011)	..ний.", "reviews": ["ФИЛЬМ, который заставит В...ило бы делать целый СЕРИАЛ, а подобное в планы...о вы сейка
https://www.kinopoisk.ru/film/4...	Дориан Грей (2009)	..уки и не читать, то ФИЛЬМ заинтересует весьма...ении Колина Ферта - САМЫЙ циничный и омерзите... характер
https://www.kinopoisk.ru/film/4...	Оно (2017)	..борную. Или сначала ФИЛЬМ предстаёт подростко...алисман фильма - не САМЫЙ талантливый монстр...чем у
https://www.kinopoisk.ru/film/3...	Воображариум доктора Парнаса (2009)	..олучается, что этот ФИЛЬМ - ирония над нашим ...что собрал наверно САМЫЙ знаменитый актерски...сера Ш
https://www.kinopoisk.ru/film/4...	Миллионер из трущоб (2008)	..как любой индийский ФИЛЬМ, даже приняв в себя...огло взять Оскар за ЛУЧШИЙ фильм (если отброси...сказат
https://www.kinopoisk.ru/film/7...	28 панфиловцев (2016)	..возможность включить ФИЛЬМ, когда вздумается. ...много. А здесь он в САМЫЙ минимум - задача вс...о шедевр
https://www.kinopoisk.ru/film/5...	Закоренение Солнца (1960)	..волны» и едва ли не САМЫЙ противоречивый кине...
https://www.kinopoisk.ru/film/3...	Бегущий по лезвию 2049 (2017)	..ток, из-за которого ФИЛЬМ смотрится тяжело. О...и создал чуть ли не ЛУЧШИЙ фильм в жанре кибер... среди в
https://www.kinopoisk.ru/film/2...	Повелитель бури (2008)	..ся смешанным: то ли ФИЛЬМ плохой, потому что ...дущее", "" - Какой САМЫЙ ЛУЧШИЙ способ обезвредить ...
https://www.kinopoisk.ru/film/4...	Турист (2010)	...["Пара странная, а ФИЛЬМ любимый!\nНасколько ...лобус - 2011!\n\n- ЛУЧШИЙ фильм (комедия или ...ойдет,
https://www.kinopoisk.ru/film/4...	Зеленая миля (1999)	...!\nВ 1999 году этот ФИЛЬМ я и увидел. Его пок... награда могли стать ЛУЧШИЙ фильм и Лучший актё... Шоушенк
https://www.kinopoisk.ru/film/3...	Трансформеры: Месть падших (2009)	.. тебя пригласили не ФИЛЬМ посмотреть, а какое...дания, сняв реально ИНТЕРЕСНЫЙ сиквел, который выс...у. V
https://www.kinopoisk.ru/film/8...	Два облика Освенцима (ТВ, 2011)	

1/3709

Замечу, что для вывода «подробностей» требуется повторный проход по документу, что замедляет поисковую выдачу. Чтобы немного сгладить ситуацию, процесс был распараллелен. Лучшим решением данной ситуации является применение координатного индекса.

Параметры

Путь к директории с корпусом D:\Мои документы\Лабы и рефераты\5 курс 1 семестр\Информационный поиск\Корпус

Путь к директории с индексом D:\Мои документы\Лабы и рефераты\5 курс 1 семестр\Информационный поиск\Корпус

Путь к директории с временными файлами D:\Мои документы\Лабы и рефераты\5 курс 1 семестр\Информационный поиск\Файлы

Число потоков 5

Также был написан автотестирующий.

2. Исходный код

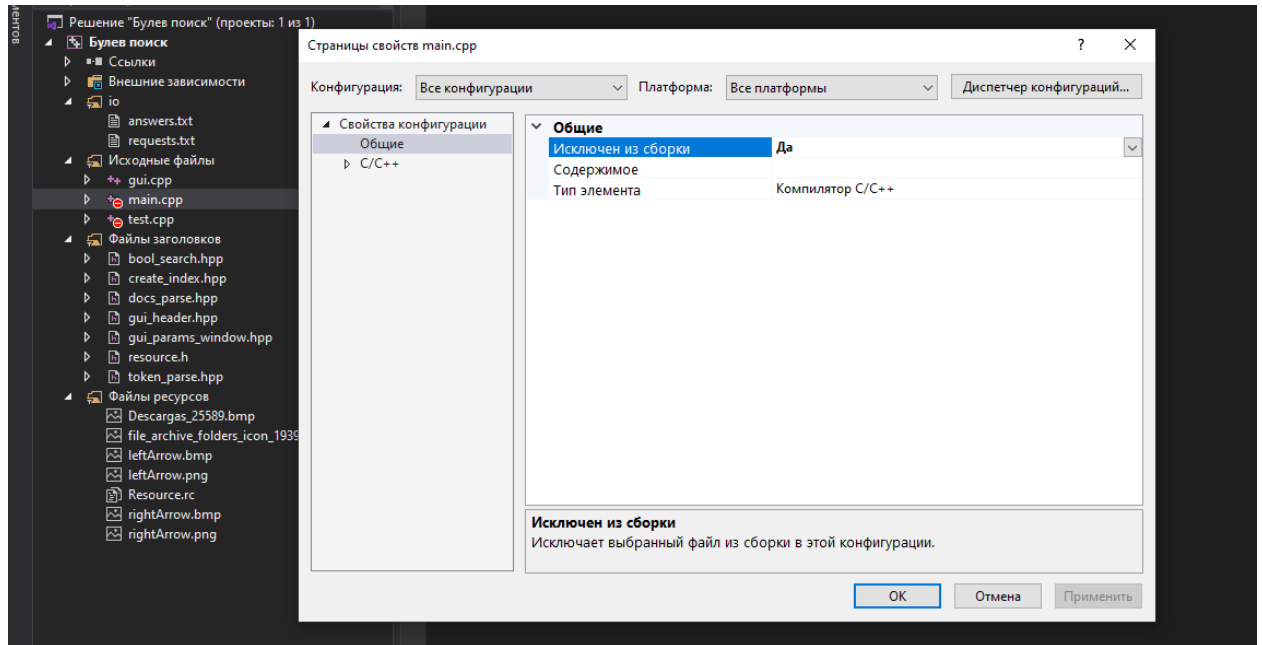
Структура проекта

- include
 - bool_search.hpp (булев поиск)
 - create_index.hpp (создание, чтение индекса)
 - docs_parse.hpp (извлечение полей из корпуса)
 - gui_header.hpp (главный хэдер для оконного интерфейса)
 - gui_params_window.hpp (окно с выбором параметров)
 - token_parse.hpp (функции для преобразования токенов в термы)
- io
 - answers.txt
 - requests.txt
- src
 - gui.cpp (точка входа в оконный интерфейс)
 - main.cpp (точка входа в консольный интерфейс)
 - test.cpp (утилита для тестирования программы)
- resources (файлы ресурсов для оконного приложения)

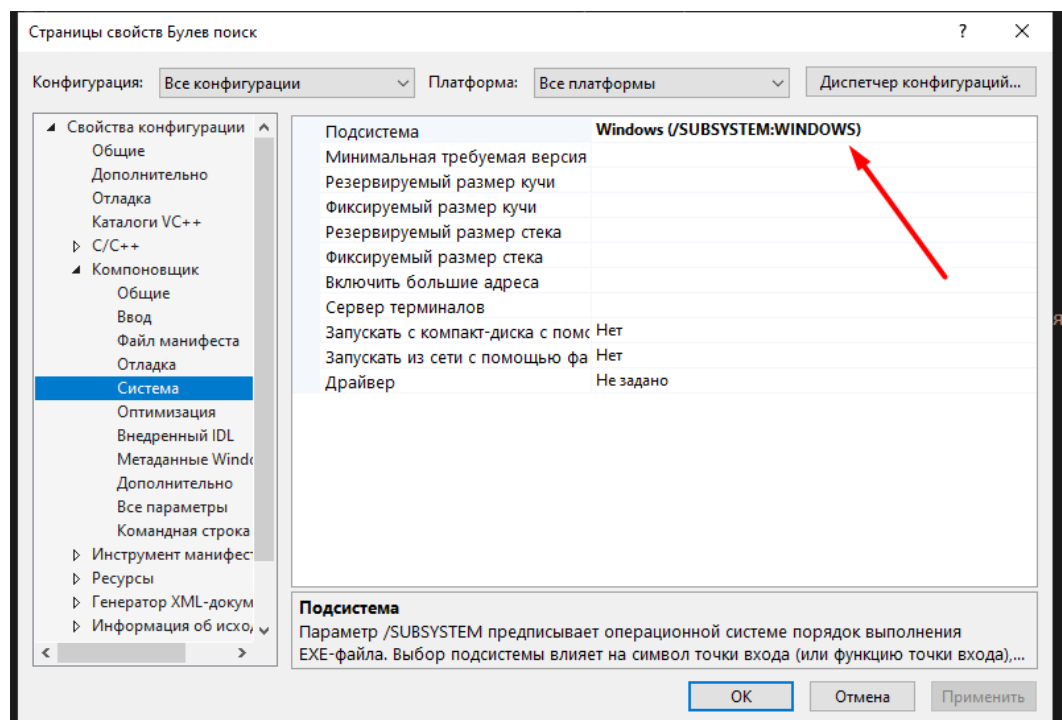
Проект был написан с помощью Microsoft Visual Studio 2019 эксклюзивно для ОС семейства Windows.

Запуск

Переключение между тремя точками входа осуществляется с помощью флага «исключить из сборки»:



Не забудь при переключении между консольными и оконными приложениями менять подсистему в настройке проекта:



Консольное приложение поддерживает флаги запуска:

- -i 'абс. путь к корпусу'
- -o 'абс. путь к индексу'
- -t 'абс. путь к директории с блочным индексом'
- -p кол-во_процессов_для_распараллеливания
- -create : создать блочный индекс
- -merge : выполнить слияние блочного индекса
- -clear : очистить папку с временными файлами после слияния
- -search : выполнить поиск

Утилита тестирования поддерживает следующие ключи:

- -i 'абс. путь к корпусу'
- -o 'абс. путь к индексу'
- -n1 число_запросов
- -n2 длина_запросы_в_термах

Пример создания индекса из корпуса:

```
$ ./Булев индекс.exe -merge -create -clear -i "D:\Мои документы\Лабы и рефераты\5 курс 1 семестр\Информационный поиск\Корпус" -o "D:\Мои документы\Лабы и рефераты\5 курс 1 семестр\Информационный поиск\Корпус_index" -t "D:\Мои документы\Лабы и рефераты\5 курс 1 семестр\Информационный поиск\ЛР3\Булев индекс\tmp"
```

```
[INFO] Создание индекса для блоков
```

```
[INFO] Thread 0 processing block 1/13 : D:\Мои документы\Лабы и рефераты\5 курс 1 семестр\Информационный поиск\Корпус\films13.txt
```

```
[INFO] Thread 1 processing block 2/13 : D:\Мои документы\Лабы и рефераты\5 курс 1 семестр\Информационный поиск\Корпус\films9.txt
```

```
[INFO] Thread 2 processing block 3/13 : D:\Мои документы\Лабы и рефераты\5 курс 1 семестр\Информационный поиск\Корпус\films10.txt
```

```
[INFO] Thread 3 processing block 4/13 : D:\Мои документы\Лабы и рефераты\5 курс 1 семестр\Информационный поиск\Корпус\films12.txt
```

```
[INFO] Block 1 has 232216 terms
```

```
<...>
```

```
[INFO] Block 10 has 597482 terms
```

[INFO] Block 11 has 669497 terms
 [INFO] Block 12 has 921883 terms
 [INFO] Block 13 has 1383148 terms
 [INFO] Создание очередей термов: 13 блок из 13
 [INFO] Слияние docs_id: 13 блок из 13
 [INFO] Слияние слопозиций термов
 [INFO] Осталось термов: 0
 [INFO] Очистка временных файлов
 [INFO] Общее число термов в словаре = 2809203

Время выполнения = 145,5 сек, размер корпуса = 2,899 Gb, документов = 186109

Средняя скорость на документ = 0,782 ms

Средняя скорость на килобайт = 0,048 ms

Пример работы тестировщика:

\$./Булев индекс.exe -n1 200 -n2 5 -search -i "..\..\Корпус" -o
 "..\..\Корпус_index" -t "tmp" <io/requests.txt >io/answers.txt

```
[INFO] Чтение термов
[INFO] Загружено 2809203 термов
Номер запроса|Документов|Время, ms|Запрос
1||0||47363||!завались-и || полубогов-наставников && !вкачал && unagi-sandжейпоп && омежному
2||97721||41163||трагичных || !draculiusваха || !фильм-эссе || !васантабалан || завертелись
3||0||49514||!пахнутаха-kui && !севадоархеолога && отцом-священнослужителем || anime-9595-amanatsu && !акирыдальше
4||7||38243||конском && !ragamuffin || shackleton || галапаго || грира
5||142308||49725||!австрийско-американскогофильм || !lloydмато && !связуяща || разгильдяи-мужья || !лапирь
6||117798||55409||тщательно-тщательно || !заступом || отпечатавшиеся || экзотике && !шафтена
7||140308||34997||!perfectionist || !размытое || !rosseti && okayмлении || feuerbach
8||0||32456||отчаянней && takahashi && фыркают && теоретике && !uc-qsijjjsnjepvcowid3sw
9||0||35578||задутся || lasky && зятанутой-перетанутой || !marey && !изображению
10||0||55991||!невежля && 24-ому && саморазрушается && djunhorдекто || евро-1988
11||0||53903||temoins || timas || !diego3000представь && милый1920x1062 && !siliang
12||0||36817||!backpacking || !москвич-2141 || проводящих && костяными || меллерукуине
13||0||45504||!репортер-плайбой && !200icq || керзон && натянутьтоксичный || !voda-i-ogon
14||0||43348||!таксандрию || минами-ловушками && братья-близнецы-они || yagakimi || !bo-ri
15||0||47784||!умничкаоднозначно && !момышулы && локдока && !predannaya-krasota && kitamuraэри
16||156072||42611||спецэффекты || !202605 || !баллы || !тщеславно-эгоцентричного || !наркодилеров-идиотов
17||0||56536||!незаземленным || !чаегиоликами && !артер850x1280850x1280850x1280850x1280850x1280850x1280797x1200
18||0||53416||попстрима || скалящая && !шер-хан && !окты && !дюбу
19||0||34299||суперзлодеях || расшалившиеся || !гинестра || veshi && !240088-obsuzhdenie-anime
20||0||40235||стороннее && разспойлерувидел && !эротико-романтический && !кот-робинзон && тушилпирогами
21||0||49357||!пазрываюсь && !спойлеротрубывать || дора && существующую && !пслоу
22||130248||30793||!28223-death-parade && !милотыкогда && !самурая-террориста || !приветствовала && !стопроцентный
23||48145||51458||!tokotokos && !anime-6444-tegamibachi && !felicitas || детское || !прошманда
24||0||51957||!friendвпрочем && !социологами && первопроходцу && !бэрона || атмосферу-ретро
25||130000||40100||!сая || !изм-тукерто || !василист || !экижаренсе || !скажана
```

```

191|| 0|| 36098|| пейзажный || клочатые && полинялых && !плагиат-подделка && переколбасил
192|| 1152|| 49506|| !704920 && фармёжки || what || наклеен || вкушали
193|| 0|| 40179|| !франском || скромной && патмор && прог-рок-группы || трёхвалентный
194|| 134339|| 72172|| боала || !бееестранная && !ботан-ботан && momose || !янетти
195|| 0|| 42298|| !неозыданно || убивавшей && !нижерадзе || !k32 && хадисы
196|| 534|| 44258|| !мужик-мачо && !архетипчик && !zmopes && !срабатывание && !87489-toki-wo-kakeru-shoujo
197|| 8936|| 44787|| !злодея-мессии || !закончились && !вот1920x10801920x1080пешипгатакое && аналогayoutube && !клензендорфа
198|| 0|| 52473|| храбрец-красавчик && !дурачащих && фильм-интрига || !тупачка || !евнухом
199|| 31139|| 56695|| !amarelo && !лисапед || !markedoneороче || !ноября704x995 && !простынейтак
200|| 0|| 323291|| !сюжетно-воспитательные || минут06 && !sonambula || !паукоубиения && !ухxxxxxxxxxдождался
Всего времени 11059715 мс, в среднем на запрос 55299 мс

```

3. Выводы

В целом процесс булевого поиска работает достаточно быстро. Самые тяжёлые запросы для системы выглядят следующим образом:

фильм && фильм && фильм ...

сериал && сериал && сериал ...

Для их обработки выполняется пересечение очень длинных одинаковых списков словопозиций. Возможное решение – символьная оптимизация запроса на уровне операторов. Также ещё раз отмечу, что для вывода блока с подробностями по запросу идёт повторный анализ документов с целью поиска термов. Это является узким местом производительности. Решения – распараллеливание (реализовано), координатный индекс.

В ходе выполнения лабораторной работы я научился выполнять булев поиск для коллекции документов. Познакомился с winapi. Научился разрабатывать оконный интерфейс на языке C/C++.

Литература

- [1] Кристофер Д.Маннинг, Прабхакар Рагхаван, Хайнрих Шютце. Введение в информационный поиск. 2020, изд. Вильямс.