

МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)

Институт №8 «Компьютерные науки и прикладная математика»

**Лабораторная работа №8
по курсу «Информационный поиск»**

«Булев индекс»

Выполнил: *Ермаков Ярослав Валерьевич*
Группа: *M8O-407Б-22*
Преподаватели: *А.А. Кухтичев*

Москва, 2025

Описание

Целью работы является реализация булевого поиска по построенному обратному индексу. Поисковая система принимает запрос с логическими операторами и возвращает документы, удовлетворяющие булевому выражению. Для восстановления результата пользователю используется прямой индекс: по doc_id выводятся URL и заголовок документа.

Журнал выполнения и решение проблем

Основной задачей было обеспечить корректную обработку логических операторов и скобок, а также согласованность обработки запроса и индекса. Для этого запрос токенизируется теми же правилами нормализации, что применялись при индексации. В ходе тестирования было учтено, что в оболочке bash символ ! интерпретируется как служебный, поэтому запросы с отрицанием корректно запускать в кавычках или экранировать!. После настройки формы запуска запросов и согласования режима обработки булев поиск стал возвращать корректные результаты.

Исходный код

Поисковый модуль реализован как CLI-утилита, которая загружает бинарные файлы индекса, разбирает запрос в булево выражение и вычисляет результат через операции над отсортированными списками doc_id. Для AND выполняется пересечение списков, для OR - объединение, для NOT - разность с универсальным множеством документов. Скобки поддерживаются за счёт преобразования выражения в постфиксную форму и последующего вычисления по стеку.

Выводы

Реализован булев поиск по собственному бинарному индексу. Поддерживаются операторы AND/OR/NOT (а также их символьные формы & | !) и скобки, выдача формируется через прямой индекс и содержит ссылки и заголовки документов.

Реализация корректно работает на полном корпусе и является базой для последующих улучшений поиска и ранжирования