

**МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)
Факультет информационных технологий и прикладной математики
Кафедра вычислительной математики и программирования**

**Отчет по курсовому проекту
по курсу
«Информационный поиск»**

Группа: М80-108М-17
Выполнил: Забарин Н.И.
Преподаватель: А.Л. Калинин

Москва, 2018

Задание

Необходимо проиндексировать англоязычную Википедию, реализовать поиск по полученному индексу.

Метод решения

В процессе выполнения лабораторных работ были разработаны инструменты позволяющие проиндексировать полный дамп Википедии. Для выполнения курсового проекта были внесены ряд правок не изменившие принципы обработки.

Статистика

Было проиндексированно 8,5 млн документов, суммарная длина всех слов около 21 миллиарда символов, на индексацию ушло 34 тысячи секунд, получено 112 индексов размера порядка 100мб. Скорость выполнения запроса из одного слова порядка 25 секунд.

Выводы

В ходе выполнения курсового проекта была реализована полноценная поисковая система. Полученная система имеет большой потенциал развития, но на ее примере я понял как работают полнотекстовые поисковые движки.