

Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт (национальный
исследовательский университет)»
Физтех-школа радиотехники и компьютерных технологий
Кафедра теоретической и прикладной информатики

Направление подготовки: 03.03.01 Прикладные математика и
физика

Направленность (профиль) подготовки: Инфокоммуникационные
системы и технологии

**СЖАТЫЕ СТРУКТУРЫ ДАННЫХ
В ЗАДАЧЕ ПОЛНОТЕКСТОВОГО ПОИСКА**
(бакалаврская диссертация)

Студент:
Соколов Вадим Андреевич

(подпись студента)

Научный руководитель:
Неганов Алексей
Михайлович

(подпись научного руководителя)

Москва 2021

Аннотация

Цели и задачи работы.

Данная работа посвящена исследованию структур данных, использующих сжатые индексы (succinct index) для хранения текстовой информации.

Целью данной работы является проверка эффективности различных методов сжатия данных. Исследуется применимость сжатых индексов на практике при работе с данными определенного типа. Производится сравнение как с традиционными решениями (suffix array), так и с более современными (radix tree), использующими подход для индексирования, отличный от исследуемых структур данных.

Полученные результаты.

Удалось получить сравнительные характеристики работы традиционных структур данных. Были измерены и проанализированы:

1. объем потребляемой памяти;
2. время, требуемое для вставки / поиска подстроки;

Содержание

1	Введение	4
2	Постановка задачи	4
3	Измерения	5
3.1	Экспериментальная платформа	5
3.2	Suffix array	5
3.3	Radix tree	5
4	Оценка результата	5
5	Выводы	5
6	Заключение	5

1 Введение

Работа с текстовыми данными находит применение в широком спектре задач современной компьютерной индустрии. Существует ряд проблем, связанных с поиском информации в поисковых сервисах. Рост количества информации в Интернете приводит к дополнительным издержкам при хранении и поиске данных. В связи с этим существует необходимость исследования различных способов уменьшения потребляемой памяти без существенных затрат на поиск данных.

Одним из возможных решений такого рода задач являются сжатые структуры данных (succinct data structures).

2 Постановка задачи

3 Измерения

3.1 Экспериментальная платформа

3.2 Suffix array

3.3 Radix tree

4 Оценка результата

5 Выводы

6 Заключение