Отчет о лабораторной работе по дисциплине Анализ Алгоритмов №1 на тему «Расстояние Левенштейна и Дамерау»

**Оглавление (АВТО)**

**Введение**

Расстояние Левенштейна (рациональное расстояние) – это минимальное количество редакторских операций, которые необходимы для превращения одной строки в другую.

Под редакторскими операциями подразумевается вставка (обозначается, как I - insert), замена (R - replace), удаление (D - delete), также сюда относится совпадение (M - match). Расстояние Левенштейна имеет широкий спектр применения, например, используется в поисковых строках, в программах, отвечающих за автоисправление, автозамену. Помимо этого, оно также применяется в биоинформатике (строение белков представляется строками, состоящими из букв ограниченного алфавита, таким образом, упрощается их анализ).

Существует много алгоритмов, рассчитывающих расстояние Левенштейна, а также их модификаций, которые и будут рассмотрены далее.

1. **Аналитическая часть**

Цель данной работы – реализовать и сравнить алгоритмы поиска расстояний Левенштейна и Дамерау-Левенштейна

Для достижения поставленной цели необходимо решить ряд следующих задач:

* дать математическое описание расстояний
* описать алгоритмы поиска расстояний
* (!!!!!!!!!!!!!) оценить затрачиваемую алгоритмами память
* реализовать эти алгоритмы
* провести замеры процессорного времени работы алгоритмов (!!!!!!!!!!!!!и затрачиваемой ими памяти) на материале серии экспериментов
* провести сравнительный анализ алгоритмов

Поиск расстояния Левенштейна можно описать разными алгоритмами:

1. матричный расчёт
2. рекурсивный расчёт по формуле
3. рекурсивный алгоритм, заполняющий незаполненные клетки матрицы

Пусть S1 и S2 – строки длиной N и M соответственно. Тогда расстояние Левенштейна можно рассчитать по следующей рекуррентной формуле:

D(i, j) =

При таком способе расчёта расстояния нужно использовать матрицу размера Len(S1) + 1 x Len(S2) + 1, элементы которого рассчитываются по формуле выше.

Что касается рекурсивного расчёта, то возникает проблема большого количества повторных вычислений. Это очень сильно влияет как на время выполнения, так и на занимаемую память.

Рекурсивный алгоритм, заполняющий незаполненные клетки матрицы, работает по аналогии с бесконечностями в алгоритме Дейкстра поиска расстояний в графе.

Расстояние Дамерау-Левенштейна дополнительно включает операцию перестановки двух соседних символов (транспозицию) и формула выглядит следующим образом:

1. **Конструкторская часть**

Входные данные: две строки

Выходные данные: вычисленное минимальное расстояние и вывод матрицы, если алгоритм был на ней построен.

СХЕМЫ!!!!!!

ТЕСТЫ!!!!!!

1. **Технологическая часть**
   1. **Выбранный язык программирования**

Для выполнения этой лабораторной работы был выбран язык программирования Python, так как есть большой навык работы с ним.

* 1. **Инструменты замеров**

Была использована библиотека time для проведения замеров затрачиваемого времени.

* 1. **Листинг**

КОД!!!!!

* 1. **Результаты тестов**

ТЕСТЫ!!!!

Все тесты пройдены успешно.

1. **Исследовательская часть**

**Заключение**

**Список литературы**