*Федеральное государственное бюджетное образовательное учреждениевысшего профессионального образования*

|  |  |
| --- | --- |
| Gerb-BMSTU_01 | ***«Московский государственный технический университет  имени Н.Э. Баумана (национальный исследовательский институт)»***  ***(МГТУ им. Н.Э. Баумана)*** |

ФАКУЛЬТЕТ Информатика и системы управления

КАФЕДРА Программное обеспечение ЭВМ и информационные технологии

**Отчёт**

**по лабораторной работе 1**

**Дисциплина: Анализ Алгоритмов**

**Тема лабораторной работы работы: Расстояния Левенштейна и Дамерау-Левенштейна**

Студентки гр. ИУ7-51б **\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Сушина А.Д.**

(Подпись, дата) (И.О. Фамилия)

Преподаватель  **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Волкова Л.Л.**

(Подпись, дата) (И.О. Фамилия)

Москва, 2019г

Оглавление

[ФАКУЛЬТЕТ Информатика и системы управления 1](#__RefHeading___Toc1029_3084784888)

# Введение

**Расстояние Левенштейна** (редакционное расстояние, дистанция редактирования) - минимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую. Измеряется для двух строк, широко используется в теории информации и компьютерной лингвистике.

**Расстояние Дамерау-Левенштейна**— это мера разницы двух строк символов, определяемая как минимальное количество операций вставки, удаления, замены и транспозиции (перестановки двух соседних символов), необходимых для перевода одной строки в другую. Является модификацией расстояния Левенштейна: к операциям вставки, удаления и замены символов, определённых в расстоянии Левенштейна добавлена операция транспозиции (перестановки) символов.

**Цель работы:** изучение метода динамического программирования на материале алгоритмов Левенштейна и Дамерау-Левенштейна.

**Задачи работы:**

1. изучение алгоритмов Левенштейна и Дамерау-Левенштейна нахождения расстояния между строками;
2. применение метода динамического программирования для матричной реализации указанных алгоритмов;
3. получение практических навыков реализации указанных алгоритмов: двух алгоритмов в матричной версии и одного из алгоритмов в рекурсивной версии;
4. сравнительный анализ линейной и рекурсивной реализаций выбранного алгоритма определения расстояния между строками по затрачиваемым ресурсам (времени и памяти);
5. экспериментальное подтверждение различий во временнóй эффективности рекурсивной и нерекурсивной реализаций выбранного алгоритма определения расстояния между строками при помощи разработанного программного обеспечения на материале замеров процессорного времени выполнения реализации на варьирующихся длинах строк;
6. описание и обоснование полученных результатов в отчете о выполненной лабораторной работе, выполненного как расчётно-пояснительная записка к работе.

# 1. Аналитическая часть

## 1.1. Описание алгоритмов

**Алгоритм вычисления расстояния Левенштейна.**

Расстояние Левенштейна или редакторское расстояние считается как минимальное количество редакторских операций, нобходимых для преобразования строки s1 в строку s2.

Редакторские операции:

* INSERT — вставкам (штраф 1)
* DELETE — удаление (штраф 1)
* REPLACE — замена (штраф 1)
* MATCH — сопадение (штраф 0)

Считается, что элементы строк нумеруются с первого, как принято в математике, а не с нулевого, как принято во многих языках программирования.

Пусть s1 и s2— две строки (длиной M и N соответственно) над некоторым афвалитом, тогда редакционное расстояние (расстояние Левенштейна) d(s1, s2) можно подсчитать по следующей рекуррентной формуле:

где операция min возвращает наименьший из аргументов.

**Алгоритм вычисление расстояния Дамерау-Левенштейна**

Вводится дополнительна операция: перестановка или транспозиция двух букв со штрафом 1.

Если индексы позволяют и если две соседние буквы , то в минимум включается перестановка.

Расстояние Левенштейна и его обобщения активно применяется:

* для исправления ошибок в слове (в поисковых системах, базах данных, при вводе текста, при автоматическом распознавании отсканированного текста или речи).
* для сравнения текстовых файлов утилитой diff и ей подобными. Здесь роль «символов» играют строки, а роль «строк»— файлы.
* в биоинформатике для сравнения генов, хромосом и белков.

# 2. Конструкторская часть

## 2.1.Разработкаалгоритмов

## 2.2.Сравнительныйанализрекурсивнойинерекурсивнойреализаций

# 3.Технологическаячасть

## 3.1.Требованиякпрограммномуобеспечению

## 3.2.Средствареализации

## 3.3.Листингкода

## 3.4.Описаниетестирования

# 4.Экспериментальнаячасть

## 4.1.Примерыработы

## 4.2.Результатытестирования

## 4.3.Постановкаэкспериментапозамерувремени [ипамяти]