**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра МО ЭВМ**

отчет

**по лабораторной работе №4**

**по дисциплине «Построение и анализ алгоритмов»**

Тема: КМП

|  |  |  |
| --- | --- | --- |
| Студент гр. 3388 |  | Сабалиров М.З. |
| Преподаватель |  | Жангиров Т.Р. |

Санкт-Петербург

2025

## Задание

Задача 1

Реализуйте алгоритм КМП и с его помощью для заданных шаблона P (|P| < 15000) и текста T (|T| < 5000000) найдите все вхождения P в T.

Вход:

Первая строка - P

Вторая строка - T

Выход:

Индексы начал вхождений P в T, разделенных запятой. Если P не входит в T, то вывести -1.

Задача 2

Заданы две строки A (|A| < 5000000) и B (|B| < 5000000).

Определить, является ли A циклическим сдвигом B (это значит, что A и B имеют одинаковую длину и A состоит из суффикса B, склеенного с префиксом B). Например, `defabc` является циклическим сдвигом `abcdef`.

Вход:

Первая строка - A

Вторая строка - B

Выход:

Если A является циклическим сдвигом B, индекс начала строки B в A, иначе вывести -1. Если возможно несколько сдвигов, вывести первый индекс.

## Выполнение работы

Для выполнения был использован алгоритм КМП. Алгоритм основан на префикс функции. Собственным префиксом строки S назовем максимальное число такое что, префикс и суффикс данной длины равны. Префикс функция строки S, есть массив, где на i-м элементе находится длина собственного префикса S[0..i]. Префиксы ищутся динамически. На каждом шаге длина очередного собственного префикса больше предыдущего не более чем на 1. Пусть N — длина S, тогда из предыдущего условия максимально возможная длина собственного префикса N — 1 (Очевидно, что она не может быть такой большой, так как префикс и суффикс должны быть равны, но для ассимтоматического сравнения это не имеет значения). Тогда суммарно за всю итерацию по строке S итераций по массиву собственных префиксов будет не больше чем N — 1. Поиск паттерна в тексте работает за O(M + N), где M и N длины паттерна и текста соответственно. Составляется строка Pattern + X + Text, где X отсутствующий символ в алфавите. Тогда очевидно, что нахождения собственного префикса длины M будет означать вхождение Pattern в Text. Причем длины собственных префиксов не более чем M — 1. Итого N + M + 1 итераций по строке и не более M — 1 итераций по массиву префиксов. Получаем искомую оценку O(M + N). Затраты по памяти соответственно O(M + N).

**Тестирование:**

|  |  |
| --- | --- |
| Input | Output |
| ab  abab | 0,2 |
| BO  ABOBOA | 1,3 |
| МОЭВМ СПБГУТ))) | -1 |

|  |  |
| --- | --- |
| Input | Output |
| defabc  abcdef | 3 |
| ABOBA  BOBAA | 1 |
| ABOBA  ABABO | -1 |

**Выводы:**

В ходе работы был разработан и протестирован алгоритм для поиска вхождений шаблона в тексте. Алгоритм использует массив собственных префиксов для эффективного поиска подстрок. Были добавлены подробные отладочные выводы для отслеживания выполнения программы, что упрощает поиск и исправление ошибок. Программа корректно обрабатывает входные данные, находит все вхождения шаблона и выводит их в порядке возрастания. Решение успешно справляется с задачей, демонстрируя высокую производительность даже на больших объемах данных.