

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: Поиск подстроки в строке.

Студент гр. 3381

Лутфулин Д.А.

Преподаватель

Жангиров Т.Р.

Санкт-Петербург

2025

Цель работы.

Реализовать алгоритм Кнута-Морриса-Пратта для поиска подстроки и алгоритм, который находит циклический сдвиг в строке.

Задание.

Кнут-Моррис-Пратт.

Реализуйте алгоритм КМП и с его помощью для заданных шаблона P ($|P| \leq 15000$) и текста T ($|T| \leq 5000000$) 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

Sample Input:

ab

abab

Sample Output:

0,2

Поиск подстроки.

Заданы две строки A ($|A| \leq 5000000$) и BB ($|B| \leq 5000000$).

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

Вход:

Первая строка - A

Вторая строка - В

Выход:

Если А является циклическим сдвигом В, индекс начала строки В в А, иначе вывести $-1-1$. Если возможно несколько сдвигов вывести первый индекс.

Sample Input:

defabc

abcdef

Sample Output:

3

Выполнение работы.

Для выполнения поставленных задач был реализован алгоритм нахождения префикс функции. Эта функция для каждой позиции строки возвращает длину наибольшего префикса, совпадающего с суффиксом подстроки, заканчивающейся в данной позиции. Если на шаге итерирования совпали очередные символы из префикса и суффикса, то к текущей длине, полученной на прошлом шаге, прибавляем единицу. В противном случае происходит переход к предыдущему потенциальному префикс-суффиксу, который равен значению префикс функции от префикс-суффикса, вычисленного на предыдущем шаге.

Кнут-Моррис-Пратт.

Функция `kmp` — алгоритм Кнута-Морриса-Пратта, который находит все вхождения строки Р в строку Т. Алгоритм применяет префикс функцию к шаблону, а затем итерируется по тексту и вычисляет текущее значение префикс функции для строки “Pattern#Text”, сохраняя последнее вычисленное k в переменную. Если на текущем шаге алгоритма значение k стало равно длине образца, то суффикс текста совпадает с образцом — значит, мы нашли вхождение образца в строке. Если не нашлось ни одной такой длины, то возвращается -1.

Поиск циклического сдвига.

В функции `find_shift` сначала идёт сравнение длин строк, если они разные, то возвращается -1, так как циклический сдвиг при таком условии не возможен.

Для строки `P` применяем префикс функцию, чтобы эффективнее искать подстроку. Далее применяем алгоритм КМП к удвоенной строке с текстом, но вместо явного удваивания используем остаток при делении на длину текста. Если в удвоенной строке нашлось вхождение образца, то образец является циклическим сдвигом.

Оценка сложности по времени.

N — длина образца, M — длина текста.

Алгоритм нахождения префикс-функции состоит из линейного прохода по образцу, длина которой N . Иногда k уменьшается, а не увеличивается на 1, но таких уменьшений не более чем n за всё время работы алгоритма. Таким образом, сложность нахождения префикс-функции $O(n)$.

Алгоритм КМП по сути вычисляет префикс-функцию образца, склеенного с текстом. Значит, сложность будет $O(N+M)$

Алгоритм поиска циклического сдвига – это КМП, в котором вычисляется префикс-функция для образца и удвоенного текста. Сложность $O(3N)$ ($N=M$ в данном случае)

Вывод.

Были разработаны и проанализированы алгоритмы Кнута-Морриса-Пратта для поиска подстрок и алгоритм поиска циклического сдвига.