**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра МО ЭВМ**

отчет

**по лабораторной работе №5**

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

Тема: Ахо-Корасик

|  |  |  |
| --- | --- | --- |
| Студентка гр. 9382 |  | Голубева В.П. |
| Преподаватель |  | Фирсов М.А. |

Санкт-Петербург

2021

## Цель работы.

## Задание 1

Разработайте программу, решающую задачу точного поиска набора образцов.

Вход:

Первая строка содержит текст (*T*, 1≤∣*T*∣≤100000).

Вторая - число n (1≤n≤3000), каждая следующая из n строк содержит шаблон из набора *P* = {*p1, …, pn*} 1≤∣pi∣≤75

Все строки содержат символы из алфавита {*A,C,G,T,N*}

Выход:

Все вхождения образцов из P

Каждое вхождение образца в текст представить в виде двух чисел - *i* *p*

Где *i* - позиция в тексте (нумерация начинается с 1), с которой начинается вхождение образца с номером p (нумерация образцов начинается с 1).

Строки выхода должны быть отсортированы по возрастанию, сначала номера позиции, затем номера шаблона.

Sample Input:

NTAG

3

TAGT

TAG

T

Sample Output:

2 2

2 3

**Задание 2**

Используя реализацию точного множественного поиска, решите задачу точного поиска для одного образца с джокером.

В шаблоне встречается специальный символ, именуемый джокером (wild card), который "совпадает" с любым символом. По заданному содержащему шаблоны образцу *P* необходимо найти все вхождения Р в текст Т.

Например, образец *аb*??*с*? с джокером ? встречается дважды в тексте *xabvccbababcax*.

Символ джокер не входит в алфавит, символы которого используются в T. Каждый джокер соответствует одному символу, а не подстроке неопределённой длины. В шаблон входит хотя бы один символ не джокер, т.е. шаблоны вида ??? недопустимы.

Все строки содержат символы из алфавита {*A,C,G,T,N*}

Вход:

Текст (*T*, 1≤∣*T*∣≤100000 )

Шаблон (*P*, 1≤∣*P*∣≤40)

Символ джокера

Выход:

Строки с номерами позиций вхождений шаблона (каждая строка содержит только один номер).

Номера должны выводиться в порядке возрастания.

Sample Input:

ACTANCA

A$$A$

$

Sample Output:

1

## Описание структур данных

**Описание алгоритма**

**Оценка сложности по памяти**

**Оценка сложности по времени**

**Тестирование**

Результаты тестирования программы можно посмотреть в приложении В.

## Выводы.

**ПРИЛОЖЕНИЕ А**

**ИСХОДНЫЙ КОД ПРОГРАММЫ**

Название файла:

**ПРИЛОЖЕНИЕ В**

**ТЕСТИРОВАНИЕ**

|  |  |
| --- | --- |
| Входные данные | Выходные данные |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |