

## Лабораторная работа № 6 “Исследование алгоритмов поиска подстроки в строке”

### Задание:

1. Разработать проект для исследования алгоритмов поиска подстроки в строке в соответствии с вариантом, приведенном в таблице 1.

2. Создать класс для исследования алгоритмов поиска подстроки в строке со следующими полями, свойствами и методами:

- образец размера M,
- текст размера N,
- метод задания образца (с параметрами алфавит и размер образца),
- метод задания текста (с параметрами алфавит и размер текста),
- метод реализации прямого поиска,
- метод реализации РК-поиска,
- метод получения хеша для РК-поиска,
- метод реализации КМП-поиска,
- метод для расчета префикс-функции (Z-функции),
- метод реализации БМ-поиска,
- метод для расчета таблицы стоп-символов,
- метод для расчета таблицы суффиксов,
- метод реализации Shift-And-поиска,
- метод для построения столбца  $U(T(j))$ ,
- метод выполнения встроенного в C# поиска,
- результат поиска образца в тексте.

3. Разработать интерфейс проекта, *позволяющий*:

- задавать образец для поиска (шаблон) задавать строку;
- задавать строку (текст);
- осуществлять выбор алгоритма поиска в строке для исследования;
- осуществлять вывод информации о результатах поиска,
- осуществлять вывод информации о результатах исследования алгоритма поиска в строке (исходную строку, пошаговую работу алгоритма поиска в строке (при небольшой размерности строки), показатели качества работы алгоритма поиска в строке).

4. Создать проект, реализующую алгоритмы поиска в строке в соответствии с вариантом. В проекте *предусмотреть*:

- поиск заданного образца и индекса его первого вхождения;
- определение количества сравнений алгоритма поиска в строке  $Sr$  (или времени работы алгоритма поиска в строке).

5. При формировании строки предусмотреть возможность задания алфавита для строки и размера строки. При формировании образца

предусмотреть возможность задания алфавита для образца и размера образца.

6. Результатом работы алгоритма поиска в строке является номер первого вхождения образца в строку.

Письменный отчет по лабораторной работе должен содержать:

1. Титульный лист. (Название лабораторной работы. Фамилия, имя, отчество, номер группы исполнителя, дата сдачи.)
2. Математическую постановку задачи поиска в строке.
3. Выполняемый вариант и его содержание.
4. Диаграмму классов.
5. Распечатку *подпрограмм алгоритмов поиска в строке и кода интерфейса* проекта для исследования алгоритмов поиска в строке (*обязательны комментарии к программе*).
6. Исследование программной реализации алгоритмов поиска в строке, содержащее следующие материалы, таблицы и графики:
  - *примеры пошаговой работы* исследуемых алгоритмов поиска в строке для небольшой размерности задачи;
  - сведенную в *таблицу* зависимость количества сравнений (или времени выполнения  $T$ ) исследуемых алгоритмов поиска в строке от размерности строки  $N$ ;

Пример таблицы:

Таблица 1. – Алгоритм прямого поиска в строке, длина образца  $M=5$

$N$	...	...	...	...	...	...	...	...	...	...
$T$										

- сведенную в *таблицу* зависимость количества сравнений (или времени выполнения  $T$ ) исследуемых алгоритмов поиска в строке от размерности образца  $M$ ;

Пример таблицы:

Таблица 2. – Алгоритм прямого поиска в строке, длина строки  $N=5000$

$M$	...	...	...	...	...	...	...	...	...	...
$T$										

- *графики* зависимости показателя качества (количества операций или времени выполнения) исследуемых алгоритмов поиска в строке от размерности строки  $N$  для фиксированного размера образца  $M$ ;
- *графики* зависимости показателя качества (количества операций или времени выполнения) исследуемых алгоритмов поиска в строке от размера образца  $M$  для фиксированной размерности строки  $N$ .

7. Привести результаты тестирования проекта.

8. *Выводы* по лабораторной работе (в выводах провести сравнительную характеристику исследованных алгоритмов поиска в строке – привести сложности алгоритмов, лучшие и худшие случаи, сравнение числа операций (или времени выполнения)).

9. *Перечень ссылок* – привести адреса источников из интернета и литературу, используемую при выполнении лабораторной работы.

#### Варианты задач по лабораторной работе:

1. Алгоритм прямого поиска в строке.
2. Алгоритм РК-поиска в строке.
3. Алгоритм КМП-поиска в строке.
4. Алгоритм БМ-поиска в строке.
5. Алгоритм Shift-And-поиска в строке.
6. Поиск с помощью встроенной функции языка C#.

Таблица 1. – Варианты исследуемых алгоритмов поиска в строке

<i>Вариант</i>	<i>Варианты задач</i>
1.	1, 2, 3, 6
2.	1, 3, 5, 6
3.	1, 4, 5, 6
4.	1, 2, 4, 6
5.	1, 2, 3, 6
6.	1, 3, 5, 6
7.	1, 4, 5, 6
8.	1, 2, 4, 6
9.	1, 2, 3, 6
10.	1, 3, 5, 6
11.	1, 4, 5, 6
12.	1, 2, 4, 6
13.	1, 2, 3, 6
14.	1, 3, 5, 6
15.	1, 4, 5, 6
16.	1, 2, 4, 6
17.	1, 2, 3, 6
18.	1, 3, 5, 6
19.	1, 4, 5, 6
20.	1, 2, 4, 6

Таблица 3. – Основные встроенные методы поиска языка C#

№	Метод поиска	Класс
1.	IndexOf()	Array
2.	LastIndexOf()	Array
3.	BinarySearch()	Array
4.	Find()	Array
5.	FindIndex()	Array
6.	FindLast()	Array
7.	FindLastIndex()	Array
8.	FindAll()	Array
9.	Exists()	Array
10.	IndexOf()	String
11.	IndexOfAny()	String
12.	LastIndexOf()	String
13.	LastIndexOfAny()	String
14.	StartsWith()	String
15.	EndsWith()	String
16.	IsMatch()	Regex

### Литература

1. Кормен Т., Лейзерсон Ч., Ривест Р. Алгоритмы: построение и анализ.- М.:МЦНМО, 2000.-960с.(с. 780 - 809)
2. Вирт Н. Алгоритмы и структуры данных.- М.: Мир, 1989. (с. 74 - 87)
3. Макконелл Дж. Анализ алгоритмов. Вводный курс.-М.: Техносфера, 2002. (с. 139 - 158)

### Контрольные вопросы

1. Какова постановка задачи поиска в строке?
2. Какие методы языка C# реализуют выполнение задачи поиска в строке?
3. В чем состоит алгоритм прямого поиска в строке? Каковы его сложность, достоинства и недостатки, лучший и худший случай?
4. В чем состоит алгоритм поиска в строке Рабина-Карпа (РК-поиск)? Каковы его сложность, достоинства и недостатки?
5. В чем состоит алгоритм поиска в строке Кнута-Морриса-Пратта (КМП-поиск)? Каковы его сложность, достоинства и недостатки?
6. Что такое префикс и суффикс строки?
7. Что такое префикс-функция строки, как она вычисляется, какова сложность ее вычисления?

8. В чем состоит алгоритм КМП-поиска (вариант 2)? Каковы его сложность, достоинства и недостатки?
9. В чем состоит алгоритм поиска в строке Боуера-Мура (БМ-поиск)? Каковы его сложность, достоинства и недостатки?
10. Что такое таблица стоп-символов, как она вычисляется и какова сложность ее вычисления, как она используется для вычисления сдвига образца?
11. Что такое эвристика совпавшего суффикса и как она используется для вычисления сдвига образца?
12. В чем состоит правило составления таблицы суффиксов? Какова сложность вычисления таблицы суффиксов?
13. Как в алгоритме БМ-поиска определяется сдвиг образца на основании эвристики стоп-символа и эвристики совпавшего суффикса?
14. Какие существуют модификации у алгоритма БМ-поиска?
15. В чем состоит алгоритм Shift-And поиска в строке? Каковы его сложность, достоинства и недостатки?
16. Как (и по какой формуле) заполняется  $(0, 1)$ -матрица  $M$  в алгоритме Shift-And поиска?
17. Какова классификация алгоритмов поиска в строке?