**Київський національний університет імені Тараса Шевченка**

**факультет комп’ютерних наук та кібернетики**

**кафедра інформаційних систем**

**Лабораторна робота № 7**

**“Реалізація узагальнення алгоритму Рабіна-Карпа для пошуку**

**в двомірному масиві”**

**Виконав студент 2 курсу**

**групи К-28**

**Руденко Юрій Олегович**

**2018**

• Зміст завдання та уточнення умови

Узагальнити алгоритм рабіна Карпа для пошуку зразка pattern[0…pRows-1, 0…pCols-1] в тексті text[0…tRols-1, 0…tCols-1].

• Опис алгоритму

1)Алгоритм Рабіна-Карпа

2)Алгоритм Кнута-Морріса-Прата

3) Поєднання двох алгоритмів вище для реалізація пошуку зразку в двомірній таблиці

Загальна схема алгоритму полягає в тому, щоб використовувати метод хеш функцій, запропонований в алгоритмі Рабіна-Карпа вертикально. Спочатку ми перетворюємо двовимірні масиви з символів TEXT і PATTERN в одновимірні масиви чисел text і pattern відповідно.

Далі робимо пошук text в pattern, використовуючи алгоритм КМП порядково.

Тобто в i-ій клітинці масиву pattern[i] (i [0…pCols-1] буде міститися значення хеш-функції для елементів таблиці pattern[0…pRows,i] тобто значення хеш-функції для елементів i-го стовпчика pattern.

Робимо пошук PATTERN в TEXT порядково. Причому row = pRows спочатку

Тоді на певному кроці маємо row(k), рядок на якому зараз відбувається пошук початку входження PATTERN в TEXT, в text[i]( i[0…tCols-1]) буде міститися значення хеш-функції для

Елементів i-го стовпчика починаючи від row(k) рядка довжиною pRows.

На кожному кроці для row(k) k від pRows до tRows ми робимо пошук pattern в text за допомогою алгоритма Кнутта-Морісса-Прата. Якщо співпали постовпчикові хеші, то перевіряємо підматрицю TEXT та PATTERN посимвольно. Таким чином це або колізія, або співпадіння.

Якщо ні і є ще рядки, на які може бути зміщений PATTERN перераховуємо hash

• Основні модулі програми:

• class Matrix\_Search з реалізованими методами, що необхідні для роботи алгоритму Рабіна-Карпа для пошуку в дмовимірному тексті. Код згрупований в клас для зручності використання.

• Метод для підрахунку значення хеш-функції в vector<int> pattern для стовпчиків(0…pCols-1) переданого шаблону довжиною pRows-1. Виконується на етамі препроцесингу, тобто перед основним циклом програми. Має асимптотику O(tCols\*pRows)

• Метод для перерахунки значення хеш-функції для стовпчиків TEXT при переході до іншого рядка. Тобто vector<int> text; Асимптотика O(tCols)

• Функція для посимвольної перевірки PATTERN з підмасивом TEXT, що починається з [row,column]. Дана функція викликається у випадку, якщо співпали постовпчикові хеші для pattern і text. Асимптотика O(pCols\*pRows)

• Метод, що будує префікс функцію для заданого шаблони(в даному випадку числового). Використовується в алгоритмі Кнутта-Моріса-Прата

• Алгоритм пошуку підрядка(в даному випадку підмасива) в іншому рядку(масиві) Кнута-Морісса-Прата. Складність в найгіршому випадку O((tRows-pRows)\*pCols)

• Головний метод класу для пошуку зразка PATTERN[][] в тексті TEXT[][].

Складність як в найгіршому, так і найкращому випадку однакова та рівна O(pRows\*pCols+tRows\*tCols)

Тобто квадратична від розміру тексту у випадку квадратних величин

• Далі в Main програмі зчитуємо TEXT та PATTERN з файлів Text.txt та Pattern.txt відповідно

Дані файли мають такий вигляд. Також реалізовані допоміжні файли для зчитування даних з файлу та виводу в консоль. Функції ReadFromFile та Print() відповідно. Та нижче приведені результати виконання тестового прикладу.

• Використані джерела:

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.113.6500&rep=rep1&type=pdf

**https://www.geeksforgeeks.org/rabin-karp-algorithm-for-pattern-searching/**

**https://ru.wikipedia.org/wiki/%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC\_%D0%A0%D0%B0%D0%B1%D0%B8%D0%BD%D0%B0\_%E2%80%94\_%D0%9A%D0%B0%D1%80%D0%BF%D0%B0**

http://e-maxx.ru/algo/export\_prefix\_function

**https://www.geeksforgeeks.org/kmp-algorithm-for-pattern-searching/**