

Київський національний університет імені Тараса  
Шевченка  
Факультет комп'ютерних наук та кібернетики

# Алгоритми та складність

Лабораторна робота №8  
“Порівняння алгоритмів пошуку підрядка в рядку”  
Звіт

Підготувала:  
Студентка групи К-29  
Поліщук Марина

**Завдання:** Реалізувати алгоритми пошуку зразка в текстовому рядку: наївний, Хорспула, Боєра-Мура, КМП та Рабіна-Карпа та порівняти їх ефективність. Виконати пошук зразків різної довжини: випадкового бінарного зразка у випадковому бінарному тексті та випадкового слова у природному тексті на цій мові.

**Наївний алгоритм:** Пошук всіх допустимих зсувів проводиться за допомогою циклу, в якому перевіряється умова  $t[s \dots s + m - 1] = p$  для кожного з  $n - m + 1$  можливих значень. Складність цього алгоритму в середньому  $O(nm)$ .

**Алгоритм Хорспула:** Цей алгоритм являє собою спрощений варіант алгоритму Боєра-Мура. Порівняння відбувається зліва направо. Якщо всі символи шаблону співпали з відповідними символами рядка, отже, підрядок знайдено і пошук закінчено. Якщо якийсь символ не співпадає, то шаблон зсувається на певну кількість символів вправо за певною логікою. Якщо є неспівпадіння, то дивимось на символ рядка над останнім символом шаблону і зсуваємо шаблон так, щоб під цим символом рядка опинився такий же символ шаблону, перше його входження справа наліво. Якщо

шаблон не містить такого символу, то він зсувається вправо на повну довжину і процедура повторюється.

**Алгоритм Боєра-Мура:** Порівняння символів починається з кінця зразка, а не з початку, тобто порівняння відбувається справа наліво. Далі за певною логікою вираховується величина зсуву вправо і знову відбувається порівняння, починаючи з кінця зразка. Складаємо таблицю для шаблону, де кожному елементу відповідає значення, рівне віддаленості відповідного символу шаблону від кінця образу. Якщо символ зустрічається більше одного разу, то йому надається значення, відповідне символу, найбільш близькому до кінця. Якщо символ в кінці шаблону зустрічається один раз - йому відповідає значення, рівне довжині образу, якщо більше одного разу - значення, відповідне символу, найбільш близькому до кінця образу. Для символів, яких немає в шаблоні, застосовується значення, рівне довжині образу.

**Алгоритм Рабіна-Карпа:** Цей алгоритм пошуку шаблону в рядку використовує хешування. Алгоритм показує високу продуктивність на практиці, а також дозволяє узагальнення на інші споріднені задачі. Ідея алгоритму полягає в заміні текстових рядків числами, порівняння яких можна виконувати значно швидше. При правильному виборі хеш-функції його середній і кращий час виконання дорівнює  $O(n)$ , в найгіршому випадку цей алгоритм працює за  $O(nm)$ . Цей алгоритм також часто застосовується у виявленні плагіату.

**Алгоритм Кнута-Морріса-Пратта:** Складається префікс-таблиця для шаблону. Зсув шаблону відбувається згідно з цією таблицею. Час роботи цього алгоритму лінійно залежить від об'єму вхідних даних, тобто розробити асимптотично більш ефективний алгоритм неможливо.