

Лабораторная работы № 5 по курсу дискретного анализа:

Суффиксные деревья

Выполнил студент группы 08-210 МАИ *Некрасов Константин*

Условие

Реализовать поиск подстрок в тексте с использованием суффиксного дерева. Суффиксное дерево можно построить за $O(n^2)$ наивным методом.

Текст располагается на первой строке, затем, до конца файла, следуют строки с образцами.

Для каждого образца, найденного в тексте, нужно распечатать строчку, начинающуюся с последовательного номера этого образца и двоеточия, за которым, через запятую, нужно перечислить номера позиций, где встречается образец в порядке возрастания.

Метод решения

Создан класс суффиксного дерева, построение которого осуществляется с помощью алгоритма Укконена.

Поиск вхождений подстрок осуществляется с помощью метода `Search`, в котором в вектор рекурсивно собираются все вхождения подстроки в текст.

Описание программ

В функции `main` происходит считывание строки, к которой добавляется терминальный символ `$`. Затем вызывается метод `Build` класса `TTrie`, который строит суффиксное дерево на основе строки.

После построения дерева осуществляется многократный поиск подстрок, вводимых из стандартного ввода. Для каждой подстроки вызывается метод `Search`, который возвращает индексы всех вхождений подстроки в исходную строку.

Дневник отладки

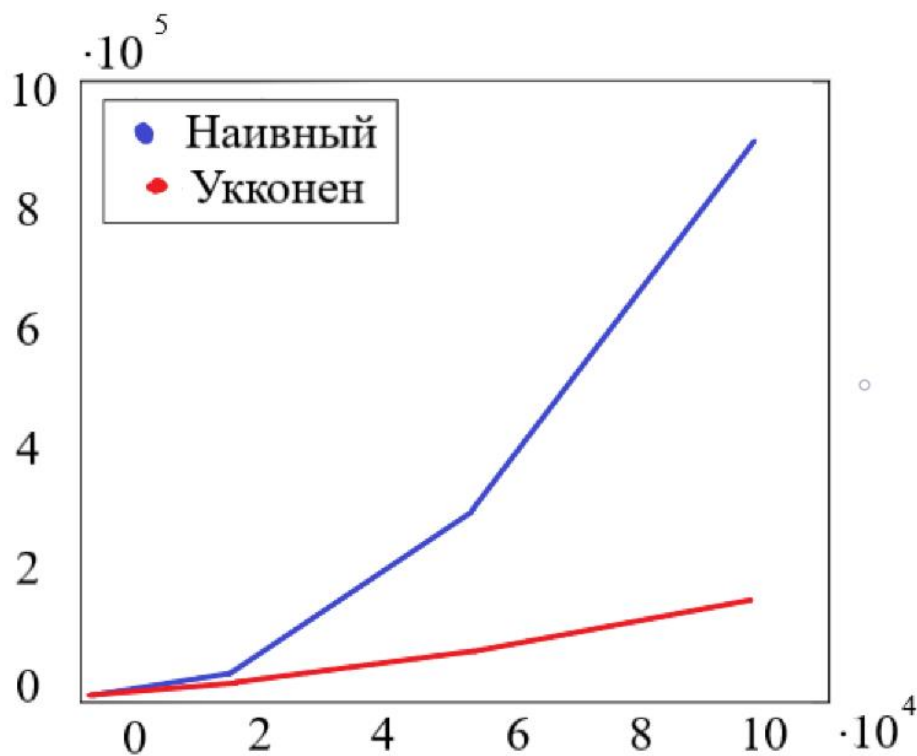
Было две посылки, обе из них успешные, но во второй исправлено оформление кода.

Тест производительности

Синим цветом обозначен наивный алгоритм построения суффиксного дерева, красный – алгоритм Укконена.

По оси X – количество символов в тексте.

По оси Y – время построения в миллисекундах.



На малых объёмах данных различия незначительны, но с увеличением входных данных, построение с помощью Укконена значительно эффективнее.

Недочёты

В процессе выполнения работы недочёты не были выявлены.

Выводы

В данной лабораторной работе я на практике применил алгоритм Укконена для построения суффиксного дерева и использовал суффиксное дерево для поиска подстрок в тексте, и убедился в эффективности дерева для этой задачи.