

Задание от 22 февраля 2013г.

Срок сдачи — начало четвёртой лекции.

Задача 1.

В этой задаче мы несколько подробнее рассмотрим идею связей выхода (output links).

а) Связь выхода в вершине v указывает на вершину, достижимую из v за наименьшее число переходов по связям неудачи. Возможно, что для каких-то вершин связей выхода не существует. Опишите, как построить все связи выхода за время $O(n)$ на этапе построения бора.

б) Покажите, что с помощью связей выхода можно найти в тексте все k вхождений образцов за $O(m + k)$.

в) Покажите, как найти самое раннее вхождение одного из образцов в текст: из всех позиций в тексте, в которых *начинается* вхождение какого-то из образцов, нужно выбрать минимальную.

Задача 2.

На лекции мы рассматривали такой способ хранения бора: в каждой вершине хранится массив размера $|\Sigma|$ указателей на переходы из этой вершины по каждому из $|\Sigma|$ символов алфавита (возможно, каких-то переходов не существует, но массив всё равно имеет фиксированный размер).

Вот ещё один способ хранения: в каждой вершине будет храниться указатель на её первого ребёнка и на её следующего брата (или сестру. То есть на следующего ребёнка у её родителя. Упорядочивание детей каждой вершины производится произвольным образом), а также символ, по которому в эту вершину осуществляется переход из её родителя.

а) Покажите, что этой информации достаточно для реализации алгоритма Ахо-Корасик.

б) Сравните два описанных способа.