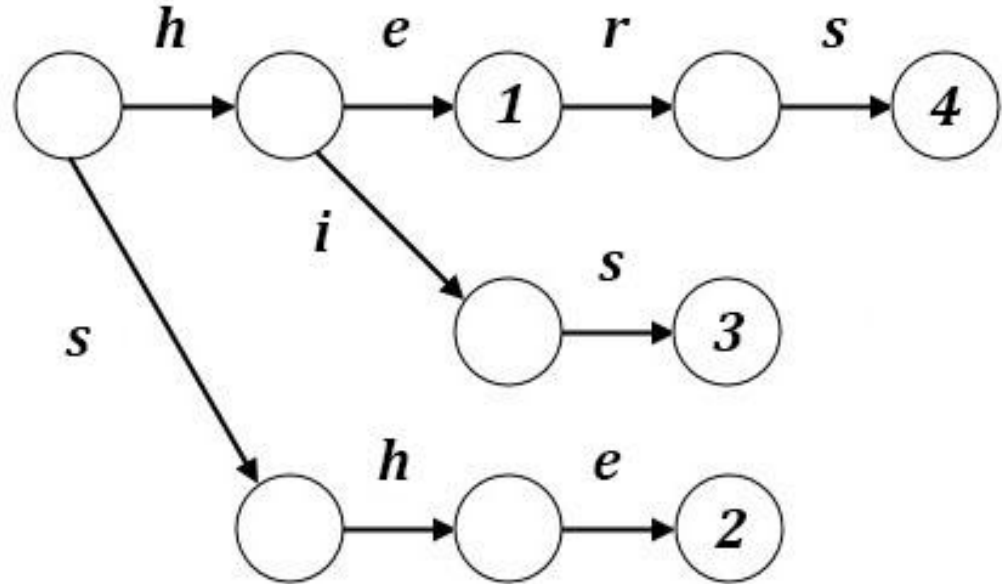


План занятия (13.10.17)

- Суффиксное дерево

Напоминание. Строковые СД

- Бор



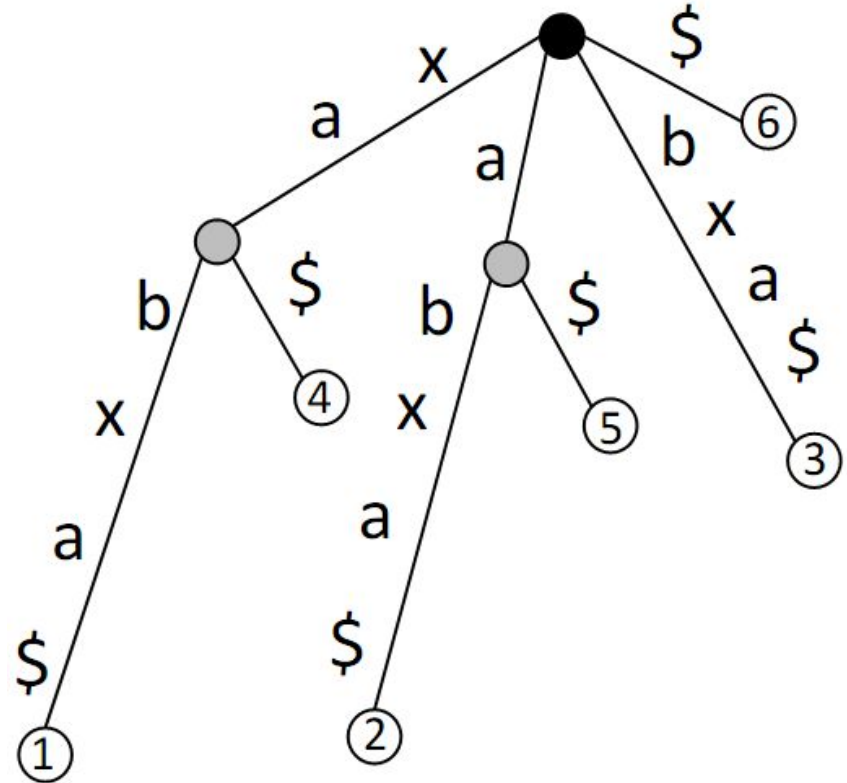
Суффиксное дерево для строки s

1. содержит все суффиксы строки $|s|$, причем каждый суффикс заканчивается в листе и только в нем
2. имеет ровно $|s|$ листьев
3. каждая внутренняя вершина имеет не менее 2 детей
4. каждое ребро помечено непустой подстрокой s
5. никакие два ребра, выходящие из одной вершины, не имеют общего начала

Суффиксное дерево. Пример

Строка *xabxa*

Добавление \$



Суффиксное дерево

- Память
 - Лемма: количество внутренних вершин не превышает количество листьев
 - $O(N)$
- Время
 - Наивный алгоритм $O(N^2)$
 - Алгоритм Укконена $O(N)$

Поиск вхождений образца в тексте

Найти *pattern* в *text*

- КМП
 - $O(|pattern|)$ препроцессинг
 - $O(|text|)$ поиск
- Суфф. дерево
 - $O(|text|)$ препроцессинг
 - $O(|pattern| + K)$ поиск, где K - число вхождений

Поиск наибольшей общей подстроки

1. Построение обобщенного суфф. дерева для строк $s1$, $s2$:
 - a. $s = s1\$s2\#$
 - b. строим суфф. дерево для s
 - c. удаляем “синтетические” суффиксы
2. Выставляем метки для внутренних вершин
3. Ищем внутреннюю вершину с двумя метками наибольшей строковой глубины

Поиск максимальных палиндромов

- Определение максимального палиндрома:
 - aabactgaaccaat
 - a**aba**ctgaaccaat, радиус 1, нечетный
 - aabactg**aacca**at, радиус 3, четный

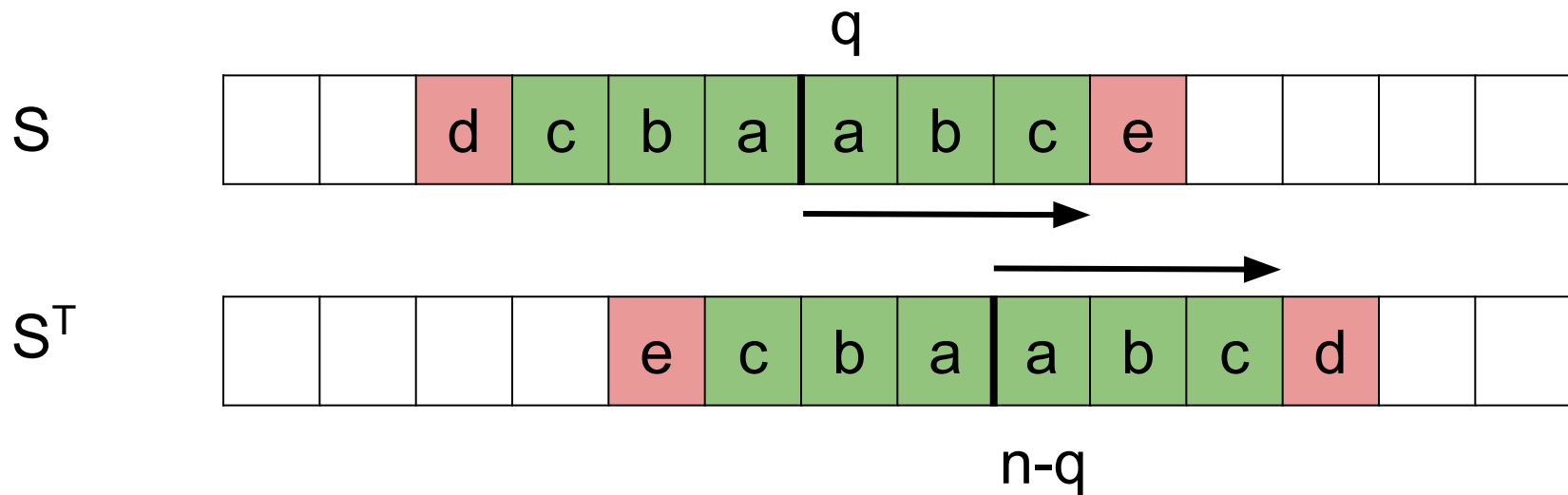
Поиск максимальных палиндромов

Разберем случай палиндромов четной длины.

Общая схема алгоритма:

1. Построение суфф. дерева для строки $S\$S^T$
2. Подготовка дерева к запросам lca
3. Нахождение палиндромов

Поиск максимальных палиндромов



Задача - научиться искать максимальное общее продолжение

Поиск максимальных палиндромов

Приходит запрос (i, j) :

найти наибольшее общее продолжение
для i -ого суффикса S и $(n-i)$ -ого суффикса S^T

Решение:

- каждому суффиксу соответствует лист \rightarrow найдем наименьшего общего предка (LCA)
- LCA \rightarrow ± 1 RMQ
- Для каждой внутренней вершины еще нужно предсчитать строковые глубины

Поиск максимальных палиндромов

Осталось только проитерироваться по q

Время работы:

- Суфф. дерево - $O(N)$ (Укконен)
- Предсчет RMQ - $O(N)$ (Фарак-Колтон-Бендер)
- Строковые глубины - $O(N)$ (Обход в глубину)
- Итерация по q - $O(N)$ (на каждой итерации $O(1)$)

Итого $O(N)$

Что почитать

1. Сходить на лекцию - послушать Укконена
2. https://neerc.ifmo.ru/wiki/index.php?title=Сжатое_суффиксное_дерево
3. Дэн Гасфилд. Строки, деревья и последовательности в алгоритмах