

Алгоритмы и структуры данных

Суффиксные деревья

Кухтичев Антон



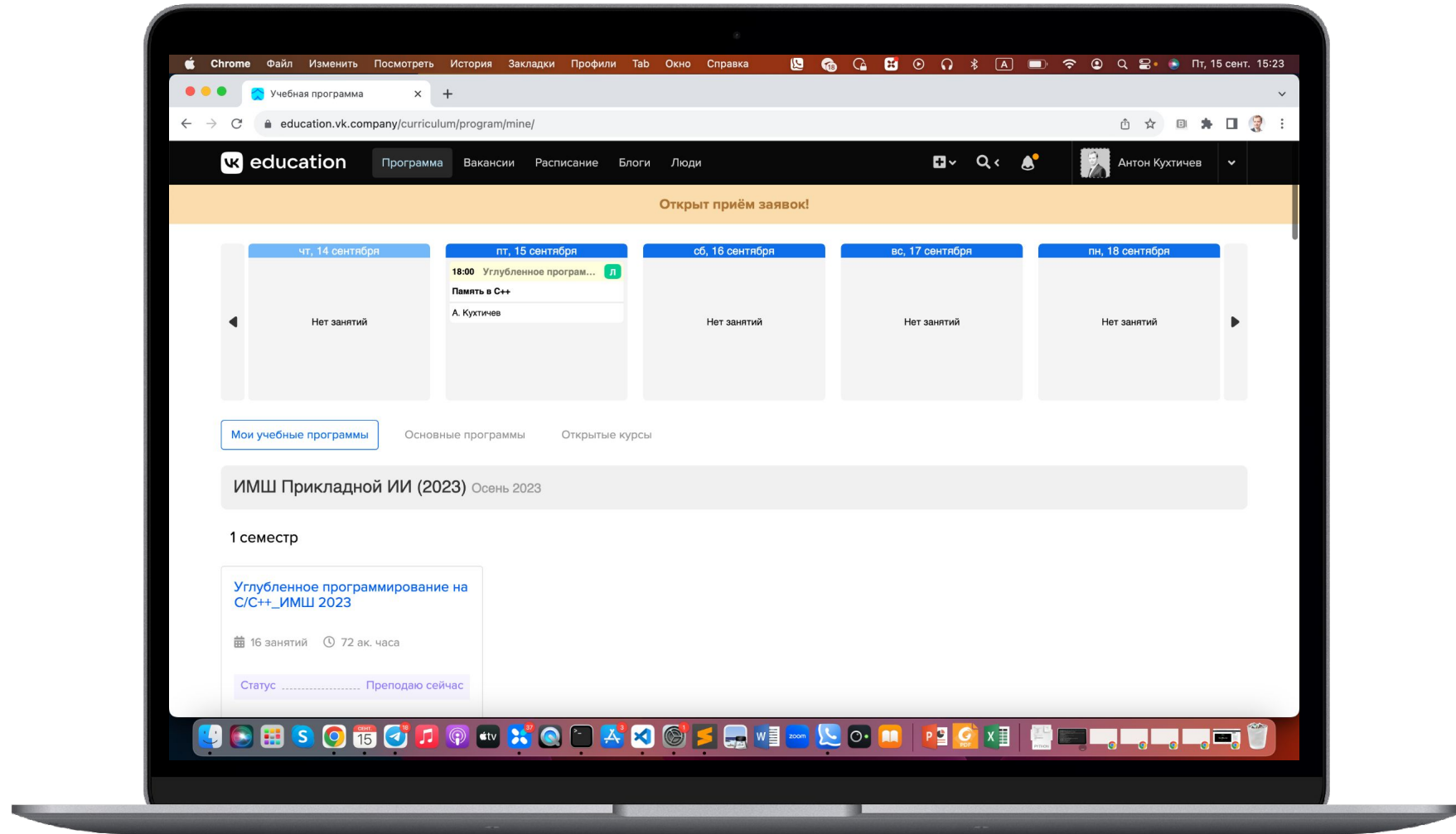
15 марта 2024 года

Содержание занятия

- Квиз
- Суффиксные деревья
- Наивный алгоритм
- Алгоритм Укконена

Напоминание отметиться на портале

и оставить отзыв
после лекции



Определение суффиксного дерева



Определение

Суффиксное дерево T для m -символьной строки S - это ориентированное дерево с корнем, в котором:

- Имеющее ровно m листьев, занумерованных от 1 до m .
- Каждая внутренняя вершина, отличная от корня, имеет не меньше двух детей, а каждая дуга помечена непустой подстрокой строки S (дуговой меткой).
- Никакие две дуги, выходящие из одной и той же вершины, не могут иметь пометок, начинающихся с одного и того же символа.

Пример суффиксного дерева

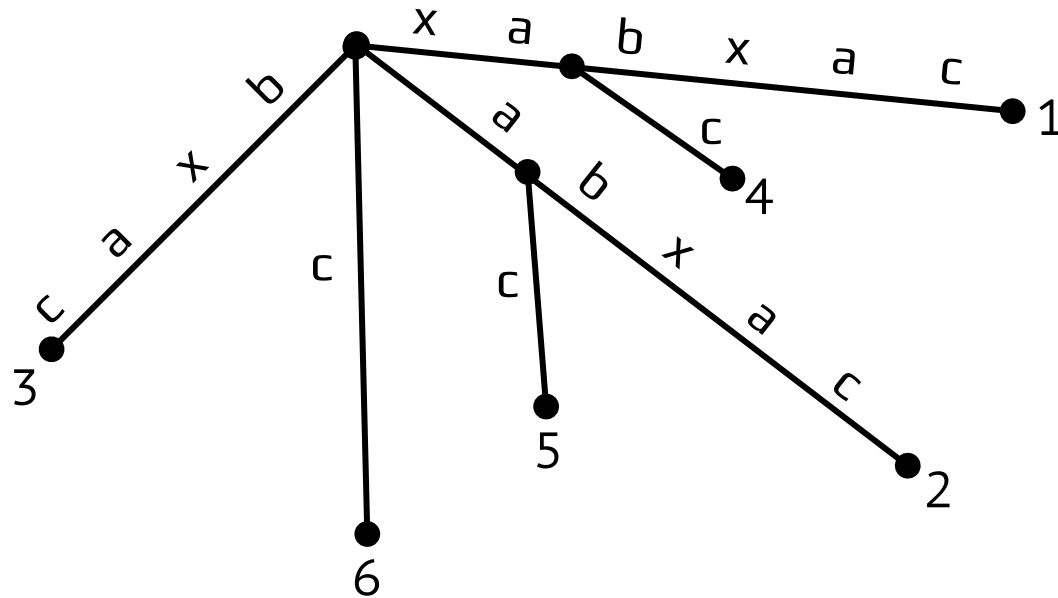
Рассмотрим строку $S = \text{“хавхас”}$

Суффиксы:

- хавхас
- авхас
- вхас
- хас
- ас
- с



Пример суффиксного дерева

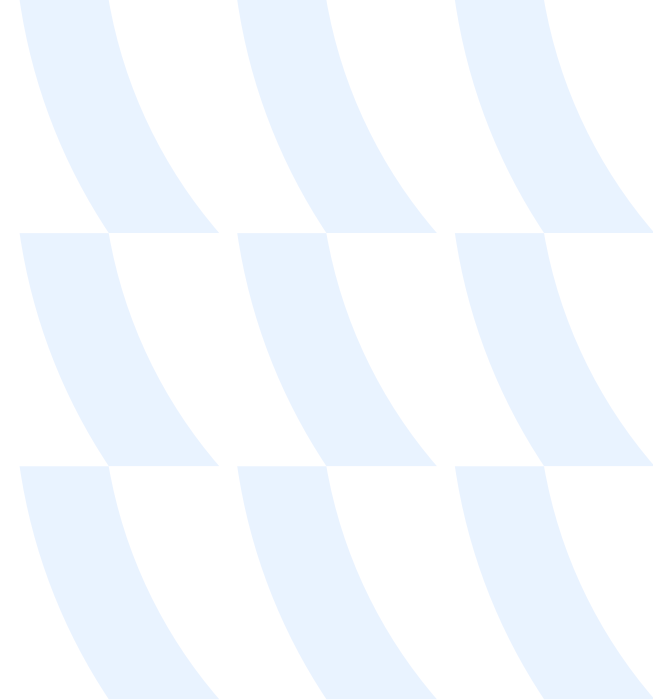


Другой пример

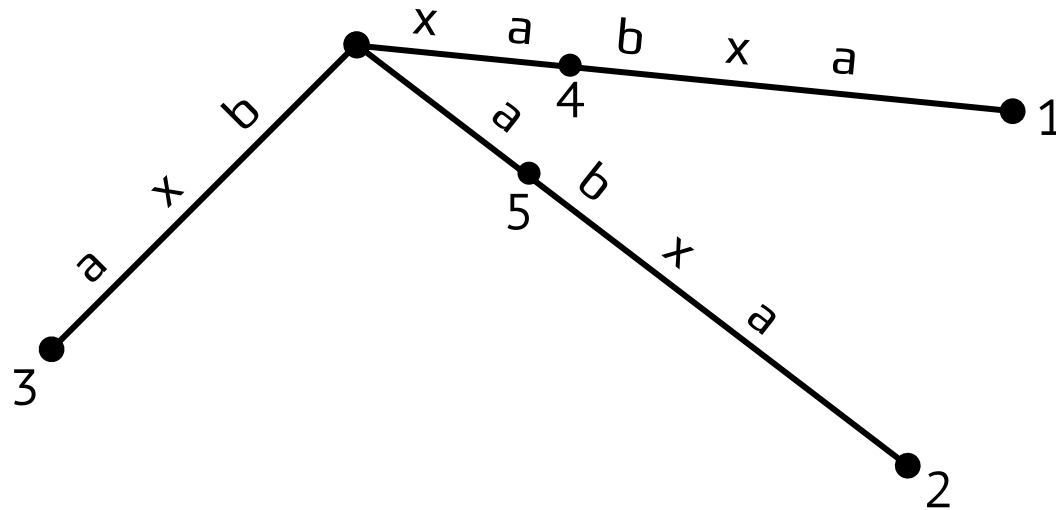
Рассмотрим строку $S = \text{“xabxa”}$

Суффиксы:

- xabxa
- abxa
- bxa
- xa
- a

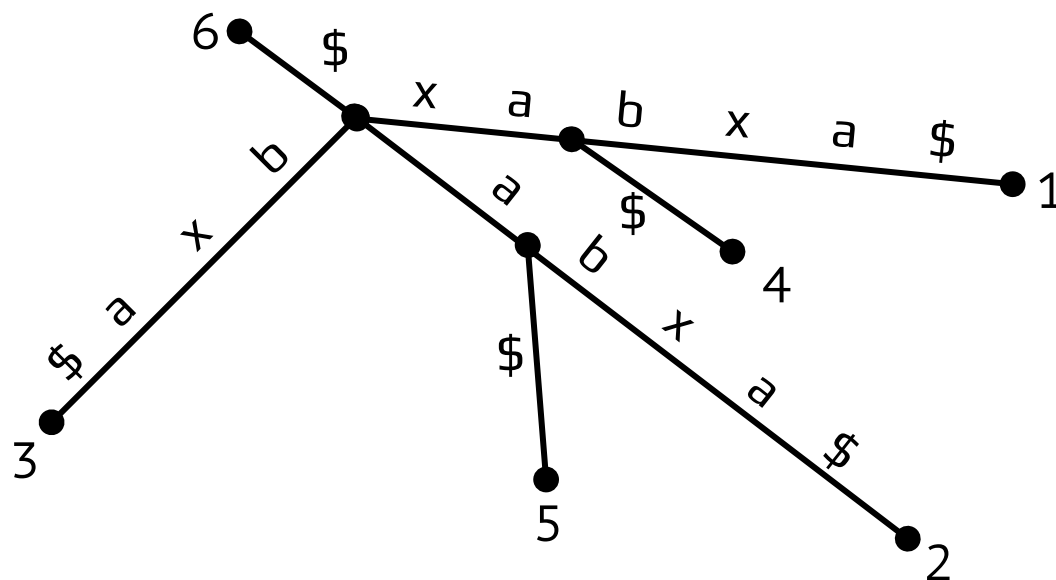


Другой пример суффиксного дерева



Неявное суффиксное дерево

Другой пример суффиксного дерева



Явное суффиксное дерево

Наивный алгоритм построения суффиксного дерева



Наивный алгоритм

- Работает за $O(n^2)$
- Построение за $O(n)$, где n - длина текста



Наивный алгоритм



Алгоритм Укконена



Алгоритм Укконен

- Работает за $O(n+m+k)$
- Построение за $O(n)$, где n - длина текста
- Алгоритм Укконена делиться на m фаз



Правила продолжения суффикса

1. В текущем дереве путь кончается в листе. Нужно добавить к концу метки этой листовой дуги символ;
2. Ни один путь из конца строки не начинается символом. но по крайней мере один начинающийся оттуда путь имеется. В этом случае должна быть создана новая листовая дуга;
3. Некоторый путь из конца строки начинается символом. Строка уже имеется в текущем дереве, так что ничего не надо делать.

Пример



Ускорение

1. Суффиксные ссылки
2. Скачок по счётчику
3. Сжатие дуговых меток
4. Заканчивать каждую фау $i+1$ после первого же использования правила продолжения 3
5. Хранение глобального символа, обозначающий “текущий конец”



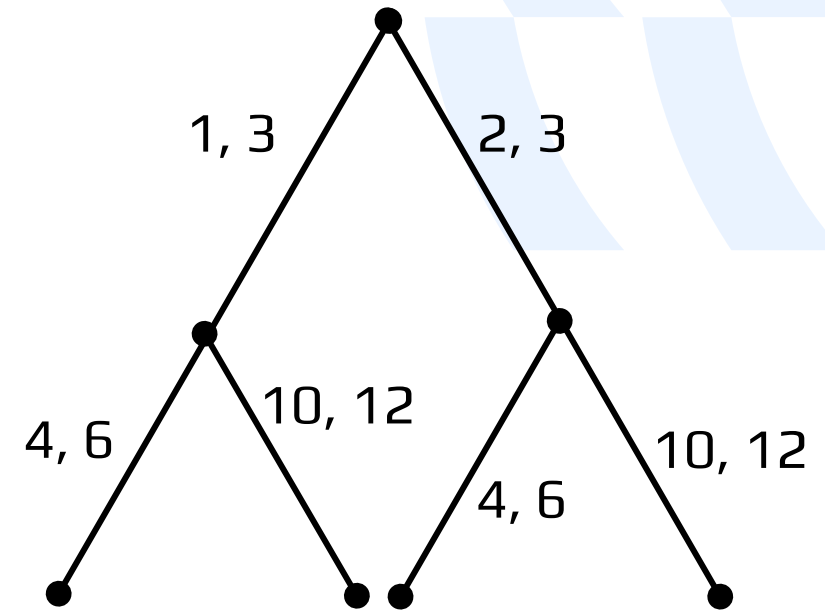
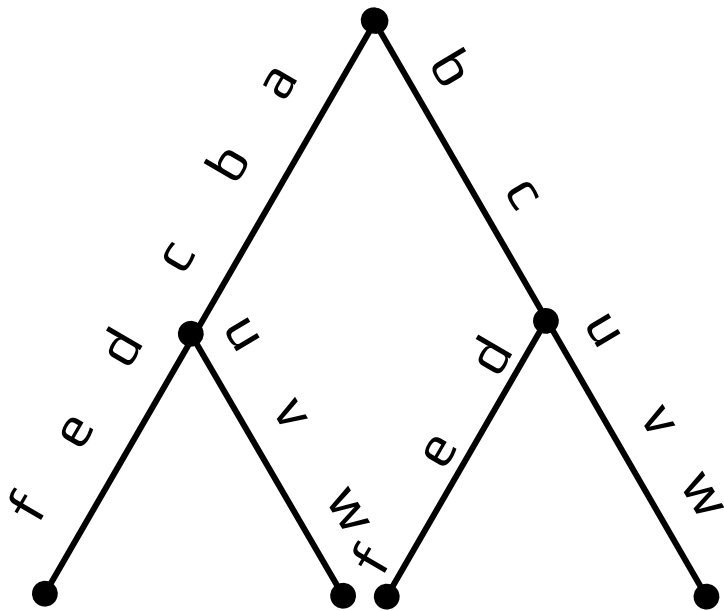
Суффиксные ссылки



Скачок по счётчику



Сжатие дуговых меток

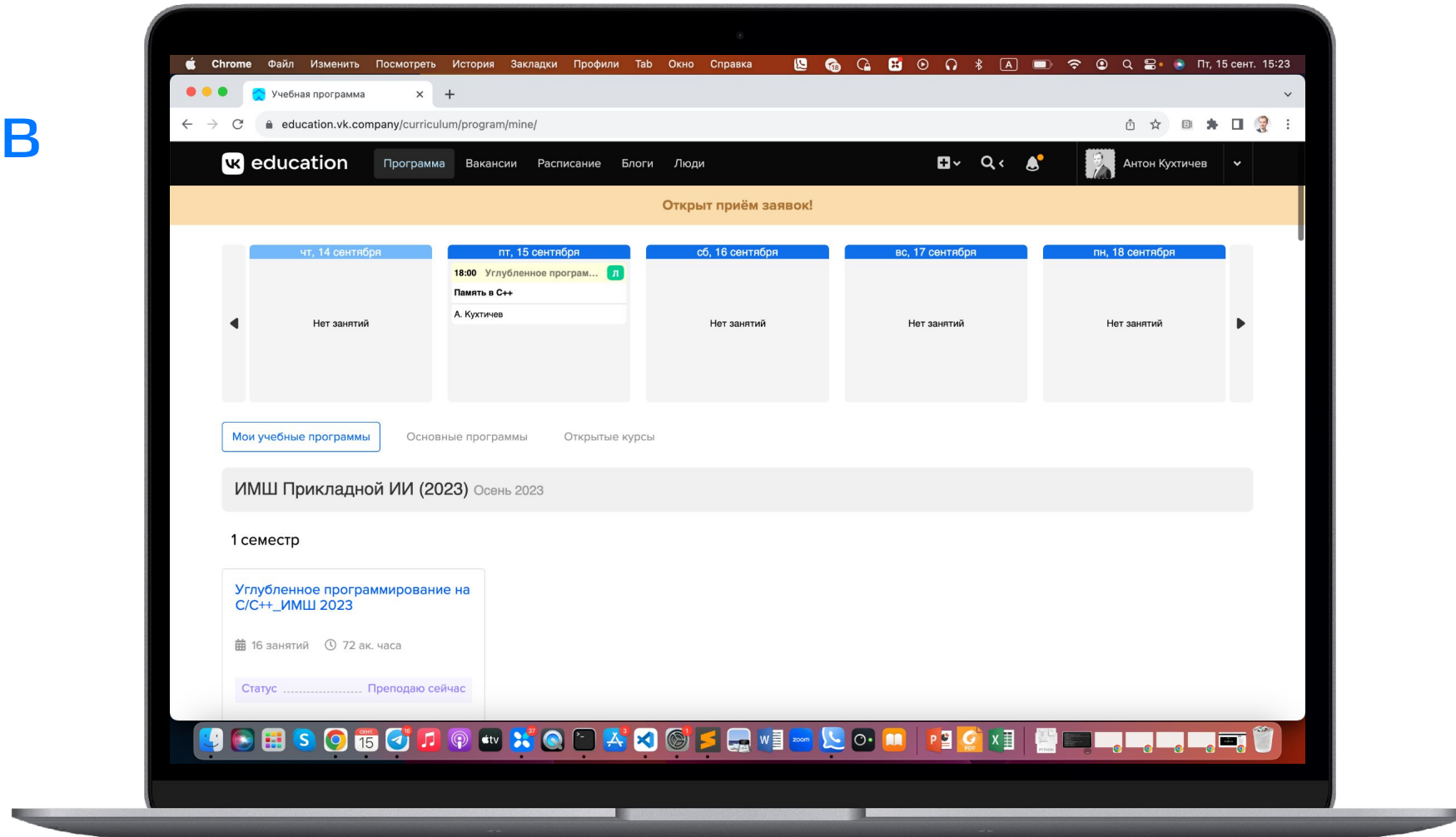


Всё вместе



Напоминание оставить отзыв

Это правда важно





Спасибо
за внимание!