

Алгоритмы и структуры данных

Суффиксные деревья

Кухтичев Антон



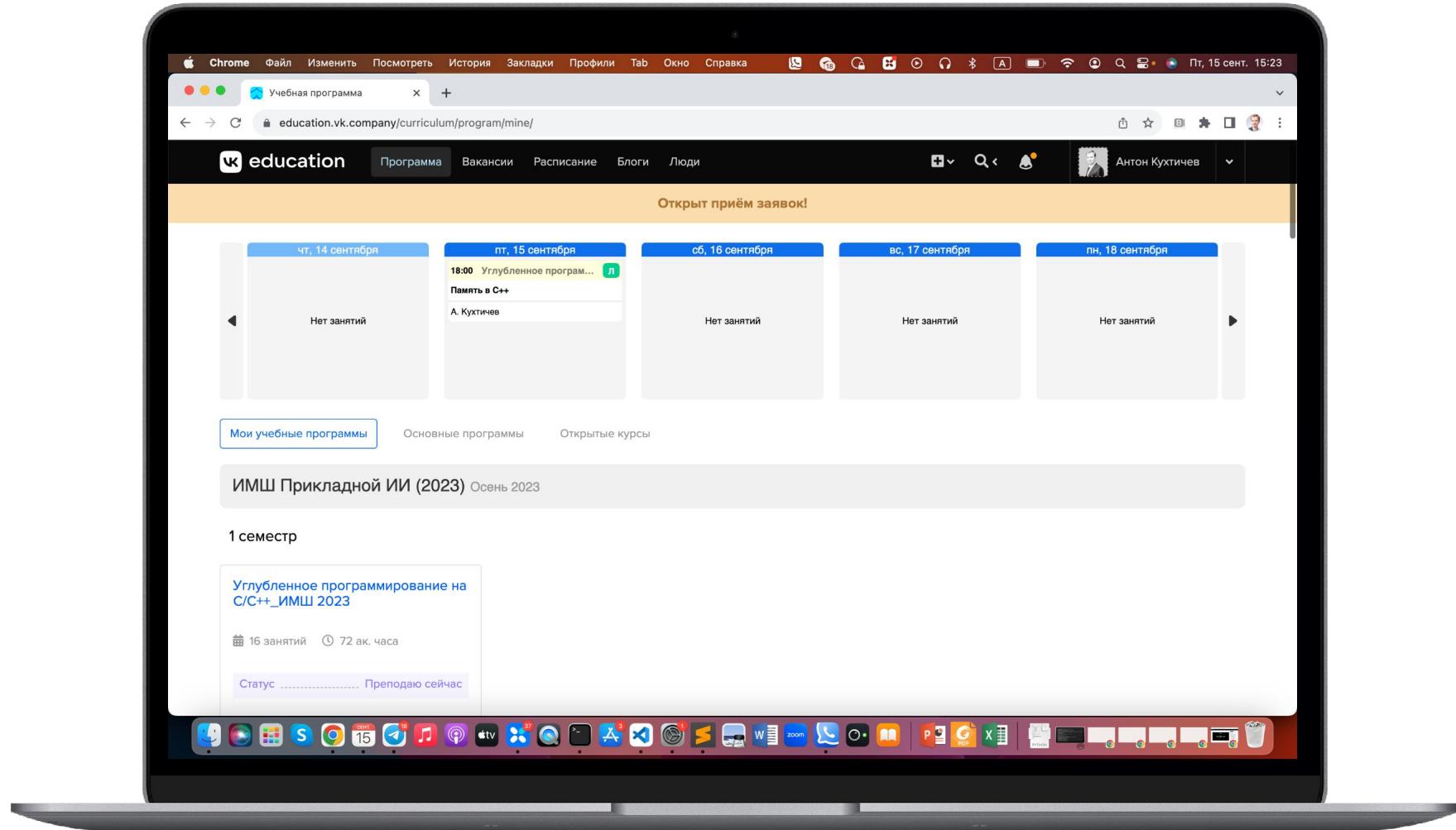
7 апреля 2025 года

Содержание занятия

- Квиз
- Сuffixные 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.
- Литература: Гасфилд: Строки, деревья и последовательности в алгоритмах

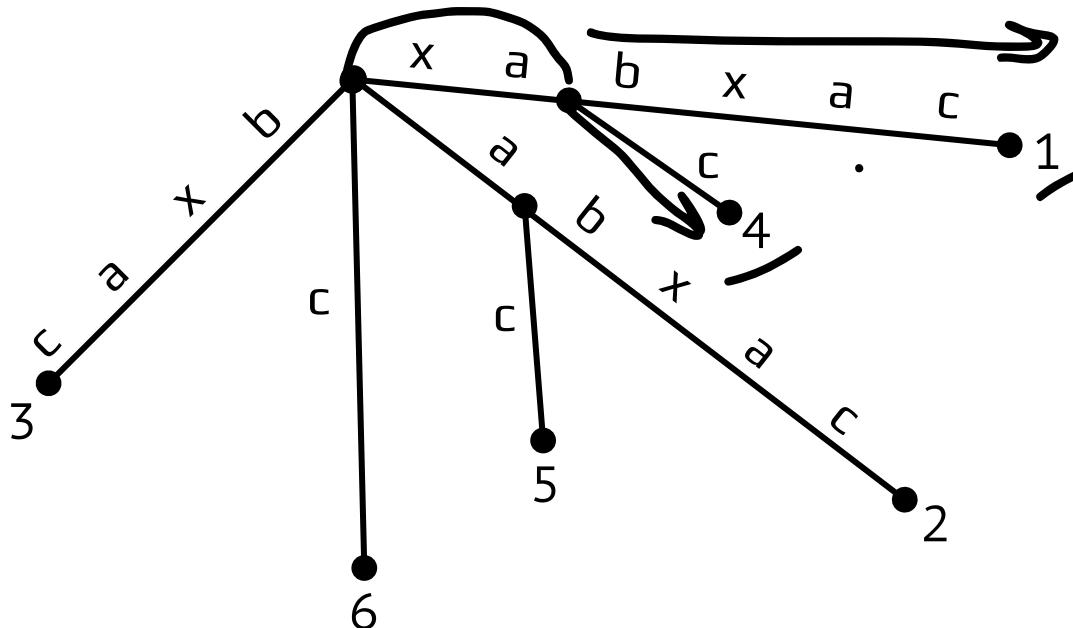
Пример суффиксного дерева

Рассмотрим строку **S** = “**хабхас**”

Суффиксы:

- хабхас
- абхас
- бхас
- хас
- ас
- с

Пример суффиксного дерева



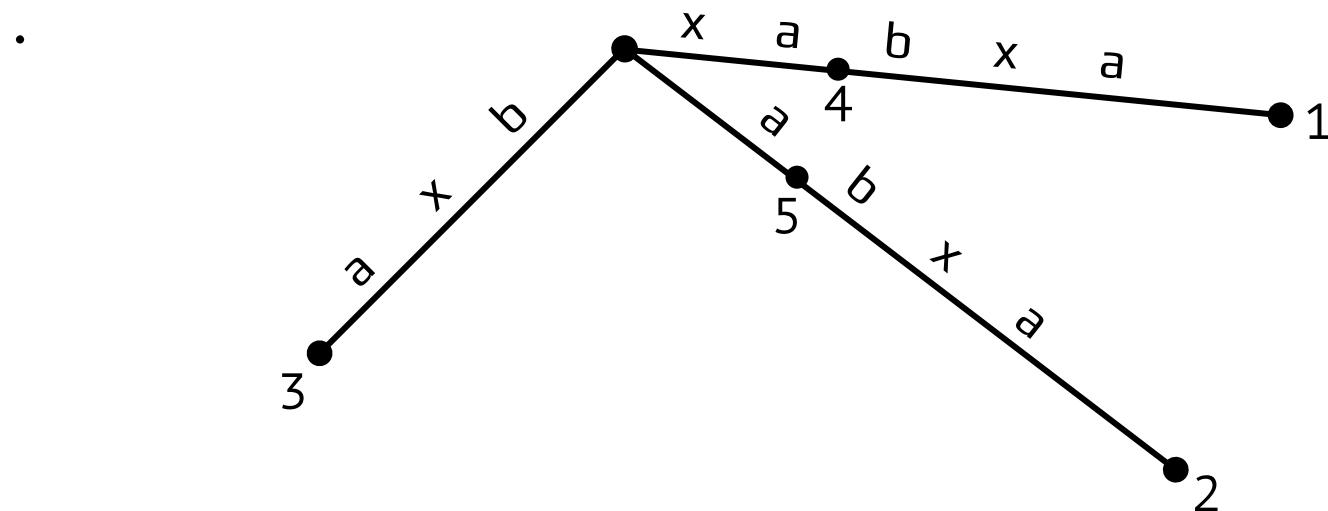
Другой пример

Рассмотрим строку **S** = “**хабха**”

Суффиксы:

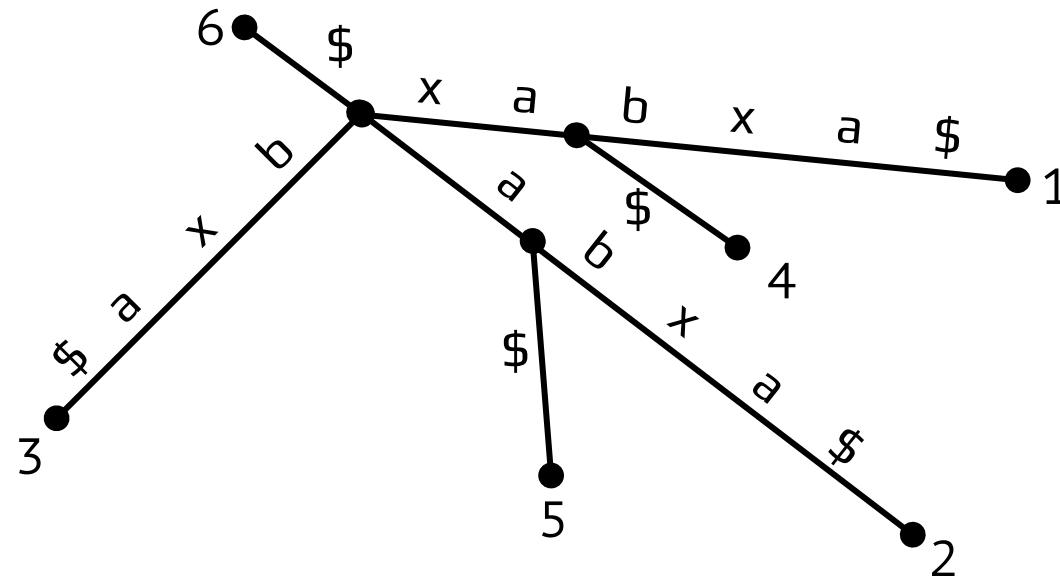
- **хабха**
- **абха**
- **бха**
- **ха**
- **а**

Другой пример суффиксного дерева



Неявное суффиксное дерево

Другой пример суффиксного дерева



Явное суффиксное дерево

Наивный алгоритм построения суффиксного дерева

• • • • •

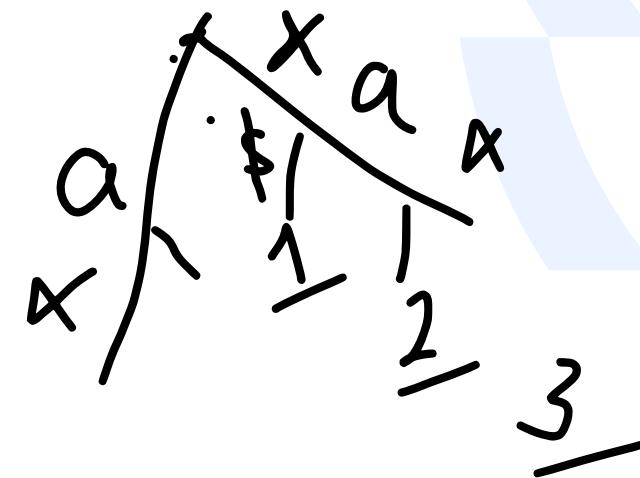
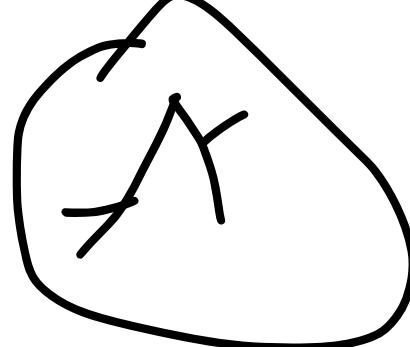
Наивный алгоритм

- Построение за $O(n^2)$
- Поиск за $O(m+k)$, где m - длина текста



Наивный алгоритм

$$\begin{aligned}
 & n \\
 & n-1 \\
 & \vdots \\
 O\left(\sum_{i=1}^n i^2\right) & = O(n^2) \\
 O(n^2 + m + k) & \xrightarrow{k \leq n} O(n^2)
 \end{aligned}$$

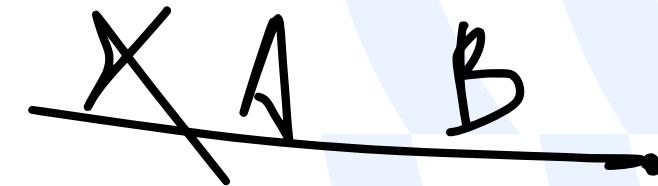


Алгоритм Укконена



Алгоритм Укконен

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



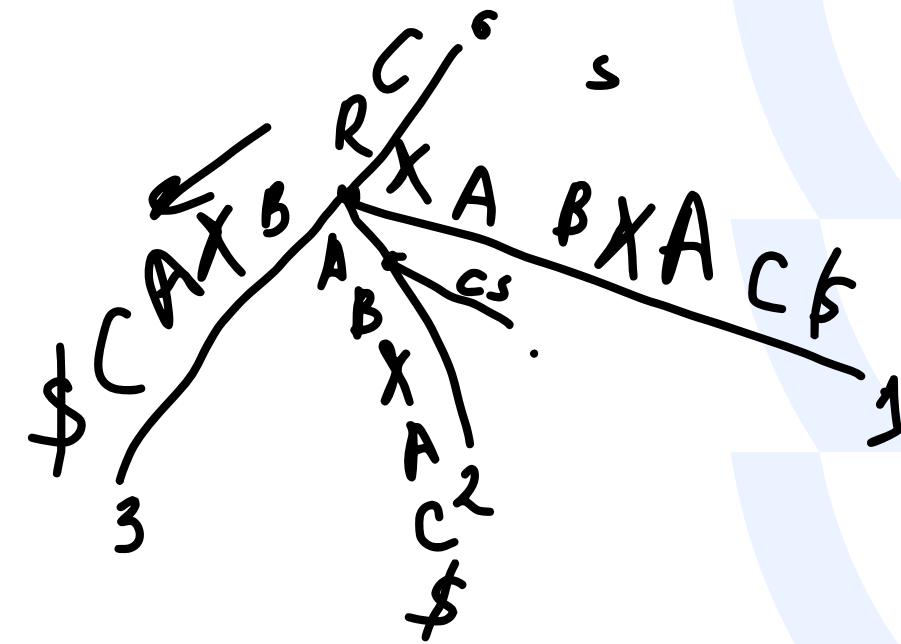
$\overbrace{X \overline{AB} XAC}$
 X $XAB \overbrace{XAC}$ } n^2
 XA $A \overbrace{B XAC}$
 A $\vdots \vdots$
 XAB
 AB
 B

Правила продолжения суффикса

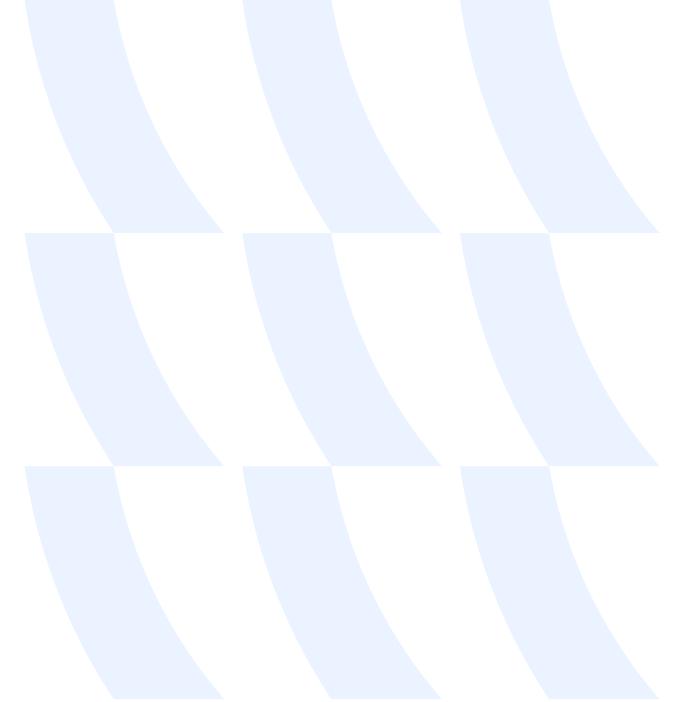
1. В текущем дереве путь кончается в листе. Нужно добавить к концу метки этой листовой дуги символ;
2. Ни один путь из конца строки не начинается символом. но по крайней мере один начинающийся оттуда путь имеется. В этом случае должна быть создана новая листовая дуга;
3. Некоторый путь из конца строки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

Пример

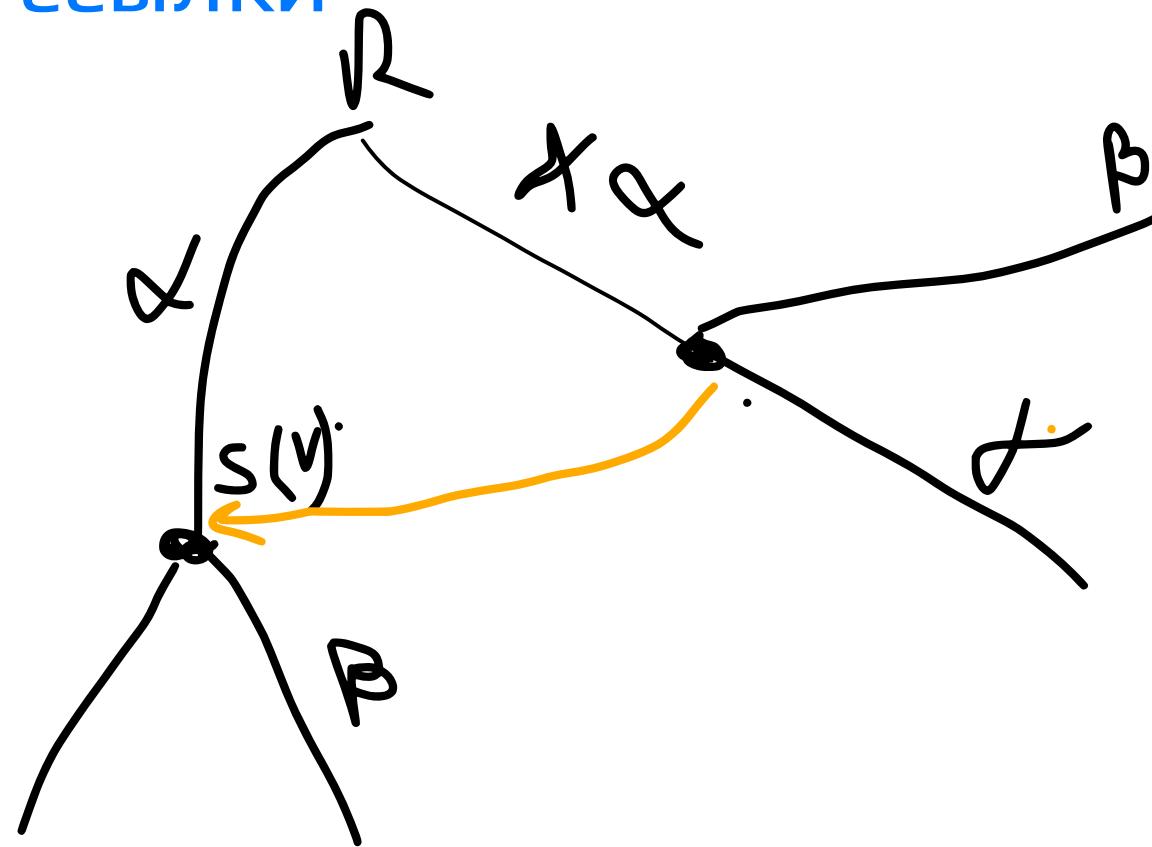
XABXAC ₽



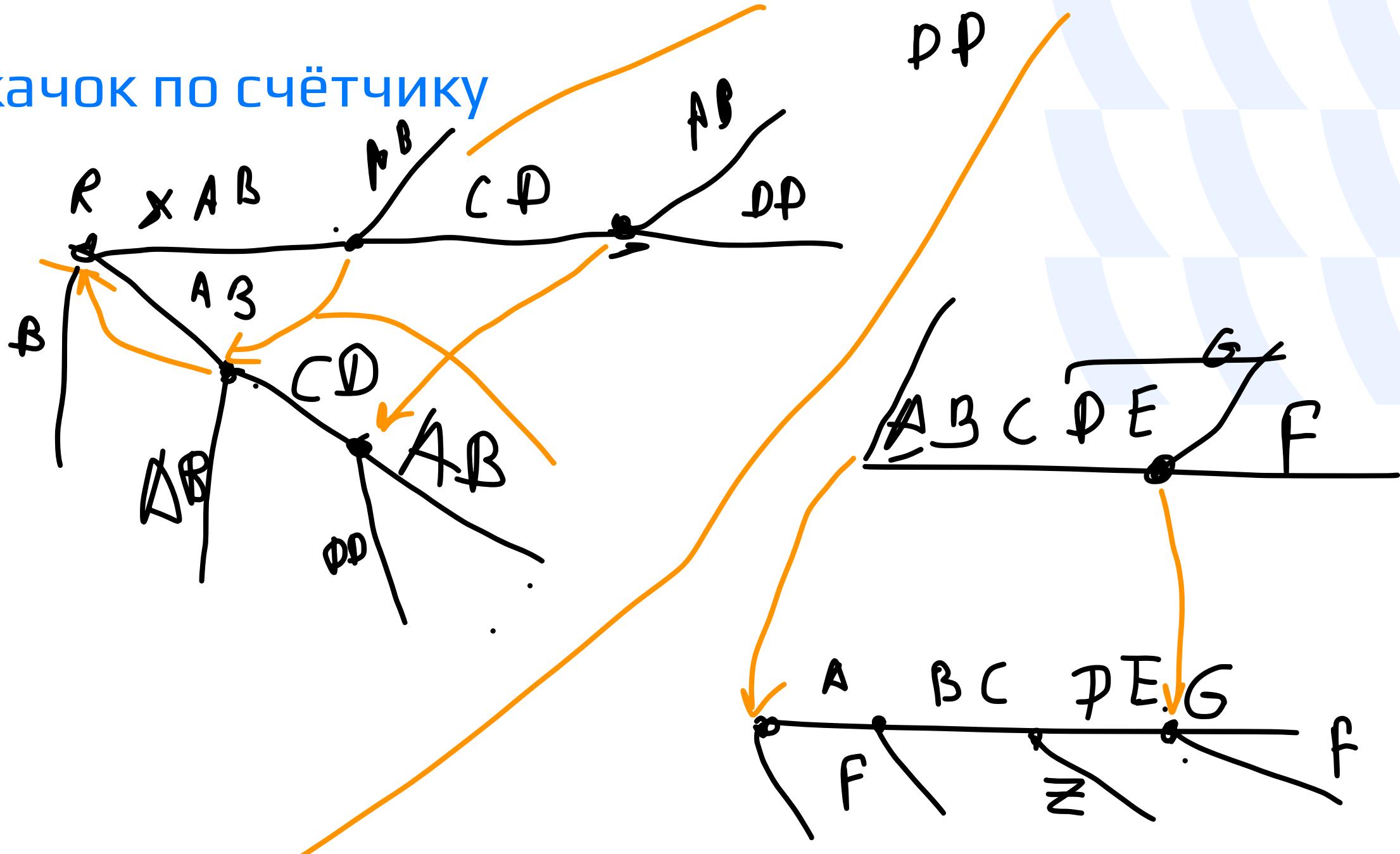
Ускорение

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

Суффиксные ссылки

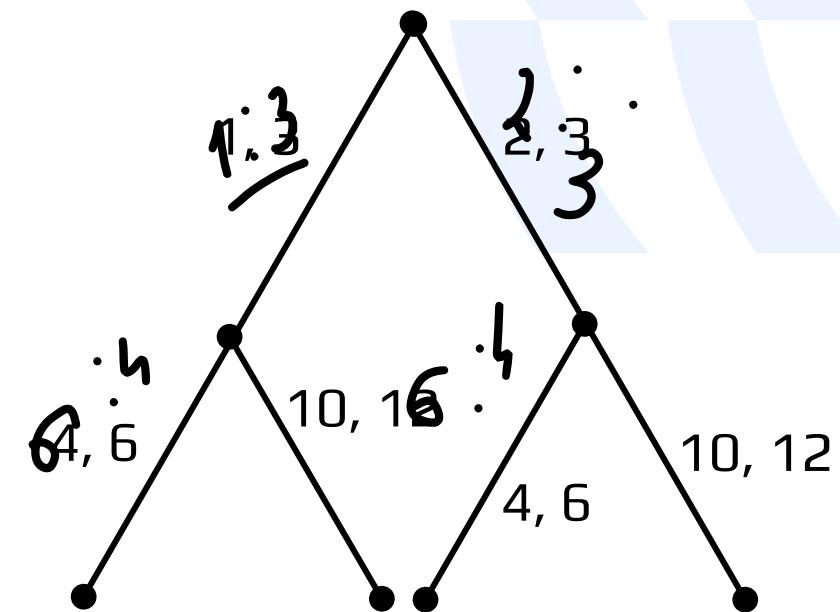
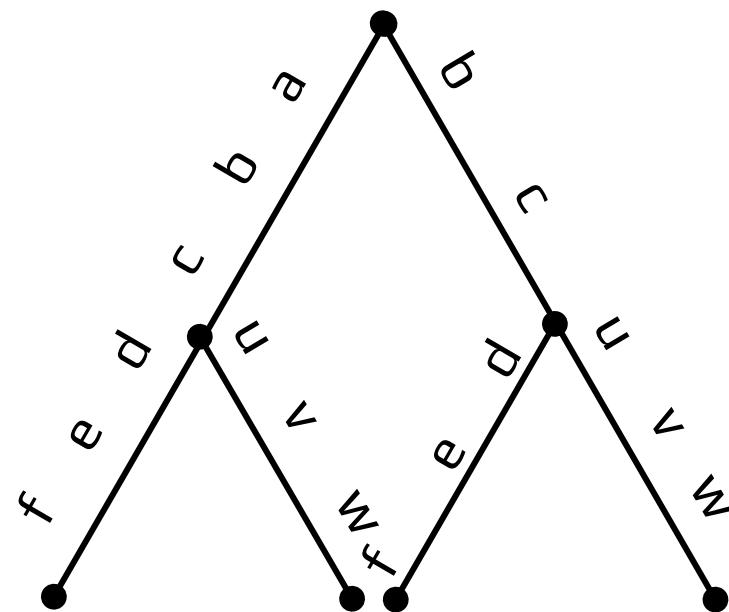


Скачок по счётчику



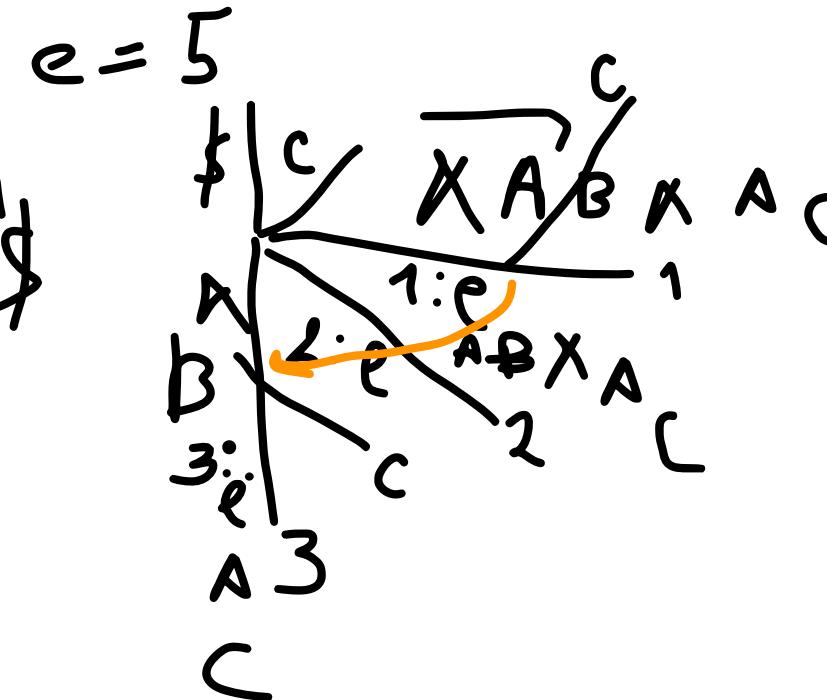
Сжатие дуговых меток

a b c d e f a b c u v w



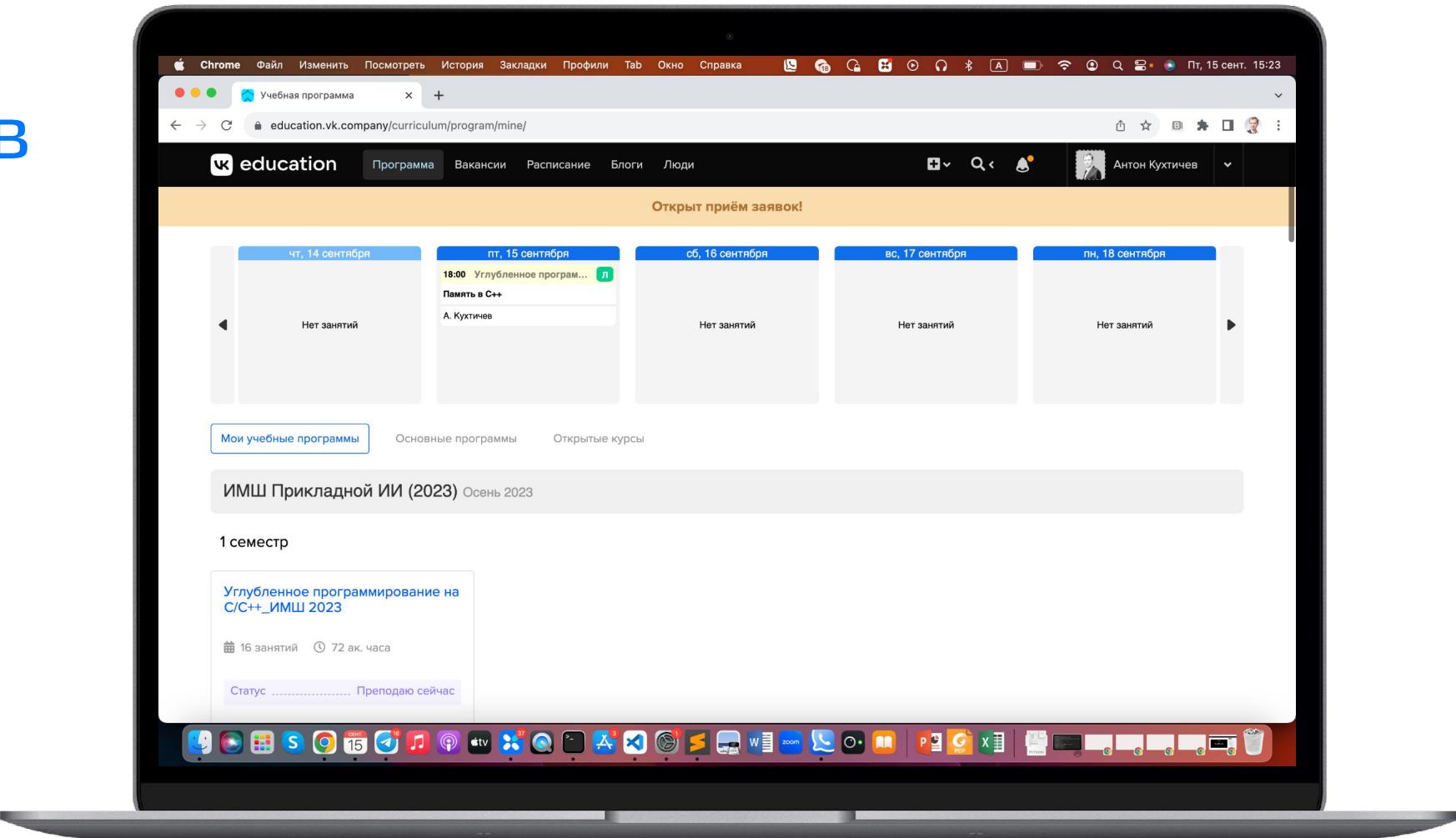
Всё вместе

XAB XAC



Напоминание оставить отзыв

Это правда важно





Спасибо
за внимание!