

# Ахо-Корасик

Гусев Илья

Московский физико-технический институт

Москва, 2018

# Содержание

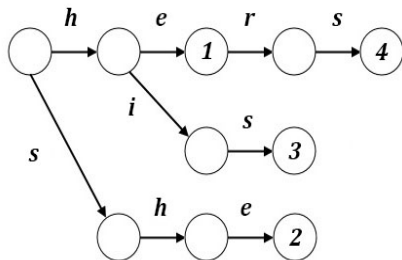
- 1 Алгоритм
  - Задача
  - Бор
  - Алгоритм
- 2 Задачи
  - Вопросыки
  - Вирусы

# Задача

Дан набор строк  $T_0, T_1, \dots, T_m$ , которые мы назовём шаблонами и строка  $S$ .  
Нужно найти все вхождения  $T_i$  в  $S$ .

# Бор

## Структура данных



Шаблоны:

- hers
- his
- she
- he

# Суффиксные ссылки

$parent(u)$  - родитель  $u$

$\pi(u) = \delta(\pi(parent(u)), c)$  - суффиксная ссылка

$\delta(u, c) = v$ , если из  $u$  есть переход в  $v$  по  $c$

$\delta(u, c) = root$ , если  $u$  - корень и из него нет перехода по  $c$

$\delta(u, c) = \delta(\pi(u), c)$ , иначе

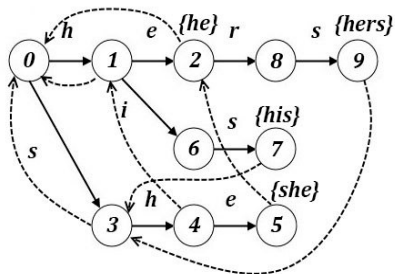
# Сжатые суффиксные ссылки

$ur(u) = \pi(u)$ , если  $\pi(u)$  - терминальная вершина

$ur(u) = \emptyset$ , если  $\pi(u)$  - корень

$ur(u) = ur(\pi(u))$ , иначе

# Автомат



Шаблоны:

- hers
- his
- she
- he

# Вопросики

Поиск шаблона ab??ceee?dg



# Вирусы

Есть вирусы - конечные пути  $V_1, \dots, V_k$ .

Существует ли бесконечно много 'здоровых' файлов?

# Полезные ссылки I



Викиконспекты: Ахо-Корасик

[http://neerc.ifmo.ru/wiki/index.php?title=Алгоритм\\_Ахо-Корасик](http://neerc.ifmo.ru/wiki/index.php?title=Алгоритм_Ахо-Корасик)



E-maxx: Ахо-Корасик

[http://www.e-maxx-ru.1gb.ru/algo/aho\\_corasick](http://www.e-maxx-ru.1gb.ru/algo/aho_corasick)