

# Алгоритм Ахо-Корасика

Занятие 21 в ФМЛ 5 г Долгопрудного.

# Задача

- Есть набор строк  $S[1..n]$
- Есть текст  $T$
- За линейное время  
( $O(\text{size}(T) + \sum_i (\text{Size} S[i]))$ )  
найти все места где строки встречаются в  
тексте.

# Алгоритм Ахо-Корасик

# Устройство алгоритма

- Построение Бора
- Построение конечного автомата из бора
- Обработка строк на конечном автомате
-

# Применение

- Grep
- Antiviruses
- Анализ генетических последовательностей

# Разберемся для начала с конечными автоматами

- Что такое конечный автомат?

# Разберемся для начала с конечными автоматами

- Что такое конечный автомат?
- 5-ка (5-tuple)

$\{ Q, \Sigma, P:(Q \times \Sigma \rightarrow Q), q_0, F \}$

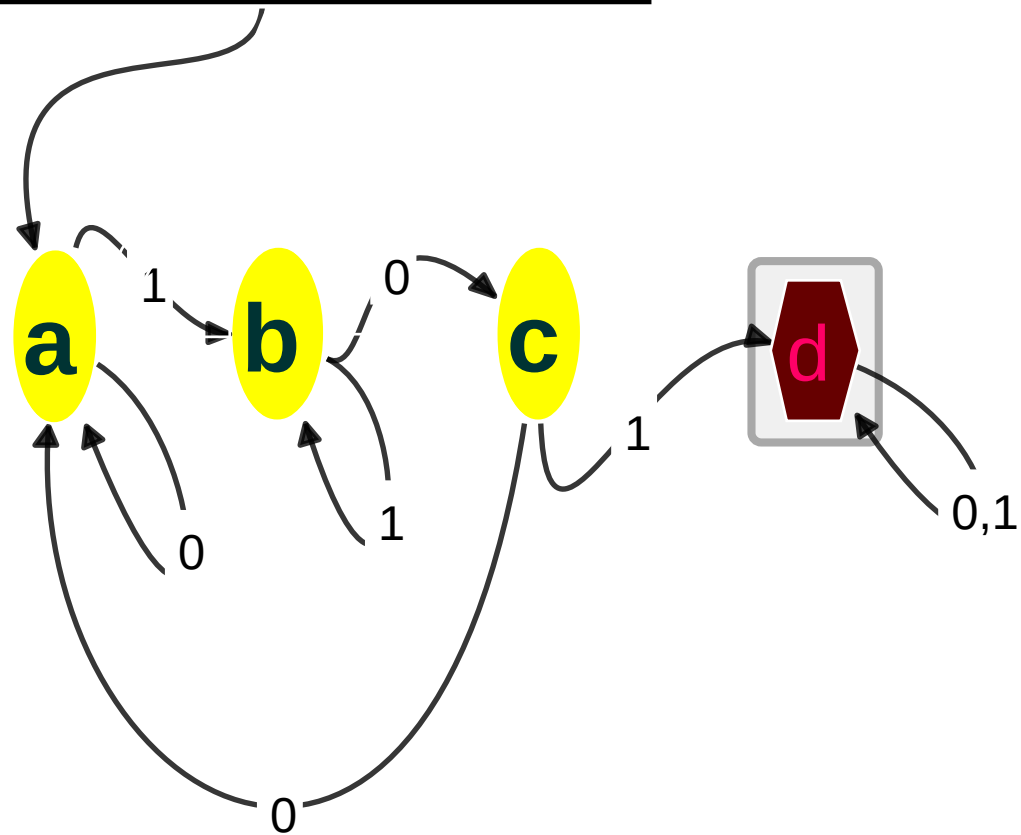
где

$Q$  - множество состояний

- $\Sigma$  — алфавит
- $P$  — функция перехода (на один символ 0-1 переход из каждого состояния)
- $q_0 \in Q$  — стартовое состояние
- $F \subseteq Q$  — множество конечных состояний

# Что делают Автомат?

? что делает конечный автомат?





—

- (ответ : 101 — подстрока)
-

# Регулярная грамматика

- Если существует конечный автомат, который принимает (то есть переходит в финальное состояние) для слов из множества, и не принимает слова не входящие в множество — то это множество регулярная грамматика.

# Упражнения

- Можно ли построить автомат для четных целых чисел?
- Является ли регулярной грамматика из бинарных чисел делящихся на 3?
- Является ли регулярной грамматика из бинарных чисел с равным числом нулей и единиц?

# Pumping lemma

- Если грамматика регулярная — то слова имеют вид  $ab^nc$ ,  $ab^2nc$ ,  $ab^3nc$ , etc  $|b| \geq 1$

?

- $\{0\dots 0\}\{1\dots 1\}$
- $\leftarrow n \rightarrow \leftarrow n \rightarrow$
- Регулярен ли язык?

# Что такое неопределенный автомат?

$(Q, \Sigma, \delta, q_0, F)$

$Q$  - конечное множество состояний

$\Sigma$  - конечный алфавит

$\delta - Q \times (\Sigma + e) \rightarrow P(Q)$

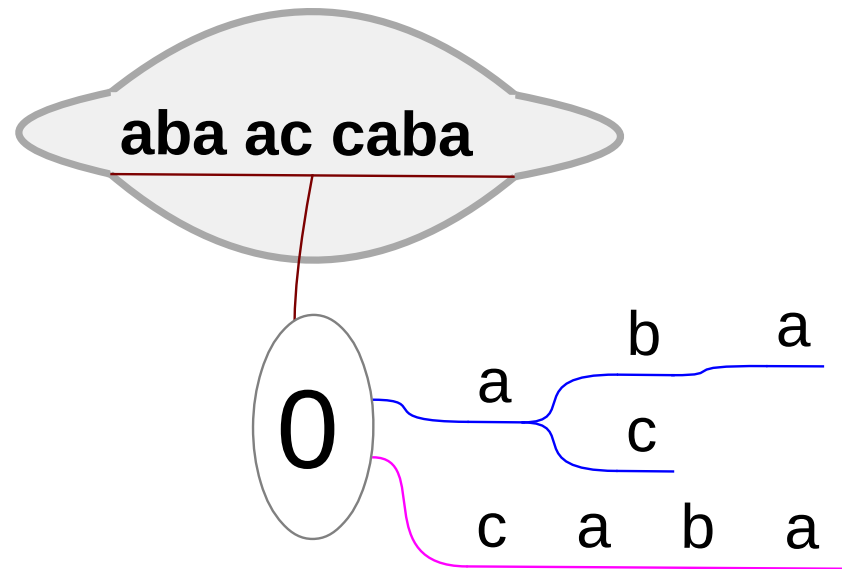
$q_0$  — начальное состояние

$F \subseteq Q$  — множество конечных состояний

# Задача

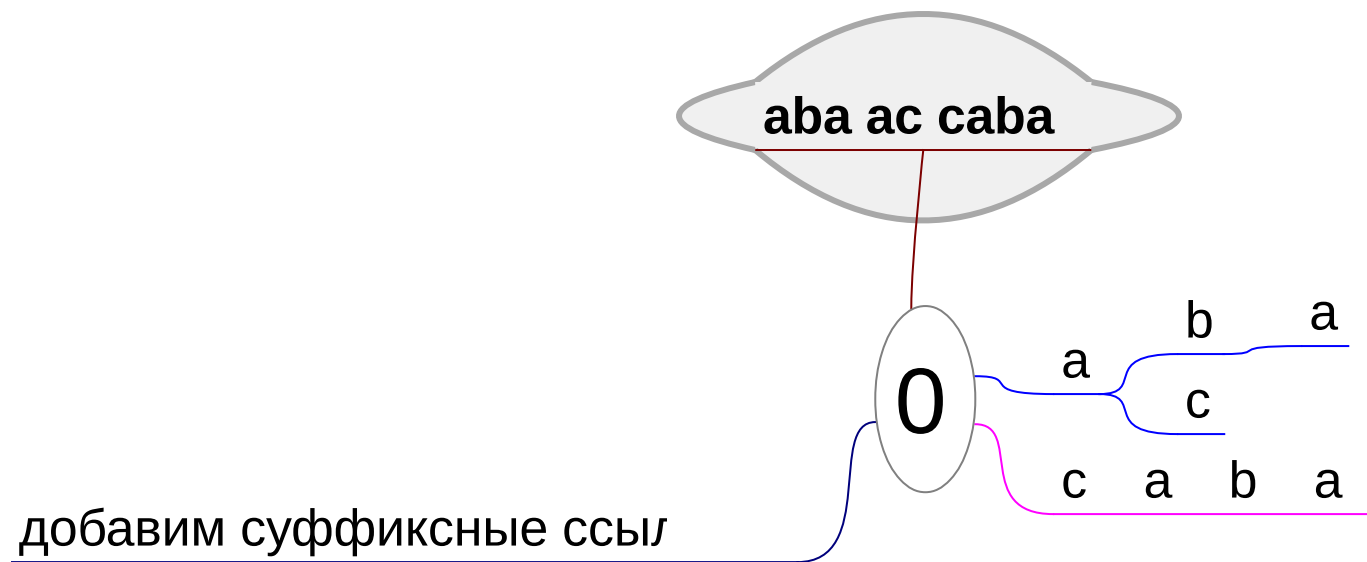
Проверяем вхождения строк в текст

# Ахо корасик

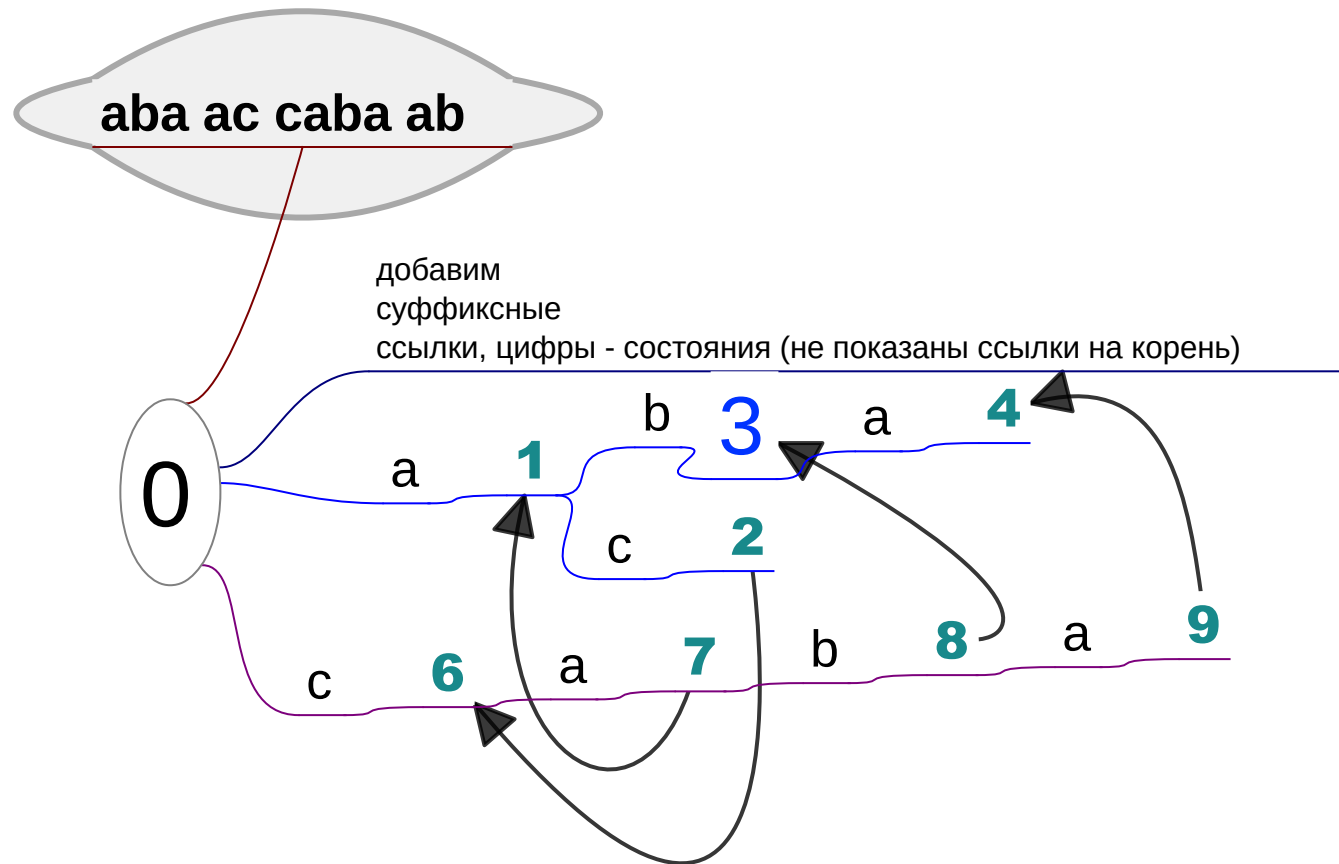




# Ахо корасик



# Ахо-Корасик



# Упражнение

- Расставить буквы на функции перехода

# Ссылки

- [https://neerc.ifmo.ru/wiki/index.php?title=Алгоритм\\_Ахо-Корасик](https://neerc.ifmo.ru/wiki/index.php?title=Алгоритм_Ахо-Корасик)  
описание алгоритма Ахо-Корасик
- <http://codeforces.com/contest/590/problem/E>
-