

---

---

# Лекция 05. Конечные автоматы



# В основе лексических анализаторов лежат регулярные грамматики

---

- **Соглашение:** в дальнейшем, если особо не оговорено, под регулярной грамматикой будем понимать левولينейную грамматику.
  - Напомним, что грамматика  $G = (V_T, V_N, P, S)$  называется *левولينейной*, если каждое правило из  $P$  имеет вид  $A \rightarrow Bt$  либо  $A \rightarrow t$ , где  $A \in V_N$ ,  $B \in V_N$ ,  $t \in V_T$ .
  
- **Соглашение:** предположим, что анализируемая цепочка заканчивается специальным символом  $\perp$  - *признаком конца цепочки*.



# Алгоритм разбора для левوليнейных грамматик (принадлежит ли цепочка $a_1a_2\dots a_n \perp$ языку грамматики)

---

- 1) первый символ исходной цепочки  $a_1a_2\dots a_n \perp$  заменяем нетерминалом  $A$ , для которого в грамматике есть правило вывода  $A \rightarrow a_1$  ("свертка" терминала  $a_1$  к нетерминалу  $A$ )
  - 2) многократно (до тех пор, пока не считаем признак конца) выполняем:  
полученный на предыдущем шаге нетерминал  $A$  и расположенный непосредственно справа от него очередной терминал  $a_i$  исходной цепочки заменяем нетерминалом  $B$ , для которого есть правило вывода  $B \rightarrow Aa_i$  ( $i = 2, 3, \dots, n$ );
- Это эквивалентно построению дерева разбора методом "снизу-вверх": на каждом шаге алгоритма строим один из уровней в дереве разбора, "поднимаясь" от листьев к корню.



# При работе алгоритма возможны следующие ситуации:

---

- 1) прочитана вся цепочка; на последнем шаге свертка произошла к символу  $S$ .  $\Rightarrow a_1 a_2 \dots a_n \perp \in L(G)$ .
- 2) прочитана вся цепочка; на последнем шаге свертка произошла к символу, отличному от  $S$ .  $\Rightarrow a_1 a_2 \dots a_n \perp \notin L(G)$ .
- 3) на некотором шаге не нашлось нужной свертки, т.е. для нетерминала  $A$  и очередного терминала  $a_i$  исходной цепочки не нашлось нетерминала  $B$ , для которого было бы правило вывода  $B \rightarrow Aa_i$ .  $\Rightarrow a_1 a_2 \dots a_n \perp \notin L(G)$ .
- 4) на некотором шаге работы алгоритма оказалось, что есть более одной подходящей свертки. Это говорит о *недетерминированности разбора*.



# Пример реализации алгоритма

---

Построим таблицу возможных сверток для грамматики

P:  $S \rightarrow C\perp$   
 $C \rightarrow Ab \mid Ba$   
 $A \rightarrow a \mid Ca$   
 $B \rightarrow b \mid Cb$

|   | a | b | $\perp$ |
|---|---|---|---------|
| C | A | B | S       |
| A | - | C | -       |
| B | C | - | -       |
| S | - | - | -       |

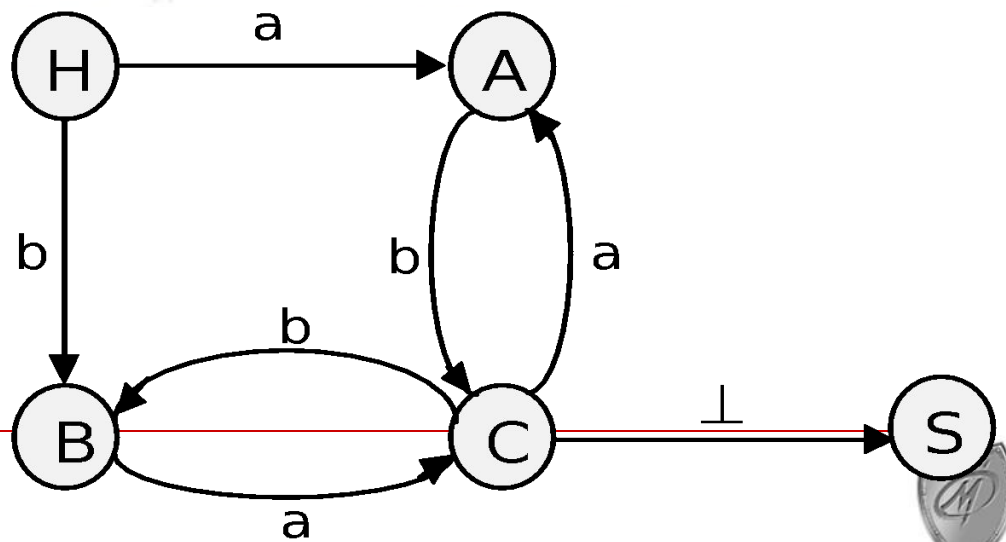


# Пример реализации алгоритма

□ Или диаграмму состояний

P:  $S \rightarrow C\perp$   
 $C \rightarrow Ab \mid Ba$   
 $A \rightarrow a \mid Ca$   
 $B \rightarrow b \mid Cb$

|   | a | b | $\perp$ |
|---|---|---|---------|
| C | A | B | S       |
| A | - | C | -       |
| B | C | - | -       |
| S | - | - | -       |



# Правила построения диаграммы

---

- 1) строим вершины графа, помеченные нетерминалами грамматики (для каждого нетерминала - одну вершину), и еще одну вершину, помеченную символом, отличным от нетерминальных (например,  $H$ ).
  - 1) Эти вершины будем называть *состояниями*.  
 $H$  - начальное состояние.
- 2) соединяем эти состояния дугами по правилам:
  - 1) для каждого правила грамматики вида  $W \rightarrow t$  соединяем дугой состояния  $H$  и  $W$  (от  $H$  к  $W$ ) и помечаем дугу символом  $t$ ;
  - 2) для каждого правила  $W \rightarrow Vt$  соединяем дугой состояния  $V$  и  $W$  (от  $V$  к  $W$ ) и помечаем дугу символом  $t$ ;



# Детерминированный конечный автомат (КА)

---

- **Определение:** *конечный автомат (КА)* - это пятерка  $(K, VT, F, H, S)$ , где
  - $K$  - конечное множество состояний;
  - $VT$  - конечное множество допустимых входных символов;
  - $F$  - отображение множества  $K \times VT \rightarrow K$ , определяющее поведение автомата;  $F$  - функция переходов;
  - $H \in K$  - начальное состояние;
  - $S \in K$  - заключительное состояние (либо конечное множество заключительных состояний).
- $F(A, t) = B$  означает, что из состояния  $A$  по символу  $t$  автомат переходит в состояние  $B$ .



# О недетерминированном разборе

---

- Для грамматики  $G = (\{a, b, \perp\}, \{S, A, B\}, P, S)$ , где
$$P: \begin{aligned} S &\rightarrow A\perp \\ A &\rightarrow a \mid Bb \\ B &\rightarrow b \mid Bb \end{aligned}$$
разбор будет недетерминированным (т.к. у нетерминалов  $A$  и  $B$  есть одинаковые правые части -  $Bb$ ).
- Такой грамматике будет соответствовать **недетерминированный конечный автомат.**



# Недетерминированный конечный автомат (НКА)

---

- **Определение:** *недетерминированный конечный автомат (НКА)* - это пятерка  $(K, VT, F, H, S)$ , где
- $K$  - конечное множество состояний;
  - $VT$  - конечное множество допустимых входных символов;
  - $F$  - отображение множества  $K \times VT$  в множество подмножеств  $K$ ;
  - $H \subset K$  - конечное множество начальных состояний;
  - $S \subset K$  - конечное множество заключительных состояний.



---

Следующая тема:

«Построение сканера»

