

# Содержание

<b>1</b>	<b>Автоматы и регулярные выражение</b>	<b>2</b>
1.1	Про НКА . . . . .	2
1.2	Про ДКА. . . . .	4
1.3	Про автоматные языки . . . . .	6
1.4	Про регулярные выражения . . . . .	7
1.5	Про эквивалентные состояния ПДКА . . . . .	8
1.6	Критерий минимальности количества состояний в ПДКА . . . . .	9
1.7	Про канонический ПДКА . . . . .	10
1.8	Алгоритмы проверок . . . . .	11
1.9	Лемма о разрастании . . . . .	12
<b>2</b>	<b>КС-грамматики и МП-автоматы</b>	<b>13</b>
2.1	Про порождающие грамматики . . . . .	13
2.2	Про праволинейные грамматики . . . . .	13
2.3	Построение конечного автомата по праволинейной грамматике . . . . .	15
2.4	Про КС-грамматики . . . . .	16
2.5	Удаление непорождающих и недостижимых символов в алгоритме примедения к нормальной форме Хомского. Асимптотика приведённых шагов. . . . .	16
2.6	Удаление длинных, смешанных правил и ерс-порождающих символов. Асимптотика приведённых шагов. . . . .	17
2.7	Обработка стартового состояния и удаление цепных правил. Асимптотика приведённых шагов. . . . .	19
2.8	Алгоритм Кока-Янгера-Касами синтаксического разбора для КС-грамматик	20
2.9	Лемма о разрастании для КС-языков . . . . .	20
2.10	МП-автомат . . . . .	21
2.11	Построение МП-автомата по КС-грамматике . . . . .	22
2.12	Построение КС-грамматики по МП-автомату . . . . .	23
2.13	Нормальная форма ГРейбах для КС-грамматик . . . . .	24
2.14	Построение МП-автомата без ерс-переходов . . . . .	26
2.15	КС-языки и теоретико-множественные операции . . . . .	27
2.16	Анализатор перенос-свёртка . . . . .	27
2.17	Операция перехода для LR-ситуаций . . . . .	27
2.18	Алгоритм построения LR-таблицы по LR-ситуациям . . . . .	27
2.19	Алгоритм разбора по LR-таблице . . . . .	27
2.20	ДМП-автоматы . . . . .	27
2.21	Приведение LR-алгоритма к распознаванию ДМП-автоматом . . . . .	29
<b>3</b>	<b>Парсеры</b>	<b>29</b>
3.1	Корректность алгоритма Эрли . . . . .	29
3.2	Полнота алгоритма Эрли . . . . .	30
3.3	Оптимальный алгоритм и обоснование сложности . . . . .	31
<b>4</b>	<b>Конечные преобразователи</b>	<b>32</b>
4.1	Конечные преобразователи и задаваемые ими преобразования . . . . .	32

# 1 Автоматы и регулярные выражение

## 1.1 Про НКА

**Недетерминированные конечные автоматы.**

**Определение 1.1.1. Алфавит** - непустое конечное множество, элементы которого называются символами. Обозначение:  $\Sigma$ .

**Замечание.** Дополнительные обозначения.

- $\Sigma^*$  - множество слов, состоящее из всех символов алфавита  $\Sigma$ .
- **Формальный язык**  $L \subset \Sigma^*$ .
- Пустое слово  $\varepsilon \in \Sigma^*$ .

**Определение 1.1.2. Конкатенацией** языков  $L_1$  и  $L_2$  называется язык:

$$L_1 L_2 := L_1 \cdot L_2 := \{uv \mid u \in L_1, v \in L_2\}$$

**Определение 1.1.3. Недетерминированный конечный автомат** - кортеж  $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$ , где:

- $Q$  - множество состояний,  $|Q| < +\infty$ .
- $\Sigma$  - алфавит.
- $\Delta \subset Q \times \Sigma^* \times Q$  - множество переходов.
- $q_0 \in Q$  - стартовое состояние.
- $F \subset Q$  - множество завершающих состояний.

**Определение 1.1.4. Конфигурация** в автомате  $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$  - элемент  $\langle q, w \rangle \in Q \times \Sigma^*$ .

**Определение 1.1.5.**  $\vdash$  - наименьшее рефлексивное транзитивное отношение над  $Q \times \Sigma^*$ , такое, что:

$$\forall w \in \Sigma^* : (\langle q_1, w \rangle \rightarrow q_2) \in \Delta \Rightarrow \forall u \in \Sigma^* : \langle q_1, wu \rangle \vdash \langle q_2, u \rangle$$

**Автоматные языки: примеры автоматных языков**

**Определение 1.1.6.** Для автомата  $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$  языком, задаваемым автоматом, называется:

$$L(M) = \{w \in \Sigma^* \mid \exists q \in F : \langle q_0, w \rangle \vdash \langle q, \varepsilon \rangle\}$$

**Определение 1.1.7.** Язык  $L$  называется **автоматным**, если существует такой НКА  $M$ , что  $L = L(M)$

**Пример.** Постройте какой-нибудь простой автомат и докажите включение в обе стороны.

## Различные варианты определений (упрощения НКА)

**Утверждение 1.1.1.** В определении автомата можно считать:  $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$ , где  $|F| = 1$ .

*Доказательство.* Пусть  $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$  – исходный НКА.

Построим  $M' = \langle Q \cup \{q_f\}, \Sigma, \Delta', q_0, \{q_f\} \rangle$ , где:

$$\Delta' = \Delta \cup \{ \langle q, \varepsilon \rangle \rightarrow q_f \mid q \in F \}$$

Покажем, что  $L(M) = L(M')$ .

В начале покажем, что  $L(M) \subset L(M')$ :

- $w \in L(M) \Rightarrow \exists q \in F : \langle q_0, w \rangle \vdash \langle q, \varepsilon \rangle$
- $(\langle q, \varepsilon \rangle \rightarrow q_f) \in \Delta' \Rightarrow \langle q, \varepsilon \rangle \vdash_{M'} \langle q_f, \varepsilon \rangle$
- Значит  $\langle q_0, w \rangle \vdash_{M'} \langle q, \varepsilon \rangle \vdash_{M'} \langle q_f, \varepsilon \rangle \Rightarrow w \in L(M')$

Теперь докажем обратное включение  $L(M') \subset L(M)$ :

- $w \in L(M') \Rightarrow \langle q_0, w \rangle \vdash_{M'} \langle q_f, \varepsilon \rangle$
- Рассмотрим цепь  $\langle q_0, w \rangle \vdash_{M'} \langle q, u \rangle \vdash_{M', 1} \langle q_f, \varepsilon \rangle$
- Тогда  $u = \varepsilon$ , а  $q \in F$  из того, как мы определили  $\Delta'$  (других переходов в  $q_f$  не существует).
- Получили  $\langle q_0, w \rangle \rightarrow q \Rightarrow w \in L(M)$ .

□

**Утверждение 1.1.2.** Для любого автоматного языка  $L$  существует НКА  $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$ , такой что  $L = L(M)$  и:

$$\forall (\langle q_1, w \rangle \rightarrow q_2) \in \Delta : |w| \leq 1$$

*Доказательство.* Разобьём все  $n$ -буквенные переходы на  $n$  однобуквенных переходов. □

**Утверждение 1.1.3.** Для любого НКА  $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$  существует НКА  $M' = \langle Q, \Sigma, \Delta', q_0, F' \rangle$ , такой, что  $L(M) = L(M')$  и:

$$\forall (\langle q_1, w \rangle \rightarrow q_2) \in \Delta' : |w| = 1$$

*Доказательство.* Обозначим  $\Delta(q, w) = \{q' \mid \langle q, w \rangle \vdash \langle q', \varepsilon \rangle\}$ , то есть вершины, достижимые по слову  $w$ .

Тогда новое множество переходов определим, как:

$$\Delta' = \{ \langle q_1, a \rangle \rightarrow q_2 \mid \exists q_3 \in \Delta(q_1, \varepsilon) : (\langle q_3, a \rangle \rightarrow q_2) \in \Delta \}$$

Новым же множеством завершающих состояний будет:

$$F' = \{q' \mid \Delta(q', \varepsilon) \cap F \neq \emptyset\}$$

Покажем же теперь, что  $L(M) = L(M')$ .

В начале докажем  $L(M) \subset L(M')$ . Тогда  $\exists q \in F : \langle q_0, w \rangle \vdash \langle q, \varepsilon \rangle$ . Б.О.О. будем считать, что  $w = w_1 w_2, w_i \in \Sigma$ . Тогда существует цепь:

- $\langle q_0, w_1 w_2 \rangle \vdash_M \langle q'_1, w_1 w_2 \rangle$  – перешли по цепочке эпсилон, где  $q'_1$  альтер-эго  $q_3$  из определения  $\Delta'$ .
- $\langle q'_1, w_1 w_2 \rangle \vdash_{M,1} \langle q_1, w_2 \rangle$  – читаем символ  $w_1$ .
- $\langle q_1, w_2 \rangle \vdash_M \langle q'_2, w_2 \rangle$  – аналогично проходимся по цепочке  $\varepsilon$ .
- $\langle q'_2, w_2 \rangle \vdash_{M,1} \langle q_2, \varepsilon \rangle$  – читаем символ  $w_2$ .
- $\langle q_2, \varepsilon \rangle \vdash_M \langle q, \varepsilon \rangle$

Теперь соберём из этих переходов из  $\Delta$  переходы из  $\Delta'$ :

$$\begin{aligned}\Delta(q_0, \varepsilon) &= q'_1, \langle q'_1, w_1 \rangle \rightarrow q_1 \Rightarrow (\langle q_0, w_1 \rangle \rightarrow q_1) \in \Delta' \\ \Delta(q_1, \varepsilon) &= q'_2, \langle q'_2, w_2 \rangle \rightarrow q_2 \Rightarrow (\langle q_1, w_2 \rangle \rightarrow q_2) \in \Delta' \\ F \ni q \in \Delta(q_2, \varepsilon) &\Rightarrow q_2 \in F'\end{aligned}$$

В итоге получим, что

$$\langle q_0, w_1 w_2 \rangle \vdash_{M'} \langle q_1, w_2 \rangle \vdash_{M'} \langle q_2, \varepsilon \rangle \Rightarrow w = w_1 w_2 \in L(M')$$

Покажем включение в обратную сторону  $L(M') \subset L(M)$ :

Пусть  $w \in L(M') \Rightarrow \exists q' \in F' : \langle q_0, w \rangle \vdash_{M'} \langle q', \varepsilon \rangle$ . Из определения  $F'$  получим:

$$\exists q \in F : \Delta(q', \varepsilon) \ni q \Rightarrow \langle q', \varepsilon \rangle \vdash_M \langle q, \varepsilon \rangle$$

Рассмотрим  $w = w_1 \cdots w_k$ . Тогда

$$\forall m \in \overline{1, k} \exists q_m : (\langle q_{m-1}, w_m \rangle \rightarrow q_m) \in \Delta' \quad (q_k := q')$$

Значит

$$\exists q'_{m-1} \in \Delta_M(q_{m-1}, \varepsilon), \langle q'_{m-1}, w_m \rangle \rightarrow q_m \Rightarrow \langle q_{m-1}, w_m \rangle \vdash_M \langle q_m, \varepsilon \rangle$$

Также найдём финальное состояние:

$$\exists q \in F : \langle q', \varepsilon \rangle \vdash_M \langle q, \varepsilon \rangle \Rightarrow \langle q_{k-1}, w_k \rangle \vdash_M \langle q, \varepsilon \rangle$$

Итоговая цепочка отношений имеет вид:

$$\langle q_0, w_1 \cdots w_k \rangle \vdash_M \langle q_1, w_2 \cdots w_k \rangle \vdash_M \langle q_2, w_3 \cdots w_k \rangle \cdots \langle q_k, \varepsilon \rangle \vdash_M \langle q, \varepsilon \rangle$$

Что даёт нам требуемое:  $w \in L(M)$ . □

## 1.2 Про ДКА.

### Детерминированные конечные автоматы (ДКА)

**Определение 1.2.1.** НКА  $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$  – детерминированный, если выполнено:

- $\forall (\langle q_1, w \rangle \rightarrow q_2) \in \Delta : |w| = 1$
- $\forall a \in \Sigma, q \in Q : |\Delta(q, a)| \leq 1$

## Эквивалентность ДКА и НКА

**Теорема 1.2.1.** Для любого НКА  $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$  существует ДКА  $M' : L(M) = L(M')$ .

*Доказательство.* Обозначим  $\Delta(S, w) = \cup_{q \in S} \Delta(q, w), w \in \Sigma^*, S \subset Q$ .

Построим ДКА  $M' = \langle 2^Q, \Sigma, \Delta', \{q_0\}, F' \rangle$ , где:

- $F' = \{S \subset Q \mid S \cap F \neq \emptyset\}$
- $\Delta' = \{\langle S, a \rangle \rightarrow \Delta(S, a) \mid S \subset Q\}$

**Лемма 1.2.1.**

$$\Delta'(\{q_0\}, w) = \Delta(\{q_0\}, w)$$

*Доказательство.* Докажем индукцией по  $|w|$  с базой  $w = \varepsilon$

• База:

- $w = \varepsilon$
- $\Delta(\{q_0\}, \varepsilon) = \{q_0\}$
- $\Delta'(\{q_0\}, \varepsilon) = \{q_0\}$

• Переход:

- $w := w'a, a \in \Sigma$
- Рассмотрим тривиальную цепочку равенств:

$$\begin{aligned} \Delta(q_0, w'a) &= \{q' \mid \langle q_0, w'a \rangle \vdash_M \langle q', \varepsilon \rangle\} = \{q' \mid \exists q'' : \langle q_0, w'a \rangle \vdash_M \langle q'', a \rangle \vdash_{M,1} \langle q', \varepsilon \rangle\} = \\ &= \{q' \mid \exists q'' \in \Delta(q_0, w') : \langle q'', a \rangle \vdash_{M,1} \langle q', \varepsilon \rangle\} = \Delta(\Delta(q_0, w'), a) \end{aligned}$$

- Воспользовавшись предположением индукции для слов длины 1 и  $n-1$ , а также свойством аддитивности множеств  $\Delta$ , получим:

$$\Delta(\Delta(\{q_0\}, w'), a) = \Delta(\Delta'(\{q_0\}, w'), a) = \Delta'(\Delta'(\{q_0\}, w'), a) = \Delta'(\{q_0\}, w'a)$$

□

Теперь напомним, что  $F' = \{S \subset Q \mid S \cap F \neq \emptyset\}$ . Тогда, используя результат предыдущей леммы, очевидно, что следующие утверждения эквивалентны:

- $w \in L(M)$
- $\exists q \in F : \langle q_0, w \rangle \vdash_M \langle q, \varepsilon \rangle$
- $\Delta(q_0, w) \cap F \neq \emptyset$
- $\Delta'(\{q_0\}, w) \cap F \neq \emptyset$
- $\Delta'(\{q_0\}, w) \in F'$
- $w \in L(M')$

□

### 1.3 Про автоматные языки

**Свойства класса автоматных языков. Замкнутость относительно булевых операций.**

**Определение 1.3.1.** Итерацией Клини для языка  $L$  называется операция:

$$L^* = \bigcup_{k=0}^{\infty} L^k$$

**Определение 1.3.2.** Полный ДКА – ДКА, для которого выполнено:

$$\forall a \in \Sigma, q \in Q : |\Delta(q, a)| = 1$$

**Теорема 1.3.1.** Автоматные языки замкнуты относительно:

- Конкатенации
- Объединения
- Пересечения
- Итерации Клини
- Дополнения

*Доказательство.* В доказательстве используем определение автоматов с ровно одним завершающим состоянием.

Доказывается рукомаханиям с рисуночками автоматов:

- Конкатенация последовательным соединением
- Объединение параллельным соединением
- Строим декартово произведение автоматов с 1-буквенными переходами, объявляя завершающими состояниями те, которые являются завершающими по обеим координатам.
- Для итерации Клини замыкаем вход автомата с выходом, а также пробрасываем переход по  $\varepsilon$  из входа в выход.
- Для дополнения строим ПДКА и меняем завершающие состояние и незавершающие между собой!

□

## 1.4 Про регулярные выражения

### Регулярные выражения

**Определение 1.4.1.** Определение рекурсивное:

RegExp (R)	Язык $L_i = L(R_i)$
0	$\emptyset$
1	$\varepsilon$
$a, a \in \Sigma$	$\{a\}$
$R_1 + R_2$	$L_1 \cup L_2$
$R_1 R_2$	$L_1 L_2$
$R^*$	$L^*$

Приоритет операций:  $*$   $\rightarrow$   $\cdot$   $\rightarrow$   $+$

### Регулярный автомат, выводимость в регулярном автомате

**Определение 1.4.2.** Регулярный автомат – НКА, в котором на рёбрах записаны регулярные выражения. Докажем утверждение для регулярных автоматов.

**Замечание.** Всякий НКА задаётся регулярным автоматом с 1 завершающим состоянием.

### Теорема Клини о совпадении классов регулярных и автоматных языков.

**Теорема 1.4.1.** Множество регулярных языков совпадает с множеством автоматных языков.

*Доказательство.* Регулярные  $\subseteq$  Автоматные.

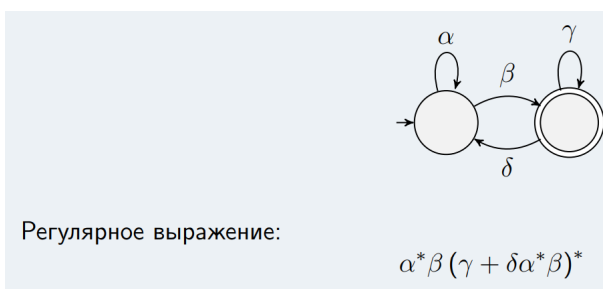
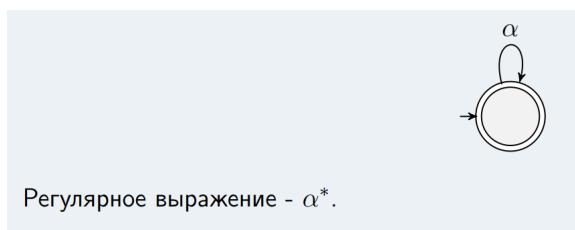
Доказываем индукцией по построению выражения. База очевидна, переход также очевидно следует из замкнутости автоматных языков относительно операций, доказанной ранее.

Автоматные  $\subseteq$  Регулярные.

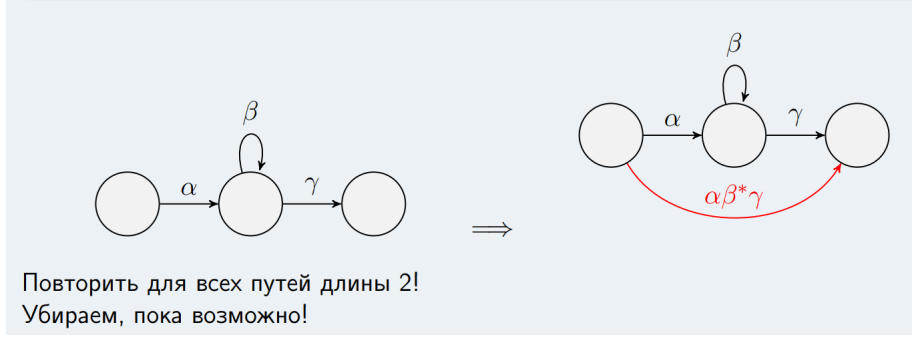
Доказываем индукцией по  $|Q|$  в регулярном автомате. □

### Алгоритм построения регулярного выражения по регулярному автомату.

- База:



- Для перехода будем удалять нестартовые и незавешающие состояния:



## 1.5 Про эквивалентные состояния ПДКА

### Эквивалентность состояний в ПДКА

**Определение 1.5.1.** Пусть  $L \subset \Sigma^*$  – автоматный язык,  $M$  – ПДКА для  $L$ . Тогда определим отношение  $\sim_L$  на  $\Sigma^*$ :

$$u \sim_L v \Leftrightarrow \forall w \in \Sigma^* : uw \in L \Leftrightarrow vw \in L$$

*Доказательство.* Проверим, что  $\sim_L$  – отношение эквивалентности:

- Рефлексивность:  $uw \in L \Leftrightarrow uw \in L$
- Симметричность:  $v \sim_L u : uw \in L \Leftrightarrow vw \in L$
- Транзитивность:
  - $u \sim_L v : uw \in L \Leftrightarrow vw \in L$
  - $v \sim_L s : vw \in L \Leftrightarrow sw \in L$
  - $uw \in L \Leftrightarrow sw \in L \Rightarrow u \sim_L s$

□

**Определение 1.5.2.** Определим  $\sim_M$  над ПДКА:

$$q_1 \sim_M q_2 \Leftrightarrow \forall w \in \Sigma^* : \Delta(q_1, w) \in F \Leftrightarrow \Delta(q_2, w) \in F$$

### Эквивалентность слов по языку

**Определение 1.5.3.** Мы можем разбить язык  $\Sigma^*$  на классы эквивалентности по отношению  $\sim_L$ :

$$\Sigma^* / \sim_L := \{ \{u \mid u \sim_L v\} \mid v \in \Sigma^* \}$$

### Оценка на минимальное количество состояний в ПДКА.

**Лемма 1.5.1.** Пусть  $L_q := \{w \mid \Delta(q_0, w) = q\}$ . Тогда каждый класс эквивалентности в  $\Sigma^* / \sim_L$  – объединение классов в  $L_q$ .



*Доказательство.* Пусть  $u, v \in L_q \Rightarrow \Delta(q_0, u) = \Delta(q_0, v) = q$ . Попробуем преобразовать множество достижимых вершин по произвольному слову  $w$ , используя это свойство:

$$\Delta(q_0, uw) = \Delta(\Delta(q_0, u), w) = \Delta(q, w) = \Delta(q_0, vw)$$

Рассмотрим цепочку эквивалентностей:

$$uw \in L \Leftrightarrow \Delta(q_0, uw) \in F \Leftrightarrow \Delta(q_0, vw) \in F \Leftrightarrow vw \in L$$

Что по определению даёт нам  $u \sim_L v$ . □

**Следствие.** Получаем оценку снизу на количество вершин в автомате:

$$|\Sigma^* / \sim_L| \leq |Q|$$

## 1.6 Критерий минимальности количества состояний в ПДКА

**Лемма 1.6.1.** Для любого автоматного языка  $L$  существует ПДКА  $M'$ , такой, что все состояния в  $M'$  попарно неэквивалентны

*Доказательство.* Построим автомат над классами  $[q] \in Q / \sim_M$ :

$$M' = \langle Q / \sim_M, \Sigma, \Delta', [q_0], F' \rangle$$

где:

- $\Delta' = \{ \langle [q], a \rangle \rightarrow [\Delta(q, a)] \}$
- $F' = \{ [q] \mid q \in F \}$

Необходимо доказать:

- Переходы согласованы
- Завершающие состояния согласованы
- Распознаваемые языки согласованы
- Состояния попарно неэквивалентны

Итак, приступим к доказательству каждого из пунктов:

Переходы согласованы, т.е.  $q_1 \in [q] \Rightarrow \Delta(q_1, a) \in [\Delta(q, a)]$ :

$$\begin{aligned} q_1 \in [q] &\Rightarrow \forall w : \Delta(q_1, w) \in F \Leftrightarrow \Delta(q, w) \in F \text{ в том числе:} \\ &\quad \forall w = au : \Delta(q_1, au) \in F \Leftrightarrow \Delta(q, au) \in F \Rightarrow \\ &\quad \forall u : \Delta(\Delta(q_1, a), u) \in F \Leftrightarrow \Delta(\Delta(q, a), u) \in F \Rightarrow \\ &\quad \Delta(q_1, a) \sim_M \Delta(q, a) \end{aligned}$$

Завершающие состояния согласованы, т.е.  $q_1 \in [q], q \in F \Rightarrow q_1 \in F$ :

$$\begin{aligned} q_1 \in [q] &\Rightarrow \forall w : \Delta(q_1, w) \in F \Leftrightarrow \Delta(q, w) \in F \text{ в том числе:} \\ &\quad (\Delta(q_1, \varepsilon) = q_1) \in F \Leftrightarrow (\Delta(q, \varepsilon) = q) \in F \end{aligned}$$

Совпадение языков, т.е.  $\forall w : \Delta([q_0], w) = [\Delta(q_0, w)]$  индукцией по  $|w|$ :

- База уже доказана в предыдущих пунктах
- Пусть  $w = ua$ :

$$\Delta([q_0], ua) = \Delta(\Delta([q_0], u), a) = \Delta([\Delta(q_0, u)], a) = [\Delta(\Delta(q_0, u), a)] = [\Delta(q_0, ua)]$$

Тогда:

$$w \in L(M) \Leftrightarrow \Delta(q_0, w) \in F \Leftrightarrow \Delta([q_0], w) \in F' \Leftrightarrow w \in L(M')$$

Осталось показать, что все состояния в получившемся автомате попарно неэквивалентны: пусть  $[q_1] \sim_{M'} [q_2]$ , тогда

$$\begin{aligned} \forall w : \Delta([q_1], w) \in F' \Leftrightarrow \Delta([q_2], w) \in F' &\Rightarrow \forall w : [\Delta(q_1, w)] \in F' \Leftrightarrow [\Delta(q_2, w)] \in F' \Rightarrow \\ \forall w : \Delta(q_1, w) \in F \Leftrightarrow \Delta(q_2, w) \in F &\Rightarrow \\ q_1 \sim_M q_2 &\Rightarrow [q_1] = [q_2] \end{aligned}$$

□

**Теорема 1.6.1.**  $M$  – минимальный ПДКА:  $L(M) = L \Leftrightarrow$  Любые два состояния попарно неэквивалентны и все состояния достижимы из стартового

*Доказательство.* ( $\Rightarrow$ ) Пусть  $M$  – минимальный ПДКА. Строим автомат  $Q / \sim_M$ : уменьшаем число состояний. Если в  $M$  есть недостижимое состояние, то удаляем его.

( $\Leftarrow$ ) Из того, что в  $M$  нет эквивалентных состояний:

$$\begin{aligned} \forall w_1, w_2 \in \Sigma^* : \Delta(q_0, w_1) \not\sim \Delta(q_0, w_2) &\Rightarrow \exists u : \Delta(\Delta(q_0, w_1), u) \notin F, \Delta(\Delta(q_0, w_2), u) \in F \Rightarrow \\ \exists u : \Delta(q_0, w_1 u) \notin F, \Delta(q_0, w_2 u) \in F &\Rightarrow \exists u : w_1 u \notin L, w_2 u \in L \Rightarrow \\ w_1 \not\sim_L w_2 \end{aligned}$$

Тогда  $|\Sigma^* / \sim_L| \geq |Q|$ , но  $\forall$  ПДКА  $M' : |\Sigma^* / \sim_L| \leq |Q'| \Rightarrow |Q| \leq |Q'| \Rightarrow M$  – минимальный.

□

## 1.7 Про канонический ПДКА

**Определение 1.7.1.**  $M_1$  и  $M_2$  изоморфны, если существует биекция  $\psi : Q_1 \rightarrow Q_2$ :

- $\psi(q_0^1) = q_0^2$
- $\psi(F_1) = F_2$
- Если  $\Delta(q_1, a) = q_2$ , то  $\Delta(\psi(q_1), a) = \psi(q_2)$

### Канонический ПДКА - корректность построения

Канонический ПДКА для языка  $L$  определим, как  $\Sigma^* / \sim_L$ :

$$M_0 = \langle \Sigma^* / \sim_L, \Sigma, \Delta, [\varepsilon], \{[w] \mid w \in L\} \rangle; \quad \Delta([u], a) = [ua], u \in \Sigma^*, a \in \Sigma$$

Для корректности необходимо показать:

- $u \sim_L v, u \in L \Rightarrow v \in L$
- $u \sim_L v \Rightarrow va \in [ua]$

## Единственность минимального ПДКА

**Лемма 1.7.1.** Пусть  $M$  – минимальный ПДКА, тогда отображение  $\psi$ :

- $\psi : Q_M \rightarrow \Sigma^* / \sim_L$
- $\psi(q) = \{w \mid \Delta_M(q_0, w) = q\}$

является изоморфизмом.

*Доказательство.* Необходимо доказать:

- $\psi(q_0) = [\varepsilon]$
- $\psi(F) = \{[w] \mid w \in L\}$
- Если  $\Delta(q_1, a) = q_2$ , то  $\Delta(\psi(q_1), a) = \psi(q_2)$

Докажем биективность  $\psi$ :  $|\Sigma^* / \sim_L| = |Q_M| \Rightarrow$  достаточно доказать инъективность:

$$\psi(q_1) = \psi(q_2) \Rightarrow \exists w : \Delta(q_0, w) = q_1, \Delta(q_0, w) = q_2 \Rightarrow q_1 = q_2$$

Докажем согласованность стартовых состояний:

$$w \in \psi(q_0) \Rightarrow \Delta_M(q_0, w) = q_0 \text{ но мы знаем, что:}$$

$$\Delta_M(q_0, \varepsilon) = q_0 \Rightarrow w \sim_L \varepsilon$$

Согласованность завершающих состояний:

$$w \in \psi(q), q \in F \Rightarrow \Delta_M(q_0, w) = q \Rightarrow w \in L \Rightarrow [w] \in F'$$

Осталось доказать согласованность переходов:

$$\begin{aligned} w \in \psi(q_1), \Delta(q_1, a) = q_2 &\Rightarrow \Delta(q_0, w) = q_1 \Rightarrow \psi(q_1) = [w], \Delta([w], a) = [wa] \Rightarrow \\ &\Delta(q_0, wa) = \Delta(q_1, a) = q_2 \Rightarrow \psi(q_2) = [wa] \end{aligned}$$

□

**Теорема 1.7.1.** МПДКА единственен с точностью до изоморфизма.

*Доказательство.* Пусть  $M_1, M_2$  – ПДКА. Построим канонические изоморфизмы  $\psi_1, \psi_2$ . Тогда  $\psi_2^{-1} \circ \psi_1$  – изоморфизм  $M_1$  в  $M_2$ . □

## 1.8 Алгоритмы проверок

### Алгоритм проверки МПДКА на эквивалентность

Необходимо получить неэквивалентные состояния, но у нас есть только слова. Как по ним понять, какие состояния попарно неэквивалентны? Ввести эквивалентность по словам малой длины.

**Определение 1.8.1.**  $q_1 \underset{n}{\sim} q_2$ , если для любого слова  $|w| \leq n$ :

$$\Delta(q_1, w) \in F \Leftrightarrow \Delta(q_2, w) \in F$$

### Лемма 1.8.1.

$$q_1 \sim q_2 \Leftrightarrow q_1 \underset{|Q|-2}{\sim} q_2$$

*Доказательство.* Покажем, если  $|Q/\underset{i}{\sim}| = |Q/\underset{i+1}{\sim}|$ , то  $|Q/\underset{i+1}{\sim}| = |Q/\underset{i+2}{\sim}|$ .

Очевидно, что  $|Q/\underset{i+1}{\sim}| \leq |Q/\underset{i+2}{\sim}|$ . По определению эквивалентности  $q_1 \underset{i+2}{\sim} q_2$ :

$$\begin{aligned} \forall w = au, |w| \leq i+2 : \Delta(q_1, au) \in F &\Leftrightarrow \Delta(q_2, au) \in F \Rightarrow \\ \forall a \in \Sigma : \Delta(q_1, a) \underset{i+1}{\sim} \Delta(q_2, a) &\overset{|Q/\underset{i}{\sim}|=|Q/\underset{i+1}{\sim}|}{\Rightarrow} \\ \forall a \in \Sigma : \Delta(q_1, a) \underset{i}{\sim} \Delta(q_2, a) &\Rightarrow q_1 \underset{i+1}{\sim} q_2 \end{aligned}$$

Теперь понятно, что, если  $|Q/\underset{i}{\sim}| = |Q/\underset{i+1}{\sim}|$ , то  $|Q/\underset{i}{\sim}| = |Q/\sim|$

Заметим, что  $|Q/\sim_0| = |\{F, Q \setminus F\}| = 2$ , но  $|Q/\sim_i|$  неубывает и стабилизируется  $\Rightarrow$

$$|Q/\sim_i| \leq |Q/\sim| \leq |Q|$$

□

### Теорема Майхилла-Нероуда

**Теорема 1.8.1.**  $L$  – автоматный  $\Leftrightarrow L$  содержит конечное количество классов эквивалентности  $\Sigma^*/\sim_L$

*Доказательство.* •  $\Rightarrow L$  – автоматный, тогда  $|\Sigma^*/\sim_L| \leq |Q| < +\infty$

•  $\Leftarrow$  построим канонический МПДКА.

□

## 1.9 Лемма о разрастании

### Лемма о разрастании для автоматных языков

**Лемма 1.9.1.** Пусть  $L$  – автоматный язык. Тогда

$$\exists P \forall w \in L : |w| \geq P \exists x, y, z : w = xyz, |xy| \leq P, |y| \neq 0 : \forall k \geq 0 : xy^kz \in L$$

*Доказательство.* Построим  $M$  – НКА с 1-буквенными переходами:  $L(M) = L$ . Тогда  $P := |Q|$ . Если  $|w| \geq P \Rightarrow$  посетили  $\geq P+1$  состояние.

Значит  $\exists q \in Q$ , которую посетили дважды, значит мы можем ходить по этому циклу любое  $k$  число раз. □

### Пример неавтоматных языков

**Пример.**

$$L = \{a^n b^n \mid n \in \mathbb{N}\}$$

*Доказательство.* Неавтоматность доказывается отрицанием леммы о разрастании. □

## 2 КС-грамматики и МП-автоматы

### 2.1 Про порождающие грамматики

#### Порождающие грамматики

**Определение 2.1.1.** Порождающая грамматика  $G = \langle N, \Sigma, P, S \rangle$ , где:

- $N$  – множество вспомогательных символов,  $|N| < +\infty$
- $\Sigma$  – алфавит – множество терминальных символов,  $|\Sigma| < +\infty, N \cap \Sigma = \emptyset$
- $S \in N$  – стартовый нетерминал
- $P \subset ((N \cup \Sigma)^+ \setminus \Sigma^*) \times (N \cup \Sigma)^*$

#### Язык, задаваемый грамматикой

**Определение 2.1.2.** Отношением выводимости  $\vdash_G$  называется наименьшее рефлексивное транзитивное отношение:

$$\forall(\alpha \rightarrow \beta) \in P, \forall \varphi, \psi \in (N \cup \Sigma)^* : \varphi \alpha \psi \vdash_G \varphi \beta \psi$$

**Определение 2.1.3.**  $w$  выводимо в грамматике  $G = \langle N, \Sigma, P, S \rangle$ , если  $S \vdash_G w$

**Определение 2.1.4.** Языком  $L$ , порождённым грамматикой  $G$  называется:

$$L(G) = L = \{w \mid S \vdash_G w\}$$

#### Иерархия Хомского порождающих грамматик

Разграничим грамматики по виду правил:

1. Порождающие грамматики: любые правила
2. Контекстно-зависимые грамматики:  $\varphi A \psi \rightarrow \varphi \alpha \psi, \alpha \neq \varepsilon$
3. Контекстно-свободные грамматики:  $A \rightarrow \alpha$
4. Праволинейные грамматики:  $A \rightarrow wB, A \rightarrow w$

При этом:

- $A \in N, B \in N$  – нетерминальные символы
- $\alpha, \varphi, \psi \in (N \cup \Sigma)^*$

### 2.2 Про праволинейные грамматики

#### Праволинейные языки

**Теорема 2.2.1.** Множество автоматных языков равно множеству языков, задаваемых праволинейными грамматиками.

*Доказательство.* Состояние в автомате – нетерминалы в грамматике + сток □

## Построение праволинейной грамматики по конечному автомату

Пусть  $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$ . Построим  $G = \langle Q, \Sigma, P, q_0 \rangle$ , где:

$$P = \{q_1 \rightarrow wq_2 \mid (\langle q_1, w \rangle \rightarrow q_2) \in \Delta\} \cup \{q \rightarrow \varepsilon \mid q \in F\}$$

*Доказательство.* Надо доказать два утверждения:

1.  $\langle q_1, w \rangle \vdash_M \langle q_2, \varepsilon \rangle \Leftrightarrow q_1 \vdash_G wq_2$
2.  $\langle q_1, w \rangle \vdash_M \langle q, \varepsilon \rangle, q \in F \Leftrightarrow q_1 \vdash_G w$

В начале докажем  $\Rightarrow$  для обоих пунктов.

Первый пункт доказывается индукцией по длине вывода в  $M$ .

- База индукции – 0 шагов:

$$\langle q_1, w \rangle \vdash_M \langle q_2, \varepsilon \rangle \Rightarrow q_1 = q_2, w = \varepsilon \Rightarrow q_1 \vdash_G \varepsilon q_2$$

- Для перехода представим произвольное слово  $w = vu, u \in \Sigma$ :

$$\langle q_1, w \rangle \vdash_M \langle q_3, v \rangle \vdash_{M,1} \langle q_2, \varepsilon \rangle \Rightarrow q_1 \vdash_G uq_3, q_3 \vdash_G vq_2 \Rightarrow q_1 \vdash_G uvq_2 = wq_2$$

Теперь второй пункт:

$$\langle q_1, w \rangle \vdash_M \langle q, \varepsilon \rangle, q \in F \Rightarrow q_1 \vdash_G wq, (q \rightarrow \varepsilon) \in P \Rightarrow q_1 \vdash_G w$$

Теперь  $\Leftarrow$  для обоих пунктов:

Первый тоже докажем индукцией по длине вывода в  $G$

- База индукции 0 шагов:

$$q_1 \vdash_{G,0} wq_2 \Rightarrow q_1 = q_2, w = \varepsilon \Rightarrow \langle q_1, \varepsilon \rangle \vdash_M \langle q_2, \varepsilon \rangle$$

- Для перехода опять разложим  $w = vu, u \in \Sigma$ :

$$q_1 \vdash_G vq_3 \vdash_G wq_2 \Rightarrow \langle q_1, v \rangle \vdash_M \langle q_3, \varepsilon \rangle, \langle q_3, u \rangle \vdash_M \langle q_2, \varepsilon \rangle \Rightarrow \langle q_1, vu \rangle \vdash_M \langle q_3, u \rangle \vdash_{M,1} \langle q_2, \varepsilon \rangle$$

Второй пункт:

$$q_1 \vdash_G wq \vdash_{G,1} w \Rightarrow \langle q_1, w \rangle \vdash_M \langle q, \varepsilon \rangle, q \vdash_{G,1} \varepsilon \Rightarrow q \in F$$

После доказательства двух пунктов нам становится очевидно, что следующие утверждения эквивалентны:

- $w \in L(M)$
- $\exists q \in F : \langle q_0, w \rangle \vdash \langle q, \varepsilon \rangle$
- $q_0 \vdash_G w$
- $w \in L_G$

□

## 2.3 Построение конечного автомата по праволинейной грамматике

Строим автомат:

$$M = \langle N \cup \{q_f\}, \Sigma, \Delta, S, \{q_f\} \rangle$$

Переходы:

1.  $\langle A, w \rangle \rightarrow B$ , если  $(A \rightarrow wB) \in P$
2.  $\langle A, w \rangle \rightarrow q_f$ , если  $(A \rightarrow w) \in P$

*Доказательство.* Надо доказать два утверждения:

1.  $\langle A, w \rangle \vdash_M \langle M, \varepsilon \rangle \Leftrightarrow A \vdash_G wB$ ;  $A, B \in N$
2.  $\langle A, w \rangle \vdash_M \langle q_f, \varepsilon \rangle \Leftrightarrow A \vdash_G w$

В начале докажем  $\Rightarrow$  для обоих пунктов:

Первый будем доказывать по индукции:

- База индукции – 0 шагов:

$$\langle A, w \rangle \vdash_{M,0} \langle B, \varepsilon \rangle \Rightarrow A = B, w = \varepsilon \Rightarrow A \vdash_G \varepsilon B$$

- Для перехода представим произвольное слово  $w = vu, u \in \Sigma$ :

$$\langle A, w \rangle \vdash_M \langle C, u \rangle \vdash_{M,0} \langle B, \varepsilon \rangle \Rightarrow A \vdash_G vC, C \vdash_G uB \Rightarrow A \vdash_G vuB = wB$$

Теперь второй пункт для произвольного  $w = vu$ :

$$\langle A, v \rangle \vdash_M \langle C, u \rangle \vdash_{M,1} \langle q_f, \varepsilon \rangle \Rightarrow A \vdash_G vC, C \vdash_G u \Rightarrow A \vdash_G vu = w$$

Перейдём к доказательству  $\Leftarrow$  для обоих пунктов:

Первый пункт также будет доказан по индукции:

- База – 0 шагов:

$$A \vdash_{G,0} wB \Rightarrow w = \varepsilon, A = B \Rightarrow \langle A, \varepsilon \rangle \vdash_M \langle B, \varepsilon \rangle$$

- Переход для произвольного  $w = uv$ :

$$A \vdash_G uC \vdash_{G,1} uvB \Rightarrow \langle A, u \rangle \vdash_M \langle C, \varepsilon \rangle, \langle C, v \rangle \vdash_M \langle B, \varepsilon \rangle \Rightarrow \langle A, uv \rangle \vdash_M \langle B, \varepsilon \rangle$$

Второй пункт для  $w = uv$ :

$$A \vdash_G uC \vdash_{G,1} uv \Rightarrow \langle A, u \rangle \vdash_M \langle C, \varepsilon \rangle, \langle C, v \rangle \vdash \langle q_f, \varepsilon \rangle \Rightarrow \langle A, uv \rangle \vdash_M \langle q_f, \varepsilon \rangle$$

После доказанных двух пунктов становится очевидным эквивалентность данных утверждений:

- $w \in L(M)$
- $\langle S, w \rangle \vdash_M \langle q_f, \varepsilon \rangle$
- $S \vdash_G w$
- $w \in L(G)$

□

## 2.4 Про КС-грамматики

### Примеры контекстно-свободных языков

Пример.

$$S \rightarrow aSb$$

$$S \rightarrow \varepsilon$$

Задаёт неавтоматный язык  $\{a^n b^n \mid n \in \mathbb{N}\}$

### Замкнутость КС-языков относительно простейших операций

**Утверждение 2.4.1.**  $L_1 \cup L_2$  является КС-языком

*Доказательство.* Построим КС-грамматику для языка  $L_1 \cup L_2$ . Для этого рассмотрим соответствующие грамматики для  $L_1, L_2$ . Пусть стартовые символы в них имеют имена  $S$  и  $T$ . Тогда стартовый символ для  $L_1 \cup L_2$  обозначим за  $S'$  и добавим правило  $S' \rightarrow S \mid T$ . Покажем, что  $S' \vdash_G w \Leftrightarrow S \vdash_G w \vee T \vdash_G w$ .

$\Leftarrow$ : Поскольку  $S \vdash_G$  и есть правило  $S' \vdash_G S$ , то по транзитивности выводимости получаем, что  $S' \vdash_G w$ . Аналогично и для  $T$ .

$\Rightarrow$ : Пусть  $S' \vdash_G w$ . Поскольку  $S' \vdash S \mid T$  – единственные правила, в которых нетерминал  $S'$  присутствует в левой части, то это означает, что либо  $S' \vdash_G S \vdash_G w$ , либо  $S' \vdash_G T \vdash_G w$ .  $\square$

**Утверждение 2.4.2.**  $L_1 L_2$  – КС-язык

*Доказательство.* Аналогично предыдущему случаю построим КС-грамматику для языка  $L_1 L_2$ . Для этого добавим правило  $S' \vdash_G ST$ , где  $S$  и  $T$  – стартовые символы языков  $L_1$  и  $L_2$  соответственно.  $\square$

**Утверждение 2.4.3.**  $L^* = \bigcup_{i=0}^{\infty} L^i$  – КС-язык.

*Доказательство.* Если  $S$  – стартовый символ КС-грамматики для языка  $L$ , то добавим в КС-грамматику для языка  $L^*$  новый стартовый символ  $S'$  и правила  $S' \vdash_G SS' \mid \varepsilon$   $\square$

## 2.5 Удаление непорождающих и недостижимых символов в алгоритме приведения к нормальной форме Хомского. Асимптотика приведённых шагов.

### Удаление непорождающих символов

**Определение 2.5.1.** Символ  $Y \in N$  называется **порождающим**, если:

$$\exists w \in \Sigma^* : Y \vdash w$$

Удаляем непорождающие символы  $Z$  и все правила, содержащие символы  $Z$  – получаем  $G_1$ .

**Утверждение 2.5.1.**

$$L(G) = L(G_1)$$



*Доказательство.* Включение  $L(G_1) \subset L(G)$  очевидно, так как мы удалили нетерминалы, которые никак не влияли на вывод слов, поэтому язык не увеличился.

Докажем  $L(G) \subset L(G_1)$ . Пусть  $w \in L(G) \setminus L(G_1)$ .

Тогда существует непорождающий  $Z$ :  $S \vdash \alpha Z \beta \vdash w$ . Тогда если  $w = w_1 u w_2$ :  $\alpha \vdash w_1, Z \vdash u, \beta \vdash w_2 \Rightarrow Z$  – порождающий, противоречие.  $\square$

## Удаление недостижимых символов

**Определение 2.5.2.** Символ  $D \in N$  называется **достижимым**, если существуют некоторые  $\varphi, \psi \in (N \cup \Sigma)^*$ , такие, что:

$$S \vdash \varphi D \psi$$

Удаляем все недостижимые символы и все содержащие их правила – получаем  $G_2$ .

### Утверждение 2.5.2.

$$L(G_1) = L(G_2)$$

*Доказательство.* Включение  $L(G_2) \subset L(G_1)$  очевидно, так как удалили все нетерминалы, которые не участвовали в выводе, поэтому язык не увеличился.

Теперь докажем  $L(G_1) \subset L(G_2) \Rightarrow \exists w \in L(G_1) \setminus L(G_2) \Rightarrow$  существует недостижимый  $U$ :  $S \vdash \alpha U \beta \vdash w \Rightarrow U$  – достижимый. Противоречие.  $\square$

### Утверждение 2.5.3. В $G_2$ не появилось непорождающих символов.

*Доказательство.* Пусть  $B$  стал новым непорождающим в  $G_2$ . Тогда:

- $B$  был достижимым в  $G_1$ .
- $B$  был порождающим в  $G$ :  $B \vdash_G u$

На пути вывода  $B \vdash u$  был недостижимый символ  $C$ . Но, тогда строим пусть  $S \rightarrow B \rightarrow C$  – противоречие!  $\square$

## Асимптотика приведённых шагов (в терминах изначальной грамматики)

1. Для поиска непорождающих нетерминалов требуется запустить  $|N|$  BFS-ов. Сложность  $O(|N|(|N| + |E|))$
2. Для поиска недостижимых нетерминалов требуется запустить один BFS из  $S$ . Сложность  $O(|N| + |E|)$

## 2.6 Удаление длинных, смешанных правил и eps-порождающих символов. Асимптотика приведённых шагов.

### Удаление длинных правил

Сделаем замену:

$$B \rightarrow A_1 A_2 \cdots A_n$$

на:

$$\begin{aligned} B &\rightarrow A_1 B_1 \\ B_1 &\rightarrow A_2 B_2 \\ &\dots \\ B_{n-2} &\rightarrow A_{n-1} A_n \end{aligned}$$

Получим грамматику  $G_3$ .

**Замечание.** Если в дереве вывода  $G_3$  появился  $B_k$ , то в нём появятся все правила, в левых и правых частях которых есть  $B_1, \dots, B_{n-2}$ .

### Удаление смешанных правил

Сделаем замену:

$$A \rightarrow A_1 b A_2 d$$

на

$$\begin{aligned} A &\rightarrow A_1 B A_2 D \\ B &\rightarrow b \\ D &\rightarrow d \end{aligned}$$

Получим грамматику  $G_4$ .

### Удаление eps-порождающих

**Определение 2.6.1.** Символ  $E$  называется  $\varepsilon$ -порождающим, если  $E \vdash \varepsilon$ .

Сделаем замену:

- Добавим правило  $A \rightarrow B$ , если  $A \rightarrow BC, C \vdash \varepsilon$ .
- Добавим правило  $A \rightarrow C$ , если  $A \rightarrow BC, B \vdash \varepsilon$ .
- Удалим правила  $A \rightarrow \varepsilon$ .

Получим грамматику  $G_5$ .

### Утверждение 2.6.1.

$$L(G_4) = L(G_5)$$

*Доказательство.*  $L(G_4) \subset L(G_5)$  докажем индукцией по длине вывода:

- База – 1 шаг:  $w = a, (A \rightarrow a) \in P_{G_5}$
- Переход  $A \vdash_{G_4, 1} \alpha \vdash_{G_4} w$ 
  - $\alpha = B \Rightarrow (A \rightarrow B) \in P_{G_5}$ .
  - $\alpha = BC \Rightarrow B \vdash_{G_4} w_1, C \vdash_{G_4} w_2$ .
  - Если  $w_1 \neq \varepsilon, w_2 \neq \varepsilon$  – применяем переход для  $B, C$ .

- Если  $w_1 = \varepsilon : A \vdash_{G_5,1} C \vdash_{G_5} w_2 = w$

$L(G_5) \subset L(G_4)$  также докажем индукцией по длине вывода:

- База – 1 шаг:  $w = a, (A \rightarrow a) \in P_{G_4}$
- Переход  $A \vdash_{G_5,1} B \vdash_{G_5} w$ :
  - $(A \rightarrow B) \in P_{G_4} \Rightarrow A \vdash_{G_4,1} B \vdash_{G_4} w$
  - $(A \rightarrow BC) \in P_{G_4}, C \vdash_{G_4} \varepsilon \Rightarrow A \vdash_{G_4,1} BC \vdash w\varepsilon = w$

□

### Асимптотика приведённых шагов (в терминах изначальной грамматики)

1. Сложность удаления длинных правил  $O(|P| \max_{p \in P} |p|)$ .
2. Сложность удаления смешанных правил можно оценить также.
3. Сложность удаления  $\varepsilon$ -порождающих нетерминалов можно оценить, как  $O(|P|h)$ , где  $h$  – максимальная глубина вывода  $\varepsilon$  для изначальной грамматики.

## 2.7 Обработка стартового состояния и удаление цепных правил. Асимптотика приведённых шагов.

### Обработка стартового состояния

Мы доказали  $L(G_5) = L(G_4) \setminus \{\varepsilon\}$ :

- Заводим новый нетерминал  $S'$ , делаем его стартовым
- Добавляем правило  $S' \rightarrow S$
- Если  $S \vdash \varepsilon$ , то добавляем  $S' \rightarrow \varepsilon$

Получили грамматику  $L(G_6)$ .

### Удаление цепных правил

Сделаем транзитивное замыкание:

$$B \rightarrow B_1 \rightarrow B_2 \rightarrow \dots \rightarrow B_n \rightarrow CD \mid a$$

заменим на:

$$B \rightarrow CD \mid a$$

Дополнительно удалим правила вида  $A \rightarrow B$ .

Получили грамматику  $L(G_7)$ .

### Утверждение 2.7.1.

$$L(G_6) = L(G_7)$$

*Доказательство.* Полностью аналогично доказательству после удаления  $\varepsilon$ -порождающих.

□

## Асимптотики

1. Для обработки стартового состояния нужно понять, было ли оно  $\varepsilon$ -порождающим, то есть достаточно запустить BFS и проверить, дошли ли мы до правила, содержащего  $\varepsilon$  в правой части –  $O(|N| + |P|)$
2. Для удаления цепных правил требуется транзитивное замыкание, которое можно построить, используя алгоритм Флойда-Фалкерсона:  $O(h^3)$ , где  $h$  – максимальная глубина вывода.

## 2.8 Алгоритм Кока-Янгера-Касами синтаксического разбора для КС-грамматик

Алгоритм построен на динамике по подотрезкам: заведём массив  $d[A][i, j] = \mathbb{I}(A \vdash w[i : j])$ , очевидно, что  $w \in L(G) \Leftrightarrow d[S][0, |w|] = \text{True}$

Индукция по длине слова:

- $A \vdash w[i : j]$ . Тогда существуют  $B, C : A \vdash_1 BC, B \vdash w[i : k], C \vdash w[k : j] \Rightarrow d[B][i : k] = \text{True}, d[C][k : j] = \text{True}$
- $d[A][i : j] = \text{True}$ . Тогда для некоторого `midPosition` сработал переключатель (см. код алгоритма). Делаем индукционный переход.

Асимптотика алгоритма –  $O(|N|^3 \cdot |P|)$

## 2.9 Лемма о разрастании для КС-языков

### Лемма о разрастании для КС-языков

**Лемма 2.9.1.** Пусть  $L$  – КС-язык. Тогда

$$\exists p : \forall w \in L : |w| \geq p : \exists x, u, y, v, z \in \Sigma^* : w = xuyvz : \\ |uv| > 0, |uyv| \leq p : \forall k \geq 0 : xu^k y v^k z \in L$$

*Доказательство.* Рассмотрим грамматику  $G$  в нормальном форме Хомского:  $L = L(G)$ . Каждый уровень дерева вывода увеличивает длину слова не более, чем вдвое. Если  $p = 2^{|N|}$ , то  $|w| \geq p \geq 2^{|N|}$ . Тогда глубина дерева разбора более  $|N|$  – воспользуемся принципом Дирихле.

Найдётся такой нетерминал  $A$ :

$$S \vdash xAz \vdash xuAvz \vdash xuyvz; \quad A \vdash uAv$$

Среди таких  $A$  рассмотрим такое, что его глубина относительно корня наибольшая. Тогда  $|uyv| \leq 2^{|N|} = p$  (иначе были бы повторения)  $\square$

### Пример не КС-языка

**Пример.** Язык

$$\{a^n b^n c^n \mid n \in \mathbb{Z}^+\}$$

не является КС-языком.

*Доказательство.* Воспользуемся отрицанием леммы о разрастании.  $\square$

## 2.10 МП-автомат

**Определение 2.10.1.** Автомат с магазинной памятью – МП-автомат:

- $Q$  – множество состояний,  $|Q| < +\infty$
- $\Sigma$  – алфавит,  $|\Sigma| < +\infty$
- $\Gamma$  – стековый алфавит,  $|\Gamma| < +\infty, \Gamma \cap \Sigma = \emptyset$
- $\Delta \subset (Q \times \Sigma^* \times \Gamma^*) \times (Q \times \Gamma^*), |\Delta| < +\infty$
- $q_0 \in Q$  – стартовое состояние.
- $F \subset Q$  – множество завершающих состояний.

**Определение 2.10.2.** Конфигурацией в МП-автомате называется тройка  $\langle q, u, \gamma \rangle \in Q \times \Sigma^* \times \Gamma^*$ .

**Определение 2.10.3.**  $\vdash$  – наименьшее рефлексивное транзитивное отношение, что

$$\forall \langle q_1, u, \theta \rangle \rightarrow \langle q_2, \beta \rangle \in \Delta : \forall v \in \Sigma^*, \eta \in \Gamma^* \langle q_1, uv, \eta\alpha \rangle \vdash \langle q_2, v, \eta\beta \rangle$$

**Определение 2.10.4.** Языком, распознаваемым МП-автоматом, называется

$$L(M) = \{w \in \Sigma^* \mid \exists q \in F : \langle q_0, w, \varepsilon \rangle \vdash \langle q, \varepsilon, \varepsilon \rangle\}$$

**Утверждение 2.10.1.** Для любого МП-автомата существует эквивалентный МП-автомат, для которого выполнено соотношение:

$$\forall \langle q_1, u, \alpha \rangle \rightarrow \langle q_2, \beta \rangle \in \Delta : |u| \leq 1, |\alpha| + |\beta| \leq 1$$

*Доказательство.* Действие "считать  $n$  букв со входа" растягиваем в "считать  $n$  раз 1 букву со входа".

Действие "снять со стека  $n$  букв и положить на него  $m$  букв" растягиваем в "снять  $n$  раз со стека 1 букву, положить  $m$  раз на стек 1 букву".  $\square$

**Утверждение 2.10.2.** Для любого МП-автомата существует эквивалентный МП-автомат, для которого выполнено соотношение:

$$\forall \langle q_1, u, \alpha \rangle \rightarrow \langle q_2, \beta \rangle \in \Delta : |u| \leq 1, |\alpha| + |\beta| = 1$$

*Доказательство.* Аналогично предыдущему утверждению, однако останутся переходы

$$\langle q_1, w, \varepsilon \rangle \rightarrow \langle q_2, \varepsilon \rangle \in \Delta$$

Вводим dummy стековый элемент  $T$  и превращаем переход выше в два последовательных перехода

$$\langle q_1, w, \varepsilon \rangle \rightarrow \langle q', T \rangle, \langle q', \varepsilon, T \rangle \rightarrow \langle q_2, \varepsilon \rangle$$

$\square$

## 2.11 Построение МП-автомата по КС-грамматике

**Теорема 2.11.1.** Для любой КС-грамматики  $G$  существует МП-автомат  $M$ :

$$L(M) = L(G)$$

*Доказательство.* Пусть  $G = \langle N, \Sigma, P, S \rangle$ , определим автомат

$$M = \langle \{q_0, q_1\}, \Sigma, N \cup \Sigma, \Delta, q_0, \{q_1\} \rangle$$

причём  $\Delta$  состоит из правил:

- $\langle q_0, \varepsilon, S \rangle \rightarrow \langle q_1, \varepsilon \rangle$
- $\langle q_0, \varepsilon, \alpha \rangle \rightarrow \langle q_0, A \rangle$ , если  $A \rightarrow \alpha \in P$
- $\langle q_0, a, \varepsilon \rangle \rightarrow \langle q_1, a \rangle$ ,  $a \in \Sigma$

Хотим доказать, что

$$\alpha \vdash w \Leftrightarrow \langle q_0, w, \varepsilon \rangle \vdash_M \langle q_0, \varepsilon, \alpha \rangle$$

Вначале для  $\Rightarrow$ :

Индукция по длине вывода:

- База  $\alpha \vdash_0 w$ . Тогда  $\alpha = a = w \Rightarrow \langle q_0, a, \varepsilon \rangle \vdash \langle q_0, \varepsilon, a \rangle$  по третьему типу правил.
- Для перехода рассмотрим  $A \vdash_1 \alpha_1 \cdots \alpha_k \vdash w_1 \cdots w_k = w$ , где  $\alpha_i$  – нетерминалы, а  $w_i$  – терминалы
- По предположению,  $\langle q_0, w_j, \varepsilon \rangle \vdash \langle q_0, \varepsilon, \alpha_j \rangle$
- Тогда

$$\langle q_0, w_1 \cdots w_k, \varepsilon \rangle \vdash \langle q_0, w_2 \cdots w_k, \alpha_1 \rangle \vdash \cdots \vdash \langle q_0, \varepsilon, \alpha_1 \cdots \alpha_k \rangle \vdash \langle q_0, \varepsilon, A \rangle$$

Теперь  $\Leftarrow$ :

Индукция по длине вывода:

- База: 1 шаг
  - $\langle q_0, a, \varepsilon \rangle \vdash \langle q_0, \varepsilon, a \rangle \Rightarrow a \vdash a$
  - $\langle q_0, \varepsilon, \varepsilon \rangle \vdash \langle q_0, \varepsilon, A \rangle \Rightarrow A \rightarrow \varepsilon \in P$
- Переход:  $\langle q_0, w, \varepsilon \rangle \vdash \langle q_0, \varepsilon, \alpha_1 \cdots \alpha_k \rangle \vdash_1 \langle q_0, \varepsilon, A \rangle$ 
  - При этом  $A \rightarrow \alpha_1 \cdots \alpha_k \in P$
  - Найдём момент, когда на стеке появится  $\alpha_1$ :  $\langle q_0, w, \varepsilon \rangle \vdash \langle q_0, w', \alpha_1 \rangle$
  - Тогда  $w = u_1 w'$ , причём  $\langle q_0, u_1, \varepsilon \rangle \vdash \langle q_0, \varepsilon, \alpha_1 \rangle$ . Тогда  $\alpha_1 \vdash u_1$ .
  - Аналогично,  $\alpha_j \vdash u_j$  и  $A \vdash \alpha_1 \cdots \alpha_k \vdash w$

Для доказательства совпадения языков остаётся заметить эквивалентность следующих фактов:

- $w \in L(M)$
- $\langle q_0, w, \varepsilon \rangle \vdash \langle q_1, \varepsilon, \varepsilon \rangle$
- $\langle q_0, w, \varepsilon \rangle \vdash \langle q_0, \varepsilon, S \rangle \vdash \langle q_1, \varepsilon, \varepsilon \rangle$
- $S \vdash w$
- $w \in L(G)$

□

## 2.12 Построение КС-грамматики по МП-автомату

**Теорема 2.12.1.** *Каждому МП-автомату  $M$  соответствует КС-грамматика  $G$ :*

$$L(M) = L(G)$$

*Доказательство.* Определим множество нетерминалов  $N$ , как

$$N = \{A_{ij} \mid q_i, q_j \in Q\} \cup S$$

Правила будем строить так:

- $A_{ij} \rightarrow uA_{st}vA_{rj}$ , если
  - $\langle q_i, u, \varepsilon \rangle \rightarrow \langle q_s, A \rangle$
  - $\langle q_t, v, A \rangle \rightarrow \langle q_r, \varepsilon \rangle$
- $A_{ij} \rightarrow \varepsilon$
- $S \rightarrow A_{0j}$ , если  $q_j \in F$

Хотим показать, что

$$A_{ij} \vdash w \Leftrightarrow \langle q_i, w, \varepsilon \rangle \vdash \langle q_j, \varepsilon, \varepsilon \rangle$$

В начале  $\Rightarrow$ :

- База:  $A_{ij} \vdash_1 w$ . Тогда  $w = \varepsilon, i = j$ . Значит,  $\langle q_i, w, \varepsilon \rangle \vdash \langle q_j, \varepsilon, \varepsilon \rangle$
- Переход:  $A_{ij} \vdash_1 uA_{st}vA_{rj} \vdash uzvy$ 
  - По построению  $\langle q_i, u, \varepsilon \rangle \rightarrow \langle q_s, A \rangle$ ,  $\langle q_t, v, A \rangle \rightarrow \langle q_r, \varepsilon \rangle$
  - По предположению:  $A_{st} \vdash z \Rightarrow \langle q_s, z, \varepsilon \rangle \vdash \langle q_t, \varepsilon, \varepsilon \rangle$  и  $A_{rj} \vdash y \Rightarrow \langle q_r, y, \varepsilon \rangle \vdash \langle q_j, \varepsilon, \varepsilon \rangle$
  - Тогда в итоге

$$\langle q_i, w, \varepsilon \rangle \vdash \langle q_i, uzvy, \varepsilon \rangle \vdash \langle q_s, zvy, A \rangle \vdash \langle q_t, vy, A \rangle \vdash \langle q_r, y, \varepsilon \rangle \vdash \langle q_j, \varepsilon, \varepsilon \rangle$$

Теперь для  $\Leftarrow$ :

- База:  $\langle q_i, w, \varepsilon \rangle \vdash_0 \langle q_j, \varepsilon, \varepsilon \rangle$ . Тогда  $i = j, w = \varepsilon, A_{ii} \rightarrow \varepsilon$
- Переход:  $\langle q_i, w, \varepsilon \rangle \vdash \langle q_j, \varepsilon, \varepsilon \rangle$

– Идя по цепочке вывода,  $\exists w = uzvy$ , причём

$$\langle q_i, w, \varepsilon \rangle \vdash \langle q_i, uzvy, \varepsilon \rangle \vdash \langle q_s, zvy, A \rangle \vdash \langle q_t, vy, A \rangle \vdash \langle q_r, y, \varepsilon \rangle \vdash \langle q_j, \varepsilon, \varepsilon \rangle$$

– По предположению:  $A_{st} \vdash z \Leftarrow \langle q_s, z, \varepsilon \rangle \vdash \langle q_t, \varepsilon, \varepsilon \rangle$  и  $A_{rj} \vdash y \Leftarrow \langle q_r, y, \varepsilon \rangle \vdash \langle q_j, \varepsilon, \varepsilon \rangle$

– Тогда

$$A_{ij} \vdash uA_{st}vA_{rj} \vdash uzvy = w$$

Для доказательства теоремы заметим эквивалентность следующих утверждений:

- $w \in L(G)$
- $S \vdash w$
- $\exists q_j \in F : S \vdash A_{0j} \vdash w$
- $\exists q_j \in F : \langle q_0, w, \varepsilon \rangle \vdash \langle q_j, \varepsilon, \varepsilon \rangle$
- $w \in L(M)$

□

## 2.13 Нормальная форма Грейбах для КС-грамматик

**Определение 2.13.1.** КС-грамматика находится в нормальной форме Грейбах, если все правила имеют такой и только такой вид:

- $A \rightarrow a, a \in \Sigma$
- $A \rightarrow aB, A \in \Sigma; B \in N; B \neq S$
- $A \rightarrow aBC, A \in \Sigma; B, C \in N; B, C \neq S$
- $S \rightarrow \varepsilon$

**Теорема 2.13.1.** Любая КС-грамматика может быть представлена в НФ Грейбах

*Доказательство.* Определим оператор левого деления

$$B \setminus A := \{w \in \Sigma^* \mid \exists x \in B : xw \in A\}$$

По исходной грамматике  $G = \langle N, \Sigma, P, S \rangle$  построим

$$G_g = \langle \{S\} \cup \{B \setminus A \mid A, B \in N\}, \Sigma, P_g, S \rangle$$

Как будут выглядеть правила новой грамматики?

- $S \rightarrow a(A \setminus S)$ , если  $A \rightarrow a$
- $A \setminus A \rightarrow \varepsilon, A \in N$
- $B \setminus A \rightarrow e(E \setminus D)(C \setminus A)$ , если  $C \rightarrow BD$  и  $E \rightarrow e$



По сути, должны доказать

$$B \setminus A \vdash w \Leftrightarrow A \vdash Bw$$

В начале  $\Rightarrow$ :

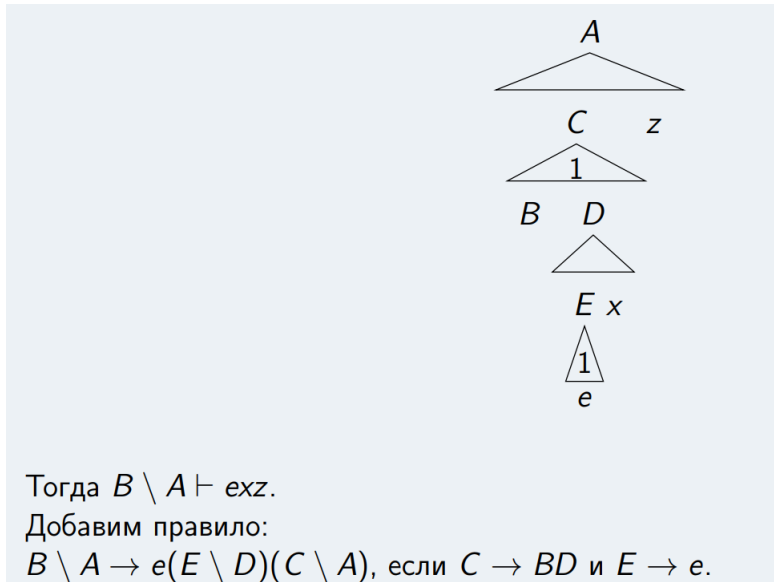
Индукция по длине вывода

- База:  $B \setminus A \vdash_1 w$ . Тогда  $A = B, w = \varepsilon$ . Значит,  $A \vdash A\varepsilon$ .
- Переход:  $B \setminus A \vdash_1 e(E \setminus D)(C \setminus A) \vdash euv$ .
  - По предположению:  $E \setminus D \vdash u \Rightarrow D \vdash Eu, C \setminus A \vdash v \Rightarrow A \vdash Cv$
  - По условию  $C \vdash BD$
  - $A \vdash Cv \vdash BDv \vdash BEuv \vdash Beuv \vdash Bw$

Теперь  $\Leftarrow$ :

Индукция по длине вывода:

- База:  $A \vdash_0 Bw$ . Тогда  $A = B, w = \varepsilon$ . Значит,  $A \setminus A \vdash \varepsilon$ .
- Переход: смотри картинку



Осталось доказать равенство языков, для этого приведём изначальную грамматику к нормальной форме Хомского  $G_h$ .

Если  $\varepsilon \in L(G_h)$ , то  $\varepsilon \in L(G_g)$  по правилу  $S \rightarrow \varepsilon$ .

Пусть  $w = au, a \in \Sigma$ . Следующие утверждения эквивалентны:

- $au \in L(G_h)$
- $\exists A \rightarrow a : S \vdash Au \vdash_1 au$
- $\exists A \rightarrow a : S \vdash a(A \setminus S) \vdash au$
- $au \in L(G_g)$

Осталось убрать правила  $A \setminus A \rightarrow \varepsilon$ : аналогично удалению  $\varepsilon$ -порождающих. □

**Замечание.** Аналогично можем определить обратную нормальную форму Грейбах с правилами вида:

- $A \rightarrow a$
- $A \rightarrow Ba$
- $A \rightarrow CBa$
- $S \rightarrow \varepsilon$

## 2.14 Построение МП-автомата без ерс-переходов

**Теорема 2.14.1.** *Для любого МП-автомата существует эквивалентный МП-автомат, для которого выполнено соотношение:*

$$\forall (\langle q_1, u, \alpha \rangle \rightarrow \langle q_2, \beta \rangle) \in \Delta : |u| = 1$$

*Доказательство.* Рассмотрим алгоритм "КС-грамматика  $\rightarrow$  МП-автомат" для обратной нормальной формы Грейбах!

Хотим доказать, что

$$\langle q_0, w, \varepsilon \rangle \vdash \langle q_0, \varepsilon, A \rangle \Leftrightarrow A \vdash w \ (A \neq S)$$

Вначале  $\Leftarrow$ :

- $w \in L(G)$
- $S \vdash \varepsilon \Leftrightarrow \langle q_0, \varepsilon, \varepsilon \rangle \vdash \langle q, \varepsilon, \varepsilon \rangle \ (q \in F)$
- $S \vdash_1 CBA \vdash vua$ :
  - $C \vdash v \Rightarrow \langle q_0, vua, \varepsilon \rangle \vdash \langle q_0, ua, C \rangle$
  - $B \vdash u \Rightarrow \langle q_0, ua, C \rangle \vdash \langle q_0, a, CB \rangle$
  - $\langle q_0, a, CB \rangle \vdash_1 \langle q_1, \varepsilon, \varepsilon \rangle$

Теперь  $\Rightarrow$ :

- $w \in L(G)$
- $S \vdash \varepsilon \Leftrightarrow \langle q_0, \varepsilon, \varepsilon \rangle \vdash \langle q, \varepsilon, \varepsilon \rangle \ (q \in F)$
- $\langle q_0, w, \varepsilon \rangle \vdash \langle q_1, \varepsilon, \varepsilon \rangle$ :
  - $w = xa, \langle q_0, a, CB \rangle \vdash_1 \langle q_1, \varepsilon, \varepsilon \rangle$
  - Тогда  $\exists x = vu : \langle q_0, ua, C \rangle \vdash \langle q_0, a, CB \rangle; \langle q_0, vua, \varepsilon \rangle \vdash \langle q_0, ua, C \rangle$
  - Значит  $B \vdash u, C \vdash v$
  - $S \vdash CBA \vdash vua = w$

□

## 2.15 КС-языки и теоретико-множественные операции

**Утверждение 2.15.1.** *КС-языки замкнуты относительно пересечения с регулярными языками*

*Доказательство.* Пусть  $M = \langle Q, \Sigma, \Delta_M, q_0, D_M \rangle, L(M) = R$  – НКА с 1-буквенными переходами.

$MP = \langle P, \Sigma, \Gamma, \Delta_P, p_0, F_P \rangle, L(MP) = L$  – МП-автомат с 1-буквенными переходами.

Построим МП-автомат как декартово произведение НКА со стеком:

$$M_i = \langle Q \times P, \Sigma, \Gamma, \Delta, (q_0, p_0), F_M \times F_P \rangle$$

где  $\Delta$  состоит из  $\langle (q_1, p_1), a, \alpha \rangle \rightarrow \langle (q_2, p_2), \beta \rangle$ , если:

- $\langle q_1, a \rangle \rightarrow q_2 \in \Delta_M$
- $\langle p_1, a, \alpha \rangle \rightarrow \langle p_2, \beta \rangle \in \Delta_P$

□

## 2.16 Анализатор перенос-свёртка

Смотри конспект Сорокина, который можно распечатать и принести на экзамен

## 2.17 Операция перехода для LR-ситуаций

Смотри конспект Сорокина, который можно распечатать и принести на экзамен

## 2.18 Алгоритм построения LR-таблицы по LR-ситуациям

Смотри конспект Сорокина, который можно распечатать и принести на экзамен

## 2.19 Алгоритм разбора по LR-таблице

Смотри конспект Сорокина, который можно распечатать и принести на экзамен

## 2.20 ДМП-автоматы

**Определение 2.20.1.** Два перехода в МП-автомате:  $\langle q_1, u_1, \alpha_1 \rangle \rightarrow \langle q'_1, \beta'_1 \rangle$  и  $\langle q_2, u_2, \alpha_2 \rangle \rightarrow \langle q'_2, \beta'_2 \rangle$  называются **совместными**, если выполняются следующие условия:

- $q_1 = q_2$
- $u_1 \sqsubset u_2$  или  $u_2 \sqsubseteq u_1$
- $\alpha_1 \sqsubseteq \alpha_2$  или  $\alpha_2 \sqsubseteq \alpha_1$

**Определение 2.20.2.** МП-автомат называется детерминированным, если в нём нет совместных переходов.

**Определение 2.20.3.** Пусть  $M$  – ДМП-автомат. Тогда языком, который задаёт данный автомат, называется

$$L(M) = \{w \mid \exists q \in F : \langle q_0, w\$, \varepsilon \rangle \vdash \langle q, \varepsilon, \varepsilon \rangle\}$$

**Лемма 2.20.1.** Любой ДМП-автомат эквивалентен ДМП автомату с переходами

$$\langle q_1, u, \alpha \rangle \rightarrow \langle q_2, \beta \rangle, |u| + |\alpha| \leq 1$$

*Доказательство.* Разбиваем переходы совместно... □

**Теорема 2.20.1.** Если  $l$  – длина самой длинной строки, которая может быть записана в ходе перехода в ДМП-автомате, то существует такая константа  $N = F(|Q|, |\Gamma|, l)$  такая, что если  $\langle q, \varepsilon, A \rangle \vdash_N \langle q', \varepsilon, \alpha \rangle, |\alpha| > 0$ , то в ДМП-автомате возникнет цикл в конфигурации.

*Доказательство.* Если длина стека в пути не больше  $|Q||\Gamma|l$ , то в качестве  $N$  возьмём

$$N = |Q|(1 + |\Gamma| + |\Gamma|^2 + \dots + |\Gamma|^{|Q||\Gamma|l})$$

где второй множитель обозначает максимальное число стеков длины  $\leq |Q||\Gamma|l$ . Тогда по принципу Дирихле найдётся одна конфигурация.

Иначе

$$\exists \langle q_0, \varepsilon, A \rangle \vdash_N \langle q', \varepsilon, \beta \rangle : |\beta| \geq |Q||\Gamma|l + 1$$

Давайте отмечать на нашем пути состояния так, чтобы длина стека у полученной подвыборки строго возрастала.

Так как за переход длина стека меняется не более чем на  $l$ , то мы отметим хотя бы  $|Q||\Gamma| + 1$  конфигураций.

Тогда по принципу Дирихле  $\exists$  два состояния:

$$\langle q_1, \varepsilon, \alpha A \rangle, \langle q_1, \varepsilon, \alpha \gamma A \rangle$$

□

**Теорема 2.20.2.** Количество Shift'ов в LR-алгоритме ограничено сверху  $|w|$ , а количество Reduce'ов  $\leq |w| \cdot N \cdot c$

*Доказательство.* Введём потенциал, который является суммой:

- Числа терминальных символов грамматики на стеке
- Числа нетерминальных символов грамматики на стеке
- Удвоенного числа непрочитанных букв в слове

При Shift имеем +1 на стеке, но -2 непрочитанных букв.

Отметим ситуации, после которых стек уменьшался – после них потенциал уменьшается хотя бы на 1.

Имея изначальный потенциал  $2|w|$  и зная, что в конце он равен 0, то получим, что количество ситуаций, разобранных выше, ограничено.

Остаётся разобрать ситуации, когда Reduce не уменьшает стек, это происходит при обработке правил

$$A \rightarrow \varepsilon; \quad A \rightarrow B$$

Очевидно, таких правил ограниченное количество. Также, мы не можем прийти в соответствующую им ситуацию несколько раз, так как тогда мы заиклимся. □

## 2.21 Приведение LR-алгоритма к распознаванию ДМП-автоматом

**Утверждение 2.21.1.** Работа  $LR(1)$  алгоритма эмулируется работой ДМП-автомата.

*Доказательство.* В качестве стека ДМП-автомата будем использовать стек LR-алгоритма (последовательность символ-вершина-...). То есть при операции Shift:

$$\langle q_i, a, \varepsilon \rangle \rightarrow \langle \text{GOTO}(q_i, a), a\text{GOTO}(q_i, a) \rangle$$

При операции Reduce снимаем со стека путь, переходим в соответствующую ситуацию из LR-таблицы и кладём на стек тот нетерминал, который считали.

Почему этот автомат будет ДМП? Заметим, что свойство детерминированности эквивалентно отсутствию конфликтов, которых у нас нет по условию.  $\square$

## 3 Парсеры

### 3.1 Корректность алгоритма Эрли

**Определение 3.1.1.** Для каждого правила  $A \rightarrow \alpha\beta$  определим ситуацию:

$$(A \rightarrow \alpha \cdot \beta, i) \in D_j, i \in [0 : |w|], j \in [0 : |w|]$$

где  $i$  отвечает за то, сколько букв было прочитано до "захода" в это правила, а  $j$  за то, сколько букв прочитано на момент символа  $\cdot$ .

На протяжении всего алгоритма используем 3 операции:

- Scan – читаем букву
- Predict – спускаемся вниз
- Complete – поднимаемся вверх

Добавим правило  $S' \rightarrow S$ , тогда стартовой ситуацией определим

$$(S' \rightarrow \cdot S, 0) \in D_0$$

а финальной будет

$$(S' \rightarrow S \cdot, 0) \in D_{|w|}$$

Корректность вывода при наличии ситуации гарантируется леммой:

**Лемма 3.1.1.** Каждой ситуации соответствует вывод:

$$(A \rightarrow \alpha \cdot \beta, i) \in D_j \Leftrightarrow \exists \psi \in (N \cup \Sigma)^* : \alpha \vdash w[i : j] \\ S' \vdash w[0 : i] A \psi \vdash_1 w[0 : i] \alpha \beta \psi$$

*Доказательство.* Докажем индукцией по количеству эффективных шагов в алгоритме:

- База
  - Появилась ситуация  $(S' \rightarrow \cdot S, 0) \in D_0$

- $S' \vdash \varepsilon w[0 : 0] S' \varepsilon = w[0 : 0] S \varepsilon, \alpha = \varepsilon = w[0 : 0]$
- Переход при Scan
  - $(A \rightarrow \alpha \cdot a\beta, i) \in D_j, w[j] = a \Rightarrow (A \rightarrow \alpha a \cdot \beta, i) \in D_{j+1}$
  - Предположение:  $S' \vdash w[0 : i] A\psi \vdash_1 w[0 : i] \alpha a\beta\psi$
  - $\alpha \vdash w[i : j], \alpha a \vdash w[i : j + 1]$
- Переход при Predict
  - $(B \rightarrow \cdot \gamma, j) \in D_j$  появилась при Predict после ситуации  $(A \rightarrow \alpha \cdot B\beta, i) \in D_j$
  - Предположение  $S' \vdash w[0 : i] A\psi \vdash_1 w[0 : i] \alpha B\beta\psi$
  - $\alpha \vdash w[i : j] \Rightarrow w[0 : i] \alpha \vdash w[0 : j]$
  - $S' \vdash w[0 : i] \alpha B\beta\psi \vdash w[0 : j] B\beta\psi \vdash_1 w[0 : j] \gamma\beta\psi$
- Переход при Complete
  - $(A \rightarrow \alpha B \cdot \beta, i) \in D_j$  появилась после ситуаций:  $\exists k : (A \rightarrow \alpha \cdot B\beta, i) \in D_k; \exists (B \rightarrow \gamma \cdot, k) \in D_j$
  - Предположение  $S' \vdash w[0 : i] A\psi \vdash_1 w[0 : i] \alpha B\beta\psi, \alpha \vdash w[i : k]$
  - Предположение 2:  $B \vdash_1 \gamma \vdash w[k : j]$
  - Итого  $\alpha B \vdash w[i : k] w[k : j] = w[i : j]$

□

### 3.2 Полнота алгоритма Эрли

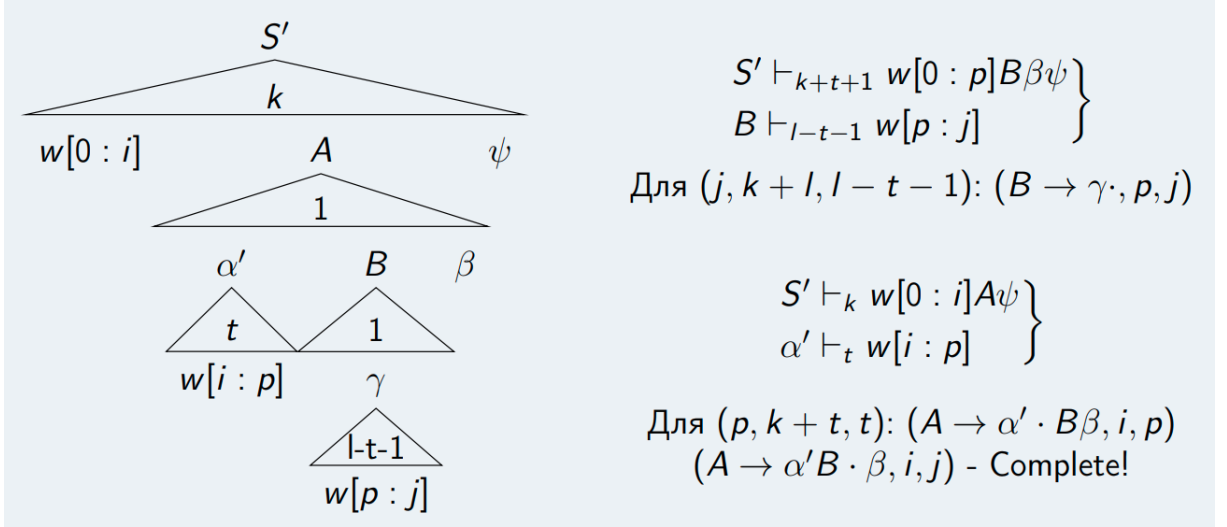
Каждому выводу соответствует ситуация:

$$S' \vdash_k w[0 : i] A\psi \vdash_1 w[0 : i] \alpha\beta\psi, \alpha \vdash_l w[i : j]$$

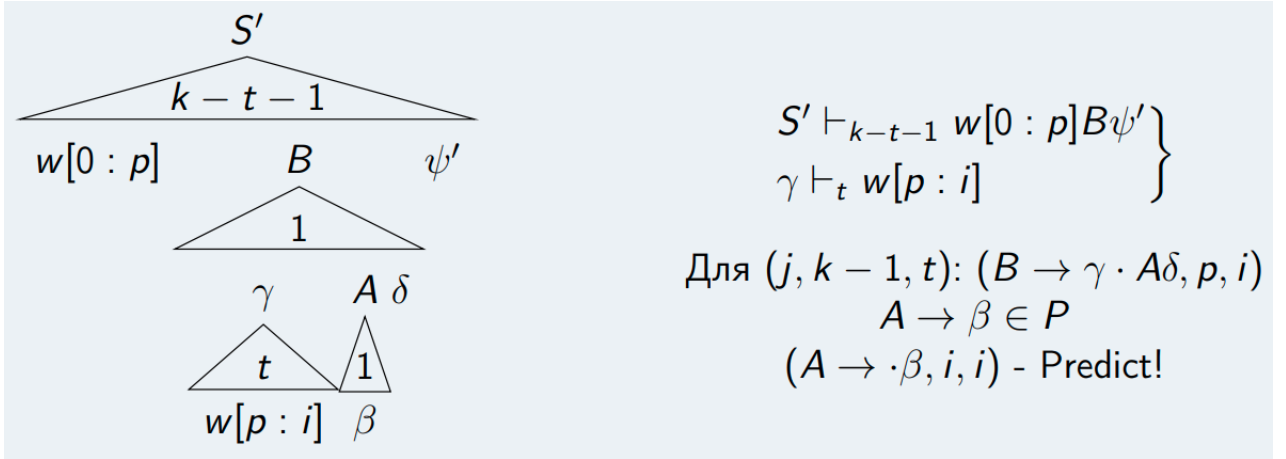
индукцией по  $(j, l + k, l)$

- База  $j = 0, k + l = 0$ 
  - $l = 0 \Rightarrow \alpha = \varepsilon$
  - $k = 0 \Rightarrow w[0 : i] = \varepsilon, A = S'$
  - $S' \rightarrow S \Rightarrow (S' \rightarrow \cdot S, 0) \in D_0$
- Переход: рассмотрим последний символ  $\alpha$ . Возможны 3 случая:
  1.  $\alpha = \alpha'b$ :
    - $\alpha' \vdash_l w[i : j - 1], w[j] = b$
    - $S' \vdash w[0 : i] \alpha'b\beta\psi$
    - Предположение  $(j - 1, k + l, k) : (A \rightarrow \alpha' \cdot bB, i) \in D_{j-1}$
    - $(A \rightarrow \alpha'b \cdot B, i) \in D_j$  по Scan

2.  $\alpha = \alpha' B$



3.  $\alpha = \varepsilon$



### 3.3 Оптимальный алгоритм и обоснование сложности

#### Оптимальный алгоритм

- В  $D_j$  надо быстро обращаться к правилам с  $\cdot B$
- Храним в виде  $D_j[B]$
- Правая часть – в виде связного списка
- Быстрее – вычислить id для каждой ситуации
- Меньше памяти – вычислить hash для каждой ситуации

#### Обоснование сложности

Пусть  $|G|$  – суммарное количество символов в правых частях правил.

- Сложность Scan

Правила из  $D_k[w[j]] \Rightarrow O(|D_j[w[j]]|) = O(|D_j|) = O(|w||G|)$

- Сложность Predict

Перебираем правила  $B \rightarrow \gamma - O(|P|)$ . Рассматриваем  $D_j[B]$  и помечаем, рассмотрена ли была ситуация с этим  $B$ .

Сложность для  $D_j$  :  $O(|D_j||P|) = O(|w||G|^2)$

- Сложность Complete

Рассматриваем правила  $(B \rightarrow \gamma \cdot, k) \in D_j$ . Делаем перебор по  $D_k[B] \Rightarrow$  количество обращений к  $D_k$  :  $O(|D_0| + |D_1| + \dots + |D_j|) = O(|w|^2|G|)$ .

Количество правил  $B \rightarrow \gamma$  :  $O(|G|) \Rightarrow$  асимптотика шага –  $O(|w|^2|G|^2)$

Количество шагов:  $O(|w|)$

Итого на Scan:  $O(|w|^2|G|)$

Итого на Predict:  $O(|w|^2|G|^2)$

Итого на Complete:  $O(|w|^3|G|^2)$

## 4 Конечные преобразователи

### 4.1 Конечные преобразователи и задаваемые ими преобразования

**Определение 4.1.1.** Конечным преобразователем (КПТелем)  $M$  называется кортеж  $M = \langle Q, \Sigma, \Gamma, \Delta, q_0, F \rangle$ , где:

- $Q$  – множество состояний,  $|Q| < +\infty$
- $\Sigma$  – входной алфавит,  $|\Sigma| < +\infty$
- $\Gamma$  – выходной алфавит,  $|\Gamma| < +\infty$
- $\Delta \subset (Q \times \Sigma^*) \times (Q \times \Gamma^*)$  – множество переходов
- $q_0$  – стартовое состояние
- $F \subset Q$  – множество завершающих состояний.

**Определение 4.1.2.** Конфигурацией КПтеля  $M$  называется тройка  $\langle q, u, v \rangle \in Q \times \Sigma^* \times \Gamma^*$ .

Неформально:

- Находимся в состоянии  $q$
- Осталось прочесть слово  $u$
- Вывели слово  $v$

**Определение 4.1.3.**  $\vdash$  – наименьшее рефлексивное транзитивное отношение, что

$$\forall \langle q_1, u \rangle \rightarrow \langle q_2, v \rangle \in \Delta$$

выполнено:

$$\forall y \in \Sigma^*, z \in \Gamma^* : \langle q_1, uv, z \rangle \vdash \langle q_2, y, zv \rangle$$



**Определение 4.1.4.** Конечным преобразованием будем называть

$$\psi = \{(u, v) \in \Sigma^* \times \Gamma^* \mid \exists q \in F : \langle q_0, u, \varepsilon \rangle \vdash \langle q, \varepsilon, v \rangle\}$$

**Утверждение 4.1.1.** Для любого КПтеля существует эквивалентный КПтель, для которого выполнено соотношение:

$$\forall (\langle q_1, u \rangle \rightarrow \langle q_2, v \rangle) \in \Delta : |u| + |v| = 1$$

*Доказательство.* 1. Раскроем все длинные переходы

2. Удалим  $\varepsilon : \varepsilon$  переходы по аналогии с НКА

□

**Теорема 4.1.1.** Нива.

Любое конечное преобразование можно представить в виде композиции:

- Обратного неудлиняющего гомоморфизма:  $\varphi^{-1}$
- Ограничения на регулярный язык  $id_R$
- Неудлиняющего гомоморфизма:  $\eta$