

Московский физико-технический институт
Физтех-школа прикладной математики и информатики

ФОРМАЛЬНЫЕ ЯЗЫКИ И ТРАНСЛЯЦИИ
III СЕМЕСТР

Лектор: *Павел Ибрагимович Ахтямов*



Автор: *Киселев Николай*
Репозиторий на Github

осень 2025

Содержание

1	Конечные Автоматы	2
1.1	Напоминание	2
1.2	Недетерминированный Конечный Автомат	2
1.3	Детерминированный Конечный Автомат	3
1.4	Регулярные выражения	5
1.5	Минимальный Полный Конечный Детерминированный Автомат?	6
2	Грамматики	9
2.1	Иерархия Хомского	9
2.2	Контекстно-свободные грамматки	11
2.3	Нормальная форма Хомского	11
2.3.1	Удаление непорождающих нетерминалов	12
2.3.2	Удаление недостижимых нетерминалов	12
2.3.3	Удаление смешанных правил	12
2.3.4	Удаление длинных цепочек	12
2.3.5	Удаление ε -переходов	13
2.3.6	Удаление одиночных правил	14
2.4	Разбор КС-грамматики	14
2.4.1	Рекурсивный спуск	14
2.4.2	Алгоритм Кока-Янгера-Касами	14

1 Конечные Автоматы

1.1 Напоминание

Определение 1.1. Алфавит — конечное непустое множество

Определение 1.2. Слово — кортеж из букв

Определение 1.3. Пусть L_1, L_2 — множества слов. $L_1 L_2 = \{uv \mid u \in L_1, v \in L_2\}$

Определение 1.4. $\Sigma^* = \bigcup_{k=0}^{\infty} \Sigma^k$ — звезда Клини

Определение 1.5. ε — пустое слово

1.2 Недетерминированный Конечный Автомат

Определение 1.6. Недетерминированный Конечный Автомат — кортеж $(Q, \Sigma, \Delta, q_0, F)$, удовлетворяющий следующим условиям:

1. Q называется множеством состояний, Q конечное
2. Σ — Алфавит
3. $\Delta \subset Q \times \Sigma^* \times Q$ — переходы, Δ конечное
4. $q_0 \in Q$ — стартовое состояние
5. $F \subset Q$ — множество завершающих состояний

Определение 1.7. Конфигурация Автомата — элемент $Q \times \Sigma^*$. обозначается $\langle q, w \rangle$

Определение 1.8. \vdash — наименьшее по включению рефлексивное транзитивное отношение над $Q \times \Sigma^*$, что $\forall u \in \Sigma^*, (\langle q_1, w \rangle \rightarrow q_2) \in \Delta \Rightarrow \langle q_1, wu \rangle \vdash \langle q_2, u \rangle$.

Определение 1.9. $L(M) = \{w \in \Sigma^* \mid \exists q_F \in F : \langle q_0, w \rangle \vdash \langle q_F, \varepsilon \rangle\}$.

Определение 1.10. Язык L называется автоматным, если $\exists M$ — автомат, такой, что $L(M) = L$

Утверждение 1.1. В определении автомата можно считать, что $|F| = 1$.

Неформально. Добавим новое состояние q_F . Из всех конечных состояний сделаем в него переходы по ε . Сделаем q_F единственным конечным. Тогда наш автомат успешно завершился \Leftrightarrow мы перешли в q_F и слово закончилось, но в него можно было прийти только по ε -переходу, значит в исходном автомате мы бы остановились в конечном состоянии и на пустом слове. \square

Доказательство. Рассмотрим автомат $M' = (Q \cup \{q_F\}, \Sigma, \Delta \cup F \times \{\varepsilon\} \times \{q_F\}, q_0, \{q_F\})$.

1. $L(M) \subset L(M')$

$$w \in L(M) \Rightarrow \exists q \in F : \langle q_0, w \rangle \vdash_M \langle q, \varepsilon \rangle$$

$$\Delta \subset \Delta' \Rightarrow \langle q_0, w \rangle \vdash_{M'} \langle q, \varepsilon \rangle \vdash_{M'} \langle q_F, \varepsilon \rangle \Rightarrow \langle q_0, w \rangle \vdash_{M'} \langle q_F, \varepsilon \rangle$$

2. $L(M) \supset L(M')$

$$w \in \text{Im } L(M') \Rightarrow \langle q_0, w \rangle \vdash_{M'} \langle q_F, \varepsilon \rangle$$

$$\exists u, q_1 \in F : \langle q_0, w \rangle \vdash_{M'} \langle q_1, u \rangle \vdash_{M',1} \langle q_F, \varepsilon \rangle \Rightarrow u = \varepsilon \Rightarrow \langle q_0, w \rangle \vdash_{M'} \langle q_1, \varepsilon \rangle$$

□

Утверждение 1.2. В определении автомата можно считать, что $\forall \langle q_1, w \rangle \rightarrow q_2 \in \Delta : |w| \leq 1$.

Доказательство. Превращаем $q_1 \rightarrow_{abc} q_2$ в $q_1 \rightarrow_a q_3 \rightarrow_b q_4 \rightarrow_c q_2$. □

Теорема 1.1. В определении автомата можно считать, что $\forall \langle q_1, w \rangle \rightarrow q_2 \in \Delta : |w| = 1$.

Неформально. Делаем две операции:

1. В цепочку $q_1 \rightarrow_\varepsilon q_2 \rightarrow_\varepsilon q_3 \rightarrow_\varepsilon \dots \rightarrow_\varepsilon q_{n-1} \rightarrow_a q_n$ удаляем ε -переходы, вместо них добавим переходы вида $q_1 \rightarrow_a q_i$.
2. В цепочке $q_1 \rightarrow_\varepsilon q_2 \rightarrow_\varepsilon q_3 \rightarrow_\varepsilon \dots \rightarrow_\varepsilon q_n$, где q_n — конечное состояние делаем все состояния конечными

Сначала до упора делаем 1, потом 2, потом удаляем оставшиеся ε -переходы. □

Доказательство. Положим $\Delta(q, w) = \{q' \in Q \mid \langle q, w \rangle \vdash \langle q', \varepsilon \rangle\}$, $M' = \{Q, \Sigma, \Delta', q_0, F'\}$, где $\Delta' = \{\langle q_1, w \rangle \rightarrow q_2 \mid \exists q_3 \in \Delta(q_1, \varepsilon) : \langle q_3, \varepsilon \rangle \rightarrow q_2 \in \Delta\}$, $F' = \{q' \mid \exists q \in F : \langle q', \varepsilon \rangle \vdash \langle q, \varepsilon \rangle\}$

1. $L(M) \subset L(M')$

$$w \in L(M), w = w_1 w_2 \dots w_k, w_i \in \Sigma$$

$$\exists q \in F \langle q_0, w \rangle \vdash \langle q, \varepsilon \rangle$$

$$\begin{array}{ccccccc} \langle q_0, w_1 w_2 \dots w_k \rangle & \vdash_M & \langle q_1, w_1 w_2 \dots w_k \rangle & \vdash_{M,1} & \langle q'_1, w_2 \dots w_k \rangle & \vdash_M & \\ & \vdash_M & \langle q_2, w_2 \dots w_k \rangle & \vdash_{M,1} & \langle q'_2, w_3 \dots w_k \rangle & \vdash_M & \\ \Rightarrow & & \vdots & & & & \\ & \vdash_M & \langle q_s, w_k \rangle & \vdash_{M,1} & \langle q'_s, \varepsilon \rangle & \vdash_M & \langle q, \varepsilon \rangle \end{array}$$

□

1.3 Детерминированный Конечный Автомат

Определение 1.11. НКА $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$ называется Детерминированным конечным автоматом, если

1. $\forall \langle q, w \rangle \rightarrow q_1 \in \Delta : |w| = 1$
2. $\forall \langle q, w \rangle \rightarrow q_1 \in \Delta : |\Delta(q, w)| \leq 1$

Определение 1.12. Если в определении выше $|\Delta(q, w)| = 1$, то такой ДКА называется полным

Теорема 1.2. \forall НКА $M \exists$ полный ДКА M' , такой, что $L(M) = l(M')$, причем $M' = \langle 2^Q, \Sigma, \Delta', \{q_0\}, F' \rangle$, где

$$\begin{aligned}\Delta(S, w) &= \bigcup_{a \in S} \Delta(a, w) \\ F' &= \{S \subset Q : S \cap F \neq \emptyset\} \\ \Delta' &= \{\langle S, a \rangle \rightarrow \Delta(S, a)\}\end{aligned}$$

Лемма 1.1. В условиях теоремы, $\Delta'(\{q_0, w\}) = \{\Delta(\{q_0\}, w)\}$

Доказательство. $\Delta'(\{q_0\}, w) = \Delta(q_0, w)$. Введем индукцию по $|w|$

База: $|w| = 0 \Rightarrow w = \varepsilon$. $\Delta(q_0, \varepsilon) = \{q_0\}$, т.к. в M все переходы однобуквенные. $\Delta'(\{q_0\}, \varepsilon) = \{S | \langle \{q_0\}, \varepsilon \rangle \vdash_{M'} \langle S, \varepsilon \rangle = \{\{q_0\}\} \Rightarrow \{\Delta(q_0, \varepsilon)\} = \{\{q_0\}\} = \Delta'(\{q_0\}, \varepsilon)$

Переход:

$$\begin{aligned}\Delta'(\{q_0\}, w'a) &= \{S | \langle \{q_0\}, w'a \rangle \vdash \langle S, \varepsilon \rangle\} = \{S | \exists T \subset Q : \langle \{q_0\}, w'a \rangle \vdash_{M'} \langle T, a \rangle \vdash \langle S, \varepsilon \rangle\} = \\ &= \{S | \exists T \subset Q : T \in \Delta'(\{q_0\}, w), \langle T, a \rangle \vdash_{M',1} \langle S, \varepsilon \rangle\} = \\ &= \{S | \exists T \subset Q : T \in \{\Delta(\{q_0\}, w')\}, \langle T, a \rangle \vdash_{M',1} \langle S, \varepsilon \rangle\} = \\ &= \{S | \exists T \subset Q : T = \Delta(\{q_0\}, w'), S = \Delta(T, a)\} = \{\Delta(q_0, w'a)\} \\ \Delta(q_0, w'a) &= \{q | \langle q_0, w'a \rangle \vdash \langle q, \varepsilon \rangle\} = \{q | \exists q' : \langle q_0, w'a \rangle \vdash \langle q', a \rangle \vdash_{M,1} \langle q, \varepsilon \rangle\} = \\ &= \{q | \exists q' \in \Delta(q_0, w') : q \in \Delta(q', a)\} = \Delta(T, a) = S\end{aligned}$$

□

Доказательство Теоремы.

$$\begin{aligned}w \in L(M) &\Leftrightarrow \exists q \in F : \langle q_0, w \rangle \vdash \langle q, \varepsilon \rangle \Leftrightarrow \Delta(q_0, w) \cap F \neq \emptyset \Leftrightarrow \Delta(\{q_0\}, w) \cap F \neq \emptyset \Leftrightarrow \\ &\Leftrightarrow \Delta(\{q_0\}, w) \in F' \Leftrightarrow \Delta'(\{q_0\}, w) = \{\Delta(\{q_0\}, w)\} \subset F' \Leftrightarrow \Delta'(\{q_0\}, w) \cap F \neq \emptyset \Leftrightarrow w \in L(M')\end{aligned}$$

□

Утверждение 1.3. $\forall M$ — ДКА $\exists M'$ — ПДКА, такой, что $L(M) = L(M')$

Идея доказательства. Добавим сток q_s , в который можно зайти, но нельзя выйти. Т.е. $M' = \langle Q \cup \{q_s\}, \Sigma, \Delta', q_0, F \rangle$, где $\Delta' = \Delta \cup \{\langle q_s, a \rangle \rightarrow a | a \in \Sigma\} \cup \{\langle q, a \rangle \rightarrow q_s | \nexists q : \langle q, a \rangle \rightarrow q'\}$. Таким образом, если мы попали в сток, то мы из него не выберемся □

Теорема 1.3. Пусть L_1, L_2 — автоматные языки. Тогда следующие языки также автоматные:

1. $L_1 L_2$
2. $L_1 \cup L_2$
3. $L_1 \cap L_2$
4. \bar{L}
5. L^*

1.4 Регулярные выражения

Определение 1.13. Пусть Σ — алфавит. Регулярное выражение — конечная последовательность из $\Sigma, *, \cdot, +, 0, 1$, определяемая индуктивно:

1. 0 — регулярное выражение
2. 1 — регулярное выражение
3. a — регулярное выражение, $a \in \Sigma$
4. $\alpha + \beta$ — регулярное выражение, где α, β — регулярные выражения
5. $\alpha \cdot \beta$ — регулярное выражение, где α, β — регулярные выражения
6. α^* — регулярное выражение, где α — регулярное выражение

Определение 1.14. $L(\alpha) : \{\text{Множество регулярных выражений}\} \rightarrow 2^\Sigma$, где α — регулярное выражение задается рекурсивно:

1. $L(0) = \emptyset$
2. $L(1) = \{\varepsilon\}$
3. $L(a) = \{a\}, a \in \Sigma$
4. $L(\alpha + \beta) = L(\alpha) \cup L(\beta)$
5. $L(\alpha \cdot \beta) = L(\alpha)L(\beta)$
6. $L(\alpha^*) = (L(\alpha))^*$

Определение 1.15. Язык L называется регулярным, если $\exists \alpha$ — регулярное выражение, т.ч. $L(\alpha) = L$

Определение 1.16. Регулярный автомат — автомат, в котором на ребрах написаны регулярные выражения.

Теорема 1.4 (Клини). *Язык L регулярен тогда и только тогда, когда он автоматный*

Доказательство. \Rightarrow Введем индукцию по построению регулярного выражения

База:

R	Автомат
0	q_0, q_F
1	$q_0 \xrightarrow{\varepsilon} q_F$
a	$q_0 \xrightarrow{a} q_F$

Переход:

R	Автомат
$\alpha \cdot \beta$	$M(\alpha) \xrightarrow{\varepsilon} M(\beta)$
$\alpha + \beta$	соединяем параллельно $M(\alpha), M(\beta)$
α^*	Зацикливаем автомат $M(\alpha)$ с переходом $\xrightarrow{\varepsilon}$

\Leftarrow Ведем индукцию по $|Q|$ в регулярном автомате

База: $Q = 1, 2$ — позже

Переход:

R	Автомат
$\alpha \cdot \beta$	$M(\alpha) \rightarrow_\epsilon M(\beta)$
$\alpha + \beta$	соединяем параллельно $M(\alpha), M(\beta)$
α^*	Заикливаем автомат $M(\alpha)$ с переходом \rightarrow_ϵ

□

Лемма 1.2 (О разрастании). Пусть L — автоматный язык. Тогда:

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

$$\forall k \in \mathbb{N} : xy^kz \in L$$

Доказательство. L — автоматный $\Rightarrow \exists M$ — НКА с однобуквенными переходами, т.ч. $L(M) = L$. Положим $P = |Q|$, где Q — множество состояний M . Рассмотрим $w = w_1w_2 \dots w_P \dots$

$$\langle q_0, w_1 \dots w_P u \rangle \vdash_1 \langle q_1, w_2 \dots w_P u \rangle \vdash \dots \vdash \langle q_P, u \rangle \vdash \langle q, \epsilon \rangle$$

Тогда $\exists k' < l' \leq P : q_{k'} = q_{l'}$ по принципу Дирихле (всего состояний P , а мы посетили больше). Положим $x = w_1 \dots w_{k'}, y = w_{k'+1} \dots w_{l'}, w = w_{l'+1} \dots w_P u$. Т.к. $k' \neq l' \Rightarrow |y| = l' - k' > 0, l' \leq P \Rightarrow |x| = l' \leq P$. Таким образом, $xy^kz \in L \forall k$. □

Замечание (Отрицание к лемме). Пусть L — некоторый язык и

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

$$\exists k \in \mathbb{N} : xy^kz \notin L$$

Тогда L — не автоматный

Пример. Рассмотрим $L = \{a^n b^n | n \in \mathbb{N}\}$. $\forall P \exists w = a^P b^P \in L$. Рассмотрим $w = xyz$. $|xy| < P \Rightarrow xy = a^k, |y| > 0 \Rightarrow y = a^l, l > 0$. Тогда $xy^2z = a^{k-l} a^{2l} b^P = a^{P+l} b^P$. Т.к. $l > 0$, то $a^{P+l} b^P \notin L \Rightarrow L$ — не автоматный.

Утверждение 1.4. Пусть L — язык, R — регулярное выражение, $L(R) \cap L$ — неавтоматный $\Rightarrow L$ — неавтоматный

Доказательство. Пусть L — автоматный, но тогда $L \cap L(R)$ — тоже. □

Пример. Рассмотрим $L = \{w | |w|_a = |w|_b\}$. Т.к. $L = \{a^k b^k | k \in \mathbb{N}\} \cap L(a^* b^*)$, то L — неавтоматный

1.5 Минимальный Полный Конечный Детерминированный Автомат?

Пусть есть два ПДКА: M_1, M_2 . Проверим, что $L(M_1) = L(M_2)$. Это равносильно тому, что $L(M_1) \Delta L(M_2) = \emptyset$. Однако есть более удобный способ это проверять.

Далее M — ПДКА, L — автоматный язык.

Определение 1.17. Будем говорить, что $u \sim_L v$, $u, v \in \Sigma^* \Leftrightarrow \forall w \in \Sigma^* uw \in L \Leftrightarrow vw \in L$.

Утверждение 1.5. \sim_L — отношение эквивалентности

Доказательство. **Рефлексивность** $u \sim_L u : uw \in L \Leftrightarrow uw \in L$

Симметричность $u \sim_L v \Leftrightarrow (uw \in L \Leftrightarrow vw \in L) \Leftrightarrow v \sim_L u$

Транзитивность $u \sim_L v, v \sim_L s \Leftrightarrow (uw \in L \Leftrightarrow vw \in L) \wedge (vw \in L \Leftrightarrow sw \in L) \Rightarrow u \sim_L s$. \square

Замечание. Σ^*/\sim_L — фактормножество Σ^* по отношению \sim_L . Класс эквивалентности слова u будем обозначать $[u]$.

Определение 1.18. Будем говорить, что $q_1 \sim_M q_2$, $q_1, q_2 \in Q \Leftrightarrow \forall w \in \Sigma^* \Delta(q_1, w) \subset F \Leftrightarrow \Delta(q_2, w) \subset F$.

Утверждение 1.6. \sim_M — отношение эквивалентности

Доказательство. **Рефлексивность** $q_1 \sim_M q_1 : \Delta(q_1, w) \subset F \Leftrightarrow \Delta(q_1, w) \subset F$

Симметричность $q_1 \sim_M q_2 \Leftrightarrow (\Delta(q_1, w) \subset F \Leftrightarrow \Delta(q_2, w) \subset F) \Leftrightarrow q_2 \sim_M q_1$

Транзитивность $q_1 \sim_M q_2, q_2 \sim_M q_3 \Leftrightarrow (\Delta(q_1, w) \subset F \Leftrightarrow \Delta(q_2, w) \subset F) \wedge (\Delta(q_2, w) \subset F \Leftrightarrow \Delta(q_3, w) \subset F) \Rightarrow q_1 \sim_M q_3$. \square

Лемма 1.3. Пусть $L_q = \{w | \Delta(q_0, w) = q\}$. Тогда каждый класс эквивалентности в Σ^*/\sim_L — объединение классов в L_q .

Доказательство. Пусть $u, v \in L_q$. Тогда $\Delta(q_0, u) = q, \Delta(q_0, v) = q$. Рассмотрим произвольное $w \in \Sigma^*$. $\Delta(q_0, uw) = q'$.

$$\langle q_0, uw \rangle \vdash \langle q_0, \varepsilon \rangle \Rightarrow \langle q_0, uw \rangle \vdash \langle q_1, w \rangle \vdash \langle q', \varepsilon \rangle \Rightarrow q' = \Delta(q, w)$$

Аналогично, $\Delta(q_0, vw) = \Delta(q, w)$. Но тогда $uw \in L \Leftrightarrow \Delta(q, w) \in F \Leftrightarrow w \in L$. \square

Следствие. $|Q| \geq |\Sigma^*/\sim_L|$

Доказательство. В каждом классе $[u]$ существует хотя бы один L_q . \square

Лемма 1.4. Пусть L — автоматный язык. Тогда \exists ПДКА M' , такой, что все состояния в M' попарно неэквивалентны.

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

$$M' = (Q/\sim_M, \Sigma, \Delta', [q_0], F')$$

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

$$F' = \{[q] | q \in F\}$$

1. Согласованность переходов:

$$q_1 \in [q] \Rightarrow \Delta(q_1, a) \in [\Delta(q, a)]$$

$$q_1 \in [q] \Rightarrow q_1 \sim q \Rightarrow \forall w \in \Sigma^* \Delta(q_1, w) \in F \Leftrightarrow \Delta(q, w) \in F$$

$$\forall w = au : \Delta(q_1, au) \in F \Leftrightarrow \Delta(q, au) \in F$$

$$\forall u : \Delta(\Delta(q_1, a), u) \in F \Leftrightarrow \Delta(\Delta(q, a), u) \in F$$

$$\Delta(q_1, a) \sim_M \Delta(q, a)$$

2. Согласованность завершающих состояний:

Пусть $q \in F$

$$q_1 \sim q \Rightarrow \forall w (\Delta(q_1, w) \in F \Leftrightarrow \Delta(q, w) \in F)$$

$$w = \varepsilon \Rightarrow ((\Delta(q_1, \varepsilon) = q_1 \in F) \Leftrightarrow (\Delta(q, \varepsilon) = q \in F)) \Rightarrow q_1 \in F$$

Покажем, что $\Delta([q_0], w) = [\Delta(q_0, w)]$ индукцией по $|w|$.

База: уже доказана

Переход: 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 состояния неэквивалентны. Пусть $[q_1] \sim_M [q_2]$.

$$\forall w : \Delta([q_1], w) \in F' \Leftrightarrow \Delta([q_2], w) \in F'$$

$$\forall w : [\Delta(q_1, w)] \in F' \Leftrightarrow [\Delta(q_2, w)] \in F'$$

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

$$q_1 \sim_M q_2$$

$$[q_1] = [q_2]$$

□

Определение 1.19. M — МПДКА, если M — ПДКА и $\nexists M'$, такой, что $L(M) = L(M'), |Q| > |Q'|$.

Теорема 1.5. M — МПДКА, такой, что $L(M) = L \Leftrightarrow$ все состояния недостижимы из стартового и никакие два неэквивалентны.

Доказательство. \Rightarrow Пусть M — МПДКА. Если в нем есть два эквивалентных состояния, то строим автомат на Q/\sim_M . Если же какое-то состояние недостижимо, то уберем его, таким образом всегда можно уменьшить число состояний

\Leftarrow Пусть в M нет эквивалентных состояний. Пусть $\Delta(q_0, w_1) \neq \Delta(q_0, w_2)$

$$\exists u : \Delta(\Delta(q_0, w_1), u) \notin F, \Delta(\Delta(q_0, w_2), u) \in F$$

$$\exists u : \Delta(q_0, w_2u) \notin F, \Delta(q_0, w_2u) \in F$$

$$\exists u : w_1u \notin L, w_2u \in L$$

$$w_1 \not\sim_L w_2$$

Тогда $|\Sigma^*/\sim_L| \geq |Q|$. Однако, $|Q| \geq |\Sigma^*/\sim_L| \Rightarrow M$ — минимальный.

□

Замечание. $|Q| \leq |\Sigma^*/\sim_L| \leq |Q| \Rightarrow |Q| = |\Sigma^*/\sim_L|$

2 Грамматики

Определение 2.1. $G = \langle N, \Sigma, P, S \rangle$, где

1. N — алфавит, называется множеством нетерминалов
2. Σ — алфавит, называется множеством терминалов
3. $P \subset (N \cup \Sigma)^+ \setminus \Sigma^* \times (N \cup \Sigma)^*$
4. S — стартовый нетерминал

Называется грамматикой

Определение 2.2. \vdash_G — наименьшее транзитивное отношение, такое, что

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

Следствие. Тогда $w \in L(G) \Leftrightarrow S \vdash w$

2.1 Иерархия Хомского

Определение 2.3 (Порождающие грамматики). Порождающие грамматики — класс вообще всех грамматик

Определение 2.4 (Контекстно-зависимые грамматики). Контекстно-зависимые грамматики — такие, в которых все правила имеют вид $\psi A \varphi \rightarrow \psi \alpha \varphi, \alpha \neq \varepsilon$

Определение 2.5 (Контекстно-свободные грамматики). Контекстно-свободные грамматики — такие, в которых все правила имеют вид $A \rightarrow \alpha$

Определение 2.6 (Праволинейные грамматики). Контекстно-свободные грамматики — такие, в которых все правила имеют вид $A \rightarrow wB, A \rightarrow w$.

Теорема 2.1. *Множество автоматных языков равно множеству языков, задаваемых праволинейными грамматиками.*

Доказательство. \subset Пусть $G = \langle N, \Sigma, P, S \rangle$ — наша грамматика, $M = \langle N \cup \{q_f\}, \Sigma, \Delta, S, \{q_f\} \rangle$, $\Delta = \{ \langle A, w \rangle \rightarrow B \mid A \rightarrow wB \} \cup \{ \langle A, w \rangle \rightarrow q_f \mid A \rightarrow w \}$. Хотим доказать два утверждения:

$$(a) \langle A, w \rangle \vdash_M \langle B, \varepsilon \rangle \Leftrightarrow A \vdash_G wB$$

$$(b) \langle A, w \rangle \vdash_M q_f, \varepsilon \Leftrightarrow A \vdash_G w$$

Докажем оба следствия вправо. Ведем индукцию по $|\vdash_M|$ — количеству шагов в выводе автомата.

База:

$$\langle A, w \rangle \vdash_0 \langle B, \varepsilon \rangle \Rightarrow A = B, w = \varepsilon \Rightarrow A \vdash_{G,0} B = A$$

$$\langle A, w \rangle \vdash_1 \langle q_f, \varepsilon \rangle \Rightarrow \langle A, w \rangle \rightarrow q_f \in \Delta \Rightarrow A \rightarrow w \in P \Rightarrow A \vdash w$$

Переход: Положим, для общности, α так, что $\langle A, w \rangle \vdash \langle \alpha, \varepsilon \rangle$ ($\alpha \in N \cup \{q_0\}$)

Положим $w = uv$

$$\langle A, uv \rangle \vdash_1 \langle C, v \rangle \vdash \langle \alpha, \varepsilon \rangle$$

$$\langle A, uv \rangle \rightarrow C \in \Delta \Rightarrow A \vdash_1 uC$$

Но, по предположению индукции:

$$\left[\begin{array}{l} C \vdash vB, B \in N(\alpha = B) \Rightarrow A \vdash_G uvB \\ C \vdash v, \alpha = q_f \Rightarrow A \vdash_G w \end{array} \right.$$

Докажем оба следствия влево. Ведем индукцию по $|\vdash_G|$ — длине вывода в грамматике

База:

$$A \vdash_0 wB \Rightarrow A = B, w = \varepsilon \Rightarrow \langle A, w \rangle \vdash_0 \langle B, \varepsilon \rangle$$

$$A \vdash_1 w \Rightarrow A \rightarrow w \in P \Rightarrow \langle A, w \rangle \vdash_1 \langle q_s, \varepsilon \rangle$$

Переход: Положим, для общности α так, что $A \vdash w\alpha, \alpha \in N \cup \{\varepsilon\}$

$$A \vdash_1 uC \vdash uv\alpha, w = uv$$

Тогда $\langle A, u \rangle \vdash \langle C, \varepsilon \rangle$, и по предположению индукции

$$\left[\begin{array}{l} \langle C, v \rangle \vdash \langle B, \varepsilon \rangle, B = \alpha \\ \langle C, v \rangle \vdash \langle q_f, \varepsilon \rangle, \alpha = \varepsilon \end{array} \right.$$

$$\text{Тогда } w \in L(G) \Leftrightarrow S \vdash w \Leftrightarrow \langle S, w \rangle \vdash \langle q_f, \varepsilon \rangle \Leftrightarrow w \in L(M)$$

⊃ Можно считать, что исходный автомат с одним завершающим состоянием, $M = \langle Q, \Sigma, \Delta, q_0, F = \{q_s\} \rangle$. Построим по нему грамматику $G = \langle Q, \Sigma, P, q_0 \rangle$, такую, что

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

Теперь, построим по грамматике G автомат M' . Тогда $M' = \langle Q, \Sigma, \Delta', q_0, \{q'_f\} \rangle$, где $\Delta' = \Delta \cup \{\langle q_f, \varepsilon \rangle \rightarrow q'_f\}$. Тогда $w \in L(M) \Leftrightarrow \langle q_0, w \rangle \vdash_M \langle q_f, \varepsilon \rangle \Leftrightarrow \langle q_0, w \rangle \vdash M' \langle q'_f, \varepsilon \rangle \vdash_1 \langle q'_f, \varepsilon \rangle \Leftrightarrow w \in L(M')$. Но тогда $L(M) = L(M')$.

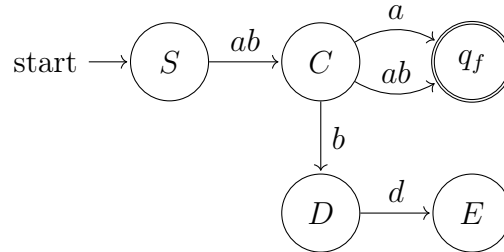
□

Пример. Пусть грамматика у нас следующая:

1. $S \rightarrow abC$
2. $C \rightarrow a$

3. $C \rightarrow ab$
4. $C \rightarrow bD$
5. $D \rightarrow dE$

Тогда для данной грамматики, автомат следующий:



2.2 Контекстно-свободные грамматики

Далее, для краткости, будем обозначать контекстно-свободные грамматики за КС

Определение 2.7. Дерево вывода — последовательность слов из $(N \cup \Sigma)^+$, каждый из элементов которой получается при переходе грамматики.

Определение 2.8. Дерево называется право(лево)сторонним, если для каждого слова, дерево его вывода единственно

Определение 2.9. Существенно неоднозначный язык — язык, которого не существует однозначной КС-грамматики, распознающей этот язык.

Задача. Доказать, что языки $\{a^k b^l c^m | k = l\}$, $\{a^k b^l c^m | m = l\}$ существенно неоднозначные.

2.3 Нормальная форма Хомского

Определение 2.10. КС грамматика G находится в нормальной форме Хомского, если все правила имеют вид:

1. $A \rightarrow BC, B \neq S, C \neq S$
2. $A \rightarrow a, a \in \Sigma$
3. $S \rightarrow \varepsilon$

Замечание.

1. $S \vdash \varepsilon \Rightarrow S \vdash_1 \varepsilon$
2. $A \vdash \varepsilon \Rightarrow A = S$

2.3.1 Удаление непорождающих нетерминалов

Определение 2.11. $D \in N$ называется порождающим, если $\exists w : D \vdash w$.

Утверждение 2.1. Пусть нам дана грамматика G_0 . Рассмотрим грамматику G_1 , такую, что $N_1 = N_0 \setminus \{D \mid D \text{ непорождающий}\}$. Тогда $L(G_0) = L(G_1)$.

Доказательство. Пусть $\exists w \in L(G_0) \setminus L(G_1)$. $S \vdash_{G_0} w \Rightarrow \exists D$ — непорождающий, такой, что $S \vdash uDv \vdash uv = w \Rightarrow D$ — порождающий, противоречие. Вложение в другую сторону очевидно. \square

Замечание. (Алгоритм для поиска непорождающих символов) Храним R_A — множество нетерминалов справа для каждого нетерминала A , создаем пустую очередь. Далее делаем следующие шаги, пока очередь не пуста:

1. Добавляем все нетерминалы с пустым R_X
2. Удаляем из всех R_Y нетерминал X

2.3.2 Удаление недостижимых нетерминалов

Определение 2.12. Нетерминал D называется достижимым, если $\exists \varphi, \psi \in (N \cup \Sigma)^*$, такие, что $S \vdash \varphi D \psi$.

Утверждение 2.2. Пусть нам дана грамматика G_1 . Рассмотрим грамматику G_2 , такую, что $N_2 = N_1 \setminus \{D \mid D \text{ — недостижимый}\}$. Тогда $L(G_1) = L(G_2)$.

Доказательство. Очевидно, что $L(G_1) \supset L(G_2)$. Пусть $w \in L(G_1) \setminus L(G_2)$. Тогда $\exists w \in \sigma^* : S \vdash \varphi D \psi \vdash w \Rightarrow D$ — достижимый, противоречие \square

Замечание. Алгоритм для поиска недостижимых символов — просто запускаем dfs.

Утверждение 2.3. Пусть нам дана грамматика G_1 без непорождающих символов. Рассмотрим грамматику G_2 , такую, что $N_2 = N_1 \setminus \{D \mid D \text{ — недостижимый}\}$. Тогда в ней нет непорождающих символов.

Доказательство. Пусть есть непорождающий символ B . Тогда есть вывод $B \vdash_{G_1} \varphi C \psi \vdash_{G_1} u$ \square

2.3.3 Удаление смешанных правил

Упражнение. Для каждого $a \in \Sigma$ заведем новый нетерминал X_a с единственным переходом $X_a \rightarrow a$. Заменим все вхождения буквы a в правила справа (слева a не может стоять) на X_a . Тогда полученная грамматика G_3 такова, что $L(G_3) = L(G_2)$

2.3.4 Удаление длинных цепочек

Упражнение. Если заменить $A \rightarrow A_1 A_2 \dots A_n$ в грамматике G_3 на $A \rightarrow A_1 A'_1, A'_1 \rightarrow A_2 \dots A_n \dots$, то $L(G_4) = L(G_3)$.

2.3.5 Удаление ε -переходов

Утверждение 2.4. Если в грамматике G_4 провести следующие операции:

$$\left[\begin{array}{l} A \rightarrow BC \\ C \rightarrow \varepsilon \end{array} \right] \Rightarrow \text{добавляем } A \rightarrow B$$

$$\left[\begin{array}{l} A \rightarrow BC \\ B \rightarrow \varepsilon \end{array} \right] \Rightarrow \text{добавляем } A \rightarrow C$$

А после этого удалить все переходы вида $A \rightarrow \varepsilon$, то множество слов полученной грамматики G_5 не изменится, кроме, быть может, пустого слова (которое удалится).

Доказательство.

1. $L(G_4) \setminus \{\varepsilon\} \subset L(G_5)$. Введем индукцию по длине слова $w \neq \varepsilon$

База: $|w| = 1$. Очевидно выводим за один шаг

Переход: пусть $A \vdash_{G_4,1} \alpha \vdash_{G_4} w$.

- i. $\alpha = B$. Тогда $A \rightarrow B \in P_4$
- ii. $\alpha = BC \Rightarrow B \vdash w_1, C \vdash w_2$.
 - A. $w_1, w_2 \neq \varepsilon \Rightarrow B \vdash_{G_5} w_1, C \vdash_{G_5} w_2 \Rightarrow A \vdash_{G_5} w$.
 - B. $w_1, w_2 \neq \varepsilon \Rightarrow B \vdash_{G_5} w_1, C \vdash_{G_5} w_2 \Rightarrow A \vdash_{G_5} w$.
 - C. $w_1 = \varepsilon, w_2 \neq \varepsilon \Rightarrow A \vdash_{G_5} C, C \vdash_{G_5} w_2 \Rightarrow A \vdash_{G_5} C \vdash_{G_6} w$.
 - D. $w_1 \neq \varepsilon, w_2 = \varepsilon$ — аналогично предыдущему

2. $L(G_4) \setminus \{\varepsilon\} \supset L(G_5)$. $A \vdash_{G_5} w \Rightarrow w \neq \varepsilon$. Докажем, что $A \vdash_{G_5} w \Rightarrow A \vdash_{G_4} w$. Введем индукцию по длине слова $| \vdash_{G_5} |$.

База: $| \vdash_{G_5} | = 1$. $A \rightarrow w \in P_5 \Rightarrow A \rightarrow w \in P_4$

Переход:

- i. $A \vdash_{G_5,1} BC \vdash_{G_5} w_1 w_2 = w \Rightarrow A \rightarrow BC \in P_5 \rightarrow A \rightarrow BC \in P_4$, но по предположению индукции $B \vdash_{G_4} w_1, C \vdash_{G_4} w_2$
- ii. $\alpha = BC \Rightarrow B \vdash w_1, C \vdash w_2$, тогда $A \vdash_{G_4} BC \vdash_{G_4} w$.
- iii. $A \vdash_{G_5,1} B \vdash_{G_5} w$.
 - A. $A \rightarrow B \in P_4$, по предположению индукции $B \vdash_{G_5} w \Rightarrow A \vdash_{G_4} B \vdash_{G_4} w$.
 - B. $A \vdash_{G_4} CB \in P_4, C \vdash_{G_4} \varepsilon \Rightarrow A \vdash_{G_4} CB \vdash_{G_4} B \vdash_{G_4} w$.
 - C. $A \vdash_{G_4} BC \in P_4, C \vdash_{G_4} \varepsilon \Rightarrow A \vdash_{G_4} BC \vdash_{G_4} B \vdash_{G_4} w$.

□

После этого добавим новый нетерминал S' , с переходом $S' \rightarrow S$, таким образом он не будет стоять нигде справа и множество слов грамматики не изменится. При этом, если $S \vdash_{G_4} \varepsilon$, то добавим еще правило $S \rightarrow \varepsilon$. Тогда новая грамматика G_6 такова, что $L(G_4) = L(G_6)$.

2.3.6 Удаление одиночных правил

Сделаем аналогично удалению ε -переходов в автомате.

Бинго! Мы получили нормальную форму Хомского

Следствие.

1. В нормальной форме Хомского дерево вывода бинарное
2. Слово длины $n > 0$ выводится за $2n - 1$ шаг

Перед нами стоит задача проверки $w \in L(G)$, и, если да, построения вывода.

2.4 Разбор КС-грамматики

2.4.1 Рекурсивный спуск

Рекурсивно спускаемся и таким образом перебираем все возможные выводы. К сожалению, может не закончиться (если есть ε -переходы).

1. Определить функцию обработки для каждого нетерминала из N .
2. Для каждого правила сгенерировать обработку:
3. Если символ слова совпадает с символом правила, то обработать символ.
4. Если следующий символ правила - нетерминал, то обрабатываем рекурсивно.
5. Если символ не совпадает - то выполняем **backtrack**.

2.4.2 Алгоритм Кока-Янгера-Касами

[См. презентацию](#)