Формальные языки и трансляции

Алексей Сорокин
Конспект подготовил Александр Васильев

ФИВТ МФТИ
1-ый семестр 2016-ого года

Введение

Основная цель математической логики, которую вы изучили на первом курсе, — формализация математических рассуждений. В определённый момент возникла идея сведениия любого объекта к последовательности некоторых символов. Не смотря на провал программы по сведению математики к чистому синтаксису, аппарат, позволяющий описывать некоторый язык с помощью строгого определения разрешённых последовательностей оказался полезен во многих других отраслях. Именно он и является предметом данного курса.

В качестве дополнительных учебных пособий можно рекомендуется использовать следующие книги:

- А. Е. Пентус, М. Р. Пентус «Теория формальных языков»
- А. Е. Пентус, М. Р. Пентус «Математическая теория формальных языков»
- А. В. Ахо «Теория синтаксического анализа, перевода и компиляции. Том 1»
- А. В. Ахо, Р. Сети, Д. Д. Ульман «Компиляторы. Принципы, технологии, инструментарий» (неофициально известна как «Dragon Book»)

Мы будем называть произвольное конечное множество Σ алфавитом, а под Σ^* понимать множество всех слов над этим алфавитом (конечных последовательностей, элементы которых лежат в Σ). Стоит заметить, что пустое слово ε, т. е. последовательность длины ноль, также лежит в Σ^*. Языком L (над алфавитом Σ) мы будем называть произвольное подмножество Σ^*.

Для двух слов $w_1 = a_1 \dots a_n$, $w_2 = b_1 \dots b_m$ определим их кокатенацию $w_1 \cdot w_2$ (точка будет часто опущена ради краткости) как дописывание одного слова к другому: $a_1 \dots a_n b_1 \dots b_m$. Это определение можно продолжить и для языков, приняв $L_1 \cdot L_2 = \{w_1 \cdot w_2 \mid w_i \in L_i\}$. Определив $L^0 = \{\varepsilon\}$ и $L^{i+1} = L^i \cdot L$ мы получим операцию возведения языка в натуральную степень, со свойствами, аналогичными возведению чисел в степень. Наконец, определим для языка операцию замыкания Клини $L^* = \bigcup_{k=0}^{\infty} L^k$. Лего видеть, что она, в некотором смысле, согласована с определением Σ^* как множества всех слов над Σ.

Оглавление

Глава 1

Конечные автоматы и регулярные выражения

1.1 Конечные автоматы

Определение 1.1. Недетерминированным конечным автомат M — это кортеж $\langle Q, \Sigma, \Delta, q_0, F\rangle$, где Q — множество состояний ($|Q| < \infty$), $\Sigma \subseteq Q \times \Sigma^* \times Q$, $|\Delta| < \infty$ ($\langle q_1, w\rangle \mapsto q_2$), $q_0 \in Q$ — стартовое состояние и $F \subseteq Q$ — завершающие состояния.

Определение 1.2. *Конфигурация* НКА — это пара $\langle q, w\rangle$, $q \in Q$, $w \in \Sigma^*$. $\vdash$ — это наименьшее рефлексивное транзитивное отношение, т. ч. $\forall (\langle q_1, w\rangle \mapsto q_2) \in \Delta\ \forall a \in \Sigma^*\ \langle q_1, wu\rangle \vdash \langle q_2, u\rangle$. *Язык, задаваемый автоматом* M — это $L(M) = \{w \mid \exists q \in F \colon \langle q_0, w\rangle \vdash \langle q, \varepsilon\rangle\}$.

Утверждение 1.1. *Можно считать, что* $|F| = 1$.

Доказательство. $F' = \{q'\}$, $Q' = Q \cup \{q'\}$, $\Delta = \Delta \cup \{\langle q, \varepsilon\rangle \mapsto q' \mid q \in F\}$. Тогда $M' = \langle Q', \Sigma, \Delta', q_o, F'\rangle$ — это НКА с одним завершающим состоянием и $L(M) = L(M')$. ■

Утверждение 1.2. *Можно считать, что* $\forall (\langle q_1, w\rangle \mapsto q_2) \in \Delta\ |w| \leqslant 1$.

Доказательство. Если $w = a_1 \dots a_n$, то добавим $n - 1$ промежуточное состояние. ■

Выражению $a(b^* c)^+ d^*$ подходит (диаграмка).
$\Delta(q, w) = \{q' \mid \langle q, w\rangle \vdash \langle q', w\rangle\}$.

Теорема 1.1. *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* $\forall (\langle q_1, w\rangle \mapsto q_2) \in \Delta'\ |w| = 1$.

Доказательство. Пользуясь утверждением, считаем что для любого перехода $|w| \leqslant 1$. Возьмём $F' = \{q' \mid \Delta(q', \varepsilon) \cap F \neq \varnothing\}$ и $\Delta' = \{\langle q_1, a\rangle \mapsto q_2 \mid \exists q_3 \in \Delta(q_1, \varepsilon) \colon \langle q_3, a\rangle \mapsto q_2\}$.

Покажем, что $L(M) = L(M')$. $w \in L(M) \Leftrightarrow \exists q \in F \colon \langle q_0, w\rangle \vdash \langle q, \varepsilon\rangle \Leftrightarrow \exists q \in F \colon \exists q_1, \ldots, q_n, q'_1, \ldots, q'_n \colon (< q_0, a_1 \ldots a_n > \vdash < q'_1, a_1 \ldots a_n\rangle \vdash < q_1, a_2 \ldots a_n > \vdash \cdots \vdash < q'_n, a_n > \vdash < q_n, \varepsilon > \vdash < q, \varepsilon >$. ∎

Утверждение 1.3. *Проблема принадлежности слова автоматному языку распознаётся за $O(|w||\Delta|)$ времени и $O(|Q|)$ памяти.*

Определение 1.3. $M = \langle Q, \Sigma, \Delta, q_0, F\rangle$ — детерминированный конечный автомат, если M имеет однобуквенные переходы и $\forall q_1 \in Q\ \forall a \in \Sigma\, |\{q_2 \mid \langle q_1, a\rangle \mapsto q_2\}| \leqslant 1$.

Для ДКА затраты памяти падают до $O(\log|Q|)$ (хранение номера состояния).

Теорема 1.2. *Для любого НКА $M = \langle Q, \Sigma, \Delta, q_0, F\rangle$ существует ДКА M' такой, что $L(M) = L(M')$.*

Доказательство. Возьмём $Q' = 2^Q$, $q'_0 = \{q_0\}$ и $F' = \{\bar{Q} \subseteq Q \mid \bar{Q} \cap F \neq \varnothing\}$. Наконец, возьмём $\Delta' = \{\langle Q_1, a\rangle \mapsto Q_2 | Q_2 = \Delta(Q_1, a) = \bigcup_{q_1 \in Q_1} \Delta(q_1, a)\}$. Тогда нам подходит $M' = \langle Q', \Sigma, \Delta', q'_0, F'\langle$.

Лемма 1.1. $\Delta'(Q_0, w) = \Delta(Q_0, w)$.

Доказательство. Введём индукцию по $|w|$. База очевидна: $\Delta(Q_0, \varepsilon) = Q_0 = \Delta'(Q_0, \varepsilon)$, т. к. у нас нет ε-переходов. Шаг индукции: пусть $w = w'a$. Тогда $\Delta(Q_0, w'a) = \Delta(\Delta(Q_0, w'), a) = \Delta(\Delta'(Q_0, w'), a) = \{q\ inQ \mid \exists q_1 \in \Delta'(Q_0, w) \colon \langle q_1, a\rangle \mapsto_M q\rangle\} = \Delta'(\Delta'(Q_0, w), a) = \Delta'(Q_0, w'a)$. ∎

$L(M) = \{w \mid \exists q \in F \colon \langle q_0, w\rangle \vdash \langle q, \varepsilon\rangle\} = \{w \mid \Delta(q_0, w) \cap F \neq \varnothing\} = \{w \mid \Delta'(q'_0, w) \cap F \neq \varnothing\} = \{w \mid \Delta'(q_0, w) \in F'\} = L(M')$. ∎

1.2 Свойства автоматных языков

Теорема 1.3. *Класс автоматных языков замкнут относительно*

1. *Конкатенации*
2. *Объединения*
3. *Пересечения*
4. *Итерации клини*
5. *Дополнения*
6. *Прямых и обратных гомоморфизмов*

Доказательство. $L = L(M_1)$, $L = L(M_2)$, $M_1 = \langle Q_1, \Sigma, \Delta_1, q_{01}, F_1\rangle$, $M_2 = \langle Q_2, \Sigma, \Delta_2, q_{02}, F_2\rangle$ — НКА с однобуквенными переходами. Тогда $L(M_1) \cap L(M_2) = \langle Q_1 \times Q_2, \Sigma, \Delta, (q_{01}, q_{02}), F_1 \times F_2\rangle$, $\Delta : \langle (q_1, q_2), a\rangle \to (q'_1, q'_2) \Leftrightarrow (\langle q_1, a\rangle \to q'_1) \in \Delta_1 \wedge (\langle q_2, a\rangle \to q'_2) \in \Delta_2$.

Утверждение 1.4. $\langle (q_1, q_2), w\rangle \vdash_M \langle (q_1', q_2'), \varepsilon\rangle \Leftrightarrow \langle q_1, w\rangle \vdash_{M_1} \langle q_1', \varepsilon\rangle \wedge \langle q_2, w\rangle \vdash_{M_2} \langle q_2', \varepsilon\rangle$

$\blacksquare$

Определение 1.4. ДКА называется *полным*, если $\forall q \in Q\, \forall a \in \Sigma\, \exists q'$: $(\langle q, a\rangle \to q') \in \Delta$.

Утверждение 1.5. *Для любого автоматного языка существует*

Утверждение 1.6. *Если M — ПДКА, то $\forall q\, \forall w\, |\Delta_M(q, w)| = 1$, $L = L(M)$, $M = \langle Q, \Sigma, \Delta, q_0, F\rangle$ — ПДКА, $\bar{L} = L(M')$, $M' = \langle Q, \Sigma, \Delta, q_0, Q \setminus F\rangle$.*

Определение 1.5. $\varphi : \Sigma^* \to \Gamma^*$ — гомоморфизм, если $\forall u, v \varphi(uv) = \varphi(u)\varphi(v)$.

Пусть $L = L(M)$, $M = \langle Q, \Sigma, \Delta, q_0, F\rangle$, тогда $\varphi(L) = L(\langle Q, \Gamma, \Delta', q_0, F\rangle)$, где $\Delta' = \{\langle q_1, \varphi(a)\rangle \to q_2 \mid \langle q_1, a\rangle \to q2 \in \Delta\}$.

$L = L(M)$, $M = \langle Q, \Gamma, \Delta, q_0, F\rangle$ — НКА с однобуквенными переходами. Тогда $\varphi^{-1}(L) = L(M')$, где $M' = \langle Q, \Sigma, \Delta', q_0, F\rangle$, $\Delta' = \{(\langle q_1, a\rangle \to q_2) \mid \langle q_{,1}, \varphi(a)\rangle \vdash_M \langle q_2, \varepsilon\rangle\}$.

$\langle q_1, w\rangle \vdash_{M'} \langle q_2, \varepsilon\rangle \Leftrightarrow \langle q_1, \varphi(w)\rangle \vdash_M \langle q_2, \varepsilon\rangle$. Индукция по $|w|$:

0. $w = \varepsilon \Leftrightarrow q_1 = q_2 \Leftrightarrow < q_1, \varepsilon > \vdash_M < q_2, \varepsilon >$ 1 $w = a \in \Sigma$ — по построению. 2. $w = ua$, то $< q_1, ua > \vdash_{M'} < q_2, \varepsilon > \Leftrightarrow \exists q_3$: $< q_1, u > \vdash_{M'} < q_3, \varepsilon >, (< q_3, a > \to q_2) \in \Delta' \Leftrightarrow \exists q_3$: $< q_1, \varphi(u) > \vdash_M \langle q_3, \varepsilon > \langle q_3, \varphi(a)\rangle \vdash_M < q_2 \varepsilon > \Leftrightarrow < q_1, \varphi(u)\varphi(a) > \vdash_M < q_2, \varepsilon >$.

1.3 Регулярные выражения

Определение 1.6. Регулярные выражения над языком Σ ($\mathrm{Reg}(\Sigma)$) — это наименьшее множество такое, что:

1. $\forall a \in \Sigma\; a \in \mathrm{Reg}(\Sigma)$
2. $0, 1 \in \mathrm{Reg}(\Sigma)$
3. $\forall \alpha, \beta \in \mathrm{Reg}(\Sigma)\; (\alpha \cdot \beta) \in \mathrm{Reg}(\Sigma), (\alpha + \beta) \in \mathrm{Reg}(\Sigma)$
4. $\forall \alpha \in \mathrm{Reg}(\Sigma)\; \alpha^* \in \mathrm{Reg}(\Sigma)$

Язык, задаваемый регулярным выражением — это:

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

Определение 1.7. Реглярный автомат — это автомат, в котором на рёбрах можно писать регулярные выражения.

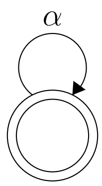
Теорема 1.4 (Клини)**.** *Классы автоматных и регулярных языков совпадают.*

Доказательство. Пусть L — регулярный язык. Тогда индукцией по построению регулярного выражения докажем существование автомата с помощью свойств автоматных языков.

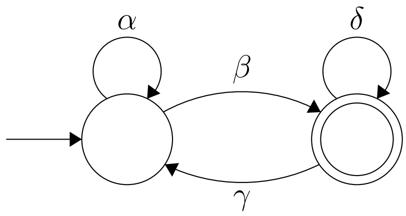
Утверждение 1.7. *Для каждого регулярного автомата можно построить регулярное выражение, задающее тот же язык.*

Доказательство. Можно считать, что в автомате одно завершающее состояние. Индукция по числу состояний:

Одно состояние.

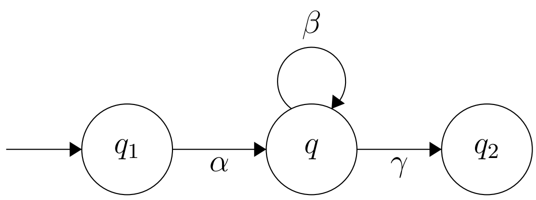


Подходит α^*.



Подходит $(\alpha^*\beta\delta^*\gamma)^*\alpha^*\beta\gamma^*$.

Пусть $|Q| > 2 \Rightarrow \exists q$: $q \notin \{q_0\} \cup F$. Тогда уберём q и для всех пар состояний q_1, q_2, таких что



Добавим переход $\langle q_1, \alpha\beta^*\gamma\rangle \to q_2$. ∎

∎

1.4 Минимизация конечных автоматов

Определение 1.8. ПДКА M — минимальный для языка $L(M)$, если для любого ПДКА M' такого, что $L(M') = L(M)$, $|Q_M| \leqslant |Q_{M'}|$.

Определение 1.9. $\sim_L$ — конгруэнция на Σ^* такая, что $u \sim_L v \Leftrightarrow \forall w \in \Sigma^* uw \in L \Leftrightarrow vw \in L$.

Утверждение 1.8. *Если $L = L(M)$, $M = \langle Q, \Sigma, \Delta, q_0, F\rangle$ — ПДКА, то* $|\Sigma^*/_{\sim_L}| \leqslant |Q|$.

Доказательство. Если $\Delta(q_0, u) = \Delta(q_0, v)$, то $u \sim_L v$. $uw \in L \Leftrightarrow \Delta(q_0, uw) \in F \Leftrightarrow \Delta(q_0, vw) \in F \Leftrightarrow vw \in L$. ■

Алгоритм минимизации полного ДКА:

1. На каждом шаге разбимваем состояния на классы эквивалентности.
2. На нулевом шаге 2 класса: F, $Q \setminus F$.
3. $q_1 \sim_{i+1} q_2 \Leftrightarrow (q_1 \sim_i q_2) \wedge (\forall a\, \Delta(q_1, a) \sim_i \Delta(q_2, a)$.

Алгорит завершается, когда классы перестают меняться.

Утверждение 1.9. *Если $\sim_i$ и $\sim_{i+1}$ совпадают, то $\sim_i = \sim_{i+1} = \sim_{i+2} = \ldots$.*

Утверждение 1.10. *Алгоритм завершается не более чем за $|Q| - 2$ шагов.*

Доказательство. На каждом шаге число классов эквивалентности либо увеличивается, либо классы не меняются. ■

Обозначим $\sim = \bigcap_{i=0}^{\infty} \sim_i$, $\tilde{M} = \langle Q/_\sim, \Sigma, \tilde{\Delta}, [q_0], \tilde{F}$, где

$$\tilde{\Delta} = \{\langle [q_1], a\rangle \to [q_2] \mid (\langle q_1, a\rangle \to q_2) \in \Delta\}$$
$$\tilde{F} = \{[q] \mid q \in F\}$$

Утверждение 1.11. $q_1 \sim_i q_2 \Leftrightarrow \forall u, |u| \leqslant i\, \Delta(q_1, w) \in F \Leftrightarrow \Delta(q_2, w) \in F$.

Доказательство. Индукция по i. База очевидна, шаг: $q_1 \sim_{i+1} q_2 \Leftrightarrow (q_1 \sim_i q_2 \wedge \forall a \in \Sigma\, \Delta(q_1, a) \sim_i \Delta(q_2, a) \Leftrightarrow q_1 \sim_i q_2 \wedge \forall a \in \Sigma\; \forall v \in \Sigma^*\; (|v| \leqslant i) \Delta(\Delta(q_1, a), v) \in F, \Delta(\Delta(q_2, a), v) \in F \Leftrightarrow (q_1 \sim_i q_2 \wedge \forall w\; (1 \leqslant |w| \leqslant i + 1 \Delta(q_1, w) \in F \Leftrightarrow \Delta(q_2, w) \in F$. ■

Утверждение 1.12. $q_1 \sim q_2 \Rightarrow \Delta(q_1, a) \sim \Delta(q_2, a)$.

Доказательство. $q_1 \sim q_2 \Leftrightarrow \forall w \Delta(q_1, w) \in F \Leftrightarrow \Delta(q_2, w) \in F \Rightarrow \forall a \forall u \Delta(\Delta(q_1, a), u) \in F \Leftrightarrow \Delta(\Delta(q_2, a), u) \in F \Rightarrow \Delta(q_1, a) \sim \Delta(q_2, a)$. ■

Утверждение 1.13. $\Delta_{\tilde{M}}([q_1], w) = [\Delta_M(q_1, w)]$

Доказательство. Индукция по $|w|$. ■

Лемма 1.2. $L(\tilde{M}) = L(M)$

Доказательство. $w \in L(M) \Leftrightarrow \exists q \in F \colon \Delta_M(q_0, w) = q \Leftrightarrow \exists q \in F \colon [\Delta_M([q_0], w) = \Delta_{\tilde{M}}([q_0], w) = [q] \Leftrightarrow \Delta_{\tilde{M}}([q_0], w) \in \tilde{F} \Leftrightarrow w \in L(\tilde{M})$ ■

Лемма 1.3. *$\tilde{M}$ — минимальный.*

Доказательство. Докажем, что если $\Delta([q_0], w_1) \neq \Delta([q_0], w_2)$, то $w_1 \not\sim_L w_2$.

$[\Delta_M(q_0, w_1)] = \Delta_{\tilde{M}}([q_0], w_1) = [q_1] \neq [q_2] = \Delta_{\tilde{M}}([q_0], w_2) = [\Delta_M(q_0, w_2)]$. $q_1 \not\sim q_2$. $\exists u$: $\Delta(q_0, w_1 u) = \Delta(q_1, u) \in F$, $\Delta(q_0, w_2 u) = \Delta(q_2, u) \notin F$ $w_1 u \in L$, $w_2 u \notin L \Rightarrow w_1 \not\sim_L w_2$.

Мы показали, каждому классу эквивалентности слов однозначно сопоставляется состояние, тогда ■

Определение 1.10. Изоморфизмом между автоматами $M_1 = \langle Q_1, \Sigma, \Delta_1, q_{01}, F_1 \rangle$ и $M_2 = \langle Q_2, \Sigma, \Delta_2, q_{02}, F_2 \rangle$ называется биекция $\varphi : Q_1 \to Q_2$ такая, что

1. $\varphi(q_{01}) = q_{02}$
2. $\varphi(q) \in F_2 \Leftrightarrow q \in F_1$
3. $\forall q\ \forall a\ \varphi(\Delta_{M_1}(q, a)) = \Delta_{M_2}(\varphi(q), a)$

Теорема 1.5. *Минимальный автомат единственнен с точностью до изоморфизма.*

Доказательство. Заметим, что если M — минимальный автомат для L, то существует биекция $\psi : Q_M \to \Sigma^*/_{\sim_L}$ такая, что $\psi(q) = \{w \mid \Delta_M(q_0, w) = q\}$. Легко видеть, что $psi(\Delta(q, a)) = [\psi(q)a]$, $\psi(q_0) = [\varepsilon]$, $\psi(q) \subseteq L \Leftrightarrow q \in F$. Тогда пусть ψ_1, ψ_2 — биекции для минимальных автоматов M_1, M_2; покажем, что $\varphi = \psi_2^{-1} \circ \psi_1$ — изоморфизм M_1 и M_2.

Во-первых, $\psi_1(q_0 1) = [\varepsilon] = \psi_2(q_{02})$. Во-втрых, $q_1 \in F \Leftrightarrow L \supseteq \psi_1(q_1) = \psi_2(q_2) \Leftrightarrow \varphi(q_1) \in F_2$. Наконец, пусть $\psi_1(q_1) = \psi_2(q_2)$, $\psi_1(\Delta_1(q_1, a)) = \psi_2(q_2')$; нужно показать, что $q_2' = \Delta_2(q_2, a)$. Действительно, $\psi_1(\Delta_1(q_1, a)) = [\psi_1(q_1)a] = [\psi_2(q_2)a] = \psi_2(\Delta_2(q_2, a))$. ■

Теорема 1.6 (Майхилл-Нероуд (Myhill-Nerode))**.** *L — автоматен $\Leftrightarrow |\Sigma^*/_{\sim_L}| < \infty$.*

Доказательство.

$\Rightarrow$ $|\Sigma^*/_{\sim_L}| \leqslant |Q_M| < \infty$, где M — ПДКА, распознающий L.

$\Leftarrow$ Рассмотрим автомат $M = \langle \Sigma^*/_{\sim_L}, \Sigma, \Delta, [\varepsilon], \{[u] \mid u \in L\} \rangle$, где $\Delta = \{\langle [w], a \rangle \to [wa] \mid w \in \Sigma^*, a \in \Sigma\}$. Корректность этого автомата доказывается аналогично корректности минимизированного автомата.

■

Утверждение 1.14. $\Delta([w], u) = [wu]$.

Доказательство. Индукция по длине $|u|$, $u = av$. $\Delta([w], av) = \Delta(\Delta([w], a), v) = \Delta([wa], v) = [wav]$. ■

$w \in L \Leftrightarrow [w] \in F \Leftrightarrow \Delta([\varepsilon], w) \in F \Leftrightarrow w \in L(M)$.

Утверждение 1.15. $L = \{a^n b^n\}$ *— неавтоматен.*

Доказательство. $\forall m \neq n\ a^m \nsim_L a^n$, ведь $a^m b^m \in L$, $a^n b^m \notin L$. Тогда $|\Sigma^*/_{\sim_L}| = \infty \Rightarrow L$ — неавтоматен. ■

Теорема 1.7 (Лемма о разрастании)**.** *Для любого автоматного* $L\ \exists p\colon \forall w(|w| \geqslant p \Rightarrow \exists x, y, z\colon (xyz = w, |y| > 0, |xy| \leqslant p \forall k\ xy^k z \in L))$.

Доказательство. Возьмём M — ДКА для L, обозначим $p = |Q_M|$. Тогда для слова с большей длинной, мы попадаем в некоторое состояние дважды по принципу Дирихле; возьмём x как слово до первого вхождения, y как слово между первым и вторым вхождением и z как оставшуюся часть. Тогда слова $xy^k z \in L(M) = L$. Если $|xy| > p$ то уже в этом слове есть цикл и мы можем повторить рассуждение. ■

Пример. Рассмотрим предыдущий язык и покажем верность отрицаний леммы о разрастании. Для произвольного p возьмём $a^p b^p \in L$. Тогда $y = a^m$, и $xy^2 z = a^{p+m} b^p \notin L$.

Глава 2

Контекстно-свободные грамматики

2.1 Порождающие грамматики

Определение 2.1. Порождающая гамматика — это кортеж $G = \langle N, \Sigma, P, S \rangle$, где

1. N — вспомогательные символы, $|N| < \infty$.
2. Σ — авфавит, $N \cap \Sigma = \varnothing$.
3. Множество правил $P \subseteq (N \cup \Sigma) + \times (N \cup \Sigma)^*$, $|P| < \infty$ (правило будем записывать $\alpha \to \beta$).
4. $S \in N$.

Определение 2.2. Отношение выводимости $\vdash \subseteq (N \cup \Sigma) + \times (N \cup \Sigma)^*$ — рефлексивное транзитивное замыкание отношения $\{\beta\alpha_1\gamma, \beta\alpha_2\gamma \mid \beta, \gamma \in (N \cup \Sigma)^*, (\alpha_1 \to \alpha_2) \in P\}$. Язык, задаваемый G $L(G) = \{w \in \Sigma^* \mid s \vdash w\}$.

Заметим, проблема принадлежности языка грамматике крайне сложна (эквивалентна чему-то там про порождение групп), поэтому мы вводим много классов грамматик.

Определение 2.3 (Иерархия Хомского, Noam Chomsky)**.** Определим классы грамматик по разрешённым в них правилах ($A, B \in N$, $\alpha, \beta_1, \beta_2 \in (N \cup \Sigma)^*$, $w \in \Sigma^*$).

1. Праволинейная грамматикка: $A \to wB$, $A \to w$.
2. Контекстно-свободные грамматики: $A \to \alpha$.
3. Контекстно-зависимые грамматики: $\beta_1 A \beta_2 \to \beta_1 \alpha \beta_2$.
4. Порождающие грамматики: любые правила.

Теорема 2.1. *Праволинейные грамматики эквивалентны конечным автоматам.*

Доказательство. Пусть дана праволинейная грамматика $G = \langle N, \Sigma, P, S\rangle$, тогда построим автомат $M_G = \langle N \cup \{q_f\}, \Sigma, \Delta, S, \{q_f\}\rangle$, где $\Delta = \{\langle q_1, w\rangle \to q_2 \mid (q_1 \to wq_2) \in P\} \cup \{\langle q_1, w\rangle \to q_f \mid (q_1 \to w) \in P\}$. Теперь покажем, что $L(M_G) = L(G)$.

Утверждение 2.1. *$\langle q_1, w\rangle \vdash_{M_G} \langle q_2, \varepsilon\rangle \Leftrightarrow q_1 \vdash_G wq_2$ $(q_1, q_2 \in N$.*

Доказательство. Индукция оп длине вывода / по числу проходимых рёбер:

$$\begin{cases} q_1 \to w_1 q_2 \\ q_2 \to w_2 q_3 \end{cases} \Rightarrow q_1 \to w_1 w_2 q_3$$

■

Утверждение 2.2. *$\langle q_1, w\rangle \vdash \rangle q_f, \varepsilon\rangle \Leftrightarrow q_1 \vdash_G w$.*

Доказательство. $\langle q_1, w\rangle \vdash \langle q_f, \varepsilon\rangle \Leftrightarrow \exists q_2, w_1, w_2 \colon (\langle q_1, w_1w_2\rangle \vdash \langle q_2, w_2\rangle \vdash \rangle q_f, \varepsilon\rangle \Leftrightarrow \exists q_2 \colon \exists w_1, w_2 \colon (q_1 \vdash_G q_2w_2, (q_2 \to w_2) \in P) \Leftrightarrow q_1 \vdash w_1w_2$. ■

$w \in L(M_G) \Leftrightarrow \langle q_0 = S, w\rangle \vdash < q_f, \varepsilon\rangle \Leftrightarrow S \vdash_G w \Leftrightarrow w \in L(G)$.

Пусть теперь нам дан автомат $M = \langle Q, \Sigma, \Delta, q_0, F\rangle$, тогда покажем, что ему эквивалентна праволинейная грамматика $G_M = \rangle Q, \Sigma, P, q_0 = S\rangle$, где $P = \{q_1 \to wq_2 \mid (\langle q_1, w\rangle \to q_2) \in \Delta\} \cup \{q \to \varepsilon \mid q \in F\}$.

Утверждение 2.3. *(1) по-прежнему верно.*

$w \in L(M) \Leftrightarrow \exists q \in F \colon (\langle q_0, w\rangle \vdash \rangle q,\varepsilon) \Leftrightarrow \exists q \colon \in F(q_0 \vdash_G wq) \Leftrightarrow q_0 \vdash_G w \Leftrightarrow w \in L(G)$. ■

Пример (контекстно-свободная грамматика)**.** Грамматика G с правилами $S \to a_i S b_i S$ $(i = 1, \ldots, n)$ и $S \to \varepsilon$ задаёт B_n — язык правильных скобочных последовательностей с n типами скобок.

Доказательство. B_n — язык слов, которые можно сократить до ε зачёркиванием рядом стоящих символов a_ib_i, т. е. $B_n = \{w \mid w =_S 1\}$, где $B = \langle a_i, b_i\rangle^* / \rangle a_ib_i = 1\rangle$.

$w \in B_n \Leftrightarrow w = a_iw_1b_iw_2$, где $w_1, w_2 \in B_n$ индукцией по длине $|w|$ легко показать, что $w \in B_n$ и $w \in L(G)$ равносильны. ■

Пример (регулярные выражения)**.** $S \to a$, $a \in \Sigma$, $S \to 0|1$, $S \to S + S|S \cdot S$, $S \to S^*$ и $S \to (S)$. Однако, для разбора слова из-за отсутствия скобок нам понадобится приоритет операций, как же закодировать его в грамматике? Добавим новые специальные символы: $S_* \to a|0|1|(S_+)|S_*|S_*^*$, $S_0 \to S_0 \cdot S_*|S_*$, $S_+ \to S_+S_0|S_+$. Тогда дерево разбора будет строиться однозначно,

2.2 Нормальная форма Хомского

Определение 2.4. Контекстно-свободная грамматика G находится в норфальной форме Хомского, если все правила грамматики имеют вид $A \to BC$, $A \to a$ и $S \to \varepsilon$ ($A \in N$, $B, C \in N \setminus \{S\}$, $a \in \Sigma$).

Замечание. Грамматика в НФ не обязательно однозначна. Более того, можно построить язык такой, что любая грамматика задающая его не будет обладать однозначным разбором.

Утверждение 2.4. *Слово длины n выводится в грамматике в НФ Хомского за $2n - 1$.*

Доказательство. Индукция по длине вывода. ∎

Теорема 2.2. *Любую КС-грамматику можно привести к НФ Хомского.*

Доказательство. 1. Уберём "смешанные правила": во всех правилах вида $A \to \alpha$, где $|\alpha| \geqslant 2$ заменим вхождения $a \in \Sigma$ на символы A_a, для которых добавим правило $A_a \to a$.

2. Убрерём длинные правила: пусть $B \to \beta$, $|\beta \geqslant 3|$. Тогда $\beta = C_1 \dots C_n$, добавим новые нетерминалы $A_2, \dots, A_{n_1}$ и правила $A_i \to C_i A_{i+1}$ для $i < n - 1$ и $A_{n-1} \to C_{n-1} C_n$. Наконец, заменим исходное правило на $B \to C_1 A_2$.

3. Уберём правила $A \to \varepsilon$. Найдём множество $N_\varepsilon = \{A \in N \mid A \vdash_G \varepsilon\}$ $((A \to \varepsilon) \in P \to A \in N_\varepsilon$ и для $B_i \in N_\varepsilon$ верно $(A \to B_1 \dots B_k) \in P \Rightarrow A \in N_\varepsilon)$. Для всех правил $A \to BC$ добавим $A \to B$, если $C \in N\varepsilon$ и $A \to C$, если $B \in N_\varepsilon$. Наконец, уберём все правла $A \to \varepsilon$ кроме $S \to \varepsilon$.

Утверждение 2.5. *Язык, задаваемый грамматикой, (почти) не изменился:*

$$\forall A \in N \ \forall u \in \Sigma^+ \ A \vdash_G \Leftrightarrow A \vdash_{G'} w.$$

Доказательство. Шаг 2: индукция по длине вывода. $A \vdash G_2 w \Rightarrow A \vdash_G W$. Пусть $A \vdash_{G_2} w$, есл и $A \to B_1 C_1 \to w_1 w_2$ или $A \to B \to w$ где $A \to B$ — "старое"правило, то $B_1 \vdash_{G_2} w_1 \Rightarrow B_1 \vdash_G w_1$ и $C_1 \vdash_{G_2} w_2 \Rightarrow C_1 \vdash_G w_2$, т.е.. $A \vdash_G w_1 w_2$. Пусть вывод начинался с "нового"правила: $A \to B \vdash_{G_2} w$, тогда $B \vdash_G w$ и либо $\exists C$: $(A \to BC) \in P$, $C \vdash_G \varepsilon$, либо $(AA \to CB) \in P$, $C \vdash_G \varepsilon \Rightarrow$

Покажем, что 3) не уменьшает множества выводимых непустых слов: пусть $A \vdash w \neq \varepsilon$. Тогда $A \to B \vdash_{G_2} w \Rightarrow$ индукция по длине вывода $A \to B \vdash_{G'} w$. Если $A \to B$ $(\vdash w_1 w_2)$, то $w_1 w_2 \neq \varepsilon \Rightarrow B \vdash_{G'} w_2$, $A \vdash_{G'} w_1 w_2$.

2. $w_2 = \varepsilon \Rightarrow C \in N_\varepsilon \Rightarrow A \to B \in P' \Rightarrow$ по п. и. $B \vdash_{G'} w_1 \Rightarrow A \vdash_{G'} w_1$.

Добавим новый ст. S' $S'' \to S$ и правило $S' \to \varepsilon$, если $S \in N_\varepsilon$. ∎

$G = \langle N, \Sigma, P, S \rangle$, $G' = \langle N, \Sigma, P', S \rangle$, где $P' = \{A \to \alpha \mid (A \to \alpha) \in P, \alpha \notin N\} \cup \{A \leftarrow \alpha \mid \exists B \in N\colon (A \vdash_G B), (B \to \alpha) \in P, \alpha \notin N\}$.

Заметим, что $L(G') \subseteq L(G)$, т. к. все новые правила можно заменить на последовательность новых. Индукция по длине вывода:

1. $(A \to \alpha) \in P$, т. к. $\alpha \notin N$, то $(A \to \alpha') \in P'$.

2. Вывод $A \vdash_G \alpha$ имеет вид $A \vdash A_1 \vdash \cdots \vdash A_n \vdash B \vdash \alpha$. Если $\beta \in \Sigma$, то $\beta = \alpha$ и $(A \to \beta) \in P' \Rightarrow A \vdash_{G'} \beta$. Путь $\beta = D$, $C \vdash_G \alpha_1$, $D \vdash_G \alpha_2$. По предположению индукции, $C \vdash_{G'} \alpha_1, D \vdash_{G'} \alpha_2$, Тогда $\beta = CD \vdash_G {}'\alpha_1\alpha_2 = \alpha$, $(A \to B) \in P' \Rightarrow A \vdash \alpha$.

$L(G) = \{w \in \Sigma^* \mid S \vdash_G w\} = \{w \in \Sigma^* \mid S \vdash_{G'} w\} = L(G')$. ∎

Пример. $S \to aSbS$, $S \to \varepsilon$. Удалим смешанные правила: $S \to ASBS$, $S \to \varepsilon$, $A \to a$, $B \to b$. Уберём длинные: $S \to AC$, $C \to SD$, $D \to BS$, $S \to \varepsilon$, $A \to b$, $B \to b$. 3. Найти ε-переход, $S \in N_\varepsilon$. Добавить правила для ПЧ од. переход $S \Rightarrow ac$, $C \to SD|D$, $D \to BS|B$, $A \to a$, $B \to b$, $S' \to S|\varepsilon$.

$C \to b$, $C \to SD$, $C \to BS$, $D \to BS$, $D \to b$, $A \to a$, $S \to AC$, $B \to b$, $S' \to \varepsilon|AC$.

Алгоритм проверки выводимости в КС-грамматике (Кок-Янгер-Кассама СУК)

Вход G — RC-грамматика в НФ Хомского, $w \in \Sigma^*$. Выход: 1, если $w \in L(G)$, 0 — иначе.

```
if w == \epsilon:
    return 1 if (s \to \epsilon) \in P else 0
int n = |w|
bool D[|N|][n + 1][n + 1] // D[A][i][j] = 1 \oTTo A \vdash w[i:j] = w_i \

for i = 0, \dots, n - 1:
    for (A \to a) \in P:
        if a == w[i] D[a][i][i + 1] = 1
for l = 2, \dots, n:
    for i = 0, \dots, n - l:
        for j = i + 1, \dots, i + l - 1:
            for (A \to BC) \in P:
                if D[B][i][j] and D[C][j][i + l]:
                    D[A][i][i+l] = 1

return D[S][0][n]
```

Корректность:

Доказательство. Докажем, что $A \vdash_G w[i:j] \Leftrightarrow D[A][i][j] = 1$. Индукция по $j-i$. $j = i+1$, $w[i:j] = a \in \Sigma$, $A \vdash w[i:j] \Leftrightarrow (A \to w[i:j]) \in P \Leftrightarrow D[A][i][j] = 1$. $j > i+1$, $A \vdash w[i:j] \Leftrightarrow \exists k$: $\exists (A \to BC) \in P$: $B \vdash w[i:k], C \vdash [w:j] \Leftrightarrow \exists k \in (i,j)$: $\exists (A \to BC) \in P$: $D[B][i][k] = 1 \wedge D[C][k][j] = 1 \Leftrightarrow D[A][i][j] = 1$. ∎

Теорема 2.3 (Лемма о разрастании для КС-грамматик, Bar-Hillel)**.** *Для любого КС-языка* L $\exists p$: $\forall w \in L\ |w| \geqslant p \Rightarrow \exists x, u, y, v, z, |uyv| \leqslant p, |uv| > 0, |xnyvz| = w$: $\forall k \in \mathbb{N}\ xu^k yv^k z \in L$.

Доказательство. $p = 2^{|N|}$, $L = L(G)$, $G = \langle N, \Sigma, P, S\rangle$ — в НФ Хомского.

Т. к. дерево вывода бинарное, то существует ветвь дерева длины хотя бы $|N| + 1$. Рассмотрим самую длинную такую ветвь, по принципу Дирихле в ней есть повторяющийся нетерминал. Рассмотрим самую нижнюю пару $S \vdash xAz \vdash xuAvz \vdash xuyvz$. Глубина $\leqslant |N| + 1 \Rightarrow |uyv| \leqslant N^{|N|+1-1} = p$. Поскольку G в НФ Хомского, то нельзя вывести $A \to A$. Больше тзтзрр шагов. $A \to uAv \Rightarrow |uv| > 0$.

$S \vdash xAz \vdash xuAvz \vdash xu^2Av^2z \vdash \dots \vdash xu^kAv^kz \vdash xu^kyv^kz$. ∎

2.3 Автоматы с магазинной памятью

Определение 2.5. МП-автомат — это кортеж $M = \langle Q, \Sigma, \Gamma, \Delta, q_0, F\rangle$., где Q — состояния, $|Q| < \infty$, Σ — алфавит, Γ — стековый алфавит, $\Delta \subseteq Q \times \Sigma^* \times (G^* \times G^*) \times Q$ — правила переходов. $q_0 \in Q$ — начальное состояние, $F \subseteq Q$ — завершающие.

Определение 2.6. Конфигурация МП-автомата — это $\langle q, u, \gamma\rangle$, где $q \in Q$, $u \in \Sigma^*$, $\gamma \in \Gamma^*$.

Отношение выводимости $\vdash$ — это наименьшее ревлексивное транзитивное отношение такое, что $\forall \langle q_1, u\rangle \to \langle q_2, \alpha, \beta\rangle\ (\forall v \in \Sigma^*\ \forall \eta \in \Gamma^*\ \langle q_1 uv, \eta\alpha\rangle \vdash \langle q_2, v, \eta\beta\rangle)$, Язык, задаваемый M $L(M) = \{w \in \Sigma^* \mid \exists q\colon \in F\langle q_0, w, \varepsilon\rangle \vdash \langle q, \varepsilon, \varepsilon\rangle\}$.

Утверждение 2.6. *Любой МП-автомат эквивалентен МП-автомату* $M = \langle Q, \Sigma, \Gamma, \Delta, q_0, F\rangle$, *где* $\forall (\langle q_1, u\rangle \to \langle q_2, \alpha, \beta\rangle) \in \Delta\ (|u| \leqslant 1, |\alpha| + |\beta| \leqslant 1$

Доказательство. ∎

Утверждение 2.7. *В условиях предыдущего утверждения можно считать, что* $|\alpha| + |\beta| = 1$.

Доказательство. ∎

Утверждение 2.8. *Для любого МП-автомата существует эквивалентный* $M = \langle Q, \Sigma, \Gamma, \Delta, q_0, F\rangle <$ *все переходы которого имеют вид* $\langle q_1, u, \alpha\rangle \to \langle q_2, \beta\rangle$, $|n| \leqslant 1$ *и* $|\alpha + \beta| = 1$.

Теорема 2.4. *Всякий язык, распознаваемый МП-автоматом —КС-язык.*

Доказательство. $L = L(M)$, $M = \langle Q, \Sigma, \Gamma, \Delta, q_0, F\rangle$ — МП-автомат, удовлетворяющий утверждению. Тогда ему соответствует грамматика $G = \langle N, \Sigma, P, S\rangle$. $N = (Q \times Q) \cup \{S\}$, $P = \{S \to (q_0, q) | q \in F\} \cup \{(q_1, q_2) \to u(q_3, q_4)v(q_5, q_2) \mid \exists A \in \Gamma\colon \langle q_1, u, \varepsilon \to \langle q_3, A\rangle \in \Delta \wedge \langle q_4, v, A\rangle \to \langle q_5, \varepsilon\rangle \in \Delta\} \cup \{(q_1, q_2) \to \varepsilon | q_1 \in Q\}$.

Заметим, что последовательность операций со стеком при распознавании слова образует ПСП (если оно принимается).

Докажем, что $(q_1,q_2)\vdash_G w \Leftrightarrow \langle q_1,w,\varepsilon\rangle \vdash_M \langle q_2,\varepsilon,\varepsilon\rangle$.

$\Rightarrow$ Докажем индукцией по длине вывода в G. $(q_1,q_2)\vdash w \Leftrightarrow \exists u,v\colon \exists q_3,q_4,q_5\colon (q_1,q_2)\to u(q_3,q_4)v(q_5,q_2)\in P$, $\exists w_1,w_2\colon (q_3,q_4)\vdash w_1,(q_5,q_2)\vdash w_2, w=uw_1vw_2 \Rightarrow \exists A\in\Gamma\colon (\langle q_1,u,\varepsilon\rangle\to\langle q_3,A\rangle)\in\Delta,(\langle q_4,v,A\rangle\to\langle q_5,\varepsilon\rangle)\in\Delta,\langle q_3,w_1,\varepsilon\rangle\vdash_M \langle q_4,\varepsilon,\varepsilon\rangle,\langle q_5,w_2,\varepsilon\rangle\vdash_M\langle q_2,\varepsilon,\varepsilon\rangle \Rightarrow \langle q_1,uw_1vw_2\varepsilon\rangle\vdash\langle q_3,w_1vw_2,A\rangle\vdash\langle_4,vw_2,A\rangle\vdash\langle q_5,w_2,\varepsilon\rangle\vdash\langle q_2,\varepsilon,\varepsilon\rangle$. База (1): в этом случае $w=\varepsilon$, $q_1=q_2$ (по построению). $\langle q_1,\varepsilon,\varepsilon\rangle\vdash_M\langle q_1,\varepsilon,\varepsilon\rangle$.

$\Leftarrow$ Шаг индукции (по числу тактов в автомате: $\langle q_1,w,\varepsilon\vdash_M\langle q_2,\varepsilon,\varepsilon\rangle$ длина больше одного $\Rightarrow$ переход можно расписать как $\langle q_1,uv_1vw_2,\varepsilon\rangle\vdash\langle q_3,w_1vw_2,A\rangle\vdash\langle q_4,vw_2,A\rangle\vdash\langle q_5,w_2,\varepsilon\rangle\vdash\langle q_2,\varepsilon,\varepsilon$. $\langle q_1,u,\varepsilon\rangle\to\langle q_3,A\rangle\in\Delta\langle q_4,u,\varepsilon\rangle\to\langle q_5,\varepsilon\rangle\in\Delta,\langle q_3,w_1,\varepsilon\rangle\vdash_M\langle q_4,\varepsilon,\varepsilon\rangle,\langle q_5,w_2,\varepsilon\rangle\vdash_M\langle q_2,\varepsilon,\varepsilon\rangle\Rightarrow(q_1,q_2)\vdash u(q_3,q_4)v(q_5,q_2)\in P,(q_3,q_4)\vdash_G w_1,(q_5,q_2)\vdash_G w_2\Rightarrow(q_1,q_2)\vdash_G uw_1vw_2=w$. База: автомат выполняет 0 тактов $\Rightarrow q_1=q_2$, $w=\varepsilon$, но $((q_1,q_2)\to\varepsilon)\in P$ (по построению), т. е. $(q_1,q_1)\vdash\varepsilon$.

Теперь покажем, что $L(M)=L(G)$: $w\in L(G)\Leftrightarrow S\vdash_G w\Leftrightarrow\exists q\in F\colon S\to(q_0,q)\vdash_G w\Leftrightarrow\exists q\in F\colon\langle q_0,w,\varepsilon\rangle\vdash\langle q,\varepsilon,\varepsilon\rangle\Leftrightarrow w\in L(M)$. ∎

Теперь мы хотим доказать, что всякий КС-язык распознаётся каким-то МП-автоматом. Посмотрим на пример: $S\to S+T|T$, $T\to T*U|U$, $U\to(S)|x|y|z$. Алгоритм похож на алгоритм стекового калькулятора: кладём операнды в стек, когда встречаем оператор — применяем его к аргументам на стеке и заменяем их на резльутат выполнения. Операндами у нас будут терминалы и нетерминалы, а операторами — замена правой части правила на левую, притом применять их мы можем в любое время, они не записаны в слове. В итоге, для выводимого слова в конце мы окажемся с пустым словом и стеком S.

x*(y+z)	ε
*(y+z)	x
*(y+z)	U
*(y+z)	T
(y+z)	T*
(y+z)	T*(

Теорема 2.5. *Всякий КС-язык распознаётся некоторым МП-автоматом.*

Доказательство. Пусть $L=L(G)$, $G=\rangle N,\Sigma,P,S\rangle$, тогда подходящим автоматом является $M=\langle\{q_0,q_1\}\Sigma,\Sigma\cup N,\Delta,q_0,\{q_1\}\rangle$, где в Δ лежат следующие правила: $\langle q_0,a,\varepsilon\rangle\to\langle q_0,a\rangle$, $a\in\Sigma$, $\langle q_0,\varepsilon,\alpha\rangle\to\langle q_0,A\rangle$, $(A\to\alpha)\in P$ и $\langle q_0,\varepsilon,S\rangle\to\langle q_1,\varepsilon\rangle$.

Теперь покажем, что $\alpha\vdash_G w\Leftrightarrow\langle q_0,w,\varepsilon\rangle\vdash_M\langle q_0,\varepsilon,\alpha\langle$ для любого $w\in\Sigma^*$ и $\alpha\in(N\cup\Sigma)^*$.

$\Rightarrow$ Индукция по длине вывода d и $|\alpha|$. База: $d=0\Rightarrow\alpha=w_1=a_1\dots a_n$, но $\langle q_0,a_1\dots a_n,\varepsilon\rangle\vdash\langle q_0,a_2\dots a_n,a_1\rangle\vdash\dots\vdash\langle q_0,\varepsilon,a_1\dots a_n\rangle$. Шаг: $\alpha=A\beta$, $A\vdash w_1$, $\beta\vdash w_2$. Тогда $<q_0,w_1,\varepsilon\rangle\vdash\langle q_0,\varepsilon,A\rangle$ и $q_0,w_2,\varepsilon\vdash\rangle q_0,\varepsilon,\beta$, $\beta\neq\varepsilon$. Но тогда

$\langle q_0, w_1w_2, \varepsilon\rangle \vdash \langle q_0, w_2, A\rangle \vdash \langle q_0, \varepsilon, AB\rangle \vdash \langle q_0, \varepsilon, \alpha\rangle$. Для данного рассуждения требуется $|\alpha| > 1$, если это не так, то вывод имеет вид $A \to \eta \vdash_G w$. По предположению индукции, $\langle q_0, w, \varepsilon\rangle \vdash_M \langle q_0, \varepsilon, \eta\rangle$, по построению $\langle q_0, \varepsilon, \eta\rangle \to \langle q_0, A\rangle \in \Delta$, тогда $\langle q_0, w, \varepsilon\rangle \vdash \langle q_0, \varepsilon, A\rangle$.

Индукция по числу шагов в автомате. База (1 шаг):

1. Шаг имеет вид $\langle q_0, a, \varepsilon\rangle \vdash \langle q_0, \varepsilon, a\rangle \Rightarrow \alpha = a = w$.
2. Шаг — свёртка: $q_1, \varepsilon, \beta\langle\vdash \langle q_0, \varepsilon, A\rangle$, $(A \to \beta) \in P$. Тогда $\alpha = A$, $w = \varepsilon$, $\beta = \varepsilon$, $(A \to \varepsilon) \in P \Rightarrow A \vdash_G \varepsilon$.

Шаг: посмотрим на послдений переход в автомате.

1. Перенос: $\langle q_0, va, \varepsilon\rangle \vdash \langle q_0, a, \beta\rangle \vdash \langle q_0, \varepsilon, \beta a\rangle \vdash \langle q_0, \varepsilon, \alpha\rangle$, по предположению индукции $\beta \vdash_G v$. $\Rightarrow \alpha = a = w$.
2. Шаг — свёртка: $q_1, \varepsilon, \beta\langle\vdash \langle q_0, \varepsilon, A\rangle$, $(A \to \beta) \in P$. Тогда $\alpha = A$, $w = \varepsilon$, $\beta = \varepsilon$, $(A \to \varepsilon) \in P \Rightarrow A \vdash_G \varepsilon$.

■

2.4 Нормальная форма Грейбах

Определение 2.7. Грамматика G находится в *нормальной форме Грейбах*, если все правила имеют следующий вид: $A \to aBC$, $A \to aB$, $A \to a$, $S \to \varepsilon$ ($a \in \Sigma$, $A \in N$, $B, C \in \Sigma\backslash\{S\}$).

Теорема 2.6. *Любая КС-грамматика приводится к нормальной форме Грейбах.*

Доказательство. Пусть $G = \langle N, \Sigma, P, S\rangle$ — грамматика в нормальной форме Хомского, порождающая L. Построим новую грамматику $G = \langle N', \Sigma, P', S\rangle$, где

$$\begin{aligned}
N' &= \{S\} \cup \{(A\backslash B) \mid A, B, \in N\} \\
P' &= \{(A\backslash A) \to \varepsilon \mid A \in N\} \cup \\
&\cup \{S \to a(A\backslash S) \mid (A \to a) \in P\} \cup \\
&\cup \{(A\backslash B) \to e(E\backslash D)(C\backslash D) \mid (E \to e) \in P, (C \to AD) \in P\}
\end{aligned}$$

Мы хотим, чтобы из $(A\backslash B)$ выводилось то, что выводилось из B кроме какого-нибудь левого поддерева с A:

$$(A\backslash B) \vdash w \Leftrightarrow B \vdash Aw,$$

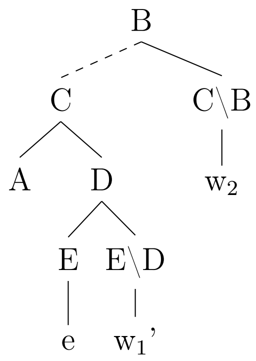
введём также аналогичное обозначение для языков:

$$L_1\backslash B = \{w \mid L_1 w = L_2\}.$$

Утверждение 2.9. $\forall A, B \in N\ \forall w \in \Sigma^*\ B \vdash_G Aw \Leftrightarrow A\backslash B \vdash_{G'} w$

Доказательство.

⇒ Индукция по длине вывода. База: вывод за 0 шагов, т. е. $B = Aw \Rightarrow A = B$, $w = \varepsilon$, и по построению $(A\backslash A) \to \varepsilon \in P'$. Пусть верно для более коротких выводов, b $B \vdash_G Aw$ более чем за 1 шаг. Т. к. A — нетерминал, и длина вывода ненулевая, $B \vdash_G Cw_2 \mapsto_G ADw_2 \vdash_G Aw_1w_2$, при этом $(C \to AD) \in P$, $w = w_1w_2$ и $|w_1| > 0$. Последнее, в свою очередь, означает, что $w_1 = ew_1'$, $E \to e$ и $D \vdash ew_1'$. Из вышеперечисленного получаем, что в P' есть правило $(B\backslash A) \to e(E\backslash D)(C\backslash B)$, а по предположению индукции $(E\backslash D) \vdash_{G'} w_1'$ и $(C\backslash B) \vdash_{G'} w_2$, а значит $(B\backslash A) \vdash_{G'} e(E\backslash D)(C\backslash B) \vdash_{G'} ew_1'w_2 = w_1w_2 = w$.



⇐ Индукция по длине вывода. База: 1 шаг, тогда было использовано правило $(A\backslash A) \to \varepsilon$, тогда $(A\backslash B) = (A\backslash A)$ и, по определению выводимости, $A \vdash A$. Пусть теперь $(A\backslash B) \vdash w$ больше, чем за один шаг. Значит, $(A\backslash B) \mapsto e(E\backslash D)(C\backslash B) \vdash ew_1w_2 = w$, где $(E\backslash D) \vdash w_1$, $(C\backslash B) \vdash w_2$. Тогда $(E \to e), (C \to AD) \in P$ и, по предположению индукции, $D \vdash_G Ew_1$, $B \vdash_G Cw_2$. Значит, $B \vdash_G Cw_2 \vdash_G ADw_2 \vdash_G Aew_1w_2 = Aw$. ■

Пусть $w = aw_1$ (значит, оно непусто). Тогда $w \in L(G) \Leftrightarrow S \vdash_G aw_1 \Leftrightarrow S \vdash_G Aw \wedge (A \to a) \in P \Leftrightarrow (A\backslash S) \vdash_{G'} w_1 \wedge (S \to a(A\backslash S)) \in P' \Leftrightarrow S \vdash_{G'} aw_1 = w \Leftrightarrow w \in L(G')$.

Теперь мы должны удалить правила $(A\backslash A) \to \varepsilon$, чтобы привести G' в НФ Грейбах. Если мы сделаем это так же, как и в доказательстве для НФ Хомского, то у нас появятся правила с правыми частями, в которых пропали некоторые *нетерминалы*, а тогда грамматика останется в НФ Грейбах!

Наконец, если нужно, добавим правило $S \to \varepsilon$. ■

Теорема 2.7. *Для любого КС-языка L существует МП-автомат, распознающий L такой, что все его переходы имеют вид*

$$\langle q_1, a, \alpha\rangle \to \langle q_2, \beta\rangle,$$

где $a \in \Sigma$.

Доказательство. Пусть $L = L(G)$, $G = \langle N, \Sigma, P, S\rangle$, и G находится в НФ Грейбах. Тогда требуемый автомат — это

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

где $F = \{q_0, q_1\}$, если $\varepsilon \in L(G)$, и $F = \{q_1\}$, иначе, а

$\Delta = \{\langle q_0, a, \varepsilon\rangle \to \langle q_1, B_n, \ldots, B_1\rangle \mid (S \to aB_1 \ldots B_n) \in P\} \cup \{\langle q_1, a, A\rangle \to \langle q_1, B_n, \ldots, B_1\rangle \mid (A \to aB_1 \ldots$

Утверждение 2.10. $\forall A \in N \smallsetminus \{S\}$ $\langle q_1, w, A\rangle \vdash_M \langle q_1, \varepsilon, \varepsilon\rangle \Leftrightarrow A \vdash_G w$.

Доказательство.

$\Rightarrow$ Индукция по длине вывода. База: 1 шаг, тогда $w = a$, $(A \to a) \in P$, и, по построению, $(\langle q_1, a, A\rangle \to \langle q_1, \varepsilon\rangle) \in \Delta$. Пусть теперь длина вывода больше единицы, тогда $A \vdash_G aB_1 \dots B_k \vdash_G aw_1 w_2 \dots w_k = w$, и $B_i \vdash_G w_i$. По построению, $(\langle q_1, a, A\rangle \to \langle q_1, B_k \dots B_1\rangle) \in \Delta$, а значит, $\langle q_1, aw_1 \dots w_k, A\rangle \mapsto_M \langle q_1, w_1 \dots w_k, B_k \dots B_1\rangle \vdash_M \langle q_1, w_2 \dots w_k, B_k \dots B_2\rangle \vdash_M \dots \vdash_M \langle q_1, w_k, B_k\rangle \vdash_M \langle q_1, \varepsilon, \varepsilon\rangle$ благодаря предположению индукции.

$\Leftarrow$ Индукция по числу переходов в автомате. База: 1 переход может соответствовать только правилу $A \to a$, и при этом $w = a$. Пусть теперь было переходов было больше одного, тогда $w = aw'$ и $\langle q_1, aw', A\rangle \mapsto_M \langle q_1, w', B_k \dots B_1\rangle$. Значит, $A \vdash aB_1 \dots B_k$, и $\exists w_1, \dots, w_k$: $B_i \vdash_G w_i land w' = w_1 \dots w_n$, и при этом $\langle q_1, w_1 \dots w_k, B_k \dots B_1\rangle \vdash_M \dots \vdash_M \langle q_1, w_k, B_k\rangle \vdash_M \langle q_1, \varepsilon, \varepsilon\rangle$, тогда $\langle q_1, w_i, B_i\rangle \vdash \langle q_1, \varepsilon, \varepsilon\rangle$, а по предположению индукции $B_i \vdash_G w_i$, т. е. $A \mapsto_G aB_1 \dots B_k \vdash_G aw_1 \dots {}^{`}w_k = w$. ∎

∎

2.5 Алгоритм перенос-свёртка

По $G = \rangle N, \Sigma, P, S\rangle$ строим МП-автомат. $M = \langle \{q_0, q_1\}, \Sigma, N \cup \Sigma, \Delta, q_0, q_1\rangle$, $\forall a \in \Sigma\, \langle q_0, a, \varepsilon\rangle \to \langle q_0, a\rangle$, $\forall (A \to \alpha) \in P\, \langle q_0, \varepsilon, \alpha\rangle \to \langle q_0, A\rangle$. Тогда $\langle q_0, w, \varepsilon\rangle \vdash \langle q_0, \varepsilon, A\rangle \Leftrightarrow A \vdash_G w$.

В разборе на стеке мы храним последовательность префиксов некторох правил, когда один из них заполняется — мы выполняем свёртку. Заглядывание на следующий символ входа можно реализовать с помощью состояний.

Определение 2.8. $\alpha \in (\Sigma \cup N)^*$ — активный префикс, если $\alpha = \alpha_1\alpha_2$, и $\exists v \in \Sigma^*$: $\exists D$: : $S \vdash_r \alpha_1 A v \to \alpha_1\alpha_2\beta_2 v$. (правосторонним, за один шаг).

Утверждение 2.11. *В процессе успешного выполнения алгоритма перенос-свёртка в стэке всегда лежат активные префиксы.*

Определение 2.9. $FIRST : (\Sigma \cup N)^* \to 2^{\Sigma \cup \{\$\}}$. $FIRST(\varepsilon) = \{\$\}$. $FIRST(\alpha) = \{a \mid \exists u \in \Sigma^*\colon (\alpha \vdash au)\} \cup \{\$ \mid \alpha \vdash \varepsilon\}$.

Определение 2.10. LR1-ситуация: $\langle A \to \alpha_2\beta_2, a\rangle$, где $(A \to \alpha_2\beta_2) \in P$, $a \in \Sigma \cup \{\$\}$.

Определение 2.11. Ситуация $\langle A \to \alpha_2\beta_2, b\rangle$ — допустимая для активного префикса $\alpha = \alpha_1\alpha_2$, если $\exists v$: $S \vdash_R \alpha_1 A v \to \alpha_1\alpha_2\beta_2 v$, $b \in FIRST(v)$.

Идея: давайте вмести с символами стека хранить допустимые ситуации для текущего активного префикса.

Следствие: при добавлении символа в стек или свёртке нужно быстро пересчитывать множество допустимых ситуаций.

Определение 2.12. Замыкание множества ситуаций I $CLOSURE(I)$ — наименьшее множество ситуаций J такое, что:

1. $I \subseteq J$
2. $\forall (A \to \alpha_2 \cdot B\gamma, b) \in J\ \forall (B \to \beta) \in P\ (B \to \circ\beta, c) \in J$ для $\forall c \in FIRST(\gamma b)$.

Утверждение 2.12. *Если I состоит только из допустимых ситуаций для активного префикса α, то в $CLOSURE(I)$ тоже будут только допустимые ситуации.*

Доказательство. Индукция по построению: пусть $(B \to \cdot\beta, c)$ получилась из ситуации $\langle A \to \alpha_2 {\cdot} B\gamma, b\rangle \in Adm(\alpha)$ (по предположению индукции). $\alpha_2' = \varepsilon$, нужно доказать, что $\exists v'$: $S \vdash \alpha B v'$, $c \in FIRST(v')$. $S \vdash \alpha_1 A v \to_\alpha \alpha_1\alpha_2 B\delta v \vdash \alpha_1\alpha_2 B v'' v$. Заметим, что $c \in FIRST(v') \Leftarrow c \in FIRST(\gamma v) \subseteq FIRST(\gamma b)$, ∎

2.6 Детерминированные автоматы с магазинной памятью

Определение 2.13. Два перехода $\langle q_1, x_1, \alpha_1\rangle \to \langle q_2\beta_2\rangle$ и $\langle q_1, x_2, \alpha_2\rangle \to \langle q_3\beta_3\rangle$ называются конфликтующими, если

1. $x_1 \sqsubseteq x_2$ или $x_2 \sqsubseteq x_1$
2. $alpha_1 \sqsupseteq alpha_2$ или $alpha_2 \sqsupseteq alpha_1$

Замечание. Если x — более длинное слово из x_1, x_2, а α — более длинное слово из α_1, α_2, то в конфигурации $\langle q_1, x, \alpha\rangle$ применимы оба конфликтующих перехода.

Определение 2.14. МП-автомат — детерминированный, если никакие два его различных перехода не конфликтуют друг с другом (кратко — ДМП-автомат).

Утверждение 2.13. *В любой конфигурации ДМП-автомата однозначно определена следующая операция.*

Следствие. *Если*

$$\begin{aligned}\langle q, w, \varepsilon\rangle &\vdash \langle q_1, \varepsilon, \alpha\rangle \\ &\vdash \langle q_2, \varepsilon, \beta\rangle,\end{aligned}$$

то либо $\langle q_1, \varepsilon, \alpha\rangle \vdash \langle q_2, \varepsilon, \beta\rangle$, либо наоборот.

Под $\$ \notin \Sigma$ мы будем понимать символ конца слова.

Определение 2.15. ДМП-автомат распознаёт язык L, если $w \in L \Leftrightarrow \exists q \in F$: $\langle q_0, w\$, \varepsilon\rangle \vdash \langle q, \varepsilon, \varepsilon\rangle$

Утверждение 2.14. *Если взять «обычное» определение распознавания ДМП-автомата, то класс распознаваемых языков меняется.*

Доказательство.

$$L = \{a^m b^n \mid m = n \vee n = 0\} = \{a^n b^n\} \cup a^*$$

Этот язык распознаётся следующим ДМП-автоматом:

Теперь покажем, что этот язык нельзя распознать без символа конца строки. Пусть это не так, тогда рассмотрим слова $a^m b^m$ и a^m. $\exists q_1 \in F$: $\langle q_0, a^m, \varepsilon\rangle \vdash \langle q_1, \varepsilon, \varepsilon\rangle$ и $\exists q_2 \in F$: $\langle q_0, a^m b^m, \varepsilon\rangle \vdash \langle q_1, b^m, \varepsilon\rangle \vdash \langle q_2, \varepsilon, \varepsilon\rangle$

По принципу Дирихле $\exists m \neq n$: $\exists q \in F$: $\langle q_0, a^m, \varepsilon\rangle \vdash \langle q, \varepsilon, \varepsilon\rangle \wedge \langle q_0, a^n, \varepsilon\rangle \vdash \langle q, \varepsilon, \varepsilon\rangle$, но $\langle q_0, a^n b^m, \varepsilon\rangle \vdash \langle q, b^m, \varepsilon\rangle$ и $\langle q_0, a^m b^m, \varepsilon\rangle \vdash \langle q, b^m, \varepsilon\rangle \vdash \langle q_1, \varepsilon, \varepsilon$ для $q_1 \in F$, а тогда и $\langle q, a^n b^m, \varepsilon\rangle \vdash \langle q_1, \varepsilon, \varepsilon$, т. е. автомат принимает $w = a^n b^m$, но $w \notin L$. ∎

Лемма 2.1. *Любой ДМП-автомат M эквивалентен некоторому ДМП-автомату M', имеющему только переходы вида $\langle q, x, \alpha\rangle \to \langle q'\beta\rangle$, где $|x| + |\alpha| \leqslant 1$.*

Доказательство. Назовём $\max\{|x| + |\alpha| - 1, 0\}$ избытком перехода, проведём индукцию по сумме избытков всех переходов.

Рассмотрим переход $\langle q, x, \alpha\rangle \to \langle q'\beta\rangle$ с положительным избытком. Возможны такие варианты:

1. Все переходы из q читают что-нибудь из входа $x = ax'$, $a \in \Sigma$, заменим все переходы вида $\langle q, ay, \alpha'\rangle \to \langle q_1\beta'\rangle$ на $\langle q, a, \varepsilon\rangle \to \langle \bar{q}, \varepsilon\rangle$ и $\langle \bar{q}, y, \alpha'\rangle \to \langle q_1\beta'\rangle$, где $\bar{q}$ — своё для каждого a.

   Новые конфликты могли появится только в $\bar{q}$. Пусть $y \sqsubseteq z$, $\alpha' \sqsupseteq \alpha''$, но тогда $ay \sqsubseteq az$ и конфликт существовал уже в исходном автомате.

2. Все переходы из q читают что-нибудь из стека. Тогда действуем аналогично, но выделяем первые символы из стека.

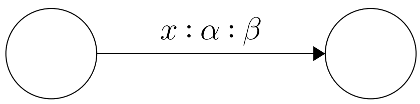
3. Существует переход $\langle q, \varepsilon, \varepsilon\rangle \to \langle q_1, \gamma\rangle$, но тогда он конфликтует с рассматриваемым переходом — противоречие.

4. Существуют переходы $\langle q, \varepsilon, \alpha_1\rangle \to \langle q_1, \beta_1\rangle$ и $\langle q, x_2, \varepsilon\rangle \to \langle q_1, \beta_1\rangle$, но тогда они конфликтуют друг с другом — противоречие.

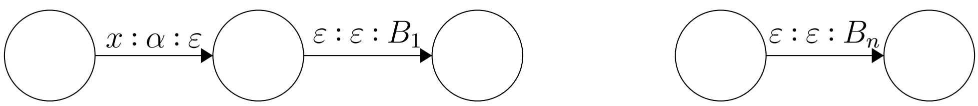
∎

Теорема 2.8. *В условиях леммы можно потребовать $|\alpha| + |\beta| = 1$ ($|x| + |\alpha| \leqslant 1$).*

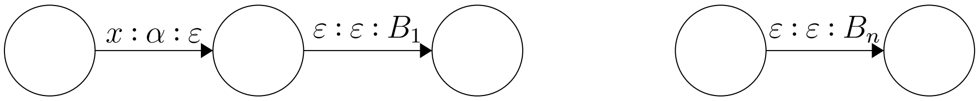
Доказательство. Пусть существует переход



где $\beta = B_1 B_2 \ldots B_n$, $n > 1$. Заменим его на цепочку



Если после этого есть переход такой, что $|\alpha| + |\beta| = 0$ по x, то заменим его на



∎

Лемма 2.2. *Всякий ДМП-автомат эквивалентен некоторому ДМП-автомату вида*

$$M = \langle Q_1 \sqcup Q_1, \Sigma, \Gamma, \Delta_{11} \sqcup \Delta_{12} \sqcup \Delta_{22}, q_0, F \rangle,$$

$q_0 \in Q_1$, $F \subseteq Q_2$ *и*

$$\Delta_{11} \subseteq (Q_1 \times \Sigma^* \times \Gamma^*) \times (Q_1 \times \Gamma^*)$$

$$\Delta_{12} \subseteq (Q_1 \times \{\$\} \times \Gamma^*) \times (Q_2 \times \Gamma^*)$$

$$\Delta_{22} \subseteq (Q_2 \times \{\varepsilon\} \times \Gamma^*) \times (Q_2 \times \Gamma^*)$$

Доказательство. Достаточно положить $Q_1 = Q \times \{1, 2\}$ и

$$\Delta_{11} = \{\langle (q, 1), x, \alpha \rangle \to \langle (q', 1), \beta \rangle \mid (\langle q, x, \alpha \rangle \to \langle q', \beta \rangle) \in \Delta, x \in \Sigma^*\},$$

$$\Delta_{12} = \{\langle (q, 1), \$, \alpha \rangle \to \langle (q', 2), \beta \rangle \mid (\langle q, \$, \alpha \rangle \to \langle q', \beta \rangle) \in \Delta\},$$

$$\Delta_{22} = \{\langle (q, 2), \varepsilon, \alpha \rangle \to \langle (q', 2), \beta \rangle \mid (\langle q, \varepsilon, \alpha \rangle \to \langle q', \beta \rangle) \in \Delta\}.$$

Корректность доказывается стандартным образом. ∎

Лемма 2.3. *В предыдущей лемме можно потребовать*

$$\Delta_{22} \subseteq (Q_2 \times \{\varepsilon\} \times \Gamma) \times (Q_2 \times \{\varepsilon\}) = \Delta_{22}^-$$

Доказательство. Идея: в Δ_{22} не нужны записи новых символов на стек, т. к. их уже всё равно придётся удалить. Заменим F на

$$F' = \{q \in Q_2 \mid \exists q' \in F \colon \langle q, \varepsilon, \varepsilon \rangle \vdash \langle q', \varepsilon, \varepsilon \rangle\},$$

а $Delta_{22}$ — на

$$\Delta'_{22} = (\Delta_{22} \cap \Delta_{22}^-) \cup \{\langle q, \varepsilon, A \rangle \to \langle q'', \varepsilon \rangle \mid \exists q' \colon \langle q, \varepsilon, \varepsilon \rangle \vdash \langle q', \varepsilon, \varepsilon \rangle \wedge \langle q', \varepsilon, A \rangle \to q'', a \rangle\}.$$

Полученный автомат обозначим M'.

Утверждение 2.15. *Если* $\alpha = A_1 \dots A_m$

$$\langle q_{21}, \varepsilon, A\alpha \rangle \vdash_M \langle q', \varepsilon, A \rangle \rangle \to \langle q_{22}, \varepsilon, \varepsilon \rangle \Leftrightarrow \langle q_{21}, \varepsilon, A\alpha \rangle \vdash_{M'} \langle q_{22}, \varepsilon, \varepsilon \rangle$$

Доказательство. Индукция по $|\alpha|$.

$\Rightarrow$ База: тривиально следует из построения. Шаг: $|\alpha| > 0$б $\alpha = A'\alpha'$. Рассмотрим удаление A' со стека, $\langle q_{21}, \varepsilon, A\alpha \rangle$

$\Leftarrow$ Всякий переход в M' дублируется последовательностью переходов в M.

■

Теперь покажем, что $L(M) = L(M')$. Рассмотрим принимающую поледовательность переходов в M:

$$\langle q_0, w\$, \varepsilon \rangle \vdash \langle q_1, \$, \alpha \rangle \mapsto \langle q_2, \varepsilon, \beta \rangle \vdash \langle q, \varepsilon, \varepsilon,$$

где $q_1 \in Q_1$, $q_2 \in Q_2$ и $q \in F$.

Если $\beta = \varepsilon$,

В обратную сторону: принимающая последоватлельность в M' имела вид

$$\langle q_0, w\$, \varepsilon \rangle \vdash \langle q, \$, \alpha \rangle \to \langle q', \varepsilon, \beta \rangle \vdash \langle q'' \in F', \varepsilon, \varepsilon \rangle$$

Тогда в M

$$\langle q_0, w\$, \varepsilon \rangle \vdash_M \langle q'', \varepsilon, \varepsilon \rangle,$$

и

$$\exists \tilde{q'} \in F\colon \langle q'', \varepsilon, \varepsilon \rangle \vdash_M \langle \tilde{q}, \varepsilon, \varepsilon \rangle,$$

а тогда

$$\langle q_0, w\$, \varepsilon \rangle \vdash_M \langle \tilde{q}, \varepsilon, \varepsilon \rangle \Rightarrow w \in L(M).$$

■

Теорема 2.9. *Любой язык L, распознающийся ДМП-автоматом можно задать однозначной КС-грамматикой.*

Доказательство. Рассмотим ДМП-автомат, удовлетворяющий условиям леммы 2, что $L(M) = L$. Применим модернизированную стандартную конструкцию строящую по МП-автомату КС-грамматику:

$$G = \langle \{S\} \cup \{A_{q_1 q_2} \mid q_1, q_2 \in Q\}, \Sigma, P, S \rangle, P = \{A_{q_1 q_2} \to x A_{q_3 q_4} y A_{q_5 q_2} \mid \exists B\colon \langle q_1, x, \varepsilon \rangle \to \langle q_1, B \rangle \land \langle q_4, y, B \rangle$$

Докажем инвариант:

Утверждение 2.16.

$$A_{q_1 q_2} \vdash_G w \Leftrightarrow \langle q_1, w, \varepsilon \rangle \vdash_M \langle q_2, \varepsilon, \varepsilon \rangle$$

Доказательство. Рж'ооооо ■

В силу детерменированности M б. о. о. $\pi_1 \subsetneq \pi_2$, т. е.

$$\langle q_1, \varepsilon, \varepsilon \rangle \vdash \langle q_2, \varepsilon, \varepsilon \rangle \Rightarrow q_1 = q_2 \Rightarrow \pi_1 = \pi_2,$$

противоречие. Значит, язык $L \cdot \{\$\}$ распознаётся однозначной грамматикой G. G' получается из G стиранием \$ во всех правилах. Заметим, что в $A_{q_1 q_2} \to \$ A_{q_3 q_4} y A_{q_5 q_2} \in P \Rightarrow q_1 \in Q_1, q_2 \in Q_2$ Заметим, что чтение доллара просиходит при переходе из q_1 в q_2, тогда слева — только q_1, а с права — q_2, эту позицию можно найти однозначно, а тогда всё ок.

■

Утверждение 2.17 (без доказательства). *Существует КС-язык, не распознаётся однозначной грамматикой.*

Пример.

$$\{a^n b^m c^r \mid n = m \vee m = r\}$$

Следствие. *ДМП-автоматы распознают не все КС-языки.*

Утверждение 2.18. *Детерминированные КС-языки замкнуты относительно дополнения.*

Утверждение 2.19. *Детерминированные КС-языки незамкнуты относительно пересечения и дополнения.*

Глава 3

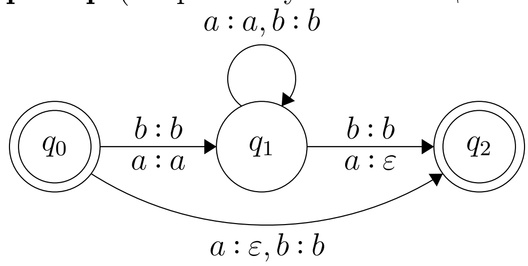
Конечные преобразователи

Определение 3.1. Конечный преобразователь — это кортеж

$$M = \langle Q, \Sigma, \Gamma, q_0, F \rangle,$$

где $q_0 \in Q$, $F \subseteq Q$, Γ — выходной алфавит и $\Delta \subseteq (Q \times \Sigma^*) \times (Q \times \Gamma^*)$ (переходы обозначаются как $\langle q_1, u \rangle \to \langle q_2, v \rangle$, где u — это то, что мы читаем, а v — что пишем).

Пример (стирание буквы в конце слова)**.**



Определение 3.2. $\rangle q, u, v \langle \in Q \times \Sigma^* \times \Gamma^*$ — конфигурация (u — осталось прочитать, v — уже написали).

Определение 3.3. $\vdash$ — наименьшее транзитивное отношоение такое, что

$$\forall \langle q_1, u \rangle \to \langle q_2, v \rangle \in \Delta \ \forall w \in \Sigma^* \ \forall x \in \Gamma^* \ \rangle q_1, uw, x \langle \vdash \rangle q_2, u, xv \langle$$

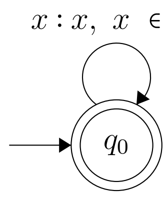
Определение 3.4. Преобразование, задаваемое преобразователем — это

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

Такое отношение, задаваемое некоторым преобразователем называется конечным (или рациональным) преобразованием.

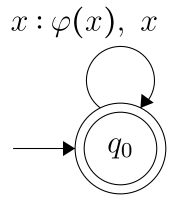
Примеры. Здесь под R понимается регулярный язык, под ψ_1, ψ_2 — конечные преобразования.

1. Тождественное преобразование (id):

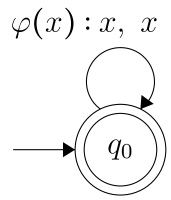


Можно также задать $\psi = id\big|_R$, $R \subset \Sigma^*$ — регулярный язык: пусть $R = L(M)$, где M — это автомат $\langle Q, \Sigma, \Delta, q_0, F\rangle$, тогда $\psi = L(T)$, где $T = \langle Q, \Sigma, \Sigma, \Delta_T, q_0, F\rangle$, $\Delta_T = \{\langle q_1, a\rangle \to \langle q_2, a\rangle | (\langle q_1, a\rangle \to q_2) \in \Delta\}$.

2. $\psi : \{\langle w, \varepsilon\rangle | w \in R\}$, где R — регулярный: заменить все рёбра $\langle q_1, a\rangle \to q_2$ на $\langle q_1, a\rangle \to \langle q_2, \varepsilon\rangle$. Симметрично строится $\psi : \{\langle \varepsilon, w\rangle | w \in R\}$,
3. Если ψ_1, ψ_2 — конечные преобразования, то $\psi_1 \cdot \psi_2 = \{\langle u_1 u_2, v_1, v_2\rangle \mid \langle u_1, v_1\rangle \in \psi_1 \wedge \langle u_2, v_2\rangle \in \psi_1\}$ — конечное преобразование: из конечных состояний ψ_1 проводим ε-переходы в начальное состояние ψ_2, конечными делаем конечные ψ_2.
4. $\psi : w \to \{wu \mid u \in R\}$, где R — регулярный: $\psi = \mathrm{id} \cdot \{langle\varepsilon, u\rangle \mid u \in R$.
5. Если ψ_1 и ψ_2 — конечные преобразования, то $\psi_1 \cup \psi_2$ — тоже. Добавляем фиктивное начальное, из которого проводим ε-переходы в начальные у исходных преобразователей, множество конечных — объединение.
6. $\psi : w \to \{w\} \cup R$ — конечное преобразование: $\psi = \mathrm{id} \cup \{\langle u, w\rangle \mid w \in R\}$.
7. Если $\varphi : \Sigma^* \to \Gamma^*$ — гомоморфизм языков ($\varphi(uv) = \varphi(u)\varphi(v)$), то φ задаётся преобразователем (т. к. гомоморфизм определяется своим действием на буквы)



Аналогично можно задать обратное преобразование:



Лемма 3.1. *Всякое конечное преобразование можно задать преобразователем $M = \langle Q, \Sigma, \Gamma, \Delta, q_0, F\rangle$, где $\forall \langle q_1, u\rangle \to \langle q_2, v\rangle \in \Delta\ |u| + |v| = 1$.*

Доказательство. Переходы вида $\langle q_1, a_1 \ldots a_m\rangle \to \langle q_2, b_1 \ldots b_n\rangle$, где $m + n > 1$ заменяются на цепочку со вспомогательными состояниями.

Осталось удалить ε-переходы. Обозначим $D(q) \coloneqq \{q' \in Q \mid \langle q, \varepsilon, \varepsilon\rangle \vdash \langle q', \varepsilon, \varepsilon\langle$, пусть в $M = \langle Q, \Sigma, \Gamma, \Delta, q_0, F\rangle$ уже провели предыдущий шаг, тогда нам подойдёт $M' = \langle Q, \Sigma, \Gamma, \Delta', q_0, F'\rangle$, где

$$F' = \{q \mid D(q) \cap F \neq \varnothing\}$$
$$\Delta' = \{\langle q_1, u\rangle \to \langle q_2, v\rangle \mid |u| + |v| = 1 \wedge \exists q \in D(q_1) \colon \langle q, u\rangle \to \langle q_2, v\rangle\}$$

Эквивалентность доказывается аналогично случаю конечных автоматов. ∎

Теорема 3.1. *1. Область определения конечного преобразования — регулярный язык*

2. Область значений конечного преобразования — регулярный язык

3. Ограничение любого конечного преобразования на регулярный язык задаётся конечным преобразователем.

Доказательство. 1. Возьмём конечный преобразователь, оставим на рёбрах только входы.

2. Аналогично, но оставим выходы.

3. Нужно заставить автомат читать только слова из языка $R = L(M)$, где $M = \langle Q, \Sigma, \Delta, q_0, F\rangle$ — ДКА, и $\psi = L(T)$, где $T = \langle Q_T, \Sigma, \Gamma, \Delta_T, q_{0T}, F_T\rangle$, где у любого перехода $|u| + |v| = 1$. Тогда требуемое $\psi_R = L(T')$, $T = \langle Q \times Q_T, \Sigma, \Gamma, \Delta', (q_0, q_{0T}), F \times F_T\rangle$, где

$$\Delta' = \{\langle (q, q_1), \varepsilon\rangle \to \langle (q, q_2), b\rangle \mid \langle q_1, \varepsilon\rangle \to \langle q_2, b\rangle \in \Delta_T\} \cup \{\langle (q, q_1), a\rangle \to \langle (q', q_2), \varepsilon\rangle \mid \langle q_1, a\rangle \to \langle q_2, \varepsilon\rangle \in$$

Утверждение 3.1.

$$\langle (q, q_1), u, \varepsilon\rangle \vdash_{T'} \langle (q', q_2), \varepsilon, v\rangle \Leftrightarrow \langle q_1, u, \varepsilon\rangle \vdash_T \langle q_2, \varepsilon, v\rangle \langle q, u\rangle \vdash_M \langle q', \varepsilon\rangle$$

Доказательство. Индукция по $|u| + |v|$. База: $|u| + |v| = 0$, значит $q = q'$, $q_1 = q_2$. Шаг: рассмотрим последний переход.

$$\langle (q, q_1), u'a, \varepsilon\rangle \vdash_{T'} \langle (q'', q_3), a, v\rangle \to \langle (q', q_2), \varepsilon, v\rangle$$

По предположению индукции $\langle q, u'\rangle_M \vdash \langle q'', \varepsilon\rangle$, также $(\langle q'', a\rangle \to q') \in \Delta$, значит $\langle q, u'a\rangle \vdash_M \langle q', \varepsilon\rangle$. Далее, $\rangle q_1, u', \varepsilon\langle \vdash_T \rangle q_3, \varepsilon, v\langle$ и $\langle q_3, a\rangle \to \langle q_2, \varepsilon\rangle \in \Delta_T$, а значит, $\langle q_1, u'a, \varepsilon\rangle \vdash_T \langle q_2, \varepsilon, v\rangle$ Аналогично рассматриваем случай, где последний шаг — запись.

В обратную сторону — также индукция по $|u| + |v|$. База — тривиальна, шаг: рассмотрим последний переход в T. Если это запись, то $\rangle q_1, u, \varepsilon\langle \vdash_T \rangle q_3, \varepsilon, v'\langle \to \rangle q_2, \varepsilon, v'a\langle$, $\langle q, u\rangle \vdash_M \langle q', \varepsilon\rangle$, значит

$$\langle (q, q_1), u, \varepsilon\rangle \vdash_{T'} \langle (q', q_3), \varepsilon, v'\rangle,$$

по построению $\langle (q', q_3), \varepsilon\rangle \to \langle (q', q_2), a\rangle \in \Delta'$, а значит,

$$\langle (q, q_1), u, \varepsilon\rangle \vdash_{T'} \langle (q', q_2), \varepsilon, v'a\rangle.$$

Если это чтение, то $\rangle q_1, u'a, \varepsilon\langle \vdash_T\rangle q_3, a, v\langle \to\rangle q_2, \varepsilon, v\langle$, и $\langle q, u'a\rangle \vdash_M \langle q'', a\rangle \to \langle q', \varepsilon\rangle$. Применим предположение индукции.

$\langle (q, q_1), u\rangle \vdash_{T'} \langle (q', q_2), v\rangle$ ■

■

Теорема 3.2. *Если φ — конечное преобразование, то φ^{-1} — тоже.*

Доказательство. $\varphi = L(M)$, $M = \langle Q, \Gamma, \Sigma, \Delta, q_0, F\rangle$, тогда $\varphi^{-1} = L(M')$ при $M = \langle Q, \Gamma, \Sigma, \Delta', q_0, F\rangle$, где

$$\Delta' = \{\langle q_1, u\rangle \to \langle q_2, v\rangle \mid \langle q_1, v\rangle \to \langle q_2, u\rangle \in \Delta\}$$

■

Теорема 3.3. *Конечные преобразования замкнуты относительно композиции.*

Доказательство. Пусть $\varphi_1 = L(M_1)$, $M_1 = \langle Q_1, H, \Gamma, \Delta_1, q_{01}, F_1\rangle$, $\varphi_2 = L(M_2)$, $M_2 = \langle Q_2, \Sigma, H, \Delta_2, q_{02}, F_2\rangle$; можно считать, что в переходах $|u| + |v| = 1$.

Тогда $\varphi_1 \circ \varphi_2 = L(M)$, где $M = \langle Q_2 \times Q_1, \Sigma, \Gamma, \Delta, (q_{02}, q_{01}), F_2 \times F_1\rangle$, где

$$\begin{aligned}\Delta =&\{\langle (q_2, q_1), a\rangle \to \langle (q_2', q_1), \varepsilon\rangle \mid \langle q_2, a\rangle \to \langle q_2', \varepsilon \in \Delta_2\rangle\}\cup\\ &\cup \{\langle (q_2, q_1), \varepsilon\rangle \to \langle (q_2, q_1'), b\rangle \mid \langle q_1, \varepsilon\rangle \to \langle q_1', b \in \Delta_1\rangle\}\cup\\ &\cup \{\langle (q_2, q_1), \varepsilon\rangle \to \langle (q_2', q_1'), \varepsilon\rangle \mid \exists a\colon \langle q_2, \varepsilon\rangle \to \langle q_2', a\rangle \in \Delta_2, \langle q_1, a\rangle \to \langle q_1', \varepsilon\rangle \in \Delta_1\}\end{aligned}$$

Утверждение 3.2. $\langle (q_2, q_1), u, \varepsilon\rangle \vdash_M \langle (q_2', q_1'), \varepsilon, v\rangle \Leftrightarrow \exists w\colon (\langle q_2, u, \varepsilon\rangle \vdash_M \langle q_2', \varepsilon, w\rangle \wedge \langle q_1, w, \varepsilon\rangle \vdash_M \langle q_1', \varepsilon, v\rangle)$.

Доказательство. $\Rightarrow$ Индукция по числу переходов в M. База: 0 переходов — $u = v = \varepsilon$, $q_2 = q_2'$, $q_1 = q_1'$, выберем $w = \varepsilon$. В шаге индукции будем разбирать последний переход.

Пусть последний переход — чтение: $\langle (q_2, q_1), u'a, \varepsilon\rangle \vdash \langle (q_2'', q_1'), a, v\rangle \to \langle (q_2', q_1'), \varepsilon, v\rangle$, по предположению индукции $\exists w'$: $(\langle q_2, u, \varepsilon\rangle \vdash_{M_2} \langle q_2'', \varepsilon, w'\rangle \wedge \langle q_1, w', \varepsilon\rangle \vdash_{M_1} \langle q_1', \varepsilon, v\rangle)$, а также $\langle q_2'', a\rangle \to \langle q_2, \varepsilon\rangle \in \Delta_2$, тогда $\langle q_2, u, \varepsilon\rangle \vdash_{M_2} \langle q_2', \varepsilon, w'a\rangle \wedge \langle q_1, w', \varepsilon\rangle \vdash_{M_1} \langle q_1', \varepsilon, v\rangle)$, значит, можно взять $w = w'$. Если последний переход —- запись, то всё симметрично.

Пусть последний переход — $(\varepsilon, \varepsilon)$. Тогда $\langle (q_2, q_1), u, \varepsilon\rangle \vdash \langle (q_2'', q_1''), , v\rangle \to \langle (q_2', q_1'), \varepsilon, v\rangle$, тогда по предположению индукции для некоторого w' $\langle q_2, u, \varepsilon\rangle \vdash_{M_2} \langle q_2'', \varepsilon, w'\rangle$, $\langle q_1, w', \varepsilon\rangle \vdash_{M_1} \langle q_1'', \varepsilon, v\rangle)$. По построению Δ $\langle q_2'', \varepsilon\rangle \to \langle q_2', a\rangle \in \Delta_2$ и $\langle q_1'', a\rangle \to \langle q_1', \varepsilon\rangle \in \Delta_1$. Выходит, $\langle q_2, u, \varepsilon\rangle \vdash_{M_2} \langle q_2', \varepsilon, w'a\rangle \wedge \langle q_1, w'a, \varepsilon\rangle \vdash_{M_1} \langle q_1', \varepsilon, v\rangle)$, тогда можно взять $w = w'a$.

$\Leftarrow$ Индукция по суммарному числу переходов, база — тривиальна.

Последний переход в M_2 — запись, в M_1 — чтение. $w = w'a$, существуют q_0'' и q_1'' такие, что $\langle q_2, u, \varepsilon\rangle \vdash_{M_2} \langle q_2'', \varepsilon, w'\rangle \to \langle q_2', \varepsilon, w'a\rangle \wedge \langle q_1, w'a, \varepsilon\rangle \vdash_{M_1} \langle q_1'', a, v\rangle) \Rightarrow \langle q_1', \varepsilon, v\rangle)$, тогда по предположению индукции $\langle (q_2, q_1), u, \varepsilon\rangle \vdash \langle (q_2'', q_1'', \varepsilon, v\rangle$ и по определению $\langle (q_2'', q_1''), \varepsilon\rangle \to \langle (q_2', q_1'), \varepsilon\rangle \in \Delta$, тогда $\langle (q_2, q_1), u, \varepsilon\rangle \vdash \langle (q_2', q_1'), \varepsilon, v\rangle$

Поседний переход в M_1 — запись. $v = v'b$, $\exists q_1''$: $\langle q_1, w, \varepsilon\rangle \vdash_{M_1} \langle q_1'', , v'\rangle) \Rightarrow \langle q_1', \varepsilon, v'b\rangle)$, и $\langle q_2, u, \varepsilon\rangle \vdash_{M_2} \langle q_2', \varepsilon, w\rangle$. Тогда по предположению индукции $\langle (q_2, q_1), u, \varepsilon\rangle \vdash \langle (q_2', q_1'), \varepsilon, v'\rangle$ и по определению Δ $\langle (q_2'', q_1'), \varepsilon\rangle \to \langle (q_2', q_1'), b\rangle \in \Delta$, тогда $\langle (q_2, q_1), u, \varepsilon\rangle \vdash \langle (q_2', q_1'), \varepsilon, v'b\rangle$. Случай, если в M_2 последний переход — чтение аналогичен.

∎

$\langle u, v\rangle \in L(M) \Leftrightarrow \exists (q_2, q_1) \in F_2 \times F_1$: $\langle (q_{02}, q_{01}), u, \varepsilon\rangle \vdash \langle (q_2, q_1), \varepsilon, v\rangle \Leftrightarrow \exists q_2 \in F$: $\exists q_1 \in F_1$: $\exists w \in \Theta^*$: $(\langle q_{02}, u, \varepsilon\rangle \vdash_M \langle q_2, \varepsilon, w\rangle \wedge \langle q_{01}, w, \varepsilon\rangle \vdash_M \langle q_1, \varepsilon, v\rangle) \Leftrightarrow \exists w$: $(u, w) \in L(M_2), (w, v) \in L(M_1) \Leftrightarrow \langle u, v\rangle \in L(M_1) \circ L(M_2)$. ∎

Теорема 3.4 (Нива). *Всякое конечное преобразование $\eta : \Sigma^* \to \Gamma^*$ представимо в виде*

$$\eta = \varphi \circ \mathrm{id}\big|_R \circ \psi^{-1},$$

где $\psi : \Theta^* \to \Gamma^*$ и $\varphi : \Theta^* \to \Sigma*$ — неудлинняющие гомоморзизмы, а $R \subseteq \Theta^*$ — регулярный язык.

Доказательство. Если $\eta = L(M)$, считаем, что все переходы выполняют только одно чтение/запись. Примем за Θ рёбра M, и определим для $t = \langle q_1, u\rangle \to \langle q_2, v\rangle$ $\varphi(t) = a$ и $\varphi(t) = b$. В качестве R возьмём язык путейв M из начального состояния в завершающие. ∎

Теорема 3.5. *Конечные преобразования сохраняют автомтность и контекстную свободность языка.*

Доказательство. Достаточно доказать для пересечения с регулярным языком и для прямых и обратных неудлинняющих гомоморфизмов. Для регулярных языков уже всё доказано.

Утверждение 3.3. *L — КС-язык, $\varphi : \Sigma \to \Sigma'$ — гомоморфзим, тогда $\varphi(L)$ — КС-язык.*

Доказательство. Пусть $L = L(G)$, где $G = \langle N, \Sigma, P, S\rangle$ — КС-грамматика в НФ Хомского. Тогда $\varphi(L) = L(G')$, где $G' = \langle N, \Sigma', P, S$, а

$$P' = \{A \to BC \mid A \to BC \in P\} \cap \{A \to \varphi(A) \mid A \to a \in P\}$$

и добавим $S \to \varepsilon$, если это правило присутствовало в P. ∎

Утверждение 3.4. *L — КС-язык, R — регулярный язык, тогда $L \cap R$ — КС-язык.*

Доказательство. Пусть $L = L(M_1)$, $M_1 = \langle Q_1, \Sigma, \Gamma, \Delta_1, q_{01}, F_1\rangle$ — МП-автомат. причём все переходы в Δ_1 имеют вид $\langle q_1, a, \alpha\rangle \to \langle q_2, \beta\rangle$, $a \in \Sigma$. Пусть также $R = L(M_2)$, $M_2 = \langle Q_2, \Sigma, \Gamma, \Delta_2, q_{02}, F_2\rangle$ — ДКА. Тогда $L \cap R = L(M)$, где $M = \langle Q_1 \times Q_2, \Sigma, \Gamma, \Delta, (q_{01}, q_{02}), F_1 \times F_2\rangle$,

$$\Delta = \{\langle (q_1, q_2), a, \alpha\rangle \to \langle (q_1', q_2'), \beta\rangle \mid \langle q_1, a, \alpha\rangle \to \langle q_1', \beta\rangle \in \Delta_1 \wedge \langle q_2, a\rangle \to \langle q_2'\rangle \in \Delta_2\}$$

Утверждение 3.5. $\langle (q_1, q_2), w, \alpha\rangle \vdash_M \langle (q_1', q_2'), \varepsilon, \beta\rangle \Leftrightarrow \langle q_1, w, \alpha\rangle \vdash_{M_1} \langle q_1', \varepsilon, \beta\rangle$, $\langle q_2, w\rangle \vdash_{M_1} \langle q_2', \varepsilon\rangle$.

Далее всё тривиально. ∎

Утверждение 3.6. *L — КС-язык, $\varphi : \Gamma \to \Sigma$ — неудлинняющий гомоморфизм, тогда $\varphi^{-1}(L)$ — КС-язык.*

Доказательство. Обозначим $\Gamma_\varepsilon = \{a \in \Gamma \mid \varphi(a) = \varepsilon\}$, $\Gamma_a = \{b \in \Gamma \mid \varphi(b) = a\}$ для $a \in \Sigma$. Тогда $\varphi^{-1}(\varepsilon) = \Gamma_\varepsilon^*$ и $\varphi^{-1}(a) = \varphi^{-1}(\varepsilon)\Gamma_a\varphi^{-1}(\varepsilon)$, а эти языки регулярны, потому мы можем так пополнить нашу грамматику, так что для любого $u \in \Sigma \cup \{\varepsilon\}$ мы добавим нетерминал S_u такой, что $S_u \vdash \varphi^{-1}(u)$. Приведём нашу грамматику в НФ Хомского, и правила вида $A \to u$ заменим на $A \to S_u$. ∎

∎

$$\mathbb{N}^k = \{[n_1, \dots, n_k] \mid n_1, \dots, n_k \in \mathbb{N}\}$$

$$[x_1, \dots, x_k] + [y_1 + \dots + y_k] = [x_1 + y_1, \dots, x_k + y_k]$$

$$X + Y = \{x + y \mid x \in X, y \in Y\}$$

$$\langle X\rangle = \{\alpha_1 x_1 + \dots + \alpha_n x_n \mid x_1, \dots, x_n \in X, \alpha_1, \dots, \alpha_n \in \mathbb{N}\}$$

Определение 3.5. $X \subseteq \mathbb{N}^k$ — *линейное*, если $X = X_1 + \langle X_2\rangle$, $|X_1| < \infty$, $|X_2| < \infty$. $X \subseteq \mathbb{N}^k$ — *полулинейное*, если X — конечное объединение линейных множеств.

$$\Sigma = \{a_1, \dots, a_n\}, \psi : \Sigma^* \mapsto \mathbb{N}^n$$

$$\psi(u) = [|u|_{a_1}, \dots, |u|_{a_n}]$$

ψ — гомоморфизм $\langle \Sigma^*, \cdot\rangle \mapsto \langle \mathbb{N}^n, +\rangle$, $\psi(x \cdot y) = \psi(x) + \psi(y)$.

Определение 3.6. Язык L — линейный (полулинейный), если его образ $\psi(L)$ линейный (полулинейный).

Теорема 3.6. *$\forall X \subseteq \mathbb{N}^{|\Sigma|}$ X полулинейно $\Rightarrow \exists R \subseteq \Sigma^*$: $\psi(R) = X$, R — регулярен.*

Доказательство. Т. к. регулярные языки замкнуты относительно объединения, достаточно показать для линейных множеств.

X — линейно, $X = \{x_1, \dots, x_m\} + \langle\{y_1, \dots, y_t\}\rangle$. Обозначим $u_i = a_1^{x_{i,1}} a_2^{x_{i,2}} \dots a_n^{x_{i,n}}$, $v_i = a_1^{y_{i,1}} \dots a_n^{y_{i,n}}$. При этом, $\psi(u_i) = x_i$ и $\psi(v_i) = y_i$, а тогда

$$X = \psi(L((u_1 + \dots + u_m)(v_1 + \dots + v_t)^*)).$$

■

Лемма 3.2 (Усиленная лемма о разрастании)**.** *Для любого КС-языка* $\exists G = \langle N, \Sigma, P, S\rangle$: $L(G) = L \forall k\ \exists p$: $\forall w \in L\ |w| \geqslant p\ \exists A \in N$: $\exists x, y, z, u_1, \dots, u_k, v_1, \dots, v_k \in \Sigma^*$: $S \vdash xAz \vdash xu_1Av_1z \vdash \dots \vdash xu_1 \dots u_kAv_k \dots v_1z \vdash xu_1 \dots u_kyv_k \dots v_1z = w$, *и* $\forall j\ |u_jv_j| > 0$, $|u_1 \dots u_kv_k \dots v_1| \leqslant p$.

Доказательство. Пусть G — грамматика в НФ Хомского, распознающая L. Выберем $p = 2^{|N| \cdot k+1}$, тогда $\forall w\ |w| \geqslant p \Rightarrow$ его дерево вывода имеет глубину $\geqslant |N| \cdot k + 1 \Rightarrow$ хотя бы один нетерминал в самой длинной ветви встретится $\geqslant k + 1$ раз. Рассмотрим самое нижнее вхождение нетерминала в эту ветвь, ниже которого есть ещё k вхождений того же нетерминала.

```
      S
     /|\
    / | \
  x__A__z
     /|\
    / | \
 u1__A__v1
     /|\
    / | \
 u2__A__v2
      .
      .
      .
     /|\
    / | \
 uk__A__vk
      |
      y
```

■

Теорема 3.7 (Пари'к, Parikh)**.** *Для любого КС-языка $\psi(L)$ полулинейно.*

Доказательство. Пусть $L = L(G)$, $G = \langle N, \Sigma, P, S\rangle$ — в НФ Хомского. Для $M \subseteq N$ обозначим,

$L_M(G) = \{w \mid |w|$ выводится в G с использованием только нетерминалов из M, причём всех$\}$.

Достаточно доказать, что $\forall M \subseteq N\ L_M(G)$ полулинейно. Пусть p — число из леммы о разрастании для L и $k = |M|$.

$$X = \{\psi(w) \mid w \in L_M(G), |w| \leqslant p\}$$

$$Y = \{\psi(uv) \mid |uv| \leqslant p, \text{ существует вывод } B \vdash uBv \text{ только из нетерминалов M}\}$$

Утверждение 3.7. $\psi(L_M(G)) = X + \langle Y \rangle$

Доказательство. Докажем, что $\psi(L_M(G)) \supseteq X + \langle Y \rangle$. Рассмотрим $x \in X + \langle Y \rangle$, $x = x_1 + \alpha_1 y_1 + \dots + \alpha_m y_m$, пусть при этом $\psi(w') = x_1$ и $\psi(w_i) = y_i$. Тогда для w' существует дерево вывода, содержащее все нетерминалы $B_1, \dots, B_m$, а для w_j существует дерево вывода только из нетерминалов M вида

```
      B_j
     /   \
    /     \
 uj__B_j__vj
```

Тогда для получения дерева вывода для w, $\psi(w) = x$, нужно α_j раз встроить деерво вывода для $u_j B_j v_j$, вместо нетерминала B_j.

Теперь докажем обратное включение. Рассмотрим $w \in L_M(G)$, $|w| \leqslant p \Rightarrow \psi(w) \in X$. Иначе, по лемме о разрастании, для $k = |N|\ \exists u_1, \dots, u_k, v_1, \dots, v_k$: . Пусть $A^{(0)}, \dots, A^{(k)}$ — выделенные вхождения A. M_i — множество нетерминалов из M встречающихся в T ниже $A^{(i)}$, не считая A. T_i — поддерево между $A^{(i-1}$ и $A^{(i)}$. Тогда $|N| \geqslant |M| > |M_0| \geqslant |M_1| \geqslant |M_2| \ldots \geqslant |M_k| \geqslant 0$, но $k = |M|$. Тогда по принципу Дирихле $\exists j$: $M_{j-1} = M_j$, т. е. если мы выкинем поддерево T_j, то дерево по пржнему будет содержать все нетерминалы из M.

Выберем w_1 как часть слева от $A^{(j-1)}$, w_3 — справа от $A^{(j-1)}$, w_2 — вывод из $A^{(j)}$. Тогда $\psi(w) = \psi(w_1 w_2 w_3) + \psi(u_j v_j)$, где $(w_1 w_2 w_3) \in L_M(G)$ а $u_j v_j \in Y$.

Тогда мы можем провести индукцию по длине слова, а $\psi(w_1 w_2 w_3) \in X + \langle Y \rangle \Rightarrow \psi(w) \in X + \langle Y \rangle$. ■

■

Теорема 3.8 (Хомский-Шютценберже, Chomsky-Schutzenberger)**.** *Для любого КС-языка существует конечный преобразователь ψ такой, что $L = \psi(B_k)$, где B_k — язык ПСП с k типами скобок.*

Доказательство. Пусть $L = L(M)$, $M = \langle Q, \Sigma, \Gamma, \Delta, q_0, F \rangle$ — МП-автомат. Считаем, что все переходы в M имеют вид $\langle q_1, x, \varepsilon \rangle \to \langle q_2, A \rangle$ или $\langle q_1, x, A \rangle \to \langle q_2, \varepsilon \rangle$.

Для каждого стекового символа добавим «противоположный», они будут образовывать пару скобок. Тогда сопоставление скобке символа, который считывался при этой операции со стеком даст требуемое. ■

Теорема 3.9 (Х-Ш, сильная форма)**.** *Для любого КС-язык L существует неудлинняющий гомоморфизм φ, регулярный язык R такой и k такие, что $=_{\epsilon}\varphi(B_k \cap R)$.*

Доказательство. По слабой форме существуют такое k и КП ψ, что $L = \psi(B_k)$. Тогда по теореме Нива $\exists \varphi_1, \varphi_2$: $\exists R_1$: $L = \varphi_2(\varphi_1^{-1}(B_k) \cap R_1)$, где φ_i — неудлинняющие гомоморфизмы, R_1 — регулярный язык. Докажем, что если $\varphi_1^{-1}(B_k)$ можно представить в виде $\varphi'(B_l)$, то теорема доказана.

$\varphi_2(\varphi'(B_l) \cap R_1) = \varphi_2 \circ \varphi'(B_l \cap \varphi'^{-1}(R_1))$, где $\varphi = \varphi 2 \circ \varphi'$, а $R = \varphi'^{-1}(R)$. ■

Приложение A

Конспекты от лектора

A.1 Алгоритм Эрли

Алгоритм Эрли представляет собой нисходящий алгоритм синтаксического разбора, то есть построение дерева разбора осуществляется сверху вниз. Хотя, как и в алгоритме Кока-Янгера-Касами, верхняя оценка на временную сложность алгоритма является кубической по длине слова, на практике константа в алгоритме Эрли значительно ниже. Кроме того, для однозначных грамматик доказана квадратичная верхняя оценка на время работы алгоритма, а для многих грамматик в реальности сложность оказывается линейной. В силу относительной простоты алгоритма это позволяет использовать его в отдельных практических приложениях (хотя в целом LR- и GLR-алгоритмы разбора являются значительно более популярными).

A.1.1 Краткое описание алгоритма.

Алгоритм Эрли получает на вход контекстно-свободную грамматику $G = \langle \Sigma, N, P, S \rangle$ и проверяет выводимость слова w в грамматике G. Для кодирования выводов в данной грамматике используются так называемые *ситуации*, имеющие вид $(A \to \alpha \cdot \beta, i)$, где $(A \to \alpha\beta) \in P$, $\cdot$ — вспомогательный символ, не принадлежащий $\Sigma \cup N$, $i \in \overline{0, |w|}$. Ситуации хранятся в множествах $D_0, \dots, D_{|w|}$, причём наличие ситуации $(A \to \alpha \cdot \beta, i)$ в множестве D_j равносильна выполнению следующих условий (через w_{kl} обозначено подслово слова w с k-ой по l-ую позицию, причём позиции нумеруются с 0):

$$\exists \gamma \in (\Sigma \cup N)^* ((S \vdash w_{0i} A \gamma) \wedge \alpha \vdash w_{ij})$$

При этом можно считать, что рассматриваются только левосторонние выводы в грамматике G. Для удобства предположим, что грамматика G содержит правило $S \to S_1$, причём S не входит в другие правила грамматики. Тогда выводимость w в грамматике G равносильна условию $(S \to S_1 \cdot, 0) \in D_{|w|}$. Алгоритм Эрли рекурсивно строит множества $D_0, \dots, D_{|w|}$, поддерживая сформулированный выше инвариант.

A.1.2 Псевдокод.

Исходные параметры: Контекстно-свободная грамматика G, слово $w \in \Sigma^*$.
Результат: **True**, если $w \in L(G)$, **False**, иначе.
Применить алгоритм 2 к входным данным.
if $(S \to S_1 \cdot, 0) \in D_{|w|}$ **then**
  return ***True***
else
  return ***False***
end

Алгоритм 1: Проверка выводимости слова w в грамматике G.

A.1.3 Доказательство корректности.

Нужно доказать, что алгоритм 2 правильно строит множества $D_0, \dots, D_{|w|}$, то есть что он поддерживает инвариант

$$(A \to \alpha \cdot \beta, i) \in D_j \leftrightarrow \exists \delta \in (\Sigma \cup N)^*((S \vdash w_{0i}A\delta) \wedge \alpha \vdash w_{ij}).$$

Докажем импликацию слева направо индукцией по построению множеств D_j. Для этого нужно разобрать, в результате применения какой из инструкций алгоритма ситуация $(A \to \alpha \cdot \beta, i)$ попадает в множество D_j. База индукции, ситуация $(S \to \cdot S_1, 0) \in D_0$, очевидным образом удовлетворяет инварианту. Докажем шаг индукции.

Пусть ситуация $(A \to \alpha \cdot \beta, i)$ попала в D_j в результате применения правила $Scan$. В этом случае имеем $\alpha = \alpha' a$, $a = w[j-1]$ и $(A \to \alpha' \cdot a\beta, i) \in D_{j-1}$. По предположению индукции имеем $S \vdash w_{0i}A\delta$, что нам и было нужно, и $\alpha' \vdash w_{i(j-1)}$, тогда в силу $a = w[j-1]$ получаем $\alpha = \alpha' a \vdash w_{i(j-1)}w[j-1] = w_{ij}$, что и требовалось.

Теперь пусть ситуация $(A \to \alpha \cdot \beta, i)$ попала в D_j в результате применения правила $Predict$. По построению получаем, что $\alpha = \varepsilon$, $i = j$, что автоматически влечёт второй пункт утверждения. Кроме того, найдутся $i' \leqslant i$ и ситуация $(A' \to \alpha' \cdot A\delta', i') \in D_i$, откуда по предположению индукции имеем $S \vdash w_{0i'}A'\delta'', \alpha' \vdash w_{i'i}$. Получаем $S \vdash w_{0i'}A'\delta'' \vdash_1 w_{0i'}\alpha' A\delta'\delta'' \vdash w_{0i'}w_{i'i}A\delta'\delta'' = w_{0i}A\delta$, что и требовалось.

Осталось разобрать случай, когда ситуация $(A \to \alpha \cdot \beta, i)$ попала в D_j в результате применения правила $Complete$. По построению $\alpha = \alpha' A'$ и найдутся i' и γ, такие что $(A \to \alpha' \cdot A'\beta, i) \in D_{i'}$ и $(A' \to \gamma \cdot, i') \in D_j$. Следовательно, $\alpha = \alpha' A' \vdash w_{ii'}w_{i'j} = w_{ij}$, что и было нужно. Кроме того, $S \vdash w_{0i}A\delta$ по предположению индукции, что и требовалось доказать.

В одну сторону равносильность доказана, докажем в противоположную. Доказательство проведём индукцией по суммарной длине вывода $w_{0i}A\delta$ из S и w_{ij} из α, после чего применим индукцию по длине вывода w_{ij} из α. Разберём несколько случаев в зависимости от последнего символа α. Если $\alpha = \alpha' a$, тогда $a = w[j-1]$, $\alpha' \vdash w_{i(j-1)}$. По предположению индукции получаем $(A \to \alpha' \cdot a\beta, i) \in D_{j-1}$. Тогда по правилу $Scan$ получаем $(A \to \alpha' a \cdot \beta, i) \in D_j$, что и требовалось.

Исходные параметры: Контекстно-свободная грамматика G, слово $w \in \Sigma^*$.
Результат: Множества ситуаций $D_0, \dots, D_{|w|}$.
Инициализация:
$D_0 \leftarrow \{(S \to \cdot S_1, 0)\}$
for $i = 1, \dots, |w|$ **do**
  $D_i \leftarrow \varnothing$;
end
Шаг работы:
for $j = 0, \dots, |w|$ **do**
  $Scan(D, j)$;
  while D_j *изменяется* **do**
    $Complete(D, j)$;
    $Predict(D, j)$;
  end
end
Function *Scan(D, j)*
  if $j = 0$ **then**
    return
  end
  for $(A \to \alpha \cdot a\beta, i) \in D_{j-1}$ **do**
    if $a = w[(j-1)]$ **then**
      D_j.add $((A \to \alpha a \cdot \beta, i))$
    end
  end
end
Function *Predict(D,j)*
  for $(A \to \alpha \cdot B\beta, i) \in D_j$ **do**
    for $(B \to \gamma) \in P$ **do**
      D_j.add $(B \to \cdot\gamma, j)$
    end
  end
end
Function *Complete(D,j)*
  for $(B \to \gamma\cdot, i) \in D_j$ **do**
    for $(A \to \alpha \cdot B\beta, k) \in D_i$ **do**
      D_j.add $(A \to \alpha B \cdot \beta, k)$
    end
  end
end

Алгоритм 2: Алгоритм построения множеств $D_0, \dots, D_{|w|}$.

Если $\alpha = \alpha' B$, тогда получаем, что существует i', такой что $\alpha' \vdash w_{ii'}$, $B \vdash w_{i'j}$. Тогда имеем $(A \to \alpha' \cdot B\beta, i) \in D_{i'}$. Кроме того, можно записать $S \vdash w_{0i} A\delta \vdash w_{0i} w_{ii'} B\beta\delta$, а также $B \vdash_1 \gamma \vdash w_{i'j}$. Применяя индукцию по второму параметру, имеем $(B \to \gamma\cdot, i') \in D_j$, откуда по правилу *Complete* получаем $(A \to \alpha' B \cdot \beta, i) \in D_j$, что и требовалось.

Пусть теперь $\alpha = \varepsilon$, тогда $i = j$. Тогда либо $i = 0$, $A = S_1$, $\delta = \varepsilon$, что доказывает базу индукции, либо вывод можно переписать в виде $S \vdash w_{0i'} A'\delta'' \vdash_1 w_{0i'} \alpha' A\delta'\delta'' \vdash w_{0i'} w_{i'i} A\delta'\delta'' = w_{0i} A\delta$ для некоторого правила $(A' \to \alpha' A\delta') \in P$ и $\alpha' \vdash w_{i'i}$. Отсюда по предположению индукции следует, что $(A' \to \alpha' \cdot A\delta', i') \in D_i$, после чего по правилу *Predict* мы получаем $(A \to \cdot\beta, i) \in D_i$, что и требовалось. Корректность доказана.

A.1.4 Анализ сложности алгоритма.

Обозначим через $|G|$ суммарную длину всех правил из множества P (очевидно, что при фиксированном стартовом нетерминале правил достаточно для задания грамматики). Для хранения всех ситуаций из каждого из множеств D_j тогда требуется не больше чем $O(|G||w|)$ памяти, что в сумме даёт пространственную сложность $O(||G||w|^2)$. Для оптимизации временных затрат для каждой ситуации $(A \to \alpha \cdot \beta, i) \in D_j$ нужно хранить ссылку на ситуации вида $(A' \to \alpha' \cdot A\beta') \in D_i$, с которыми данная ситуация может «взаимодействовать» при применении инструкции *Complete*. Для этого достаточно либо поддерживать соответствующие обратные ссылки при появлении новых ситуаций, либо просто обеспечить быстрый доступ ко всем ситуациям со вторым индексом i в множестве D_j. Второе легко достигается при хранении множества D_j с помощью булева массива или массива множеств, индексированного возможными вторыми индексами.

Рассмотрим временную сложность при хранении D_j в виде булева массива. В массиве для D_j следует выделить $(j+1) \cdot |G|$ ячеек, обозначим эту величину за $|D_j|$. Тогда на каждой итерации алгоритма на процедуру *Scan* тратится $O(|D_{j-1}|)$, а на процедуру *Predict* — $O(|D_j|)$ времени, что даёт затраты $O(|G||w|)$ на каждой итерации. При операции *Complete* максимально возможное затраченное время на каждую ситуацию $(A \to \alpha \cdot \beta, i) \in D_j$ равно $O(|D_i|)$, что в сумме приводит к максимальным затратам на итерации на процедуру *Complete*, равным $O(|G|^2|w|^2)$. Суммарные затраты времени, таким образом, равны $O(|G|^2|w|^2)$ на одну итерацию и $O(|G|^2|w|^3)$ на весь алгоритм.

Данная оценка является сильно огрублённой, поскольку в большинстве случаев множества D_j оказываются существенно меньше. Кроме того, если для грамматики удаётся доказать, что число появлений каждой ситуации ограничено некоторым числом C, то суммарная сложность при правильной реализации не превысит C на число ситуаций, то есть будет квадратичной.

A.2 Свойства замкнутости.

Теорема A.1.

1. *Автоматные языки замкнуты относительно конечных преобразований.*
2. *Контекстно-свободные языки замкнуты относительно конечных преобразований.*

Доказательство.

1. Пусть L — автоматный язык, задаваемый конечным автоматом $A = \langle Q, \Sigma, \Delta, q_0, F\rangle$ с однобуквенными переходами. Пусть $T = \langle Q', \Sigma, \Sigma', \Delta', q_0', F'\rangle$ — конечный преобразователь, ψ — задаваемое им преобразование, причём $L' = \psi(L)$. Докажем автоматность языка L'. Можно считать, что для всякого перехода $(\langle q_1, x\rangle \to \langle q_2, x'\rangle) \in \Delta'$ выполняется соотношение $|x| + |x'| = 1$.

   Рассмотрим автомат $M = \langle Q \times Q', \Sigma', \Delta'', (q_0, q_0'), F \times F'\rangle$, где множество Δ'' задаётся следующим образом:

   - Если $(\langle q_1, x\rangle \to q_2) \in \Delta$ и $(\langle q_1', x\rangle \to \langle q_2', \varepsilon\rangle) \in \Delta'$, $x \in \Sigma$, то $(\langle (q_1, q_1'), \varepsilon\rangle \to (q_2, q_2')) \in \Delta''$.
   - Если $(\langle q_1', \varepsilon\rangle \to \langle q_2', x'\rangle \in \Delta'$, $x' \in \Sigma'$, то для всех $q \in Q$ выполняется $(\langle (q, q_1'), x'\rangle \to (q, q_2')) \in \Delta''$.

   Утверждение A.1. $\langle (q_1, q_1'), u'\rangle \vdash_M \langle (q_2, q_2'), \varepsilon\rangle \leftrightarrow \exists u \in \Sigma^*(\langle q_1, u\rangle \vdash_A \langle q_2, \varepsilon\rangle \wedge \langle q_1', u, \varepsilon\rangle \vdash_T \langle q_2', \varepsilon, u'\rangle)$

   Доказательство. $\Rightarrow$ Индукция по числу переходов. База очевидна. Пусть последний переход имел вид $\langle (q_3, q_3'), \varepsilon\rangle \to (q_2, q_2')$, тогда имеем $\langle (q_1, q_1'), u'\rangle \vdash_M \langle (q_3, q_3'), \varepsilon\rangle$. По предположению индукции получаем, что найдётся слово v, такое что $\langle q_1, v\rangle \vdash_A \langle q_3, \varepsilon\rangle$ и $\langle q_1', v, \varepsilon\rangle \vdash_T \langle q_3', \varepsilon, u'\rangle$. Кроме того, по построению имеем, что найдётся $a \in \Sigma$, такое что $(\langle q_3, a\rangle \to q_2) \in \Delta$ и $(\langle q_3', a\rangle \to \langle q_2', \varepsilon\rangle) \in \Delta'$. Отсюда вытекает, что $\langle q_1, va\rangle \vdash_A \langle q_2, \varepsilon\rangle$ и $\langle q_1', va, \varepsilon\rangle \vdash_T \langle q_2', \varepsilon, u'\rangle$, то есть в качестве v можно взять ua.

   Пусть теперь последний переход имел вид $\langle (q_3, q_3'), x'\rangle \to (q_2, q_2')$, тогда по построению имеем $q_3 = q_2$ и $\langle (q_1, q_1'), v'\rangle \vdash_M \langle (q_2, q_3'), \varepsilon\rangle$, где $v'x' = u'$, а также $(\langle q_3', \varepsilon\rangle \to \langle q_2', x'\rangle) \in \Delta'$. По предположению индукции получаем, что найдётся слово v, такое что $\langle q_1, v\rangle \vdash_A \langle q_2, \varepsilon\rangle$ и $\langle q_3', v, \varepsilon\rangle \vdash_T \langle q_2', \varepsilon, v'\rangle$. Отсюда следует, что $\langle q_1, v\rangle \vdash_A \langle q_2, \varepsilon\rangle$ и $\langle q_1', v, \varepsilon\rangle \vdash_T \langle q_2', \varepsilon, v'x'\rangle$, то есть можно взять $u = v$.

   $\Leftarrow$ Индукция по $|u| + |v|$ и $|u|$. База индукции очевидна, на шаге индукции нужно рассмотреть последний переход в преобразователе T, после чего воспользоваться предположением индукции и определением автомата M. Детали предоставляются читателю. ■

Таким образом, условие $\exists(q,q') \in F \times F'(\langle\langle(q_0,q_0'),u'\rangle \vdash_M \langle(q,q'),\varepsilon\rangle\rangle$ равносильно конъюнкции условий $\exists q \in F((\langle q_0,u\rangle \vdash_A \langle q,\varepsilon\rangle)$ и $\exists q' \in F'((\langle q_0', u, \varepsilon\rangle \vdash_A \langle q', \varepsilon, u'\rangle)$ для некоторого слова u. Отсюда очевидным образом следует доказываемое утверждение.

2. Пусть L —контекстно-свободный язык, задаваемый контекстно-свободной грамматикой $G = \langle N, \Sigma, P, S\rangle$ в нормальной форме Хомского. Пусть $T = \langle Q, \Sigma, \Sigma', \Delta,$
$q_0, F\rangle$ — конечный преобразователь, ψ — задаваемое им преобразование, причём $L' = \psi(L)$. Докажем, что L' является контекстно-свободным языком. Можно считать, что для всякого перехода $(\langle q_1, x\rangle \to \langle q_2, x'\rangle) \in \Delta'$ выполняется соотношение $|x| + |x'| = 1$.

   Положим $G' = \langle Q \times N \times Q \cup S', \Sigma', P', S'\rangle$, где P' задаётся следующим определением:

   - Для каждого правила $(A \to BC) \in P$ и произвольных состояний q_1, q_2, q_3 множество P' содержит правило $\langle q_1, A, q_2\rangle \to \langle q_1, B, q_3\rangle\langle q_3, C, q_2\rangle$.
   - Для каждого правила $(A \to a), a \in \Sigma$ и каждого перехода $\langle q_1, a\rangle \to \langle q_2, \varepsilon\rangle$ множество P' содержит правило $\langle q_1, A, q_2\rangle \to \varepsilon$.
   - Для каждого перехода $\langle q_1, \varepsilon\rangle \to \langle q_2, a'\rangle$ и произвольных нетерминала A и состояния q множество P' содержит правило $\langle q, A, q_2\rangle \to \langle q, A, q_1\rangle a'$
   - Для всех завершающих состояний $q \in F$ множество P' содержит правило $S' \to \langle q_0, S, q\rangle$.

   Утверждение A.2. $\langle q_1, A, q_2\rangle \vdash_{G'} u' \Leftrightarrow \exists u((A \vdash_G u) \wedge \langle q_1, u, \varepsilon\rangle \vdash_T \langle q_2, \varepsilon, u'\rangle$.

   Доказательство. $\Rightarrow$ Индукция по длине вывода. База индукции следует из определения грамматики G'. На шаге индукции нужно разобрать самое верхнее правило вывода в дереве вывода u' из $\langle q_1, A, q_2\rangle$.

   $\Leftarrow$ Индукция по длине слов u и u'. База индукции для случая $|u| = |u'| = 0$ легко проверяется. Теперь докажем шаг индукции, пусть вначале $|u| = 1$, тогда при $|u'| = 0$ утверждение следует из определения грамматики G'. Пусть $|u| = 1, |u'| > 0$, тогда без ограничения общности можно считать, что $u' = v'a'$ и $\langle q_1, u, \varepsilon\rangle \vdash_T \langle q_3, \varepsilon, v'\rangle$, а также $(\langle q_3, \varepsilon\rangle \to \langle q_2, a'\rangle) \in \Delta$. По предположению индукции и построению грамматики G' получаем, что $\langle q_1, A, q_3\rangle \vdash_{G'} v'$ и $\langle q_1, A, q_2\rangle \to \langle q_1, A, q_3\rangle a'$, что приводит к требуемому утверждению. В случае $|u| \geqslant 2$ имеем, что найдётся такое представление $u = u_1u_2, u' = u_1'u_2'$, состояние q_3 и правило $(A \to BC) \in P$, что $B \vdash_G u_1, C \vdash_G u_2, \langle q_1, u_1, \varepsilon\rangle \vdash_T \langle q_3, \varepsilon, u_1'\rangle$ и $\langle q_3, u_2, \varepsilon\rangle \vdash_T \langle q_2, \varepsilon, u_2'\rangle$. Отсюда следует, что $(\langle q_1, A, q_2\rangle \to \langle q_1, B, q_3\rangle\langle q_3, C, q_2\rangle) \in P'$, а также $\langle q_1, B, q_3\rangle \vdash_{G'} u_1'$ и $\langle q_3, C, q_2\rangle \vdash_{G'} u_2'$. Вывод $\langle q_1, A, q_2\rangle \vdash \langle q_1, B, q_3\rangle\langle q_3, C, q_2\rangle \to v_1v_2$ даёт требуемое утверждение. ∎

Теперь утверждение теоремы следует из четвёртого пункта определения грамматики G' и определения контекстно-свободной грамматики и рационального преобразования. Данная теорема может быть доказана и проще с использованием теоремы Нива.

■

Обозначим через $\mathcal{B}_k$ язык правильных скобочных последовательностей с k типами скобок. Формально, язык $\mathcal{B}_k$ задаётся грамматикой $S \to a_1 S \overline{a}_1 S \mid \ldots \mid a_k S \overline{a}_k S \mid \varepsilon$.

Теорема A.2 (Хомский-Щютценберже, слабая форма)**.** *Всякий контекстно-свободный язык можно получить рациональным преобразованием из языка $\mathcal{B}_k$ для некоторого натурального k.*

Доказательство. Пусть L — контекстно-свободный язык, распознаваемый МП-автоматом $M = \langle Q, \Sigma, \Gamma, \Delta, q_0, F\rangle$. Можно считать, что для всякого перехода $(\langle q_1, x, \alpha\rangle \to \langle q_2, \beta\rangle) \in \Delta$ выполняется условие $|\alpha|+|\beta| = 1$. Положим $k = |\Gamma|$ и будем считать, что $\Gamma = \{A_1, \ldots, A_k\}$. Обозначим $\overline{\Gamma} = \{\overline{A}_i \mid A_i \in \Gamma\}$.

Построим конечный преобразователь $T = \langle Q, \Gamma \cup \overline{\Gamma}, \Sigma, \Delta', q_0, F\rangle$, где Δ' строится по следующему алгоритму:

- Если $(\langle q_1, x, A_i\rangle \to \langle q_2, \varepsilon\rangle) \in \Delta$, то $(\langle q_1, \overline{A}_i\rangle \to \langle q_2, x\rangle) \in \Delta'$.
- Если $(\langle q_1, x, \varepsilon\rangle \to \langle q_2, A_i\rangle) \in \Delta$, то $(\langle q_1, A_i\rangle \to \langle q_2, x\rangle) \in \Delta'$.

Нетрудно доказать, что этот преобразователь является искомым. Действительно, если заменить в автомате с магазинной памятью переходы $(\langle q_1, x, A_i\rangle \to \langle q_2, \varepsilon\rangle$ на $(\langle q_1, x\rangle \to \langle q_2, \overline{A}_i\rangle$, а $(\langle q_1, x, A_i\rangle \to \langle q_2, \varepsilon\rangle$ на $(\langle q_1, x\rangle \to \langle q_2, A_i\rangle$, то мы получим преобразователь, отображающий слова в последовательности операций над ними в автомате с магазинной памятью. В частности, слова из языка будут переходить в правильные скобочные последовательности, а слова не из языка — в последовательности другого вида. Значит, преобразователь T осуществляет обратную операцию, то есть преобразует правильные скобочные последовательности в слова из языка. Формально нужно проверить, что условия $\langle q_1, u, \varepsilon\rangle \vdash_M \langle q_2, \varepsilon, \varepsilon\rangle$ и $\exists w \in \mathcal{B}_k \ (\langle q_1, w, \varepsilon\rangle \vdash_T \langle q_2, \varepsilon, u\rangle)$ являются равносильными, что делается индукцией по числу переходов в первом случае и по длине слова — во втором.

Отсюда получаем, что $S_T(\mathcal{B}_k) = L(M)$, что и требовалось. ■

Теорема A.3 (Хомский-Щютценберже, сильная форма)**.** *Всякий контекстно-свободный язык $L \in \Sigma^*$ представим в виде $L = \chi(\mathcal{B}_k \cap R)$, где χ — гомоморфизм, а R — автоматный язык.*

Доказательство.

Лемма A.1. *Пусть* $\varphi_1 : \Sigma^* \to \Sigma'^*, \varphi_2 : \Sigma'^* \to \Sigma''^*$ *— гомоморфизмы,* $R' \subseteq \Sigma'^*$ *— автоматный язык. Тогда найдутся автоматный язык* $R \subseteq \Sigma^*$ *и гомоморфизм* $\varphi : \Sigma^* \to \Sigma''^*$*, такие что для всякого языка* $L \subseteq \Sigma^*$ *выполняется соотношение* $\varphi(L \cap R) = \varphi_2(\varphi_1(L) \cap R')$*.*

Доказательство. Достаточно положить $\varphi = \varphi_2 \circ \varphi_1, R = \varphi_1^{-1}(R')$. ∎

Введём скобочный алфавит $\mathbb{A} = \{a_1, \overline{a}_1, \ldots, a_k, \overline{a}_k\}$. В силу теоремы Нива и слабой формы теоремы Хомского-Шютценберже найдутся алфавит Γ, автоматный язык $R' \subseteq \Gamma^*$ и неудлиняющие гомоморфизмы $\psi_1 : \Gamma^* \to \mathbb{A}^*, \psi_2 : \Gamma^* \to \Sigma^*$, такие что $L = \psi_2(\psi_1^{-1}(\mathcal{B}_k) \cap R')$. По доказанной лемме достаточно подобрать такое натуральное число l, автоматный язык R'' и гомоморфизм ψ, что $\psi_1^{-1}(\mathcal{B}_k) = \psi(\mathcal{B}_l)$.

Пусть $\Gamma = \{a_{11}, \ldots, a_{1m_1}, b_{11}, \ldots, b_{1n_1}, \ldots, a_{k1}, \ldots, a_{km_k}, b_{k1}, \ldots, b_{kn_k}, c_1, \ldots, c_r\}$ и $\psi_1(a_{ij}) = a_i, \psi_1(b_{ij}) = \overline{a}_i, \psi_1(c_j) = \varepsilon$. Положим $\mathbb{A}' = \{a_{ijk}, \overline{a}_{ijk} \mid i \leqslant k, j \leqslant m_i, k \leqslant n_i\} \cup \{d_i, \overline{d}_i \mid i \leqslant r\}$, $\psi(a_{ijk}) = a_{ij}, \psi(\overline{a_{ijk}}) = b_{ik}, \psi(d_i) = c_1, \psi(\overline{d_i}) = \varepsilon$.

Обозначим через $\mathcal{B}_l$ язык скобочных последовательностей над алфавитом $\mathbb{A}'$ и докажем, что $\psi(\mathcal{B}_l) = \psi_1^{-1}(\mathcal{B}_k)$. Включение слева направо очевидно следует из того, что при композиции $\psi_1 \circ \varphi$ скобки одного и того же типа переходят либо в скобки одного и того же типа, либо одновременно переходят в пустое слово. Проверим обратное включение.

Действительно, $\psi_1^{-1}(\mathcal{B}_k)$ содержит слова, которые после вычёркивания символов c_i можно сократить до пустого удалением пар вида $a_{ij}b_{ik}$. Подберём для каждого такого слова w его прообраз под действием отношения ψ, принадлежащий $\mathcal{B}_l$. Если символы a_{ik}, b_{il} сокращались друг с другом, то заменим их на a_{ijk} и $\overline{a}_{ijk}$. Все вхождения символа c_i заменим на $d_i\overline{d}_i$. Нетрудно проверить, что получится правильная скобочная последовательность, сокращаемая до пустого ровно теми же операциями, что и w, принадлежность языку R также очевидна.

Теорема доказана. ∎