Лекция A9
КС-языки: лемма о накачке, операции на них

Вадим Пузаренко

30 октября 2023 г.

Высота дерева разбора

Предложение A9.1.

Пусть дано дерево разбора, соответствующее НФХ-грамматике $\mathfrak{G} = (V, \Sigma, P, S)$, и пусть кроной дерева является терминальная цепочка α. Если n — наибольшая длина пути от корня к листьям, то $\mathrm{lh}(\alpha) \leqslant 2^{n-1}$.

Высота дерева разбора

Предложение А9.1.

Пусть дано дерево разбора, соответствующее НФХ-грамматике $\mathfrak{G} = (V, \Sigma, P, S)$, и пусть кроной дерева является терминальная цепочка α. Если n — наибольшая длина пути от корня к листьям, то $\mathrm{lh}(\alpha) \leqslant 2^{n-1}$.

Доказательство.

Простой индукцией по n.

Базис. Дерево с максимальной длиной пути 1 состоит из корня и листа, отмеченного терминалом. Цепочка α — терминал и $\mathrm{lh}(\alpha) = 1 = 2^0 = 2^{1-1}$.

Индукция. Предположим, что самый длинный путь имеет длину $n > 1$. Тогда корень дерева имеет продукцию вида $A \longrightarrow BC$. Все пути в поддеревьях с корнями, отмеченными B и C, имеют длину $\leqslant n-1$, поскольку в путях исключено ребро от корня A к сыну (B или C). По предположению индукции, эти поддеревья имеют кроны $\leqslant 2^{n-2}$. Таким образом, крона всего дерева имеет длину $\leqslant 2^{n-2} + 2^{n-2} = 2^{n-1}$. $\square$

Лемма о накачке

Теорема A9.1.

Пусть L — КС-язык; тогда существует $n_0 = n_0(L) \geqslant 1$ такое, что выполняется следующее: если $\zeta \in L$ таково, что $\mathrm{lh}(\zeta) \geqslant n_0$, то оно представляет собой $\zeta = \alpha\hat{}\beta\hat{}\gamma\hat{}\delta\hat{}\eta$, удовлетворяющее следующим условиям.

1. $\mathrm{lh}(\beta\hat{}\gamma\hat{}\delta) \leqslant n_0$;
2. $\beta\hat{}\delta \neq \varepsilon$;
3. $\alpha\hat{}\beta^l\hat{}\gamma\hat{}\delta^l\hat{}\eta \in L$ для всех $l \geqslant 0$.

Лемма о накачке

Теорема A9.1.

Пусть L — КС-язык; тогда существует $n_0 = n_0(L) \geqslant 1$ такое, что выполняется следующее: если $\zeta \in L$ таково, что $\mathrm{lh}(\zeta) \geqslant n_0$, то оно представляет собой $\zeta = \alpha\hat{}\beta\hat{}\gamma\hat{}\delta\hat{}\eta$, удовлетворяющее следующим условиям.

1. $\mathrm{lh}(\beta\hat{}\gamma\hat{}\delta) \leqslant n_0$;
2. $\beta\hat{}\delta \neq \varepsilon$;
3. $\alpha\hat{}\beta^l\hat{}\gamma\hat{}\delta^l\hat{}\eta \in L$ для всех $l \geqslant 0$.

Доказательство.

Если $L \subseteq \{\varepsilon\}$, то слово $\zeta \in L$ с $\mathrm{lh}(\zeta) > 0$ отсутствует, поэтому можно считать, что $L \setminus \{\varepsilon\} \neq \varnothing$.
По теореме A8.9, существует НФХ-грамматика $\mathfrak{G} = (V, \Sigma, P, S)$, порождающая язык $L \setminus \{\varepsilon\}$. Положим $m = |V|$ и $n_0 = 2^m$.

Лемма о накачке

Доказательство (продолжение).

Предположим, что $\zeta \in L$ имеет длину $\geqslant n_0$. По предложению A9.1, любое дерево разбора, в котором наибольшая длина пути $\leqslant m$, должно иметь крону $\leqslant 2^{m-1} = \frac{n_0}{2}$. Такое дерево разбора не может иметь крону ζ, поскольку $\mathrm{lh}(\zeta) \geqslant n_0 > \frac{n_0}{2}$. Тем самым, любое дерево T разбора с кроной ζ имеет путь длиной $k+1 \geqslant m+1$. Пусть $S = A_0, A_1, \ldots, A_k$ — вершины данного пути, отмеченные переменными. Так как $m = |V|$, найдутся $k - m \leqslant i < j \leqslant k$ такие, что $A_i = A_j$. Определим представление ζ следующим образом.

- γ — крона дерева, корень которого помечен A_j.
- $\beta\hat{}\gamma\hat{}\delta$ — крона дерева, корень которого помечен A_i.
- ζ — крона дерева, корень которого помечен S.

Докажем, что данное представление удовлетворяет всем требуемым условиям.

Лемма о накачке

Доказательство (окончание).

1) Действительно, $k - i \leqslant m$, поэтому самый длинный путь в поддереве с корнем A_i не превосходит $m + 1$, а по предложению А9.1, длина кроны $\mathrm{lh}(\beta\hat{}\gamma\hat{}\delta) \leqslant 2^m = n_0$.

2) Дерево с корнем, помеченным A_i, содержит в качестве собственного поддерева дерево с корнем, помеченным A_j; поэтому $\beta\hat{}\delta \neq \varepsilon$.

3) При $l = 1$ случай очевиден.

Пусть $l = 0$; тогда достаточно рассмотреть дерево $T_{T(A_j)}^{T(A_i)}$.

Индукцией по $l \geqslant 1$ докажем, что $\alpha\hat{}\beta^l\hat{}\gamma\hat{}\delta^l\hat{}\eta \in L$. О базе сказано выше. Пусть T_0 — дерево разбора, кроной которого является $\alpha\hat{}\beta^l\hat{}\gamma\hat{}\delta^l\hat{}\eta$. Пусть также $A (= A_i)$ — вершина, кроной поддерева $T_0(A)$ для которой является цепочка γ. Тогда кроной дерева $T_0{}_{T(A_i)}^{T_0(A)}$ будем цепочка $\alpha\hat{}\beta^{l+1}\hat{}\gamma\hat{}\delta^{l+1}\hat{}\eta$. $\square$

Лемма о накачке

1. Мы выбираем язык L, желая доказать, что он не контекстно-свободный.
2. Наш "противник" выбирает заранее неизвестное $n_0 \geqslant 1$, поэтому мы должны рассчитывать на любое возможное значение.
3. Мы выбираем слово ζ с $\mathrm{lh}(\zeta) \geqslant n_0$.
4. "Противник" предоставляет разбиение $\zeta = \alpha\hat{}\beta\hat{}\gamma\hat{}\delta\hat{}\eta$, причём $\beta\hat{}\delta \neq \varepsilon$ и $\mathrm{lh}(\beta\hat{}\gamma\hat{}\delta) \leqslant n_0$.
5. Мы "выигрываем", если можем выбрать $l \in \omega$ так, что $\alpha\hat{}\beta^l\hat{}\gamma\hat{}\delta^l\hat{}\eta \notin L$.

Лемма о накачке

Пример А9.1.

Пусть $L = \{0^n 1^n 2^n | n \in \omega\}$. Допустим, что язык L контекстно-свободный. Тогда существует n_0 из леммы о накачке. Выберем $\zeta = 0^{n_0} 1^{n_0} 2^{n_0}$. Пусть дано представление $\zeta = \alpha\hat{}\beta\hat{}\gamma\hat{}\delta\hat{}\eta$, где $\beta\hat{}\delta \neq \varepsilon$ и $\mathrm{lh}(\beta\hat{}\gamma\hat{}\delta) \leqslant n_0$. Так как $\mathrm{lh}(01^{n_0}2) = n_0 + 2 > n_0$, цепочка $\xi = \beta\hat{}\gamma\hat{}\delta$ не содержит нулей или двоек.

1. ξ не содержит нулей; тогда $\alpha\hat{}\gamma\hat{}\eta \notin L$, поскольку количество нулей в нём равняется n_0, а суммарное количество единиц и двоек $< 2n_0$.
2. ξ не содержит двоек; тогда $\alpha\hat{}\gamma\hat{}\eta \notin L$, поскольку количество двоек в нём равняется n_0, а суммарное количество единиц и нулей $< 2n_0$.

Лемма о накачке

Пример А9.2.

Пусть $L = \{0^i 1^j 2^i 3^j | i, j \in \omega\}$. Допустим, что язык L контекстно-свободный. Тогда существует n_0 из леммы о накачке. Выберем $\zeta = 0^{n_0} 1^{n_0} 2^{n_0} 3^{n_0}$. Пусть дано представление $\zeta = \alpha\hat{\ }\beta\hat{\ }\gamma\hat{\ }\delta\hat{\ }\eta$, где $\beta\hat{\ }\delta \neq \varepsilon$ и $\mathrm{lh}(\beta\hat{\ }\gamma\hat{\ }\delta) \leqslant n_0$. Как и выше, доказывается, что $\xi = \beta\hat{\ }\gamma\hat{\ }\delta$ не может содержать одновременно представителей трёх символов.

1. ξ не содержит нулей и троек; тогда $\alpha\hat{\ }\gamma\hat{\ }\eta \notin L$, поскольку количество нулей и троек в нём равняется n_0, а суммарное количество единиц и двоек $< 2n_0$.
2. ξ не содержит двоек и троек; тогда $\alpha\hat{\ }\gamma\hat{\ }\eta \notin L$, поскольку количество двоек и троек в нём равняется n_0, а суммарное количество единиц и нулей $< 2n_0$.
3. ξ не содержит нулей и единиц; тогда $\alpha\hat{\ }\gamma\hat{\ }\eta \notin L$, поскольку количество нулей и единиц в нём равняется n_0, а суммарное количество двоек и троек $< 2n_0$.

Лемма о накачке

Пример А9.3.

Пусть $L = \{\alpha\hat{}\alpha | \alpha \in \{0;1\}^*\}$. Допустим, что L контекстно-свободный. Тогда существует n_0 из леммы о накачке. Выберем $\zeta = 0^{n_0}1^{n_0}0^{n_0}1^{n_0}$. Пусть дано представление $\zeta = \alpha\hat{}\beta\hat{}\gamma\hat{}\delta\hat{}\eta$, где $\beta\hat{}\delta \neq \varepsilon$ и $\mathrm{lh}(\beta\hat{}\gamma\hat{}\delta) \leqslant n_0$. Докажем, что $\alpha\hat{}\gamma\hat{}\eta \notin L$. Так как $\mathrm{lh}(\beta\hat{}\gamma\hat{}\delta) \leqslant n_0$, имеем $\mathrm{lh}(\alpha\hat{}\gamma\hat{}\eta) \geqslant 3n_0$. Таким образом, если $\alpha\hat{}\gamma\hat{}\eta = \xi\hat{}\xi$, то $\mathrm{lh}(\xi) \geqslant \dfrac{3n_0}{2}$. Возможны несколько случаев.

- Предположим, что $\beta\hat{}\gamma\hat{}\delta$ находится в пределах первых групп нулей и единиц (скажем, $\mathrm{lh}(\beta\hat{}\delta) = k > 0$). Тогда $\mathrm{lh}(\alpha\hat{}\gamma\hat{}\eta) = 4n_0 - k$ и, следовательно, начало длины $2n_0 - \dfrac{k}{2}$ заканчивается нулем, а само слово заканчивается единицей.
- Предположим, что $\beta\hat{}\gamma\hat{}\delta$ находится в пределах последних групп нулей и единиц (скажем, $\mathrm{lh}(\beta\hat{}\delta) = k > 0$). Тогда $\mathrm{lh}(\alpha\hat{}\gamma\hat{}\eta) = 4n_0 - k$ и, следовательно, начало длины $2n_0 - \dfrac{k}{2}$ начинается нулем, а конец слова той же длины начинается единицей.

Лемма о накачке

Пример A9.3 (окончание).

- Предположим, что $\beta\hat{}\gamma\hat{}\delta$ находится в пределах второй и третьей групп (скажем, $\beta\hat{}\delta = 1^{k_0}0^{k_1}$). Тогда $\alpha\hat{}\gamma\hat{}\eta = 0^{n_0}1^{n_0-k_0}0^{n_0-k_1}1^{n_0}$ и, следовательно, начало длины $2n_0 - \dfrac{k_0+k_1}{2}$ имеет n_0 нулей, а конец слова той же длины — $\leqslant n_0 - k_1 < n_0$ нулей.

Однобуквенный алфавит

Замечание А9.1.

Хотя в общем случае леммы о накачке для КС- и регулярных языков содержат только необходимые условия, для однобуквенных алфавитов являются и достаточными.

Однобуквенный алфавит

Замечание A9.1.

Хотя в общем случае леммы о накачке для КС- и регулярных языков содержат только необходимые условия, для однобуквенных алфавитов являются и достаточными.

Теорема A9.2.

Любой контекстно-свободный язык $L \subseteq \{0\}^*$ является регулярным.

Однобуквенный алфавит

Замечание A9.1.

Хотя в общем случае леммы о накачке для КС- и регулярных языков содержат только необходимые условия, для однобуквенных алфавитов являются и достаточными.

Теорема A9.2.

Любой контекстно-свободный язык $L \subseteq \{0\}^*$ является регулярным.

Доказательство.

Без ограничения общности можно считать, что L — бесконечный язык. По теореме A9.1 о накачке, существует $n_0 \geqslant 1$, для которого выполняется следующее:

Однобуквенный алфавит

Доказательство (продолжение).

1. для всех $\zeta \in L$ таких, что $\mathrm{lh}(\zeta) \geqslant n_0$,
2. найдутся α, β, γ, δ, η $(\in \{0\}^*)$, удовлетворяющие следующим условиям:
   - $\zeta = \alpha\hat{}\beta\hat{}\gamma\hat{}\delta\hat{}\eta$,
   - $\beta\hat{}\delta \neq \varepsilon$,
   - $\mathrm{lh}(\beta\hat{}\gamma\hat{}\delta) \leqslant n_0$;
3. такие что $\alpha\hat{}\beta^l\hat{}\gamma\hat{}\delta^l\hat{}\eta \in L$ для всех $l \in \omega$.

Однобуквенный алфавит

Доказательство (продолжение).

1. для всех $\zeta \in L$ таких, что $\mathrm{lh}(\zeta) \geqslant n_0$,
2. найдутся α, β, γ, δ, η $(\in \{0\}^*)$, удовлетворяющие следующим условиям:
   - $\zeta = \alpha\hat{}\beta\hat{}\gamma\hat{}\delta\hat{}\eta$,
   - $\beta\hat{}\delta \neq \varepsilon$,
   - $\mathrm{lh}(\beta\hat{}\gamma\hat{}\delta) \leqslant n_0$;
3. такие что $\alpha\hat{}\beta^l\hat{}\gamma\hat{}\delta^l\hat{}\eta \in L$ для всех $l \in \omega$.

Положим
$L_0 = \{\zeta \in L \mid \mathrm{lh}(\zeta) < n_0\}$, $L_1 = \{\zeta \in L \mid \mathrm{lh}(\zeta) \geqslant n_0\}$;
тогда $L = L_0 \uplus L_1$.

Однобуквенный алфавит

Доказательство (продолжение).

1. для всех $\zeta \in L$ таких, что $\mathrm{lh}(\zeta) \geqslant n_0$,
2. найдутся α, β, γ, δ, η $(\in \{0\}^*)$, удовлетворяющие следующим условиям:
   - $\zeta = \alpha\hat{\ }\beta\hat{\ }\gamma\hat{\ }\delta\hat{\ }\eta$,
   - $\beta\hat{\ }\delta \neq \varepsilon$,
   - $\mathrm{lh}(\beta\hat{\ }\gamma\hat{\ }\delta) \leqslant n_0$;
3. такие что $\alpha\hat{\ }\beta^l\hat{\ }\gamma\hat{\ }\delta^l\hat{\ }\eta \in L$ для всех $l \in \omega$.

Положим
$L_0 = \{\zeta \in L \mid \mathrm{lh}(\zeta) < n_0\}$, $L_1 = \{\zeta \in L \mid \mathrm{lh}(\zeta) \geqslant n_0\}$;
тогда $L = L_0 \uplus L_1$.
Так как любой конечный язык является регулярным (см. следствие A1.2), L_0 является таковым. По теореме A4.2, достаточно показать, что L_1 представляется в виде объединения конечного числа языков, задаваемых арифметическими прогрессиями.

Однобуквенный алфавит

Доказательство (продолжение).

Пусть $\zeta \in L_1$; тогда найдутся слова α, β, γ, δ, η $(\in \{0\}^*)$, удовлетворяющие следующим условиям:

1. $\zeta = \alpha\hat{\ }\beta\hat{\ }\gamma\hat{\ }\delta\hat{\ }\eta$;
2. $\mathrm{lh}(\beta\hat{\ }\gamma\hat{\ }\delta) \leqslant n_0$;
3. $\beta\hat{\ }\delta \neq \varepsilon$;
4. $\alpha\hat{\ }\beta^{l+1}\hat{\ }\gamma\hat{\ }\delta^{l+1}\hat{\ }\eta = 0^{\mathrm{lh}(\zeta)+l\cdot\mathrm{lh}(\beta\hat{\ }\delta)} \in L_1$ для всех $l \in \omega$.

Однобуквенный алфавит

Доказательство (продолжение).

Пусть $\zeta \in L_1$; тогда найдутся слова α, β, γ, δ, η $(\in \{0\}^*)$, удовлетворяющие следующим условиям:

1. $\zeta = \alpha\hat{\ }\beta\hat{\ }\gamma\hat{\ }\delta\hat{\ }\eta$;
2. $\mathrm{lh}(\beta\hat{\ }\gamma\hat{\ }\delta) \leqslant n_0$;
3. $\beta\hat{\ }\delta \neq \varepsilon$;
4. $\alpha\hat{\ }\beta^{l+1}\hat{\ }\gamma\hat{\ }\delta^{l+1}\hat{\ }\eta = 0^{\mathrm{lh}(\zeta)+l\cdot\mathrm{lh}(\beta\hat{\ }\delta)} \in L_1$ для всех $l \in \omega$.

Положим $d = \mathrm{lh}(\beta\hat{\ }\gamma)$ и $L_{\zeta,d} = \{0^{\mathrm{lh}(\zeta)+d\cdot l} \mid l \in \omega\}$. Заметим, что $L_{\zeta,d} \subseteq L_1$ и для каждого $\zeta \in L_1$ существует $d > 0$ такое, что язык $L_{\zeta,d}$ определён и не пуст.

Если же $L_{\zeta,d}$ не определён к этому моменту, положим $L_{\zeta,d} = \varnothing$ (здесь $\zeta \in L_1$ и $0 < d \leqslant n_0$). Тем самым, $L_{\zeta,d} \subseteq L_1$ для всех $\zeta \in L_1$ и $0 < d \leqslant n_0$.

Однобуквенный алфавит

Доказательство (продолжение).

Докажем, что $L_1 = \bigcup\limits_{\zeta \in L_1,\, 0 < d \leqslant n_0} L_{\zeta,d}$.

Однобуквенный алфавит

Доказательство (продолжение).

Докажем, что $L_1 = \bigcup_{\zeta\in L_1,\, 0<d\leqslant n_0} L_{\zeta,d}$.

$(\subseteq)$ Пусть $\phi \in L_1$; тогда из вышесказанного следует существование $d_0 > 0$ $(d_0 \leqslant n_0)$ такого, что $0^{\mathrm{lh}(\phi)+d_0\cdot l} \in L_1$ для всех $l \in \omega$; в частности, $\phi = 0^{\mathrm{lh}(\phi)+d_0\cdot 0} \in L_1$; таким образом, $\phi \in L_{\phi,d_0} \subseteq \bigcup_{\zeta\in L_1,\, 0<d\leqslant n_0} L_{\zeta,d}$.

Однобуквенный алфавит

Доказательство (продолжение).

Докажем, что $L_1 = \bigcup\limits_{\zeta \in L_1,\, 0<d\leqslant n_0} L_{\zeta,d}$.

($\subseteq$) Пусть $\phi \in L_1$; тогда из вышесказанного следует существование $d_0 > 0$ ($d_0 \leqslant n_0$) такого, что $0^{\mathrm{lh}(\phi)+d_0\cdot l} \in L_1$ для всех $l \in \omega$; в частности, $\phi = 0^{\mathrm{lh}(\phi)+d_0\cdot 0} \in L_1$; таким образом, $\phi \in L_{\phi,d_0} \subseteq \bigcup\limits_{\zeta \in L_1,\, 0<d\leqslant n_0} L_{\zeta,d}$.

($\supseteq$) Так как $L_{\zeta,d} \subseteq L_1$ для всех $\zeta \in L_1$ и $0 < d \leqslant n_0$, имеем $\bigcup\limits_{\zeta \in L_1,\, 0<d\leqslant n_0} L_{\zeta,d} \subseteq L_1$.

Однобуквенный алфавит

Доказательство (продолжение).

Докажем, что $L_1 = \bigcup\limits_{\zeta \in L_1,\, 0<d\leqslant n_0} L_{\zeta,d}$.

($\subseteq$) Пусть $\phi \in L_1$; тогда из вышесказанного следует существование $d_0 > 0$ ($d_0 \leqslant n_0$) такого, что $0^{\mathrm{lh}(\phi)+d_0\cdot l} \in L_1$ для всех $l \in \omega$; в частности, $\phi = 0^{\mathrm{lh}(\phi)+d_0\cdot 0} \in L_1$; таким образом, $\phi \in L_{\phi,d_0} \subseteq \bigcup\limits_{\zeta \in L_1,\, 0<d\leqslant n_0} L_{\zeta,d}$.

($\supseteq$) Так как $L_{\zeta,d} \subseteq L_1$ для всех $\zeta \in L_1$ и $0 < d \leqslant n_0$, имеем $\bigcup\limits_{\zeta \in L_1,\, 0<d\leqslant n_0} L_{\zeta,d} \subseteq L_1$.

Положим теперь $S_j^d = \{\zeta \in L_1 \mid \mathrm{lh}(\zeta) \equiv j (\mathrm{mod}\, d),\ L_{\zeta,d} \neq \varnothing\}$, где $0 < d \leqslant n_0$ и $0 \leqslant j < d$. Нетрудно видеть, что выполняется соотношение

$$L_1 = \bigcup_{d=1}^{n_0} \bigcup_{j=0}^{d-1} S_j^d.$$

Однобуквенный алфавит

Доказательство (продолжение).

Выберем в каждом непустом множестве S_j^d слово μ_j^d наименьшей длины. Так как $\mu_j^d \in S_j^d$, выполняются следующие соотношения: $\mu_j^d \in L_1$, $\mathrm{lh}(\mu_j^d) \equiv j (\mathrm{mod}\, d)$ и $L_{\mu_j^d, d} \neq \varnothing$. Докажем теперь, что $L_{\mu_j^d, d} = S_j^d$.

Однобуквенный алфавит

Доказательство (продолжение).

Выберем в каждом непустом множестве S_j^d слово μ_j^d наименьшей длины. Так как $\mu_j^d \in S_j^d$, выполняются следующие соотношения: $\mu_j^d \in L_1$, $\mathrm{lh}(\mu_j^d) \equiv j (\mathrm{mod}\, d)$ и $L_{\mu_j^d,d} \neq \varnothing$. Докажем теперь, что $L_{\mu_j^d,d} = S_j^d$.

($\subseteq$) Пусть $\phi \in L_{\mu_j^d,d}$; тогда $\phi \in L_1$ и $\phi = 0^{\mathrm{lh}(\mu_j^d)+d\cdot l_0}$ для подходящего $l_0 \in \omega$; следовательно, $\mathrm{lh}(\phi) \equiv j (\mathrm{mod}\, d)$. Остаётся проверить только, что $L_{\phi,d} \neq \varnothing$. В самом деле, если $\psi = 0^{\mathrm{lh}(\phi)+d\cdot k}$, где $k \in \omega$, то $\psi = 0^{\mathrm{lh}(\mu_j^d)+(l_0+k)\cdot d} \in L_{\mu_j^d,d} \subseteq L_1$. Таким образом, $\phi \in S_j^d$.

Однобуквенный алфавит

Доказательство (продолжение).

Выберем в каждом непустом множестве S_j^d слово μ_j^d наименьшей длины. Так как $\mu_j^d \in S_j^d$, выполняются следующие соотношения: $\mu_j^d \in L_1$, $\mathrm{lh}(\mu_j^d) \equiv j(\mathrm{mod}\, d)$ и $L_{\mu_j^d,d} \neq \varnothing$. Докажем теперь, что $L_{\mu_j^d,d} = S_j^d$.

($\subseteq$) Пусть $\phi \in L_{\mu_j^d,d}$; тогда $\phi \in L_1$ и $\phi = 0^{\mathrm{lh}(\mu_j^d)+d\cdot l_0}$ для подходящего $l_0 \in \omega$; следовательно, $\mathrm{lh}(\phi) \equiv j(\mathrm{mod}\, d)$. Остаётся проверить только, что $L_{\phi,d} \neq \varnothing$. В самом деле, если $\psi = 0^{\mathrm{lh}(\phi)+d\cdot k}$, где $k \in \omega$, то $\psi = 0^{\mathrm{lh}(\mu_j^d)+(l_0+k)\cdot d} \in L_{\mu_j^d,d} \subseteq L_1$. Таким образом, $\phi \in S_j^d$.

($\supseteq$) Пусть теперь $\phi \in S_j^d$; тогда $\mathrm{lh}(\phi) \equiv j(\mathrm{mod}\, d)$ и, следовательно, $\phi = 0^{\mathrm{lh}(\mu_j^d)+l\cdot d}$ для подходящего $l \in \omega$, поскольку μ_j^d имеет наименьшую длину в S_j^d. Таким образом, $\phi \in L_{\mu_j^d,d}$.

Однобуквенный алфавит

Доказательство (окончание).

В конечном итоге, $L_1 = \bigcup\{L_{\mu_j^d,d} \mid S_j^d \neq \varnothing\}$. $\square$

Подстановки

Пусть Σ_1, Σ_2 — алфавиты и пусть каждому $a \in \Sigma_1$ сопоставляется КС-язык $s(a) \subseteq \Sigma_2^*$. Если $\alpha = a_1 a_2 \dots a_k \in \Sigma_1^*$, то полагаем конкатенацию языков $s(\alpha) = s(a_1)s(a_2)\dots s(a_k)$ (в случае, когда $\alpha = \varepsilon$, имеем $s(\alpha) = \varepsilon$). Выбор языков выше определяют функцию s (называемую **подстановкой** на (Σ_1, Σ_2)).

Подстановки

Пусть Σ_1, Σ_2 — алфавиты и пусть каждому $a \in \Sigma_1$ сопоставляется КС-язык $s(a) \subseteq \Sigma_2^*$. Если $\alpha = a_1 a_2 \ldots a_k \in \Sigma_1^*$, то полагаем конкатенацию языков $s(\alpha) = s(a_1)s(a_2)\ldots s(a_k)$ (в случае, когда $\alpha = \varepsilon$, имеем $s(\alpha) = \varepsilon$). Выбор языков выше определяют функцию s (называемую **подстановкой** на (Σ_1, Σ_2)).

Теорема А9.3.

Если L — КС-язык в алфавите Σ_1 и s — подстановка на (Σ_1, Σ_2), то $s(L)$ также КС-язык.

Подстановки

Пусть Σ_1, Σ_2 — алфавиты и пусть каждому $a \in \Sigma_1$ сопоставляется КС-язык $s(a) \subseteq \Sigma_2^*$. Если $\alpha = a_1 a_2 \ldots a_k \in \Sigma_1^*$, то полагаем конкатенацию языков $s(\alpha) = s(a_1)s(a_2)\ldots s(a_k)$ (в случае, когда $\alpha = \varepsilon$, имеем $s(\alpha) = \varepsilon$). Выбор языков выше определяют функцию s (называемую **подстановкой** на (Σ_1, Σ_2)).

Теорема А9.3.

Если L — КС-язык в алфавите Σ_1 и s — подстановка на (Σ_1, Σ_2), то $s(L)$ также КС-язык.

Доказательство.

Пусть $\mathfrak{G} = (V, \Sigma_1, P, S)$ и $\mathfrak{G}_a = (V_a, \Sigma_2, P_a, S_a)$ $(a \in \Sigma_1)$ таковы, что $L(\mathfrak{G}) = L$ и $L(\mathfrak{G}_a) = s(a)$ для любого $a \in \Sigma_1$. Без ограничения общности, будем предполагать, что множества переменных (нетерминальных символов) попарно не пересекаются. Определим грамматику $\mathfrak{G}' = (V', \Sigma_2, P', S)$ следующим образом.

Подстановки

Доказательство (продолжение).

- $V' = V \cup \bigcup\limits_{a \in \Sigma_1} V_a$;
- P' содержит продукции P_a для любого $a \in \Sigma_1$;
- P' содержит результаты подстановок продукций из P, в которых $a \in \Sigma_1$ заменяется на S_a для любого терминального символа;
- P' не содержит других продукций, кроме описанных выше.

Докажем, что $\alpha \in L(\mathfrak{G}') \Leftrightarrow \alpha \in s(L)$ для всех $\alpha \in \Sigma_2^*$.
($\Leftarrow$) Пусть $\alpha \in s(L)$; тогда существует цепочка $a_1 a_2 \ldots a_k \in L$ и $\alpha_i \in s(a_i)$, $1 \leqslant i \leqslant k$, таковы, что $\alpha = \alpha_1 \hat{}\, \alpha_2 \hat{}\, \ldots \hat{}\, \alpha_k$. Согласно конструкции, $S \Rightarrow^*_{\mathfrak{G}'} S_{a_1} S_{a_2} \ldots S_{a_k}$ и, кроме того, $S_{a_i} \Rightarrow^*_{\mathfrak{G}'} \alpha_i$. Таким образом, $S \Rightarrow^*_{\mathfrak{G}'} \alpha_1 \hat{}\, \alpha_2 \hat{}\, \ldots \hat{}\, \alpha_k = \alpha$.

Подстановки

Доказательство (окончание).

($\Rightarrow$) Пусть $\alpha \in L(\mathfrak{G}')$. Пусть T — дерево разбора для α в $\mathfrak{G}'$. Так как множества нетерминалов грамматик попарно не пересекаются, данное дерево может быть получено только как результат подстановки в дереве разбора для $\mathfrak{G}$ с кроной $a_1 a_2 \ldots a_k (\in L)$ вместо листьев a_i деревьев с корнем S_{a_i}. Пусть их крона равняется α_i. Таким образом, $\alpha = \alpha_1 \hat{\ } \alpha_2 \hat{\ } \ldots \hat{\ } \alpha_k \in s(a_1)s(a_2)\ldots s(a_k) \subseteq s(L)$. $\square$

Операции

Теорема A9.4.

Контекстно-свободные языки замкнуты относительно следующих операций:

1. объединения;
2. конкатенации;
3. звёздочки Клини и операции $\cdot \mapsto \cdot^{+}$;
4. гомоморфных образов.

Операции

Теорема A9.4.

Контекстно-свободные языки замкнуты относительно следующих операций:

1. объединения;
2. конкатенации;
3. звёздочки Клини и операции $\cdot \mapsto \cdot^{+}$;
4. гомоморфных образов.

Доказательство.

Воспользуемся теоремой A9.3.
1. Пусть L_1 и L_2 — КС-языки. Тогда $L_1 \cup L_2 = s(L)$, где $L = \{1; 2\}$ и $s(1) = L_1$, $s(2) = L_2$.
2. Пусть L_1 и L_2 — КС-языки. Тогда $L_1 L_2 = s(L)$, где $L = \{12\}$ и $s(1) = L_1$, $s(2) = L_2$.

Операции

Доказательство (окончание).

3. Пусть L_1 — КС-язык. Тогда $L_1^* = s(L)$, где $L = \{1\}^*$ и $s(1) = L_1$; $L_1^+ = s(L)$, где $L = \{1\}^+$ и $s(1) = L_1$.
4. Пусть L — КС-язык над алфавитом Σ и пусть h — гомоморфизм на Σ. Пусть также s — подстановка, осуществляющая замену символа $a \in \Sigma$ на $\{h(a)\}$; тогда $h(L) = s(L)$. □

Обращение

Теорема А4.17.

Если L — КС-язык, то и L^R также КС-язык.

Обращение

Теорема A4.17.

Если L — КС-язык, то и L^R также КС-язык.

Доказательство.

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика такова, что $L = L(\mathfrak{G})$. Тогда $L^R = L(\mathfrak{G}^R)$, где грамматика $\mathfrak{G}^R = (V, \Sigma, P', S)$ определена следующим образом: $P' = \{(A, \alpha) | P(A, \alpha^R)\}$. Достаточно только доказать, что $\alpha \in L(\mathfrak{G}) \Longrightarrow \alpha^R \in L(\mathfrak{G}^R)$ для всех $\alpha \in \Sigma^*$ (**упражнение!!!**). $\square$

Пересечение

Замечание A4.3.

КС-языки не замкнуты относительно операции пересечения.

Пересечение

Замечание А4.3.

КС-языки не замкнуты относительно операции пересечения.

Пример А4.6.

$L = \{0^n 1^n 2^n | n \in \omega\}$ не является контекстно-свободным, однако $L = L_1 \cap L_2$, где $L_1 = \{0^n 1^n 2^i | n, i \in \omega\}$ и $L_2 = \{0^i 1^n 2^n | n, i \in \omega\}$.
При этом L_1 и L_2 контекстно-свободны.

Пересечение с регулярным языком

Теорема A4.18.

Пусть L — КС-язык, а R — регулярный язык; тогда $L \cap R$ также КС-язык.

Пересечение с регулярным языком

Теорема А4.18.

Пусть L — КС-язык, а R — регулярный язык; тогда $L \cap R$ также КС-язык.

Доказательство.

Пусть $\mathcal{P} = (Q_P, \Sigma, \Gamma; \delta_P, q_P, Z_0, F_P)$ и $\mathcal{M} = (Q_M, \Sigma, \delta_M, q_M, F_M)$ — МП-автомат и ДКА соответственно таковы, что $L = L(\mathcal{P})$ и $R = L(\mathcal{M})$. Воспользуемся конструкцией произведения автоматов. А именно, определим МП-автомат $\mathcal{P}' = (Q_P \times Q_M, \Sigma, \Gamma; \delta_P \times \delta_M, \langle q_P, q_M \rangle, Z_0, F_P \times F_M)$, где $\delta_P \times \delta_M = \{((\langle q_1, q_2 \rangle, a, X), (\langle q_1', q_2' \rangle, \alpha)) | \delta_P((q_1, a, X), (q_1', \alpha)), \ \delta_M((q_2, a), q_2')\}$. Индукцией по числу переходов в МП-автоматах доказывается, что $(q_P, \beta, Z_0) \vdash_{\mathcal{P}}^* (q, \varepsilon, \gamma) \iff (\langle q_P, q_M \rangle, \beta, Z_0) \vdash_{\mathcal{P}'}^* (\langle q, \delta_M^*(q_M, \beta) \rangle, \varepsilon, \gamma)$ для всех $\beta \in \Sigma^*$ (упражнение !!!) Далее, $\beta \in L(\mathcal{P}')$, если и только если $\delta_M^*(q_M, \beta) \in F_M$ и $(q_P, \beta, Z_0) \vdash_{\mathcal{P}}^* (q, \varepsilon, \gamma)$ для некоторого $q \in F_P$ ($\Leftrightarrow \beta \in L(\mathcal{P}) \cap L(\mathcal{M})$). $\square$

Дополнение

Теорема А4.19.

Пусть L_1, L_2 и L — КС-языки, а R — регулярный язык. Тогда справедливы следующие условия:

1. $L \setminus R$ — КС-язык;
2. $\overline{L}$ может не быть КС-языком;
3. $L_1 \setminus L_2$ может и не быть КС-языком.

Дополнение

Теорема А4.19.

Пусть L_1, L_2 и L — КС-языки, а R — регулярный язык. Тогда справедливы следующие условия:

1. $L \setminus R$ — КС-язык;
2. $\overline{L}$ может не быть КС-языком;
3. $L_1 \setminus L_2$ может и не быть КС-языком.

Доказательство.

1) $L \setminus R = L \cap \overline{R}$ и $\overline{R}$ — регулярный язык.
2) Если бы КС-языки были бы замкнуты относительно дополнения, то они были бы замкнуты относительно пересечения: $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$.
3) $\overline{L} = \Sigma^* \setminus L$, при этом Σ^* и L — КС-языки. $\square$

Дополнение

Теорема А4.19.

Пусть L_1, L_2 и L — КС-языки, а R — регулярный язык. Тогда справедливы следующие условия:

1. $L \setminus R$ — КС-язык;
2. $\overline{L}$ может не быть КС-языком;
3. $L_1 \setminus L_2$ может и не быть КС-языком.

Доказательство.

1) $L \setminus R = L \cap \overline{R}$ и $\overline{R}$ — регулярный язык.
2) Если бы КС-языки были бы замкнуты относительно дополнения, то они были бы замкнуты относительно пересечения: $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$.
3) $\overline{L} = \Sigma^* \setminus L$, при этом Σ^* и L — КС-языки. □

Упражнение А4.7

Укажите явно пример КС-языка L, для которого $\overline{L}$ не является КС-языком.

Спасибо за внимание.