Лекция А4
Грамматики (доп.)

Вадим Пузаренко

16 ноября 2022 г.

ДМП-автоматы: определение

Хотя МП-автоматы по определению не детерминированы, их детерминированный случай чрезвычайно важен. В частности, синтаксические анализаторы в целом ведут себя как детерминированные МП-автоматы, поэтому класс языков, допускаемых этими автоматами, углубляет понимание конструкций, пригодных для языков программирования. Интуитивно МП-автомат является детерминированным, если в любой ситуации у него нет возможности выборов перехода. Эти выборы имеют два вида. Если $\Delta(q, a, X)$ содержит более одной пары, то МП-автомат безусловно не является детерминированным, поскольку можно выбирать из этих двух пар. Однако если $\Delta(q, a, X)$ всегда одноэлементно, все равно остаётся возможность выбора между чтением входного символа и совершением ε-перехода.

ДМП-автоматы: определение

Определение А4.1.

МП-автомат $\mathcal{P} = (Q, \Sigma, \Gamma; \Delta, q_0, Z_0, F)$ называется **детерминированным (ДМП-автоматом)**, если выполняются следующие условия:

1. $|\Delta(q, a, X)| \leqslant 1$ для каждых $q \in Q$, $a \in \Sigma \cup \{\varepsilon\}$ и $X \in \Gamma$;
2. если $\Delta(q, a, X) \neq \varnothing$ для некоторого $a \in \Sigma$, то $\Delta(q, \varepsilon, X)$ должно быть пустым.

ДМП-автоматы: определение

Определение А4.1.

МП-автомат $\mathcal{P} = (Q, \Sigma, \Gamma; \Delta, q_0, Z_0, F)$ называется **детерминированным (ДМП-автоматом)**, если выполняются следующие условия:

1. $|\Delta(q, a, X)| \leqslant 1$ для каждых $q \in Q$, $a \in \Sigma \cup \{\varepsilon\}$ и $X \in \Gamma$;
2. если $\Delta(q, a, X) \neq \varnothing$ для некоторого $a \in \Sigma$, то $\Delta(q, \varepsilon, X)$ должно быть пустым.

Примеры А4.1.

1. КС-язык $\{\alpha\hat{}\alpha^R | \alpha \in \{0; 1\}^*\}$ не распознаётся никаким ДМП-автоматом.
2. КС-язык $\{\alpha\hat{}c\hat{}\alpha^R | \alpha \in \{0; 1\}^*\}$ распознаётся некоторым ДМП-автоматом (здесь $c \notin \{0; 1\}$).

ДМП-автоматы: определение

Определение А4.1.

МП-автомат $\mathcal{P} = (Q, \Sigma, \Gamma; \Delta, q_0, Z_0, F)$ называется **детерминированным (ДМП-автоматом)**, если выполняются следующие условия:

1. $|\Delta(q, a, X)| \leqslant 1$ для каждых $q \in Q$, $a \in \Sigma \cup \{\varepsilon\}$ и $X \in \Gamma$;
2. если $\Delta(q, a, X) \neq \varnothing$ для некоторого $a \in \Sigma$, то $\Delta(q, \varepsilon, X)$ должно быть пустым.

Примеры А4.1.

1. КС-язык $\{\alpha\hat{}\alpha^R | \alpha \in \{0;1\}^*\}$ не распознаётся никаким ДМП-автоматом.
2. КС-язык $\{\alpha\hat{}c\hat{}\alpha^R | \alpha \in \{0;1\}^*\}$ распознаётся некоторым ДМП-автоматом (здесь $c \notin \{0;1\}$).

Упражнение А4.1.

Обосновать примеры А4.1.

Регулярные языки и ДМП-автоматы

Теорема А4.1.

Если L — регулярный язык, то $L = L(\mathcal{P})$ для некоторого ДМП-автомата $\mathcal{P}$.

Регулярные языки и ДМП-автоматы

Теорема А4.1.

Если L — регулярный язык, то $L = L(\mathcal{P})$ для некоторого ДМП-автомата $\mathcal{P}$.

Доказательство.

Пусть $\mathcal{A} = (Q, \Sigma; \delta, q_0, F)$ — ДКА такой, что $L = L(\mathcal{A})$. Положим ДМП-автомат $\mathcal{P} = (Q, \Sigma, \{Z_0\}; \Delta, q_0, Z_0, F)$, определив $\Delta(q, a, Z_0) = \{(\delta(q, a), Z_0)\}$ для всех $q \in Q$ и $a \in \Sigma$. Утверждается, что $(q, \alpha, Z_0) \vdash^*_{\mathcal{P}} (p, \varepsilon, Z_0) \Leftrightarrow \delta^*(q, \alpha) = p$. Доказывается в обе стороны индукцией по $\mathrm{lh}(\alpha)$. Таким образом, $\alpha \in L(\mathcal{A}) \Leftrightarrow \delta^*(q_0, \alpha) \in F \Leftrightarrow (q_0, \alpha, Z_0) \vdash^*_{\mathcal{P}} (p, \varepsilon, Z_0)$ для некоторого $p \in F \Leftrightarrow \alpha \in L(\mathcal{P})$. $\square$

ДМП-автомат и пустой магазин

Определение А4.2.

Говорят, что язык L имеет **префиксное свойство**, если в L нет двух различных цепочек α и β, где $\alpha \sqsubset_{beg} \beta$.

ДМП-автомат и пустой магазин

Определение А4.2.

Говорят, что язык L имеет **префиксное свойство**, если в L нет двух различных цепочек α и β, где $\alpha \sqsubset_{beg} \beta$.

Примеры А4.2.

1) $\{\alpha\hat{\ }c\hat{\ }\alpha^R | \alpha \in \{0;1\}^*\}$ имеет префиксное свойство.

2) $\{0\}^*$ не имеет префиксного свойства, однако является регулярным языком.

ДМП-автомат и пустой магазин

Определение А4.2.

Говорят, что язык L имеет **префиксное свойство**, если в L нет двух различных цепочек α и β, где $\alpha \sqsubset_{beg} \beta$.

Примеры А4.2.

1) $\{\alpha\hat{\ }c\hat{\ }\alpha^R | \alpha \in \{0;1\}^*\}$ имеет префиксное свойство.
2) $\{0\}^*$ не имеет префиксного свойства, однако является регулярным языком.

Теорема А4.2.

$L = N(\mathcal{P})$ для некоторого ДМП-автомата $\mathcal{P}$, если и только если L имеет префиксное свойство и $L = L(\mathcal{P}')$ для некоторого ДМП-автомата $\mathcal{P}'$.

ДМП-автомат и пустой магазин

Определение A4.2.

Говорят, что язык L имеет **префиксное свойство**, если в L нет двух различных цепочек α и β, где $\alpha \sqsubset_{beg} \beta$.

Примеры A4.2.

1) $\{\alpha\hat{\ }c\hat{\ }\alpha^R | \alpha \in \{0;1\}^*\}$ имеет префиксное свойство.
2) $\{0\}^*$ не имеет префиксного свойства, однако является регулярным языком.

Теорема A4.2.

$L = N(\mathcal{P})$ для некоторого ДМП-автомата $\mathcal{P}$, если и только если L имеет префиксное свойство и $L = L(\mathcal{P}')$ для некоторого ДМП-автомата $\mathcal{P}'$.

Упражнение A4.2.

Доказать теорему A4.2 и обосновать примеры A4.2.

ДМП-автоматы и неоднозначность

Замечание А4.1.

Языки, распознаваемые ДМП-автоматами, имеют однозначную КС-грамматику. Однако класс языков, распознаваемых ДМП-автоматами, не совпадает с классом КС-языков, не являющихся существенно неоднозначными. Например, язык $\{\alpha\hat{\ }\alpha^R | \alpha \in \{0;1\}^*\}$ имеет однозначную КС-грамматику $S \longrightarrow \varepsilon|0S0|1S1$, хотя и не распознаётся никаким ДМП-автоматом.

ДМП-автоматы и неоднозначность

Замечание А4.1.

Языки, распознаваемые ДМП-автоматами, имеют однозначную КС-грамматику. Однако класс языков, распознаваемых ДМП-автоматами, не совпадает с классом КС-языков, не являющихся существенно неоднозначными. Например, язык $\{\alpha\hat{}\alpha^R|\alpha \in \{0;1\}^*\}$ имеет однозначную КС-грамматику $S \longrightarrow \varepsilon|0S0|1S1$, хотя и не распознаётся никаким ДМП-автоматом.

Теорема А4.3.

Если $L = N(\mathcal{P})$ для некоторого ДМП-автомата $\mathcal{P}$, то L имеет однозначную КС-грамматику.

ДМП-автоматы и неоднозначность

Доказательство.

Докажем, что конструкция теоремы А3.7 по ДМП-автомату задаёт однозначную КС-грамматику. По теореме А3.3, достаточно доказать, что каждое $\alpha \in L$ имеет уникальное левое порождение. Предположим, что $\mathcal{P}$ распознаёт α по пустому магазину. Тогда он это делает с помощью единственной последовательности переходов, поскольку он детерминирован и прекращает работу, когда опустошается магазин.

Правило автомата $\mathcal{P}$, на основании которого применяется продукция, всегда одно. Но правило, скажем, $\Delta(q, a, X) = \{(r, Y_1 Y_2 \ldots Y_k)\}$, может порождать много продукций грамматики $\mathfrak{G}$, с различными состояниями в позициях, отражающих состояния $\mathcal{P}$ после удаления каждого из Y_1, Y_2, ..., Y_k. Однако, в силу детерминированности $\mathcal{P}$, осуществляется только одна из этих последовательностей переходов, поэтому только одна из этих продукций в действительности ведет к порождению α. $\square$

ДМП-автоматы и неоднозначность

Теорема A4.4.

Если $L = L(\mathcal{P})$ для некоторого ДМП-автомата $\mathcal{P}$, то L имеет однозначную КС-грамматику.

ДМП-автоматы и неоднозначность

Теорема А4.4.

Если $L = L(\mathcal{P})$ для некоторого ДМП-автомата $\mathcal{P}$, то L имеет однозначную КС-грамматику.

Доказательство.

Пусть \$ — "концевой маркер", отсутствующий в цепочках языка L и пусть $L' = L\hat{\ }\$$. Тогда L' имеет префиксное свойство, а по теореме А4.2, $L' = N(\mathcal{P}')$ для некоторого ДМП-автомата $\mathcal{P}'$. По теореме А4.3, L' имеет однозначную КС-грамматику (скажем, $\mathfrak{G}' = (V, \Sigma, P, S)$). Теперь по грамматике $\mathfrak{G}'$ построим грамматику $\mathfrak{G} = (V \cup \{\$\}, \Sigma \setminus \{\$\}, P', S)$, для которой $L = L(\mathfrak{G})$. Для этого определим $P' = P \cup \{\$ \longrightarrow \varepsilon\}$. Так как $L(\mathfrak{G}') = L'$, имеем $L(\mathfrak{G}) = L$. Докажем, что $\mathfrak{G}$ однозначна. В самом деле, левые порождения в $\mathfrak{G}$ совпадают с левыми порождениями в грамматике $\mathfrak{G}'$, за исключением последнего шага в $\mathfrak{G}$ — замены \$ на ε. Таким образом, если бы слово α имело бы два различных левых порождения в $\mathfrak{G}$, то $\alpha\hat{\ }\$$ имело бы два различных левых порождения в $\mathfrak{G}'$, противоречие. □

Основные понятия

Определение А4.3.

Говорят, что КС-грамматика $\mathfrak{G} = (V, \Sigma, P, S)$ находится в **нормальной форме Хомского (НФХ)**, если её продукции имеют вид $A \longrightarrow BC$, $A \longrightarrow a$, где $A, B, C \in V$, $a \in \Sigma$.

Основные понятия

Определение А4.3.

Говорят, что КС-грамматика $\mathfrak{G} = (V, \Sigma, P, S)$ находится в **нормальной форме Хомского (НФХ)**, если её продукции имеют вид $A \longrightarrow BC$, $A \longrightarrow a$, где $A, B, C \in V$, $a \in \Sigma$.

Основная цель.

Для любого непустого КС-языка, не содержащего ε, построить грамматику, находящуюся в НФХ.

Основные понятия

Определение А4.3.

Говорят, что КС-грамматика $\mathfrak{G} = (V, \Sigma, P, S)$ находится в **нормальной форме Хомского (НФХ)**, если её продукции имеют вид $A \longrightarrow BC$, $A \longrightarrow a$, где $A, B, C \in V$, $a \in \Sigma$.

Основная цель.

Для любого непустого КС-языка, не содержащего ε, построить грамматику, находящуюся в НФХ.

Алгоритмы.

1. Удалить бесполезные символы.
2. Удалить ε-продукции ($A \longrightarrow \varepsilon$).
3. Удалить цепные продукции ($A \longrightarrow B$).

Бесполезные символы

Определение А4.4.

Символ $X \in V \cup \Sigma$ называется **полезным** в грамматике $\mathfrak{G} = (V, \Sigma, P, S)$, если существует некоторое порождение $S \Rightarrow^* \alpha\hat{}X\hat{}\beta \Rightarrow^* \gamma$, где $\alpha, \beta, \gamma \in \Sigma^*$. Символ X называется **бесполезным**, если он не является полезным.

Бесполезные символы

Определение А4.4.

Символ $X \in V \cup \Sigma$ называется **полезным** в грамматике $\mathfrak{G} = (V, \Sigma, P, S)$, если существует некоторое порождение $S \Rightarrow^* \alpha\hat{\ }X\hat{\ }\beta \Rightarrow^* \gamma$, где $\alpha, \beta, \gamma \in \Sigma^*$. Символ X называется **бесполезным**, если он не является полезным.

Свойства полезных символов.

1. Символ X называется **порождающим**, если $X \Rightarrow^* \alpha$ для некоторого $\alpha \in \Sigma^*$. Заметим, что $X \in \Sigma \cup \{\varepsilon\}$ — порождающий символ.
2. Символ X называется **достижимым**, если $S \Rightarrow^* \alpha\hat{\ }X\hat{\ }\beta$ для некоторых $\alpha, \beta \in \Sigma^*$.

Любой полезный символ является одновременно и порождающим, и достижимым.

Удаление бесполезных символов

Теорема А4.5.

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ такова, что $L(\mathfrak{G}) \neq \varnothing$, и пусть $\mathfrak{G}_1 = (V_1, \Sigma_1, P_1, S)$ — грамматика, полученная с помощью следующих двух шагов:

1. вначале удаляются все непорождающие символы и продукции, их содержащие (в результате получим грамматику $\mathfrak{G}_2 = (V_2, \Sigma_2, P_2, S)$);
2. затем удаляются все символы, не достижимые в $\mathfrak{G}_2$.

Тогда $\mathfrak{G}_1$ не имеет бесполезных символов и $L(\mathfrak{G}) = L(\mathfrak{G}_1)$.

Удаление бесполезных символов

Теорема А4.5.

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ такова, что $L(\mathfrak{G}) \neq \varnothing$, и пусть $\mathfrak{G}_1 = (V_1, \Sigma_1, P_1, S)$ — грамматика, полученная с помощью следующих двух шагов:

1. вначале удаляются все непорождающие символы и продукции, их содержащие (в результате получим грамматику $\mathfrak{G}_2 = (V_2, \Sigma_2, P_2, S)$);
2. затем удаляются все символы, не достижимые в $\mathfrak{G}_2$.

Тогда $\mathfrak{G}_1$ не имеет бесполезных символов и $L(\mathfrak{G}) = L(\mathfrak{G}_1)$.

Доказательство.

Пусть $X \in V_1 \cup \Sigma_1$; тогда $X \Rightarrow^*_{\mathfrak{G}} \alpha$ для некоторого $\alpha \in \Sigma^*$. Кроме того, каждый символ, использованный в порождении α из X, также является порождающим. Таким образом, $X \Rightarrow^*_{\mathfrak{G}_2} \alpha$.

Удаление бесполезных символов

Доказательство (продолжение).

Так как X не был удален после второго шага, найдутся $\beta \in \Sigma^*$ и $\gamma \in \Sigma^*$, для которых $S \Rightarrow^*_{\mathfrak{G}_2} \beta\hat{}X\hat{}\gamma$. Кроме того, каждый символ, использованный в этом порождении, достижим, поэтому $S \Rightarrow^*_{\mathfrak{G}_1} \beta\hat{}X\hat{}\gamma$.

Известно, что каждый символ в цепочке $\beta\hat{}X\hat{}\gamma$ достижим, и что все эти символы принадлежат $V_2 \cup \Sigma_2$, поэтому каждый из них является порождающим в $\mathfrak{G}_2$. Порождение терминальной цепочки $\beta\hat{}X\hat{}\gamma \Rightarrow^*_{\mathfrak{G}_2} \beta\hat{}\alpha\hat{}\gamma$ содержит только символы, достижимые из S, поскольку они достижимы из символов цепочки $\beta\hat{}X\hat{}\gamma$. Таким образом, это порождение есть также порождение в $\mathfrak{G}_1$, т.е. $S \Rightarrow^*_{\mathfrak{G}_1} \beta\hat{}X\hat{}\gamma \Rightarrow^*_{\mathfrak{G}_1} \beta\hat{}\alpha\hat{}\gamma$.

Итак, X полезен в $\mathfrak{G}_1$. Ввиду произвольности X в $\mathfrak{G}_1$, заключаем, что $\mathfrak{G}_1$ не содержит бесполезных символов.

Удаление бесполезных символов

Доказательство (окончание).

Докажем теперь, что $L(\mathfrak{G}_1) = L(\mathfrak{G})$.

$(\subseteq)$ Очевидно, поскольку все символы и продукции $\mathfrak{G}_1$ входят и в $\mathfrak{G}$.

$(\supseteq)$ Пусть $\alpha \in L(\mathfrak{G})$; тогда $S \Rightarrow^*_{\mathfrak{G}} \alpha$. Следовательно, каждый символ в этом порождении является порождающим, поэтому $S \Rightarrow^*_{\mathfrak{G}_2} \alpha$. Кроме того, все символы данного порождения являются достижимыми в $\mathfrak{G}_2$ и, следовательно, $S \Rightarrow^*_{\mathfrak{G}_1} \alpha$; таким образом, $\alpha \in L(\mathfrak{G}_1)$. $\square$

Алгоритм нахождения порождающих

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика.
Базис. Если $a \in \Sigma$, то a — порождающий.
Индукция. Если $A \longrightarrow \alpha$ и цепочка α состоит только из порождающих (возможно, $\alpha = \varepsilon$), то A — порождающий.

Алгоритм нахождения порождающих

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика.
Базис. Если $a \in \Sigma$, то a — порождающий.
Индукция. Если $A \longrightarrow \alpha$ и цепочка α состоит только из порождающих (возможно, $\alpha = \varepsilon$), то A — порождающий.

Теорема А4.6.

Данный алгоритм находит в точности все порождающие грамматики $\mathfrak{G}$.

Алгоритм нахождения порождающих

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика.
Базис. Если $a \in \Sigma$, то a — порождающий.
Индукция. Если $A \longrightarrow \alpha$ и цепочка α состоит только из порождающих (возможно, $\alpha = \varepsilon$), то A — порождающий.

Теорема А4.6.

Данный алгоритм находит в точности все порождающие грамматики $\mathfrak{G}$.

Упражнение А4.3.

Докажите теорему А4.6.

Алгоритм нахождения достижимых символов

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика.
Базис. S — достижимый символ.
Индукция. Если $A \longrightarrow \alpha$ и A — достижимый символ, то все символы цепочки α также достижимы.

Алгоритм нахождения достижимых символов

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика.
Базис. S — достижимый символ.
Индукция. Если $A \longrightarrow \alpha$ и A — достижимый символ, то все символы цепочки α также достижимы.

Теорема А4.7.

Данный алгоритм находит в точности все достижимые символы грамматики $\mathfrak{G}$.

Алгоритм нахождения достижимых символов

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика.
Базис. S — достижимый символ.
Индукция. Если $A \longrightarrow \alpha$ и A — достижимый символ, то все символы цепочки α также достижимы.

Теорема А4.7.

Данный алгоритм находит в точности все достижимые символы грамматики $\mathfrak{G}$.

Упражнение А4.4.

Докажите теорему А4.7.

Удаление ε-продукций

Определение A4.4.

Переменная A называется **ε-порождающей** в $\mathfrak{G}$, если $A \Rightarrow^*_{\mathfrak{G}} \varepsilon$.

Удаление ε-продукций

Определение А4.4.

Переменная A называется ***ε-порождающей*** в $\mathfrak{G}$, если $A \Rightarrow^*_{\mathfrak{G}} \varepsilon$.

Алгоритм.

Базис. Если $A \longrightarrow \varepsilon$ — продукция в $\mathfrak{G}$, то A — ε-порождающий.

Индукция. Если в $\mathfrak{G}$ есть продукция $B \longrightarrow C_1 C_2 \ldots C_k$, в которой C_i, $i = 1, \ldots, k$ — ε-порождающие, то и B также ε-порождающий.

Удаление ε-продукций

Определение А4.4.

Переменная A называется **ε-порождающей** в $\mathfrak{G}$, если $A \Rightarrow^*_{\mathfrak{G}} \varepsilon$.

Алгоритм.

Базис. Если $A \longrightarrow \varepsilon$ — продукция в $\mathfrak{G}$, то A — ε-порождающий.
Индукция. Если в $\mathfrak{G}$ есть продукция $B \longrightarrow C_1 C_2 \dots C_k$, в которой C_i, $i = 1, \dots, k$ — ε-порождающие, то и B также ε-порождающий.

Теорема А4.8.

В грамматике $\mathfrak{G}$ являются в точности ε-порождающими переменные, найденные вышеприведённым алгоритмом.

Удаление ε-продукций

Определение А4.4.

Переменная A называется **ε-порождающей** в $\mathfrak{G}$, если $A \Rightarrow^*_{\mathfrak{G}} \varepsilon$.

Алгоритм.

Базис. Если $A \longrightarrow \varepsilon$ — продукция в $\mathfrak{G}$, то A — ε-порождающий.
Индукция. Если в $\mathfrak{G}$ есть продукция $B \longrightarrow C_1 C_2 \dots C_k$, в которой C_i, $i = 1, \dots, k$ — ε-порождающие, то и B также ε-порождающий.

Теорема А4.8.

В грамматике $\mathfrak{G}$ являются в точности ε-порождающими переменные, найденные вышеприведённым алгоритмом.

Упражнение А4.5.

Докажите теорему А4.8.

Удаление ε-продукций

Конструкция.

1. Удаляем все продукции вида $A \longrightarrow \varepsilon$.
2. Если $A \longrightarrow \alpha\hat{\ }B\hat{\ }\gamma$ — продукция, в которой B — ε-порождающий, а $\mathrm{lh}(\alpha\hat{\ }\gamma) > 0$, то добавляем продукцию $A \longrightarrow \alpha\hat{\ }\gamma$. Повторяем данный пункт, пока это возможно.

Удаление ε-продукций

Конструкция.

1. Удаляем все продукции вида $A \longrightarrow \varepsilon$.
2. Если $A \longrightarrow \alpha\hat{}B\hat{}\gamma$ — продукция, в которой B — ε-порождающий, а $\mathrm{lh}(\alpha\hat{}\gamma) > 0$, то добавляем продукцию $A \longrightarrow \alpha\hat{}\gamma$. Повторяем данный пункт, пока это возможно.

Теорема А4.9.

Если грамматика $\mathfrak{G}_1$ построена по грамматике $\mathfrak{G}$ с помощью описанной выше конструкции удаления ε-продукций, то $L(\mathfrak{G}_1) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Удаление ε-продукций

Конструкция.

1. Удаляем все продукции вида $A \longrightarrow \varepsilon$.
2. Если $A \longrightarrow \alpha\hat{\ }B\hat{\ }\gamma$ — продукция, в которой B — ε-порождающий, а $\mathrm{lh}(\alpha\hat{\ }\gamma) > 0$, то добавляем продукцию $A \longrightarrow \alpha\hat{\ }\gamma$. Повторяем данный пункт, пока это возможно.

Теорема А4.9.

Если грамматика $\mathfrak{G}_1$ построена по грамматике $\mathfrak{G}$ с помощью описанной выше конструкции удаления ε-продукций, то $L(\mathfrak{G}_1) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Доказательство.

Необходимо доказать, что $\alpha \in L(\mathfrak{G}_1)$, если и только если $\alpha \in L(\mathfrak{G})$, для любого $\alpha \in \Sigma^* \setminus \{\varepsilon\}$. Докажем более общее утверждение $(\alpha \in (V \cup \Sigma)^*)$: $A \Rightarrow^*_{\mathfrak{G}_1} \alpha \Longleftrightarrow [A \Rightarrow^*_{\mathfrak{G}} \alpha \& (\alpha \neq \varepsilon)]$.

Удаление ε-продукций

Доказательство (продолжение).

($\Rightarrow$) Пусть $A \Rightarrow^*_{\mathfrak{G}_1} \alpha$; тогда $\alpha \neq \varepsilon$, поскольку $\mathfrak{G}_1$ не имеет ε-продукций. Покажем индукцией по длине порождения, что $A \Rightarrow^*_{\mathfrak{G}} \alpha$.

Базис. В $\mathfrak{G}_1$ имеется продукция $A \longrightarrow \alpha$; согласно конструкции, $A \longrightarrow \alpha'$ — продукция в $\mathfrak{G}$, где α получается из α' удалением ε-порождающих переменных. Следовательно, $A \Rightarrow_{\mathfrak{G}} \alpha' \Rightarrow^*_{\mathfrak{G}} \alpha$.

Индукция. Пусть в порождении $A \Rightarrow^*_{\mathfrak{G}_1} \alpha$ имеется $n > 1$ шагов. Тогда оно имеет вид $A \Rightarrow_{\mathfrak{G}_1} X_1X_2\ldots X_k \Rightarrow^*_{\mathfrak{G}_1} \alpha$. Цепочку α можно представить в виде $\alpha_1\hat{\ }\alpha_2\hat{\ }\ldots\hat{\ }\alpha_k$, где $X_i \Rightarrow^*_{\mathfrak{G}_1} \alpha_i$. По предположению индукции, $X_i \Rightarrow^*_{\mathfrak{G}} \alpha_i$. Согласно конструкции, в $\mathfrak{G}$ имеется продукция $A \longrightarrow Y_1Y_2\ldots Y_m$, где $X_1X_2\ldots X_k$ получена из $Y_1Y_2\ldots Y_m$ удалением ε-порождающих символов. Таким образом,

$A \Rightarrow_{\mathfrak{G}} Y_1Y_2\ldots Y_m \Rightarrow^*_{\mathfrak{G}} X_1X_2\ldots X_k \Rightarrow^*_{\mathfrak{G}} \alpha_1\hat{\ }\alpha_2\hat{\ }\ldots\hat{\ }\alpha_k = \alpha$.

Удаление ε-продукций

Доказательство (окончание).

($\Leftarrow$) Пусть $A \Rightarrow^*_{\mathfrak{G}} \alpha$ и $\alpha \neq \varepsilon$; как и прежде, будем доказывать индукцией по длине порождения.

Базис. $A \longrightarrow \alpha$ — продукция в $\mathfrak{G}$; так как $\alpha \neq \varepsilon$, $A \longrightarrow \alpha$ будет продукцией и в $\mathfrak{G}_1$. В частности, $A \Rightarrow^*_{\mathfrak{G}_1} \alpha$.

Индукция. Пусть в порождении $A \Rightarrow^*_{\mathfrak{G}} \alpha$ имеется $n > 1$ шагов. Тогда оно имеет вид $A \Rightarrow_{\mathfrak{G}} Y_1 Y_2 \ldots Y_m \Rightarrow^*_{\mathfrak{G}} \alpha$. Цепочку α можно представить в виде $\alpha_1\hat{\ }\alpha_2\hat{\ }\ldots\hat{\ }\alpha_m$, где $Y_i \Rightarrow^*_{\mathfrak{G}} \alpha_i$. Пусть цепочка $X_1 X_2 \ldots X_k$ получена из $Y_1 Y_2 \ldots Y_m$ удалением Y_j таких, что $\alpha_j = \varepsilon$. По предположению индукции, $Y_i \Rightarrow^*_{\mathfrak{G}_1} \alpha_i$. Согласно конструкции, $A \longrightarrow X_1 X_2 \ldots X_k$ — продукция в $\mathfrak{G}_1$. Таким образом, $A \Rightarrow_{\mathfrak{G}_1} X_1 X_2 \ldots X_k \Rightarrow^*_{\mathfrak{G}_1} \alpha_1\hat{\ }\alpha_2\hat{\ }\ldots\hat{\ }\alpha_m = \alpha$.

Для завершения доказательства осталось положить $A = S$. $\square$

Удаление цепных продукций

Определение A4.5.

Пара нетерминалов (A, B) назовём **цепной**, если $A \Rightarrow^* B$, причём в порождении используются только цепные продукции.

Удаление цепных продукций

Определение А4.5.

Пара нетерминалов (A, B) назовём **цепной**, если $A \Rightarrow^* B$, причём в порождении используются только цепные продукции.

Алгоритм.

1. (A, A) — цепная пара.
2. Если (A, B) — цепная пара и $B \longrightarrow C$ — цепная продукция, то (A, C) — цепная пара.

Удаление цепных продукций

Определение А4.5.

Пара нетерминалов (A, B) назовём **цепной**, если $A \Rightarrow^* B$, причём в порождении используются только цепные продукции.

Алгоритм.

1. (A, A) — цепная пара.
2. Если (A, B) — цепная пара и $B \longrightarrow C$ — цепная продукция, то (A, C) — цепная пара.

Теорема А4.10.

Вышеприведённый алгоритм находит в точности все цепные пары грамматики $\mathfrak{G}$.

Удаление цепных продукций

Определение А4.5.

Пара нетерминалов (A, B) назовём **цепной**, если $A \Rightarrow^* B$, причём в порождении используются только цепные продукции.

Алгоритм.

1. (A, A) — цепная пара.
2. Если (A, B) — цепная пара и $B \longrightarrow C$ — цепная продукция, то (A, C) — цепная пара.

Теорема А4.10.

Вышеприведённый алгоритм находит в точности все цепные пары грамматики $\mathfrak{G}$.

Упражнение А4.6.

Докажите теорему А4.10.

Удаление цепных продукций

Конструкция.

1. Найдём все цепные пары грамматики $\mathfrak{G}$.
2. Для каждой цепной пары (A, B) добавляем продукцию $A \longrightarrow \alpha$ в P_1, если $B \longrightarrow \alpha$ — нецепная продукция в P.

Удаление цепных продукций

Конструкция.

1. Найдём все цепные пары грамматики $\mathfrak{G}$.
2. Для каждой цепной пары (A, B) добавляем продукцию $A \longrightarrow \alpha$ в P_1, если $B \longrightarrow \alpha$ — нецепная продукция в P.

Замечание А4.2.

Заметим, что в случае, когда $A = B$, в P_1 помещаются все нецепные продукции из P.

Удаление цепных продукций

Конструкция.

1. Найдём все цепные пары грамматики $\mathfrak{G}$.
2. Для каждой цепной пары (A, B) добавляем продукцию $A \longrightarrow \alpha$ в P_1, если $B \longrightarrow \alpha$ — нецепная продукция в P.

Замечание А4.2.

Заметим, что в случае, когда $A = B$, в P_1 помещаются все нецепные продукции из P.

Теорема А4.11.

Если $\mathfrak{G}_1$ построена по $\mathfrak{G}$ согласно конструкции, описанной выше, то $L(\mathfrak{G}_1) = L(\mathfrak{G})$.

Удаление цепных продукций

Доказательство.

Докажем, что $\alpha \in L(\mathfrak{G}) \Leftrightarrow \alpha \in L(\mathfrak{G}_1)$.

Удаление цепных продукций

Доказательство.

Докажем, что $\alpha \in L(\mathfrak{G}) \Leftrightarrow \alpha \in L(\mathfrak{G}_1)$.
$(\Leftarrow)$ Пусть $S \Rightarrow^*_{\mathfrak{G}_1} \alpha$. Так как каждая продукция в $\mathfrak{G}_1$ эквивалентна последовательности из нескольких цепных продукций, за которыми следует одна нецепная продукция из $\mathfrak{G}$, имеем $S \Rightarrow^*_{\mathfrak{G}} \alpha$.

Удаление цепных продукций

Доказательство.

Докажем, что $\alpha \in L(\mathfrak{G}) \Leftrightarrow \alpha \in L(\mathfrak{G}_1)$.

($\Leftarrow$) Пусть $S \Rightarrow^*_{\mathfrak{G}_1} \alpha$. Так как каждая продукция в $\mathfrak{G}_1$ эквивалентна последовательности из нескольких цепных продукций, за которыми следует одна нецепная продукция из $\mathfrak{G}$, имеем $S \Rightarrow^*_{\mathfrak{G}} \alpha$.

($\Rightarrow$) Пусть $\alpha \in L(\mathfrak{G})$; по теореме А3.2, $S \underset{l}{\Rightarrow}^* \alpha$. Где бы в левом порождении ни использовалась цепная продукция, переменная её тела остаётся крайней слева. Тем самым, левое порождение в $\mathfrak{G}$ можно разбить на последовательность "шагов", в которых несколько цепных продукций сопровождаются нецепной. Заметим, что любая нецепная продукция, перед которой нет цепных, сама по себе образует такой "шаг". Но по построению грамматики $\mathfrak{G}_1$, каждый из этих шагов может быть выполнен одной её продукцией. Таким образом, $S \Rightarrow^*_{\mathfrak{G}_1} \alpha$. $\square$

Вспомогательные конструкции

Теорема А4.12.

Пусть $\mathfrak{G}$ — КС-грамматика, у которой $L(\mathfrak{G}) \setminus \{\varepsilon\} \neq \varnothing$. Тогда можно построить КС-грамматику $\mathfrak{G}_1$, в которой отсутствуют бесполезные символы, ε-продукции и цепные продукции, такую что $L(\mathfrak{G}_1) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Вспомогательные конструкции

Теорема А4.12.

Пусть $\mathfrak{G}$ — КС-грамматика, у которой $L(\mathfrak{G}) \setminus \{\varepsilon\} \neq \varnothing$. Тогда можно построить КС-грамматику $\mathfrak{G}_1$, в которой отсутствуют бесполезные символы, ε-продукции и цепные продукции, такую что $L(\mathfrak{G}_1) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Доказательство.

Сначала, по теореме А4.9, удалим ε-продукции; затем удалим цепные продукции, по теореме А4.11 (заметим, что в этом случае будут отсутствовать также и ε-продукции); в конечном итоге, применим конструкцию теоремы А4.5 (поскольку все продукции полученной грамматики содержатся во множестве продукций, полученной на предыдущем шаге, грамматика также не будет содержать ε- и цепных продукций). $\square$

Нормальная форма

Конструкция.

Опишем один шаг конструкции, в котором в грамматике происходят изменения только с одной продукцией. Пусть $A \longrightarrow X_1X_2\ldots X_k$, где либо $k > 2$, либо $k = 2$, но X_1 и X_2 не являются одновременно переменными. Для каждого слова X_1 и $X_2\ldots X_k$ вводим новую переменную Y_1 и Y_2, если оно не является нетерминалом. Возможны следующие случаи:

1. X_1 является нетерминалом; тогда рассматриваемую продукцию заменяем на следующий список: $A \longrightarrow X_1Y_2$, $Y_2 \longrightarrow X_2\ldots X_k$;
2. $X_2\ldots X_k = X_2$ является нетерминалом; тогда $A \longrightarrow Y_1X_2$, $Y_1 \longrightarrow X_1$;
3. X_1 является терминалом и $k > 2$; тогда $A \longrightarrow Y_1Y_2$, $Y_1 \longrightarrow X_1$, $Y_2 \longrightarrow X_2\ldots X_k$.

Конструкция завершится на конечном шаге (почему?)

Нормальная форма

Теорема А4.13.

Если $\mathfrak{G}$ — КС-грамматика, порождающая хотя бы одну непустую цепочку, то существует НФХ $\mathfrak{G}_1$ такая, что $L(\mathfrak{G}_1) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Нормальная форма

Теорема A4.13.

Если $\mathfrak{G}$ — КС-грамматика, порождающая хотя бы одну непустую цепочку, то существует НФХ $\mathfrak{G}_1$ такая, что $L(\mathfrak{G}_1) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Доказательство.

По теореме A4.12, можно построить КС-грамматику $\mathfrak{G}_2 = (V_2, \Sigma, P_2, S)$, свободную от бесполезных символов, ε-продукций и цепных продукций, для которой $L(\mathfrak{G}_2) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.
Докажем, что $\alpha \in L(\mathfrak{G}_2) \Leftrightarrow \alpha \in L(\mathfrak{G}_1)$, где $\mathfrak{G}_1 = (V_1, \Sigma, P_1, S)$ строится по грамматике $\mathfrak{G}_2$ согласно конструкции.
$(\Rightarrow)$ Непосредственно вытекает из того, что $P_2(A, X_1X_2\dots X_k)$ влечёт $A \Rightarrow^*_{\mathfrak{G}_1} X_1X_2\dots X_k$.
$(\Leftarrow)$ Доказывать будем по длине вывода следующую импликацию: $A \Rightarrow^*_{l,\mathfrak{G}_1} \alpha \Longrightarrow A \Rightarrow^*_{l,\mathfrak{G}_2} \alpha$, для всех $A \in V_2$ и $\alpha \in \Sigma^*$.

Нормальная форма

Доказательство (продолжение).

Базис. Если $P_1(A, \alpha)$ и $A \in V_2$, то $\alpha \in \Sigma$ и $P_2(A, \alpha)$, поскольку $\mathfrak{G}_2$ не имеет цепных правил и ε-продукций.

Индукция. Пусть $A \Rightarrow^*_{l,\mathfrak{G}_1} \alpha$, и данное порождение имеет длину $n > 1$. Тогда $A \longrightarrow BC \Rightarrow^*_{l,\mathfrak{G}_1} \alpha$ и, следовательно, $B \Rightarrow^*_{l,\mathfrak{G}_1} \beta$, $C \Rightarrow^*_{l,\mathfrak{G}_1} \gamma$, причём $\alpha = \beta\hat{}\gamma$. Разберём несколько случаев.

- $B \notin V_2$; согласно конструкции, $P_1(B, \beta)$, $\beta \in \Sigma$ и $P_2(A, \beta X_2 \ldots X_k)$, причём эта продукция задаётся единственным образом. В частности, $A \Rightarrow^*_{l,\mathfrak{G}_2} \beta X_2 \ldots X_k$.
- $B \in V_2$; согласно конструкции, $P_2(A, BX_2 \ldots X_k)$, причём эта продукция задаётся единственным образом; по предположению индукции, $B \Rightarrow^*_{l,\mathfrak{G}_2} \beta$. В частности, $A \Rightarrow^*_{l,\mathfrak{G}_2} \beta X_2 \ldots X_k$.

Докажем теперь индукцией по k, что существует последовательность цепочек $\gamma_2, \ldots, \gamma_k$ такая, что $\gamma = \gamma_2\hat{}\ldots\hat{}\gamma_k$ и $X_i \Rightarrow^*_{l,\mathfrak{G}_2} \gamma_i$, $2 \leqslant i \leqslant k$.

Нормальная форма

Доказательство (окончание).

Базис. Возможны несколько случаев.

- $C \in V_2$; по индукционному предположению, $C = X_2 \Rightarrow^*_{l,\mathfrak{G}_2} \gamma$.
- $P_1(C, \gamma)$; тогда $\gamma \in \Sigma$ и, согласно конструкции, $P_2(A, B\gamma)$; в частности, $X_2 = \gamma \Rightarrow^*_{l,\mathfrak{G}_2} \gamma$.

Индукция. Пусть $P_1(C, C_1C_2)$. Разберём снова несколько случаев.

- $C_1 \in V_2$; тогда $X_2 = C_1 \Rightarrow^*_{l,\mathfrak{G}_2} \gamma_2$, а по предположению индукции, найдутся цепочки $\gamma_3, \ldots, \gamma_k$, удовлетворяющие условию.
- $C_1 \notin V_2$; тогда $P_2(C_1, \gamma_2)$, $\gamma_2 \in \Sigma$ и $X_2 = \gamma_2 \Rightarrow^*_{l,\mathfrak{G}_2} \gamma_2$; по предположению индукции, найдутся цепочки $\gamma_3, \ldots, \gamma_k$, удовлетворяющие условию.

Для завершения доказательства заметим, что $S \in V_2$ и, по доказанному, $\alpha \in L(\mathfrak{G}_2) \Leftrightarrow S \Rightarrow^*_{l,\mathfrak{G}_2} \alpha \Leftrightarrow S \Rightarrow^*_{l,\mathfrak{G}_1} \alpha \Leftrightarrow \alpha \in L(\mathfrak{G}_1)$. □

Высота дерева разбора

Предложение А4.1.

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 А4.1.

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 А4.14.

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 А4.14.

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 А4.13, 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 А4.1, 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 А4.1, 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 $\alpha\hat{}\beta^{l+1}\hat{}\gamma\hat{}\delta^{l+1}\hat{}\eta$. □

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 А4.3.

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 A4.4.

Пусть $L = \{0^i 1^j 2^i 3^j \mid i, j \in \omega\}$. 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 А4.5.

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 А4.5 (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 А4.3.

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 А4.3.

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

Теорема А4.16.

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

Однобуквенный алфавит

Замечание А4.3.

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

Теорема А4.16.

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

Доказательство.

Без ограничения общности можно считать, что L — бесконечный язык. По теореме А4.15 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 А1.2), L_0 является таковым. По теореме А2.4, 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 $L_1 = \bigcup\limits_{\zeta \in L_1,\, 0 < d \leqslant n_0} L_{\zeta,d}$.

($\subseteq$) 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, $\phi \in L_{\phi,d_0} \subseteq \bigcup\limits_{\zeta \in L_1,\, 0 < d \leqslant n_0} L_{\zeta,d}$.

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 $S_j^d = \{\zeta \in L_1 \mid \mathrm{lh}(\zeta) \equiv j (\bmod d),\ L_{\zeta,d} \neq \varnothing\}$, 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 А4.15.

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 $\alpha = a_1 a_2 \dots a_k \in \Sigma_1^*$, 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)).

Теорема A4.15.

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 А4.16.

Контекстно-свободные языки замкнуты относительно следующих операций:

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

Операции

Теорема А4.16.

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 А4.15.

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 А4.17.

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 А4.3.

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 А4.18.

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. □

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.