Лекции по теории формальных языков

Лекция 4.
Контекстно-свободные грамматики и языки: определения и примеры.
Лемма о накачке

Александр Сергеевич Герасимов
http://gas-teach.narod.ru

Кафедра математических и информационных технологий
Санкт-Петербургского академического университета
Российской академии наук.
Весенний семестр 2010/11 учебного года

4 марта 2011 г.

План

1. Контекстно-свободные грамматики и языки: определения и примеры
2. Лемма о накачке

План

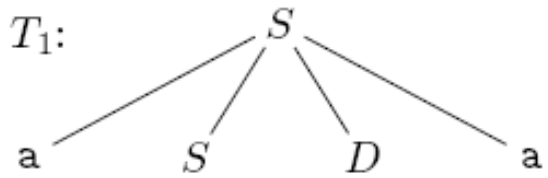
Определения КС-грамматики и КС-языка. Примеры

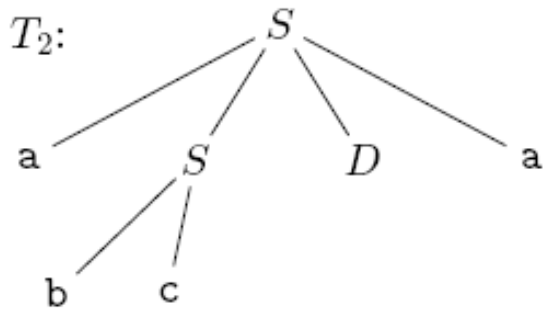
- Грамматика называется *контекстно-свободной* (*КС-грамматикой*), если каждое её правило имеет вид $A \to \alpha$.
- Язык называется *контекстно-свободным* (*КС-языком*), если некоторая КС-грамматика его порождает.
- Если $G = (\Sigma, \Gamma, P, S)$ — грамматика, $S \Rightarrow_G^* \beta \Rightarrow_G^* w$, то цепочку β будем называть *формой*.
- $\{S \to \mathrm{a}S\mathrm{b} | \mathrm{ab}\}$ — КС-грамматика, она порождает КС-язык $\{\mathrm{a}^n\mathrm{b}^n | n \geqslant 1\}$.
- Язык $\{\mathrm{a}^n\mathrm{b}^n\mathrm{a}^m | m, n \geqslant 1\}$ контекстно-свободен, поскольку порождается КС-грамматикой $\{S \to CD,\ C \to \mathrm{a}C\mathrm{b} | \mathrm{ab},\ D \to \mathrm{a}D | \mathrm{a}\}$.
- Грамматика $\{S \to \mathrm{a}SD\mathrm{a},\ S \to \mathrm{aba},\ \mathrm{a}D \to D\mathrm{a},\ \mathrm{b}D \to \mathrm{bb}\}$ не являющаяся контекстно-свободной, порождает язык $\{\mathrm{a}^n\mathrm{b}^n\mathrm{a}^n \mid n \geqslant 1\}$ (см. лекцию 3). Существует ли КС-грамматика, порождающая этой язык? Ответ — в текущей лекции.

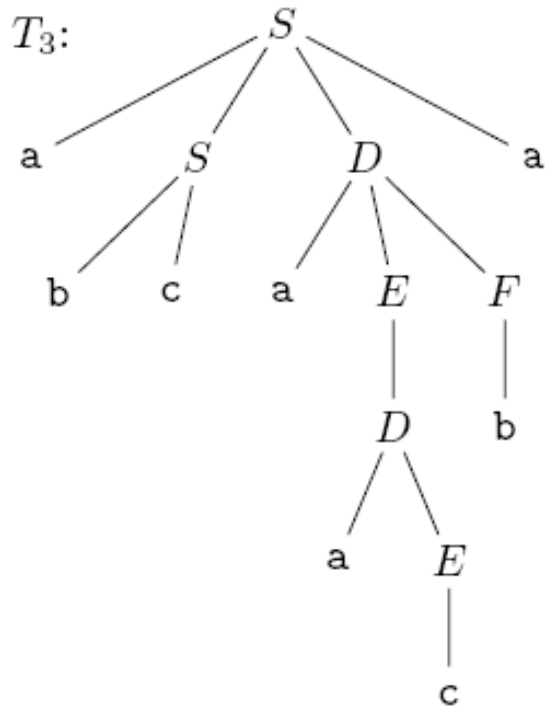
Дерево вывода: пример

- Вывод в КС-грамматике удобно представлять в виде дерева.
- Для примера рассмотрим КС-грамматику G_1 с правилами $\{S \to \mathtt{a}S D\mathtt{a} | \mathtt{bc},\ D \to \mathtt{a}E | \mathtt{a}EF,\ E \to D | \mathtt{c},\ F \to \mathtt{b}\}$ и вывод

$$\underline{S} \Rightarrow \mathtt{a}\underline{S}D\mathtt{a} \Rightarrow \mathtt{abc}\underline{D}\mathtt{a} \Rightarrow \mathtt{abca}\underline{E}F\mathtt{a} \Rightarrow \mathtt{abca}\underline{D}F\mathtt{a} \Rightarrow$$
$$\mathtt{abcaa}\underline{E}F\mathtt{a} \Rightarrow \mathtt{abcaac}\underline{F}\mathtt{a} \Rightarrow \mathtt{abcaacba} = w\ .$$







Дерево вывода: определение

Деревом вывода (или *деревом разбора*) цепочки $w \in \Sigma^*$ в грамматике $G = (\Sigma, \Gamma, P, S)$ называется такое упорядоченное дерево, что

- его корень помечен S;
- если его внутрений узел помечен $A \in \Gamma$ и $X_1, \ldots, X_k \in \Sigma \cup \Gamma$ — перечисленные слева направо пометки всех сыновей этого внутреннего узла, то $A \to X_1 \ldots X_k \in P$;
- если его внутрений узел помечен $A \in \Gamma$ и ε — пометка единственного сына этого внутреннего узла, то $A \to \varepsilon \in P$;
- $w = \mathrm{a}_1 \ldots \mathrm{a}_m$, где $\mathrm{a}_1, \ldots, \mathrm{a}_m \in \Sigma \cup \{\varepsilon\}$ — перечисленные слева направо пометки всех листьев этого дерева.

На слайде 5 дерево T_3 — дерево вывода цепочки w.

Стандартное поддерево дерева вывода. Представление вывода деревом вывода

- Поддерево T' дерева вывода T называется *стандартным*, если корень T является корнем T', и для любого узла x дерева T' либо x является листом T', либо все сыновья узла x в дереве T также лежат в T'.
- Например, на слайде 5 деревья T_1 и T_2 — стандартные поддеревья дерева вывода T_3.
- Будем говорить, что вывод $S \Rightarrow \alpha_1 \Rightarrow \ldots \Rightarrow \alpha_n = w$ в грамматике G *представлен* деревом вывода T, если существуют такие стандартные поддеревья $T_1 \subset \ldots \subset T_n$ дерева T, что для каждого $i = 1, \ldots, n$ конкатенация пометок перечисленных слева направо листьев поддерева T_i есть форма α_i.
  Также мы будем говорить, что вывод формы α_i *представлен* деревом T_i.
- Например, вывод на слайде 5 представлен деревом вывода T_3.

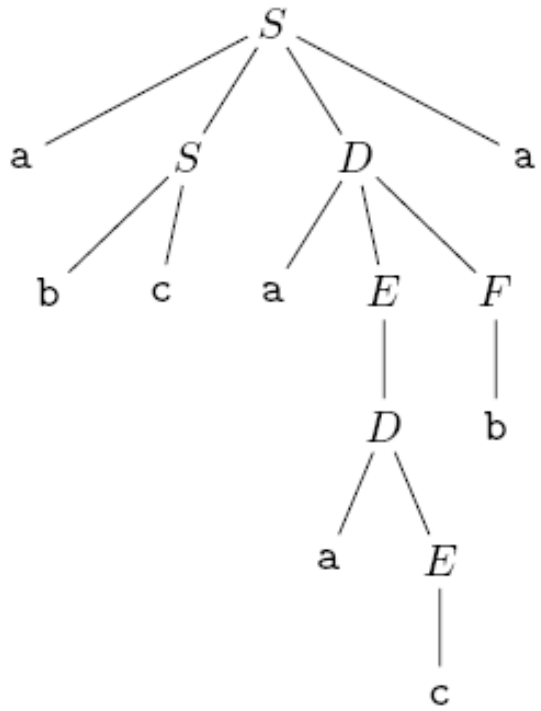
Разные выводы одной цепочки I

Продолжаем рассматривать КС-грамматику G_1 с правилами $\{S \to \mathrm{a}S D\mathrm{a} | \mathrm{bc},\ D \to \mathrm{a}E | \mathrm{a}EF,\ E \to D | \mathrm{c},\ F \to \mathrm{b}\}$ со слайда 5.

Вывод

$$(1)\quad \underline{S} \Rightarrow \mathrm{a}\underline{S}D\mathrm{a} \Rightarrow \mathrm{abc}\underline{D}\mathrm{a} \Rightarrow \mathrm{abca}\underline{E}F\mathrm{a} \Rightarrow \mathrm{abca}\underline{D}F\mathrm{a} \Rightarrow$$
$$\mathrm{abcaa}\underline{E}F\mathrm{a} \Rightarrow \mathrm{abcaac}\underline{F}\mathrm{a} \Rightarrow \mathrm{abcaacba} = w$$

представлен следующим деревом вывода:



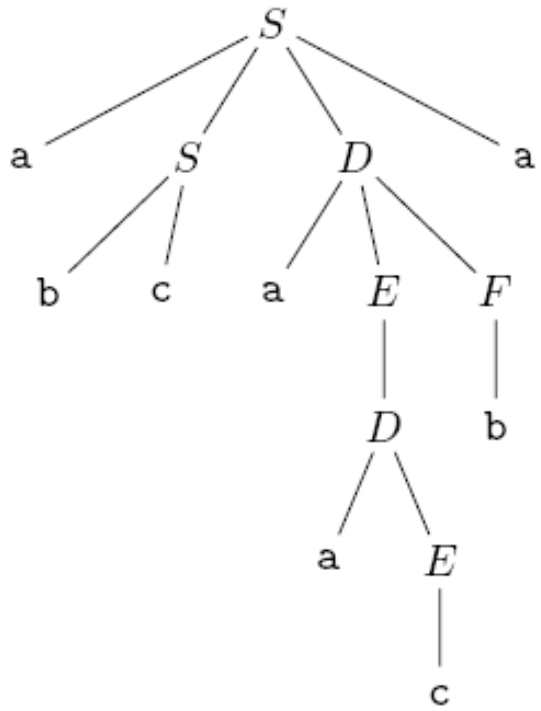
Разные выводы одной цепочки II

Продолжаем рассматривать КС-грамматику G_1 с правилами $\{S \to \mathrm{a}S D\mathrm{a} | \mathrm{bc},\ D \to \mathrm{a}E | \mathrm{a}EF,\ E \to D | \mathrm{c},\ F \to \mathrm{b}\}$ со слайда 5.

Вывод

$$(2)\quad \underline{S} \Rightarrow \mathrm{a}S\underline{D}\mathrm{a} \Rightarrow \mathrm{a}S\mathrm{a}E\underline{F}\mathrm{a} \Rightarrow \mathrm{a}S\mathrm{a}\underline{E}\mathrm{ba} \Rightarrow \mathrm{a}S\mathrm{a}\underline{D}\mathrm{ba} \Rightarrow$$
$$\mathrm{a}S\mathrm{aa}\underline{E}\mathrm{ba} \Rightarrow \mathrm{a}\underline{S}\mathrm{aacba} \Rightarrow \mathrm{abcaacba} = w$$

представлен тем же деревом вывода, что и на слайде 8:



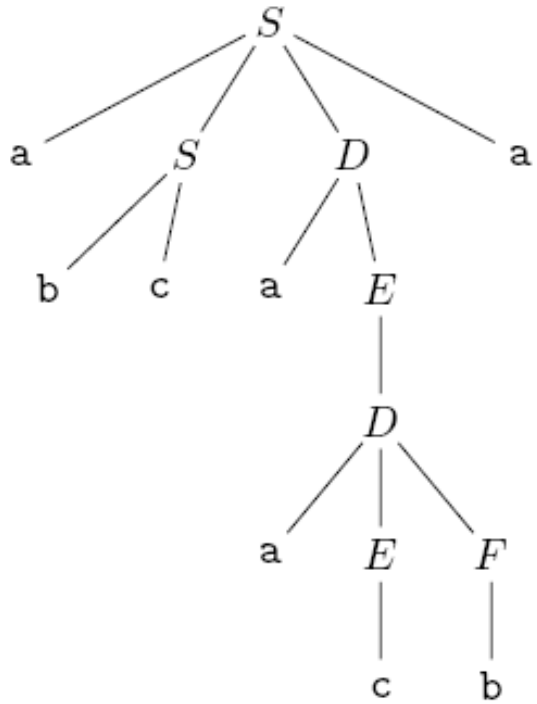
Разные выводы одной цепочки III

Рассматриваем КС-грамматику G_1 с правилами $\{S \to \mathrm{a}S D\mathrm{a} | \mathrm{bc},\ D \to \mathrm{a}E | \mathrm{a}EF,\ E \to D | \mathrm{c},\ F \to \mathrm{b}\}$ со слайда 5.

Вывод

$$(3)\quad \underline{S} \Rightarrow \mathrm{a}\underline{S}D\mathrm{a} \Rightarrow \mathrm{abc}\underline{D}\mathrm{a} \Rightarrow \mathrm{abca}\underline{E}\mathrm{a} \Rightarrow \mathrm{abca}\underline{D}\mathrm{a} \Rightarrow$$
$$\mathrm{abcaa}\underline{E}F\mathrm{a} \Rightarrow \mathrm{abcaac}\underline{F}\mathrm{a} \Rightarrow \mathrm{abcaacba} = w$$

представлен следующим деревом вывода:

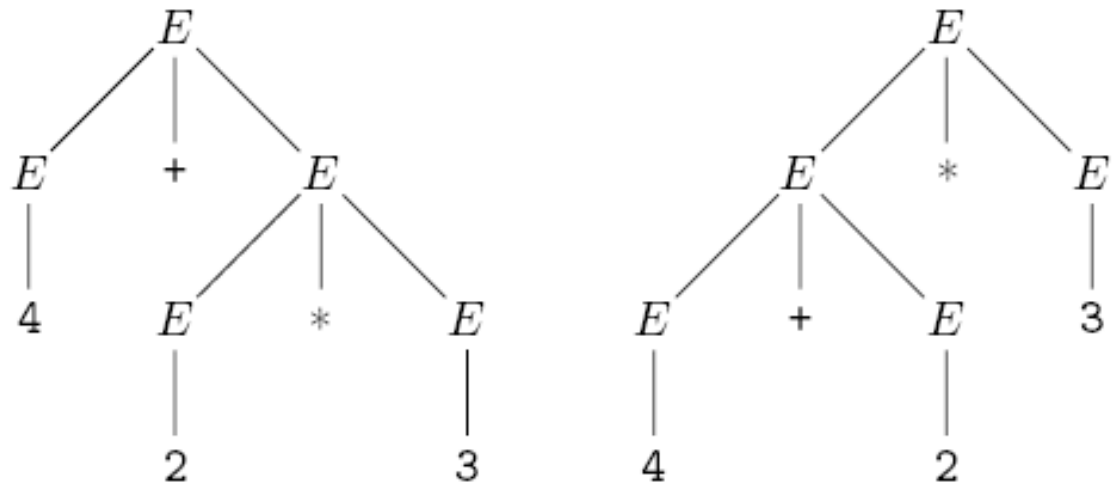


Левый и правый выводы

- Вывод в КС-грамматике называется *левым*, или *левосторонним*, (соответственно, *правым*, или *правосторонним*), если правило применяется к самому левому (соответственно, правому) вхождению нетерминала в каждую форму этого вывода.
- Произвольную форму из левого (соответственно, правого) вывода будем называть *l-формой* (соответственно, *r-формой*).
- Например, вывод (1) на слайде 8 и вывод (3) на слайде 10 являются левыми, а вывод (2) на слайде 9 — правым.
- Любое дерево вывода цепочки представляет единственный левый и единственный правый вывод этой цепочки.
- Цепочка имеет два или более различных дерева вывода тогда и только тогда, когда эта цепочка имеет два или более различных левых (соответственно, правых) вывода.

Однозначные и неоднозначные грамматики

- КС-грамматика G называется *однозначной*, если каждая цепочка из $L(G)$ имеет единственное дерево вывода. В противном случае КС-грамматика G называется *неоднозначной*.
- КС-грамматика G однозначна тогда и только тогда, когда каждая цепочка из $L(G)$ имеет единственный левый вывод.
- Рассматривавшаяся выше КС-грамматика G_1 неоднозначна.
- КС-грамматика $\{E \to E + E | E * E | 0 | 1 | \ldots | 9\}$ неоднозначна:



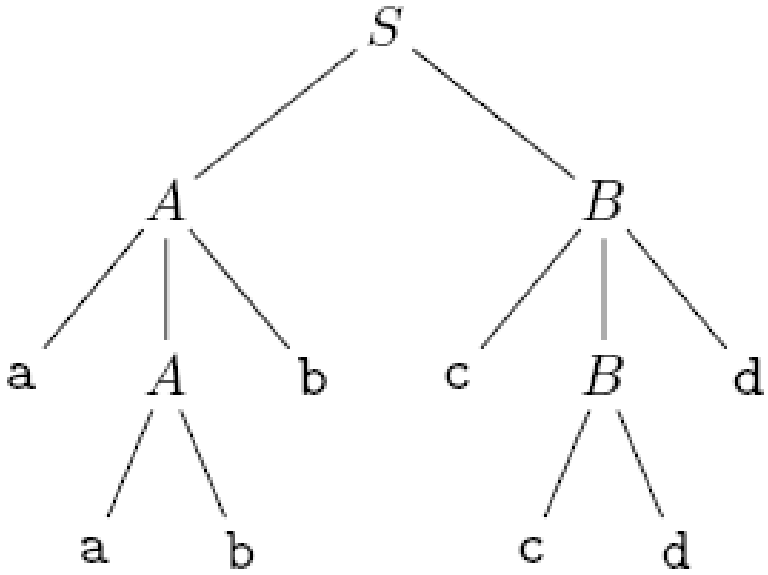
- Желательно, чтобы грамматика, определяющая язык программирования, была однозначной. Иначе смысл некоторых программ можно понимать по-разному.

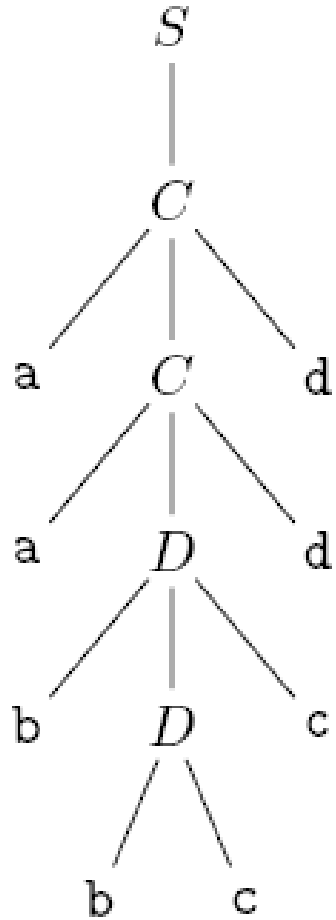
Неразрешимость проблемы однозначности. Однозначные и неоднозначные языки

- Доказано, что не существует алгоритма, определяющего по произвольной КС-грамматике, однозначна она или нет.
- КС-язык называется *однозначным*, если некоторая однозначная КС-грамматика порождает его. В противном случае КС-язык называется *неоднозначным*.
- Установлено, что КС-язык $L = \{\mathtt{a}^m\mathtt{b}^m\mathtt{c}^n\mathtt{d}^n \mid m, n \geqslant 1\} \cup \{\mathtt{a}^m\mathtt{b}^n\mathtt{c}^n\mathtt{d}^m \mid m, n \geqslant 1\}$ неоднозначен.
- КС-язык L порождается, например, КС-грамматикой с правилами $S \to AB|C$, $A \to \mathtt{a}A\mathtt{b}|\mathtt{ab}$, $B \to \mathtt{c}B\mathtt{d}|\mathtt{cd}$, $C \to \mathtt{a}C\mathtt{d}|\mathtt{a}D\mathtt{d}$, $D \to \mathtt{b}D\mathtt{c}|\mathtt{bc}$.

Два дерева вывода одной цепочки

Цепочка $\mathrm{a}^2\mathrm{b}^2\mathrm{c}^2\mathrm{d}^2 \in L$ (см. предыдущий слайд) имеет два дерева вывода в грамматике с правилами $S \to AB|C$, $A \to \mathrm{a}A\mathrm{b}|\mathrm{ab}$, $B \to \mathrm{c}B\mathrm{d}|\mathrm{cd}$, $C \to \mathrm{a}C\mathrm{d}|\mathrm{a}D\mathrm{d}$, $D \to \mathrm{b}D\mathrm{c}|\mathrm{bc}$:





КС-грамматики, порождающие скобочный язык LB

- Алфавит языка LB: $\{[,\ ]\}$.
  Определение языка LB:
  - $\varepsilon \in LB$;
  - если $u \in LB$ и $v \in LB$, то $[u] \in LB$ и $uv \in LB$.
- Граматика $GB_1 = \{S \to \varepsilon \mid [S] \mid SS\}$ неоднозначна, поскольку цепочка ε имеет два различных левых вывода:
  - $S \Rightarrow \varepsilon$ и
  - $S \Rightarrow SS \Rightarrow \varepsilon S \Rightarrow \varepsilon\varepsilon = \varepsilon$.
- Граматика $GB_2 = \{S \to \varepsilon \mid S[S]\}$ однозначна, но содержит правило вида $A \to A\alpha$ (так называемое леворекурсивное правило).
- Граматика $GB_3 = \{S \to \varepsilon \mid [S]S\}$ однозначна.

КС-грамматики, порождающие язык списков LL

- Алфавит языка LL: $\{\mathrm{a},\ ;,\ [,\ ]\}$.
  Определение языка LL:
  - $\mathrm{a} \in LL$;
  - если $u \in LL$ и $v \in LL$, то $[u] \in LL$ и $u; v \in LL$.
- Грамматика $GL_1 = \{S \to \mathrm{a} \mid [S] \mid S; S\}$ неоднозначна, поскольку цепочка $\mathrm{a}; \mathrm{a}; \mathrm{a}$ имеет два различных левых вывода:
  - $S \Rightarrow S; S \Rightarrow \mathrm{a}; S \Rightarrow \mathrm{a}; S; S \Rightarrow \mathrm{a}; \mathrm{a}; S \Rightarrow \mathrm{a}; \mathrm{a}; \mathrm{a}$ и
  - $S \Rightarrow S; S \Rightarrow S; S; S \Rightarrow \mathrm{a}; S; S \Rightarrow \mathrm{a}; \mathrm{a}; S \Rightarrow \mathrm{a}; \mathrm{a}; \mathrm{a}$.
- Граматика $GL_2 = \{S \to S; L \mid L,\quad L \to \mathrm{a} \mid [S]\}$ однозначна.
- Граматика $GL_3 = \{S \to L; S \mid L,\quad L \to \mathrm{a} \mid [S]\}$ однозначна.

КС-грамматика, порождающая язык описаний типов LD

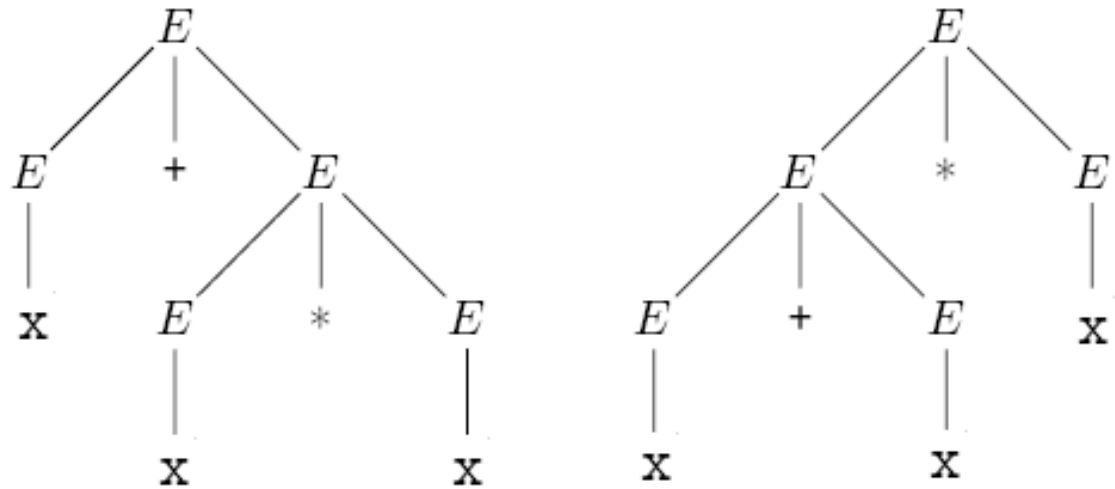
- Алфавит языка LD: $\{\texttt{i},\ :,\ ;,\ \texttt{int},\ \texttt{real}\}$.
  Каждая цепочка языка LD имеет вид:
  - непустой список букв `i`, разделённых точкой с запятой,
  - затем двоеточие,
  - наконец, один из символов `int` или `real`.
- Грамматика $GD = \{D \to L : T,\quad L \to L; \texttt{i} \mid \texttt{i},\quad T \to \texttt{int} \mid \texttt{real}\}$.

КС-грамматики, порождающие языки двоичных чисел LN и LN'

- Алфавит языка LN: $\{0,\ 1,\ .\}$.
  Каждая цепочка языка LN имеет вид
  - u (где $u \in \{0,\ 1\}^+$),
  - $.v$ (где $v \in \{0,\ 1\}^+$) или
  - $u.v$ (где $u, v \in \{0,\ 1\}^+$).
- $GN = \{S \to L \mid .L \mid L.L,\quad L \to B \mid LB,\quad B \to 0 \mid 1\}$.
- Язык LN' получаем, если дополнительно потребуем, чтобы целая часть (u) числа начиналась с 1, дробная часть (v) заканчивалась на 1.
- $GN' = \{S \to L \mid .R \mid L.R,\quad L \to 1 \mid L0 \mid L1,\quad R \to 0R \mid 1R \mid 1\}$.

КС-грамматики, порождающие язык арифметических выражений LA

- Алфавит языка LA: $\{\mathtt{x},\ +,\ *,\ (,\ )\}$.
  Определение языка LA:
  - $\mathtt{x} \in LA$;
  - если $u \in LA$ и $v \in LA$, то $(u) \in LA$, $u + v \in LA$ и $u * v \in LA$.
- Грамматика $GA_1 = \{E \to E + E \mid E * E \mid (E) \mid \mathtt{x}\}$ неоднозначна:



- Грамматика
  $GA_2 = \{E \to T \mid E + T,\quad T \to F \mid T * F,\quad F \to (E) \mid \mathtt{x}\}$
  однозначна, она учитывает естественный приоритет операций и их левую ассоциативность.

План

Лемма о накачке

Лемма (о накачке)

Для любого КС-языка L существуют числа n и m такие, что каждая цепочка $w \in L$ при $|w| > n$ представима в виде $w = xuzvy$, где

- $uv \neq \varepsilon$,
- $|uzv| \leqslant m$,
- $xu^k zv^k y \in L$ *для любого* $k \geqslant 0$.

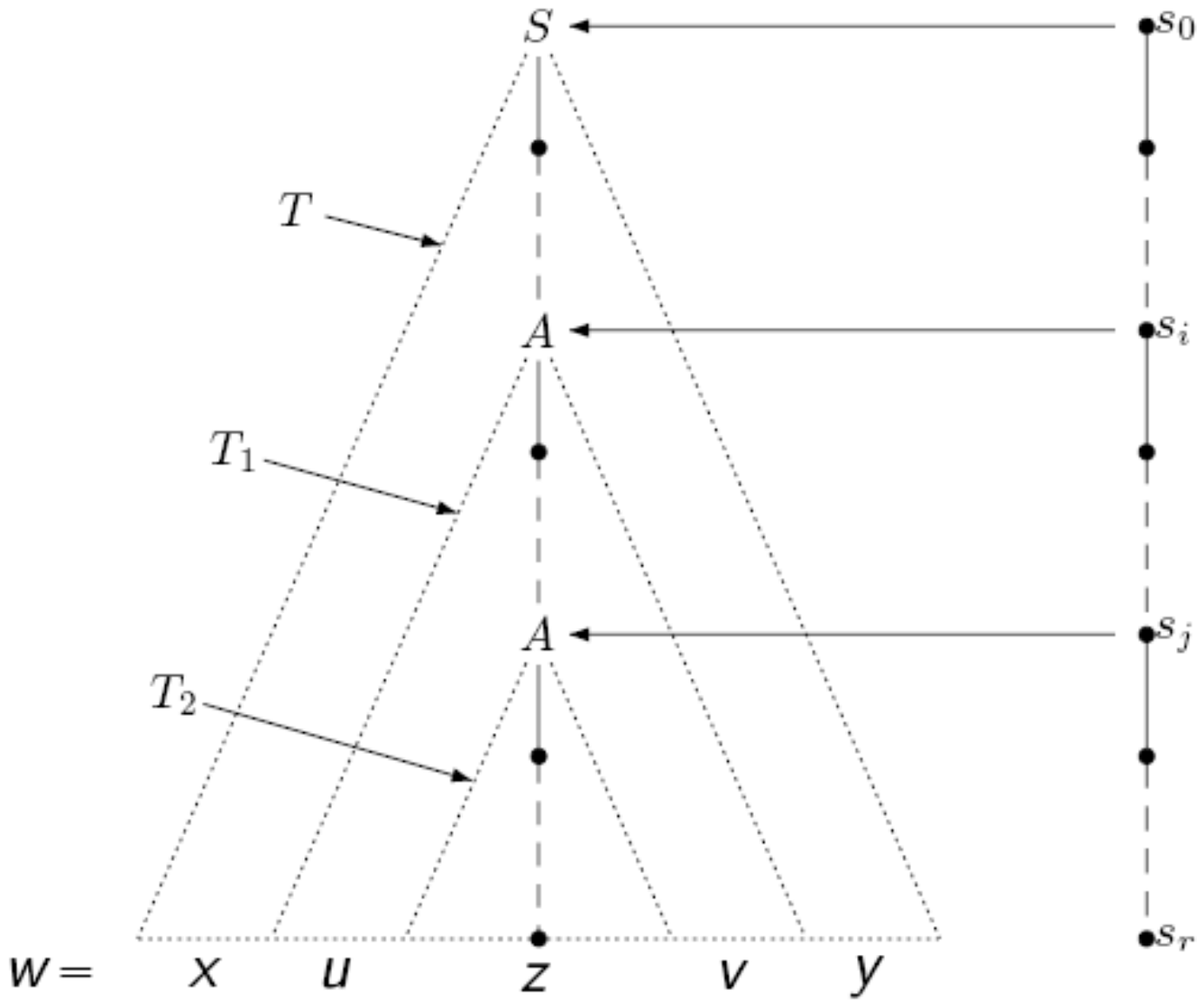
Доказательство.

- Для $L = \emptyset$ лемма верна. Далее считаем, что $L \neq \emptyset$.
- Пусть КС-грамматика $G = (\Sigma, \Gamma, P, S)$ порождает язык L.
- Для каждой цепочки $w \in L$ можно выбрать её дерево вывода в G с минимальным числом узлов; такое дерево вывода мы будем называть минимальным.

Лемма о накачке: продолжение доказательства

- Пусть n — максимальная длина цепочки с минимальным деревом вывода высоты не более $|\Gamma|$. (Хотя бы одна такая цепочка существует, так как $L \neq \emptyset$.)
- Рассмотрим цепочку $w \in L$, $|w| > n$, а также минимальное дерево вывода T цепочки w.
- Высота дерева T больше $|\Gamma|$, поэтому в длиннейшем пути от корня до листа найдутся 2 различных узла, помеченные одним и тем же нетерминалом.
- Пусть этот путь имеет вид $s_0, \ldots, s_i, \ldots, s_j, \ldots, s_r$, а узлы s_i и s_j $(i < j)$ помечены нетерминалом A. Можно считать, что $r - i \leqslant |\Gamma| + 1$.
- Введём обозначения следующих поддеревьев дерева T:
  - T_1, состоящее из узла s_i и всех его потомков,
  - T_2, состоящее из узла s_j и всех его потомков,
  - T', получаемое из T удалением всех потомков узла s_i,
  - T'_1, получаемое из T_1 удалением всех потомков узла s_j.

Лемма о накачке: продолжение доказательства



Лемма о накачке: продолжение доказательства

- Дерево T представляет вывод $S \Rightarrow^* w$.
- Стандартное поддерево T' дерева T представляет вывод $S \Rightarrow^* xAy$.
- Поддерево T_1 с корнем s_i дерева T представляет вывод $A \Rightarrow^* uzv$.
- Стандартное поддерево T_1' дерева T_1 представляет вывод $A \Rightarrow^* uAv$.
- Поддерево T_2 с корнем s_j дерева T представляет вывод $A \Rightarrow^* z$.
- Докажем, что представление $w = xuzvy$ — искомое.
- $S \Rightarrow^* xAy \Rightarrow^* xuAvy \Rightarrow^* xu^2Av^2y \Rightarrow^* \ldots \Rightarrow^* xu^kAv^ky \Rightarrow^* xu^kzv^ky$, так что $xu^kzv^ky \in L$ для любого $k \geqslant 0$.
- Если бы $uv = \varepsilon$, то дерево T_1' представляло бы вывод $A \Rightarrow^* A$. Тогда, заменив в дереве вывода T поддерево T_1 на T_2, мы получили бы дерево вывода цепочки w с меньшим, чем в T, числом узлов; это противоречило бы минимальности T. Следовательно, $uv \neq \varepsilon$.

Лемма о накачке: окончание доказательства

- Положим $m = M^{|\Gamma|+1}$, где M — наибольшая длина правой части правила из P, и покажем, что $|uzv| \leqslant m$.
- Высота дерева T_1 равна $(r - i)$. Поэтому в T_1 не более M^{r-i} листьев.
- Таким образом, $|uzv| \leqslant M^{r-i} \leqslant M^{|\Gamma|+1} = m$.

Язык $L_1 = \{a^n b^n a^n \mid n \geqslant 1\}$ не является КС-языком

- Предположим, что L_1 является КС-языком.
- Тогда по лемме о накачке для некоторого m верно $a^m b^m a^m = xuzvy$, $uv \neq \varepsilon$ и $xu^2zv^2y \in L_1$.
- Мы будем говорить, что цепочка $a^n b^n a^n$ состоит из трёх блоков a^n, b^n и a^n.
- Если u или v пересекается с двумя блоками, то в цепочке xu^2zv^2y более двух перемен букв. Поэтому как u, так и v входит лишь в один блок.
- Но тогда в xu^2zv^2y один из блоков короче другого. Противоречие.

□

Язык $L_2 = \{wcw \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$ не является КС-языком: начало доказательства

- Предположим, что L_2 является КС-языком.
- В силу леммы о накачке цепочка $w_1 = \mathtt{a}^m\mathtt{b}^m\mathtt{c}\mathtt{a}^m\mathtt{b}^m \in L_2$ при некотором m представима в виде $w_1 = xuzvy$, где $uv \neq \varepsilon$, $|uzv| \leqslant m$ и $w_2 = xu^2zv^2y \in L_2$.
- Буква c не входит ни в u, ни в v, так как иначе эта буква будет входить в w_2 дважды.
- Цепочки u и v располагаются в w_1 по разные стороны от буквы c, поскольку в противном случае с одной стороны от c в цепочке w_2 будет больше букв, чем с другой.
- Теперь очевидно, что буква c входит в z.
- Поскольку $|uzv| \leqslant m$, то
  - u входит в подцепочку $\mathtt{b}^m$ цепочки w_1 слева от буквы c,
  - v входит в подцепочку $\mathtt{a}^m$ цепочки w_1 справа от буквы c.

Язык $L_2 = \{wcw \mid w \in \{a, b\}^*\}$ не является КС-языком: окончание доказательства

- Тогда
  (1) при $u \neq \varepsilon$ в w_2 число вхождений b слева от c больше числа вхождений b справа от c,
  (2) при $v \neq \varepsilon$ в w_2 число вхождений a справа от c больше числа вхождений a слева от c.

  Так как мы имеем $uv \neq \varepsilon$ и в каждом из случаев (1) и (2) заключаем, что $w_2 \notin L_2$, то мы получили противоречие.

Теорема о языках в однобуквенных алфавитах и периодических множествах

Множество $M \subseteq \{0, 1, 2, \ldots\}$ называется *периодическим*, если существуют целые положительные числа n_0 (*индекс*) и d (*период*) такие, что для любого $n \geqslant n_0$ условие $n \in M$ влечёт $n + d \in M$.

Теорема

Для произвольного языка L над алфавитом $\{\mathrm{a}\}$ следующие утверждения эквивалентны:

(1) *L — КС-язык;*

(2) *L — регулярный язык;*

(3) *$M = \{n \geqslant 0 \mid \mathrm{a}^n \in L\}$ — периодическое множество.*

Теорема: начало доказательства

(1) влечёт (3).

- По лемме о накачке для КС-языка L существуют числа n и m такие, что каждая цепочка $w \in L$ при $|w| > n$ представима в виде $w = xuzvy$, где
  - $uv \neq \varepsilon$,
  - $|uzv| \leqslant m$,
  - $xu^k zv^k y \in L$ для любого $k \geqslant 0$.
- Пусть $j = |uv|$; очевидно, $0 < j \leqslant m$. Для любой цепочки $w \in L$ при $|w| > n$ имеем $w = xzy\mathrm{a}^j$ и $xu^{k'+1}zv^{k'+1}y = w\mathrm{a}^{jk'} \in L$ для любого $k' \geqslant 0$.
- Положим $n_0 = n + 1$ и $d = m!$. Тогда для любой цепочки $w \in L$ при $|w| \geqslant n_0$ выполняется $w\mathrm{a}^d \in L$.
- Следовательно, n_0 и d можно взять в качестве индекса и периода множества M соответственно.

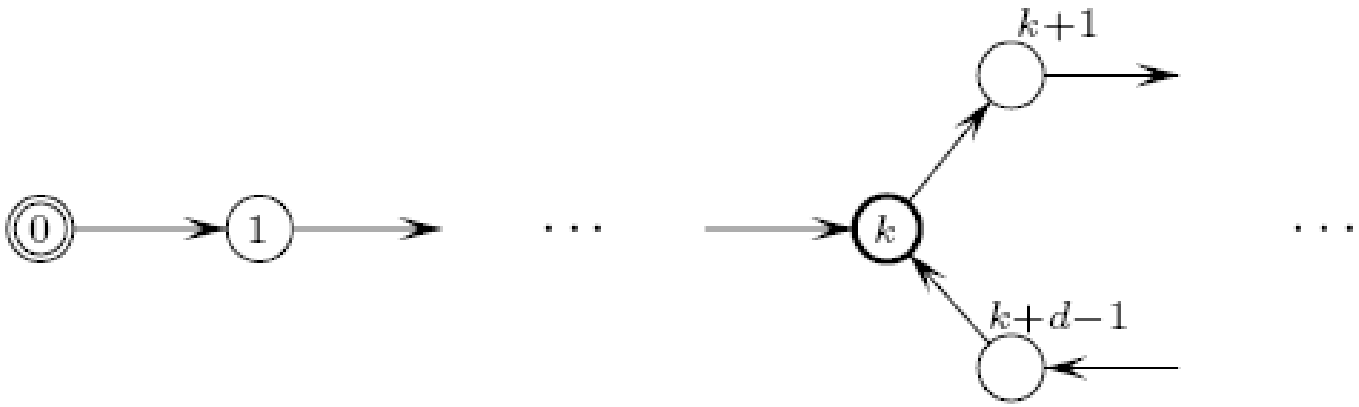
Теорема: продолжение доказательства

(3) влечёт (2).

- Если множество M конечно, то язык L конечен и потому регулярен. Далее считаем, что множество M бесконечно.
- Пусть n_0 и d являются соответственно индексом и периодом множества M.
- Для каждого i такого, что $0 \leqslant i < d$, определим (если оно существует) минимальное число k_i, удовлетворяющее условиям $k_i \in M$, $k_i \geqslant n_0$ и $k_i \equiv i \pmod{d}$. Хотя бы одно из k_i определено, поскольку M бесконечно.
- Обозначим через k наибольшее из всех определённых k_i.

Теорема: продолжение доказательства

- Построим ДКА $\mathcal{A}$, распознающий язык L. На диаграмме переходов автомата $\mathcal{A}$ каждое ребро помечено символом a.



- Из начального состояния 0 до состояния n при $n < k$ читается только цепочка a^n, а при $n \geqslant k$ — множество всех цепочек вида a^{n+md} для каждого $m \geqslant 0$.
- Каждое состояние n такое, что $a^n \in L$, объявим заключительным.
- Таким образом, $L = L(\mathcal{A})$, значит, язык L является регулярным.

Теорема: окончание доказательства

(2) влечёт (1). Индукция по числу операций (объединения, произведения и итерации) в регулярном выражении, обозначающем язык L.

- База индукции. Языки $\emptyset$, $\{\varepsilon\}$ и $\{a\}$, обозначаемые регулярными выражениями без операций, контекстно-свободны.
- Индукционный переход. Пусть языки L_1 и L_2 порождаются КС-грамматиками $G_1 = (\Sigma_1, \Gamma_1, P_1, S_1)$ и $G_2 = (\Sigma_2, \Gamma_2, P_2, S_2)$. Можно считать, что $\Gamma_1 \cap \Gamma_2 = \emptyset$. Достаточно построить КС-грамматики, порождающие языки $L_1 \cup L_2$, $L_1 L_2$ и L_1^*. Выберем новый нетерминал $S \notin \Gamma_1 \cup \Gamma_2$.
  - $L_1 \cup L_2$ порождается КС-грамматикой с правилами $\{S \to S_1 | S_2\} \cup P_1 \cup P_2$.
  - $L_1 L_2$ порождается КС-грамматикой с правилами $\{S \to S_1 S_2\} \cup P_1 \cup P_2$.
  - L_1^* порождается КС-грамматикой с правилами $\{S \to S_1 S | \varepsilon\} \cup P_1$.

Литература

Основная литература

- Замятин А. П., Шур А. М. Языки, грамматики, распознаватели: Учебное пособие. Екатеринбург : Изд-во Урал. ун-та, 2007 (электронный вариант книги — на `http://elar.usu.ru`, поиск).

Дополнительная литература

- Ахо А., Лам М., Сети Р., Ульман Дж. Компиляторы: принципы, технологии и инструментарий. М.: ООО "И.Д. Вильямс", 2008.
- Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. М.: Мир, 1978.
- Мартыненко Б. К. Языки и трансляции: Учеб. пособие. СПб.: Издательство С.-Петербургского университета, 2004 (электронный вариант книги — на `http://www.math.spbu.ru/user/mbk`).