Лекции по теории формальных языков

Лекция 8.
Синтаксический анализ для LL(1)-грамматик

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

1 апреля 2011 г.

План

1. LL(1)-грамматики
2. Синтаксический анализ для LL(1)-грамматик
3. Вычисление множеств выбора правил (начало)

План

1. LL(1)-грамматики
2. Синтаксический анализ для LL(1)-грамматик
3. Вычисление множеств выбора правил (начало)

Наводящие соображения

- Далее считаем все рассматриваемые грамматики приведёнными, а выводы — левыми, если не будет оговорено иное.
- Расширим класс разделённых грамматик, сохранив возможность моделирования (Д)МП-автоматом левого вывода.
- Пусть G — грамматика с аксиомой S, $w = \mathrm{a}_1 \dots \mathrm{a}_n \in L(G)$.
- Тогда существует левый вывод цепочки w:

$$S = \alpha_0 \Rightarrow \alpha_1 \Rightarrow \dots \Rightarrow \alpha_m = w.$$

- Пусть в ходе синтаксического анализа цепочки w прочитан её префикс $\mathrm{a}_1 \dots \mathrm{a}_j$ и получена $\alpha_i = \mathrm{a}_1 \dots \mathrm{a}_j A\beta$.
- Когда можно однозначно определить правило (для перехода $\alpha_i \Rightarrow \alpha_{i+1}$) по $\mathrm{a}_1 \dots \mathrm{a}_j$, A и следующему входному символу a_{j+1}?

Функция FIRST

- Пусть $G = (\Sigma, \Gamma, P, S)$ — КС-грамматика, $\alpha \in (\Sigma \cup \Gamma)^*$. Тогда

$$\textsc{first}(\alpha) = \{\mathtt{a} \in \Sigma \mid \exists \beta \ (\alpha \Rightarrow^* \mathtt{a}\beta)\} \cup \{\varepsilon \mid \alpha \Rightarrow^* \varepsilon\}.$$

- Пример. Грамматика G_1: $S \to AC$, $A \to \mathtt{ab}C | \mathtt{b}B$, $B \to \mathtt{b}$, $C \to \mathtt{c}|\varepsilon$.
    - $\textsc{first}(\mathtt{ab}) = \{\mathtt{a}\}$;
    - $\textsc{first}(\varepsilon) = \{\varepsilon\}$;
    - $\textsc{first}(\mathtt{b}AC) = \{\mathtt{b}\}$;
    - $\textsc{first}(A\mathtt{c}C) = \{\mathtt{a}, \mathtt{b}\}$;
    - $\textsc{first}(C) = \{\mathtt{c}, \varepsilon\}$;
    - $\textsc{first}(CA) = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$.
- Расширим определение функции FIRST. Для КС-грамматики $G = (\Sigma, \Gamma, P, S)$ и $L \subseteq (\Sigma \cup \Gamma)^*$ положим

$$\textsc{first}(L) = \{u \mid \exists \alpha \in L \ (u \in \textsc{first}(\alpha))\}.$$

Функция FOLLOW

- Пусть $G = (\Sigma, \Gamma, P, S)$ — КС-грамматика, $A \in \Gamma$, $\dashv \notin \Sigma \cup \Gamma$. Тогда

  $\mathrm{FOLLOW}(A) = \{\mathtt{a} \in \Sigma \mid \exists \alpha, \beta \; (S \Rightarrow^* \alpha A \mathtt{a} \beta)\} \cup \{\dashv \mid \exists \alpha \; (S \Rightarrow^* \alpha A)\}$,

  где выводы необязательно левые.
- Пример. Грамматика G_1: $S \to AC$, $A \to \mathtt{ab}C | \mathtt{b}B$, $B \to \mathtt{b}$, $C \to \mathtt{c} | \varepsilon$.
  - $\mathrm{FOLLOW}(A) = \{\mathtt{c}, \dashv\}$;
  - $\mathrm{FOLLOW}(B) = \{\mathtt{c}, \dashv\}$;
  - $\mathrm{FOLLOW}(C) = \{\mathtt{c}, \dashv\}$;
  - $\mathrm{FOLLOW}(S) = \{\dashv\}$.

Определения множества выбора правила и LL(1)-грамматики

- Пусть $A \to \alpha$ — правило вывода КС-грамматики. Тогда *множеством выбора* этого правила называется множество

$$\textsc{select}(A \to \alpha) = \textsc{first}(\alpha\textsc{follow}(A)).$$

- Ясно, что

$$\textsc{select}(A \to \alpha) = \begin{cases} \textsc{first}(\alpha), & \text{если } \varepsilon \notin \textsc{first}(\alpha), \\ (\textsc{first}(\alpha) \setminus \{\varepsilon\}) \cup \textsc{follow}(A), & \text{если } \varepsilon \in \textsc{first}(\alpha). \end{cases}$$

- КС-грамматика называется *LL(1)-грамматикой*, если для любого нетерминала A и любых его различных альтернатив α и β выполняется

$$\textsc{select}(A \to \alpha) \cap \textsc{select}(A \to \beta) = \emptyset.$$

Множества выбора правил: пример

Грамматика GA_3 (порождающая язык арифметических выражений):

- $E \to TE', \quad E' \to +TE' | \varepsilon,$
- $T \to FT', \quad T' \to *FT' | \varepsilon,$
- $F \to (E) \mid \mathtt{x}.$

Правило	Множество выбора
$F \to (E)$	$\{(\}$
$F \to \mathtt{x}$	$\{\mathtt{x}\}$
$T' \to *FT'$	$\{*\}$
$E' \to +TE'$	$\{+\}$
$T \to FT'$	$\{(, \mathtt{x}\}$
$E \to TE'$	$\{(, \mathtt{x}\}$
$E' \to \varepsilon$	$\textsc{Follow}(E') = \{\dashv,)\}$
$T' \to \varepsilon$	$\textsc{Follow}(T') = \{\dashv, +,)\}$

GA_3 является LL(1)-грамматикой.

Грамматика, не являющаяся LL(1)-грамматикой

Грамматика GA_2 (порождающая язык арифметических выражений):

- $E \to T \mid E + T$,
- $T \to F \mid T * F$,
- $F \to (E) \mid \mathtt{x}$.

GA_2 не является LL(1)-грамматикой, поскольку

$$\textsc{select}(E \to T) = \{(, \mathtt{x}\} = \textsc{select}(E \to E + T).$$

Теорема о равносильных определениях LL(1)-грамматики

Теорема

КС-грамматика G с аксиомой S является LL(1)-грамматикой тогда и только тогда, когда из существования двух выводов

(1) $S \Rightarrow^* wA\alpha \Rightarrow w\beta\alpha \Rightarrow^* wx$ *и*

(2) $S \Rightarrow^* wA\alpha \Rightarrow w\gamma\alpha \Rightarrow^* wy$,

для которых $\mathrm{FIRST}(x) = \mathrm{FIRST}(y)$, *следует, что* $\beta = \gamma$.

Д о к а з а т е л ь с т в о.

- Пусть G является LL(1)-грамматикой. Рассмотрим выводы (1) и (2), для которых $\mathrm{FIRST}(x) = \mathrm{FIRST}(y)$. Но предположим, что $\beta \neq \gamma$.

Теорема о равносильных определениях LL(1)-грамматики: окончание доказательства

- Так как $\beta\alpha \Rightarrow^* x$, то $\mathrm{FIRST}(x) \subseteq \mathrm{FIRST}(\beta\alpha)$. Аналогично $\mathrm{FIRST}(y) \subseteq \mathrm{FIRST}(\gamma\alpha)$.
- Тогда

$$\mathrm{FIRST}(x) = \mathrm{FIRST}(y) \subseteq \mathrm{FIRST}(\beta\alpha) \cap \mathrm{FIRST}(\gamma\alpha) \neq \emptyset.$$

- Вспоминая, что $\mathrm{SELECT}(A \to \beta) = \mathrm{FIRST}(\beta\mathrm{FOLLOW}(A))$, $\mathrm{SELECT}(A \to \gamma) = \mathrm{FIRST}(\gamma\mathrm{FOLLOW}(A))$ и $S \Rightarrow^* wA\alpha$, получаем

$$\mathrm{SELECT}(A \to \beta) \cap \mathrm{SELECT}(A \to \gamma) \neq \emptyset.$$

  Противоречие с тем, что G является LL(1)-грамматикой.
- Доказательство в обратную сторону — упражнение.

$\square$

Теорема о непринадлежности леворекурсивной грамматики классу LL(1)-грамматик

Теорема

Никакая леворекурсивная КС-грамматика не является LL(1)-грамматикой.

Д о к а з а т е л ь с т в о.

- Пусть G — леворекурсивная КС-грамматика с аксиомой S.
- Покажем, что существуют выводы
  - $S \Rightarrow^* wA\alpha \Rightarrow w\beta\alpha \Rightarrow^* wx$ и
  - $S \Rightarrow^* wA\alpha \Rightarrow w\gamma\alpha \Rightarrow^* wy$

  для которых $\mathrm{FIRST}(x) = \mathrm{FIRST}(y)$, но $\beta \neq \gamma$. Тогда по предыдущей теореме G не будет LL(1)-грамматикой.

Теорема: окончание доказательства

- Существует вывод $B \Rightarrow^+ B\eta$:

$$B = B_0 \Rightarrow B_1\eta_1 \Rightarrow B_2\eta_2\eta_1 \Rightarrow \ldots \Rightarrow B_m\eta_m \ldots \eta_1 = B\eta \ .$$

- Для некоторого j имеются различные правила $B_j \to B_{j+1}\eta_{j+1} \mid \beta$.
- Существует вывод

$$S \Rightarrow^* wB\mu \Rightarrow^* wB\eta\mu \Rightarrow^* w \underbrace{B_j}_{A} \underbrace{\eta_j \ldots \eta_1\eta\mu}_{\alpha} \ ,$$

  который можно продолжить двумя путями:

  (1) $\Rightarrow w\beta \underbrace{\eta_j \ldots \eta_1\eta\mu}_{\alpha} \Rightarrow^* w \underbrace{vu_j \ldots u_1 uz}_{x}$ и

  (2) $\Rightarrow w \underbrace{B_{j+1}\eta_{j+1}}_{\gamma} \underbrace{\eta_j \ldots \eta_1\eta\mu}_{\alpha} \Rightarrow^* w\beta\eta_j \ldots \eta_1\eta^2\mu \Rightarrow^* w \underbrace{vu_j \ldots u_1 u^2 z}_{y}$,

  где $\beta \Rightarrow^* v$, $\eta_i \Rightarrow^* u_i$ $(i = 1, \ldots, j)$, $\eta \Rightarrow^* u$, $\mu \Rightarrow^* z$.
- Итак, имеем $\mathrm{FIRST}(x) = \mathrm{FIRST}(y)$, но $\beta \neq \gamma$.

□

Замечание о непринадлежности некоторых грамматик классу LL(1)-грамматик

Если грамматика содержит правила $A \to \alpha\beta_1 | \alpha\beta_2$, $\beta_1 \neq \beta_2$ и существует терминал $\mathsf{a} \in \mathrm{FIRST}(\alpha)$, то эта грамматика не является LL(1)-грамматикой, поскольку

$$\mathsf{a} \in \mathrm{SELECT}(A \to \alpha\beta_1) \cap \mathrm{SELECT}(A \to \alpha\beta_2).$$

Пример преобразования грамматики в эквивалентную LL(1)-грамматику

- Граматика GL_3 (порождающая язык списков):
  - $S \to L; S \mid L$,
  - $L \to \mathtt{a} \mid [S]$.

  GL_3 не является LL(1)-грамматикой (почему?).
- Произведём левую факторизацию граматики GL_3 и получим грамматику GL'_3:
  - $S \to LS'$,
  - $S' \to ; S \mid \varepsilon$,
  - $L \to \mathtt{a} \mid [S]$.

Правило	Множество выбора
$L \to \mathtt{a}$	$\{\mathtt{a}\}$
$L \to [S]$	$\{[\}$
$S \to LS'$	$\{\mathtt{a}, [\}$
$S' \to ; S$	$\{;\}$
$S' \to \varepsilon$	$\{\dashv,]\}$

GL'_3 является LL(1)-грамматикой.

План

1. LL(1)-грамматики
2. Синтаксический анализ для LL(1)-грамматик
3. Вычисление множеств выбора правил (начало)

Алгоритм построения нисходящего синтаксического анализатора для LL(1)-грамматики

В выводе

$$S \Rightarrow^* w\underline{A}\alpha \Rightarrow w\beta\alpha \Rightarrow^* w\mathtt{b}\chi$$

$\underline{A}$ заменяется по такому правилу $A \to \beta$, что $\mathtt{b} \in \mathrm{SELECT}(A \to \beta)$.

Вход. LL(1)-грамматика $G = (\Sigma, \Gamma, P, S)$.
Выход. МПА $\mathcal{M} = (\widehat{\Sigma}, \widehat{\Gamma}, \delta, \gamma_0)$ такой, что $L(G) = L(\mathcal{M})$.

1. $\widehat{\Sigma} := \Sigma \cup \{\dashv\}$; $\widehat{\Gamma} := \Sigma \cup \Gamma \cup \{\nabla\}$; $\gamma_0 := S$;
2. $\delta := \{(\dashv, \nabla) \to \checkmark\}$;
3. **для каждого** $(B \to \gamma) \in P$
4. **для каждого** $\mathtt{a} \in \mathrm{SELECT}(B \to \gamma)$
5. $\delta := \delta \cup \{(\mathtt{a}, B) \to (\gamma, \_)\}$;
6. **для каждого** $\mathtt{a} \in \Sigma$
7. $\delta := \delta \cup \{(\mathtt{a}, \mathtt{a}) \to (\varepsilon, \rightarrow)\}$

Почему построенный МПА $\mathcal{M}$ детерминированный?

Пример построения синтаксического анализатора для LL(1)-грамматики

Грамматика GA_3:

- $E \to TE', \quad E' \to +TE' | \varepsilon,$
- $T \to FT', \quad T' \to *FT' | \varepsilon,$
- $F \to (E) \mid \mathrm{x}.$

Правило	Мн. выб.
$E \to TE'$	$\{(, \mathrm{x}\}$
$E' \to +TE'$	$\{+\}$
$E' \to \varepsilon$	$\{\dashv,)\}$
$T \to FT'$	$\{(, \mathrm{x}\}$
$T' \to *FT'$	$\{*\}$
$T' \to \varepsilon$	$\{\dashv, +,)\}$
$F \to (E)$	$\{(\}$
$F \to \mathrm{x}$	$\{\mathrm{x}\}$

	x	$+$	$*$	$($	$)$	$\dashv$
E	TE'			TE'		
E'		$+TE'$			ε	ε
T	FT'			FT'		
T'		ε	$*FT'$		ε	ε
F	x			(E)		

Теорема о корректности алгоритма построения синтаксического анализатора для LL(1)-грамматики

Теорема

Для любой LL(1)-грамматики $G = (\Sigma, \Gamma, P, S)$ алгоритм на слайде 17 строит МПА $\mathcal{M}$ такой, что $L(G) = L(\mathcal{M})$.

Д о к а з а т е л ь с т в о.

- Пусть автомат $\mathcal{M}$ обрабатывает цепочку w в течение n тактов, после чего отвергает её или выполняет команду допуска.
- Для каждого $k = 0, 1 \ldots, n$ через w_k обозначим прочитанный за первые k тактов префикс цепочки w, а через β_k — содержимое стека сразу после k-го такта.
- Положим $\alpha_k = w_k \beta_k$. Имеем $w_0 = \varepsilon$, $\beta_0 = S$, $\alpha_0 = S$.
- Из последовательности $\alpha_0, \alpha_1, \ldots, \alpha_n$ вычеркнем (рекурсивно) все элементы α_k такие, что $\alpha_{k-1} = \alpha_k$. Получим последовательность П: $\alpha_0 = \alpha_{i_0}, \alpha_{i_1}, \ldots, \alpha_{i_n} = \alpha_n$.

Теорема: продолжение доказательства

- П — (левый) вывод цепочки α_n в грамматике G:
    - если на k-м такте применена команда вида $(\mathtt{a}, \mathtt{a}) \to (\varepsilon, {}^{\rightarrow})$, то $w_{k-1}\mathtt{a} = w_k$, $\beta_{k-1} = \mathtt{a}\beta_k$ и $\alpha_{k-1} = \alpha_k$;
    - если же на k-м такте применена команда вида $(\mathtt{a}, B) \to (\gamma, \_)$, то $w_k = w_{k-1}$, $\beta_{k-1} = B\beta'$, $\beta_k = \gamma\beta'$ (для некоторой β') и $(B \to \gamma) \in P$. Тогда $\alpha_{k-1} = w_k B\beta' \Rightarrow w_k\gamma\beta' = \alpha_k$.
- Если $w \in L(\mathcal{M})$, то $w_n = w$, $\beta_n = \varepsilon$ и $\alpha_n = w$, поэтому $w \in L(G)$.
- Обратно, пусть $w \in L(G)$. Тогда существует (левый) вывод $S = \gamma_0, \gamma_1, \ldots, \gamma_s = w$.
- Индукцией покажем, что для любого $i \leqslant s$ найдётся такое k, что $\gamma_i = \alpha_k$.
- База индукции верна: $\gamma_0 = S = \alpha_0$.
- Индукционный переход. Пусть для некоторого $i < s$ верно $\gamma_i = \alpha_k = w_k\beta_k$.
    - γ_{i+1} получена из γ_i заменой самого левого вхождения нетерминала по некоторому правилу $A \to \gamma$.

Теорема: продолжение доказательства

- $\beta_k = \mathrm{a}_1 \dots \mathrm{a}_m A\beta'$ $(m \geqslant 0)$, $\gamma_i = \alpha_k = w_k\beta_k = w_k \mathrm{a}_1 \dots \mathrm{a}_m A\beta'$.
- $w_k \mathrm{a}_1 \dots \mathrm{a}_m$ — префикс входной цепочки w. На тактах с $(k+1)$-го по $(k+m)$-й автомат $\mathcal{M}$ прочитает и удалит из стека $\mathrm{a}_1 \dots \mathrm{a}_m$.
- $w_{k+m} = w_k \mathrm{a}_1 \dots \mathrm{a}_m$, $\beta_{k+m} = A\beta'$, $\gamma_i = \alpha_k = \alpha_{k+m}$.
- Обозначим через w' непрочитанный после $(k+m)$ тактов суффикс входной цепочки w, а через c — первый символ w' при $w' \neq \varepsilon$ или $\dashv$ при $w' = \varepsilon$. Покажем, что $\mathrm{c} \in \mathrm{SELECT}(A \to \gamma)$.
- $\gamma_i = w_k \mathrm{a}_1 \dots \mathrm{a}_m A\beta' \Rightarrow \gamma_{i+1} = w_k \mathrm{a}_1 \dots \mathrm{a}_m \gamma\beta' \Rightarrow^* w = w_k \mathrm{a}_1 \dots \mathrm{a}_m w'$, следовательно, $\gamma\beta' \Rightarrow^* w'$.
- Если вывод $\gamma\beta' \Rightarrow^* w'$ имеет вид $\gamma\beta' \Rightarrow^* \mathrm{a}\gamma'\beta' \Rightarrow^* w'$, то $\mathrm{c} = \mathrm{a} \in \mathrm{FIRST}(\gamma) \setminus \{\varepsilon\} \subseteq \mathrm{SELECT}(A \to \gamma)$.
- Если же вывод $\gamma\beta' \Rightarrow^* w'$ имеет вид $\gamma\beta' \Rightarrow^* \beta' \Rightarrow^* w'$, то $\mathrm{c} \in \mathrm{FOLLOW}(A) \subseteq \mathrm{SELECT}(A \to \gamma)$.
- Таким образом, $\mathrm{c} \in \mathrm{SELECT}(A \to \gamma)$. Значит, автомат $\mathcal{M}$ имеет команду $(\mathrm{c}, A) \to (\gamma, \_)$, которую выполнит на $(k+m+1)$-м такте. Тогда имеем
  $\alpha_{k+m+1} = w_{k+m+1}\beta_{k+m+1} = w_{k+m}\gamma\beta' = w_k \mathrm{a}_1 \dots \mathrm{a}_m \gamma\beta' = \gamma_{i+1}$.

Теорема о корректности алгоритма построения синтаксического анализатора для LL(1)-грамматики: окончание доказательства

- Найдётся такое $\bar{k}$, что $w = \gamma_s = \alpha_{\bar{k}}$.
- Обозначим через w' непрочитанный после $\bar{k}$ тактов суффикс цепочки w.
- Тогда $w_{\bar{k}}\beta_{\bar{k}} = \alpha_{\bar{k}} = w = w_{\bar{k}}w'$, откуда $\beta_{\bar{k}} = w'$.
- Следовательно, автомат $\mathcal{M}$ прочтёт w' и опустошит стек, выполняя команды вида $(\mathrm{a}, \mathrm{a}) \to (\varepsilon, \vec{\ })$, а затем выполнит команду допуска.
- Итак, $w \in L(\mathcal{M})$.

□

Оценка числа шагов МПА для LL(1)-грамматики

Теорема

Число шагов (тактов), выполняемых МПА, который построен алгоритмом на слайде 17, линейно зависит от длины входной цепочки.

Д о к а з а т е л ь с т в о.

- Пусть n — длина входной цепочки.
- Рассматриваемая LL(1)-грамматика нелеворекурсивна в силу теоремы на слайде 12.
- Тогда длина любого вывода вида $B_0\alpha_0 \Rightarrow B_1\alpha_1 \Rightarrow \ldots \Rightarrow B_m\alpha_m$ меньше, чем число c нетерминалов в данной грамматике.
- Значит, построенный МПА, воспроизводя вывод входной цепочки (см. доказательство предыдущей теоремы), до любого из $\leqslant n$ тактов со сдвигом по входной цепочке, а также до такта, на котором выполняется команда допуска, выполняет менее c тактов без сдвига по входной цепочке.
- Таким образом, этот МПА выполняет всего $O(n)$ тактов.

□

План

1. LL(1)-грамматики
2. Синтаксический анализ для LL(1)-грамматик
3. Вычисление множеств выбора правил (начало)

Вычисление $\textsc{select}((A \to \alpha)$ и $\textsc{first}(X)$

Пусть дана LL(1)-грамматика $G = (\Sigma, \Gamma, P, S)$.

$$\textsc{select}(A \to \alpha) = \begin{cases} \textsc{first}(\alpha), & \text{если } \varepsilon \notin \textsc{first}(\alpha), \\ (\textsc{first}(\alpha) \setminus \{\varepsilon\}) \cup \textsc{follow}(A), & \text{если } \varepsilon \in \textsc{first}(\alpha). \end{cases}$$

Для вычисления $\textsc{select}(A \to \alpha)$ достаточно уметь вычислять $\textsc{first}(\alpha)$ и $\textsc{follow}(A)$.

Как вычислить $\textsc{first}(X)$?

- $\textsc{first}(\mathtt{a}) = \{\mathtt{a}\}$.
- $\textsc{first}(A) = \bigcup\limits_{(A \to \alpha) \in P} \textsc{first}(\alpha)$.
- Пусть $L \subseteq_{\mathtt{t}} M$ означает, что $L \cap \Sigma \subseteq M$.
- Рассмотрим $\alpha = X_1 \ldots X_n$, $n > 0$.
  - $\textsc{first}(X_1) \subseteq_{\mathtt{t}} \textsc{first}(\alpha)$;
  - $\textsc{first}(X_2) \subseteq_{\mathtt{t}} \textsc{first}(\alpha)$, если $\varepsilon \in \textsc{first}(X_1)$;
  - $\textsc{first}(X_3) \subseteq_{\mathtt{t}} \textsc{first}(\alpha)$, если $\varepsilon \in \textsc{first}(X_1) \cap \textsc{first}(X_2)$; ...;
  - $\textsc{first}(X_n) \subseteq_{\mathtt{t}} \textsc{first}(\alpha)$, если $\varepsilon \in \textsc{first}(X_1) \cap \ldots \cap \textsc{first}(X_{n-1})$;
  - $\varepsilon \in \textsc{first}(\alpha)$, если $\varepsilon \in \textsc{first}(X_1) \cap \ldots \cap \textsc{first}(X_n)$.

Алгоритмы вычисления $\mathrm{FIRST}(A)$ и $\mathrm{FIRST}(\alpha)$

Вход. LL(1)-грамматика $G = (\Sigma, \Gamma, P, S)$.
Выход. Массив множеств $\mathrm{FIRST}(A)$ для каждого $A \in \Gamma$.

1. **для каждого** $\mathtt{a} \in \Sigma$ $\mathrm{FIRST}(\mathtt{a}) := \{\mathtt{a}\}$;
2. **для каждого** $A \in \Gamma$
3.   **если** $((A \to \varepsilon) \in P)$ $\mathrm{FIRST}(A) := \{\varepsilon\}$;
4.   **иначе** $\mathrm{FIRST}(A) := \emptyset$;
5. **пока** все множества $\mathrm{FIRST}(A)$ не стабилизировались, **повторять**
6.   **для каждого** $(A \to X_1 \dots X_n) \in P$, где $n > 0$, **выполнить**
7. $i := 1$;
8. $\mathrm{FIRST}(A) := \mathrm{FIRST}(A) \cup (\mathrm{FIRST}(X_i) \cap \Sigma)$;
9.   **если** $(\varepsilon \in \mathrm{FIRST}(X_i))$
10.   **если** $(i < n)$
11. $i := i + 1$; **перейти на** 8;
12.   **иначе** $\mathrm{FIRST}(A) := \mathrm{FIRST}(A) \cup \{\varepsilon\}$

Вычисление $\mathrm{FIRST}(\alpha)$ для $\alpha = X_1 \dots X_n$ $(n > 0)$: инициализировать $\mathrm{FIRST}(\alpha) := \emptyset$ и выполнить строки 7–12, где A заменён на α.

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`).