

Лекции по теории формальных языков

Лекция 10.

Грамматики простого предшествования

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

15 апреля 2011 г.

План

- 1 Отношения простого предшествования
- 2 Грамматики простого предшествования

План

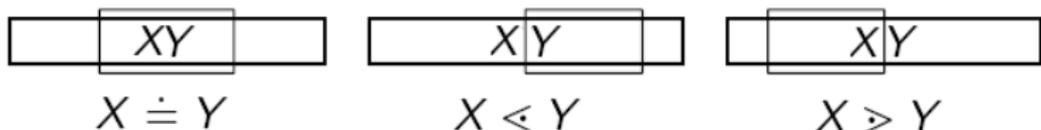
- 1 Отношения простого предшествования
- 2 Грамматики простого предшествования

Отношения простого предшествования: определение

Пусть $G = (\Sigma, \Gamma, P, S)$ — КС-грамматика, $X, Y \in \Sigma \cup \Gamma$.

Определим отношения простого предшествования $\dot{=}$, \triangleleft и \triangleright :

- $X \dot{=} Y$, если XY входит в основу некоторой r -формы;
- $X \triangleleft Y$, если основа некоторой r -формы начинается с символа Y , перед которым стоит символ X ;
- $X \triangleright Y$, если основа некоторой r -формы заканчивается символом X , после которого стоит символ Y ($Y \in \Sigma$?).



Теорема об отношениях простого предшествования

Теорема

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

- (1) $X \doteq Y \Leftrightarrow$ существует правило вида $A \rightarrow \alpha XY\beta$;
- (2) $X \triangleleft Y \Leftrightarrow$ существует правило вида $A \rightarrow \alpha XZ\beta$ такое, что $Z \Rightarrow^+ Y\gamma$;
- (3) $X \triangleright Y \Leftrightarrow Y \in \Sigma$ и существует правило вида $A \rightarrow \alpha Z_1 Z_2 \beta$ такое, что $Z_1 \Rightarrow^+ \gamma_1 X$ и $Z_2 \Rightarrow^* Y\gamma_2$.

Доказательство. Докажем (1).

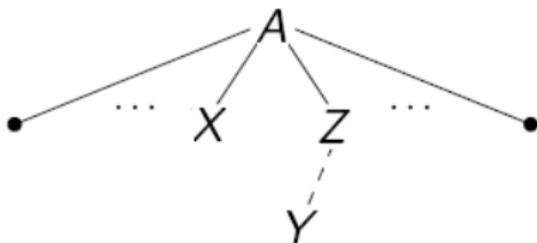
- Если $X \doteq Y$, то основа некоторой формы имеет вид $\alpha XY\beta$ и потому существует правило вида $A \rightarrow \alpha XY\beta$.
- Если существует правило вида $A \rightarrow \alpha XY\beta$, то оно применяется в некотором выводе из S некоторой цепочки $w \in \Sigma^*$:

$$S \Rightarrow^* \xi Au \Rightarrow \xi \alpha XY\beta u \Rightarrow^* w.$$

- Основой формы $\xi \alpha XY\beta u$ является $\alpha XY\beta$, следовательно, $X \doteq Y$.

Теорема: продолжение доказательства

Докажем (2).



- Пусть $X \leq Y$. Тогда существует форма $\mu XY\nu$, основа которой начинается с Y .
- Рассмотрим дерево T , представляющее вывод формы $\mu XY\nu$. Пусть x и y — листья этого дерева, соответствующие выделенным вхождениям X и Y в $\mu XY\nu$.
- Лист y является также листом самого левого куста дерева T , а лист x не является листом никакого куста.
- Следовательно, x и y не братья.

Теорема: продолжение доказательства

- Все левые братья узла x — листья. Тогда узел x имеет правого брата. Обозначим ближайшего правого брата узла x через z .
- Узел z является внутренним; пусть он помечен нетерминалом Z .
- Существует правило вида $A \rightarrow \alpha X Z \beta$.
- Поскольку грамматика G неукорачивающая, в рассматриваемом выводе формы $\mu X Y \nu$ имеем $Z \Rightarrow^+ \eta \neq \varepsilon$; следовательно, $\eta = Y \gamma$.
- Обратно, пусть существует правило вида $A \rightarrow \alpha X Z \beta$ и $Z \Rightarrow^+ Y \gamma$.
- Тогда существует вывод

$$S \Rightarrow^* \xi A u \Rightarrow \xi \alpha X Z \beta u \Rightarrow^* \xi \alpha X Z \nu u. \quad (*)$$

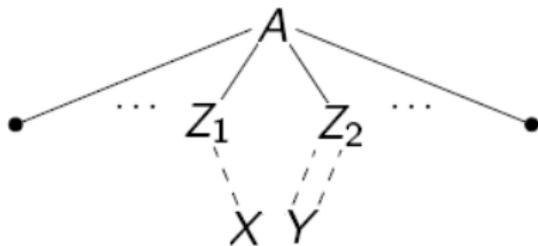
- Пусть $Y \gamma'$ — первая цепочка в выводе $Z \Rightarrow^+ Y \gamma$, которая начинается с Y . Значит, правая часть правила, применённого последним в выводе $Z \Rightarrow^+ Y \gamma'$, начинается с Y .
- Продолжим вывод (*):

$$\xi \alpha X Z \nu u \Rightarrow^+ \xi \alpha X Y \gamma' \nu u \Rightarrow^* w.$$

Основа формы $\xi \alpha X Y \gamma' \nu u$ начинается с Y , так что $X \triangleleft Y$.

Теорема: продолжение доказательства

Докажем (3).



- Пусть $X \succ Y$. Тогда $Y \in \Sigma$ и существует форма $\mu XY\nu$, основа которой заканчивается символом X .
- Рассмотрим дерево T , представляющее вывод формы $\mu XY\nu$. Пусть x и y — листья этого дерева, соответствующие выделенным вхождением X и Y в $\mu XY\nu$.
- Обозначим через z ближайшего общего предка узлов x и y , а через z_1 и z_2 — сыновей узла z , потомками которых являются узлы x и y соответственно.

Теорема: продолжение доказательства

- Поскольку грамматика G неукорачивающая, z_1 и z_2 — стоящие рядом братья.
- Следовательно, имеется правило вида $A \rightarrow \alpha Z_1 Z_2 \beta$, где A, Z_1, Z_2 — метки узлов z, z_1, z_2 соответственно.
- x — самый правый лист поддерева (дерева T) с корнем z_1 , поэтому $Z_1 \Rightarrow^* \gamma_1 X$.
- Аналогично $Z_2 \Rightarrow^* Y \gamma_2$.
- Покажем, что вместо $Z_1 \Rightarrow^* \gamma_1 X$ можно записать $Z_1 \Rightarrow^+ \gamma_1 X$.
 - ▶ Предположим, что $z_1 = x$.
 - ▶ В случае $Z_2 \Rightarrow^+ Y \gamma_2$ имеем: узел z_2 является братом узла x и имеет потомка $y \neq z_2$. Получили противоречие с тем, что узел x является листом куста дерева T .
 - ▶ А в случае $z_2 = y$ имеем: узел y является ближайшим правым братом узла x , что невозможно.

Теорема: окончание доказательства

- Обратно, пусть $Y \in \Sigma$ и существует правило вида $A \rightarrow \alpha Z_1 Z_2 \beta$ такое, что $Z_1 \Rightarrow^+ \gamma_1 X$ и $Z_2 \Rightarrow^* Y \gamma_2$.
- Тогда существует вывод

$$S \Rightarrow^* \xi A u \Rightarrow \xi \alpha Z_1 Z_2 \beta u \Rightarrow^* \xi \alpha Z_1 Z_2 v u \Rightarrow^* \xi \alpha Z_1 Y \gamma_2 v u \Rightarrow^* \xi \alpha Z_1 Y v' v u. \quad (**)$$

- Пусть $\gamma'_1 X$ — первая цепочка в выводе $Z_1 \Rightarrow^+ \gamma_1 X$, которая заканчивается на X . Поэтому правая часть правила, применённого последним в выводе $Z_1 \Rightarrow^+ \gamma'_1 X$, заканчивается на X .
- Продолжим вывод (**):

$$\xi \alpha Z_1 Y v' v u \Rightarrow^+ \xi \alpha \gamma'_1 X Y v' v u \Rightarrow^* w.$$

Основа формы $\xi \alpha \gamma'_1 X Y v' v u$ заканчивается на X , следовательно, $X \succ Y$.



Вычисление отношений простого предшествования

- Пусть $G = (\Sigma, \Gamma, P, S)$ — КС-грамматика, $B \in \Gamma$. Тогда положим

$$\text{FIRST}'(B) = \{X \in \Sigma \cup \Gamma \mid \exists \beta (B \Rightarrow^+ X\beta)\},$$

$$\text{LAST}'(B) = \{X \in \Sigma \cup \Gamma \mid \exists \beta (B \Rightarrow^+ \beta X)\}.$$

- Упражнение. Разработать алгоритмы вычисления функций FIRST' и LAST' .
- $X \doteq Y \Leftrightarrow$ существует правило вида $A \rightarrow \alpha XY\beta$.
- $X \triangleleft Y \Leftrightarrow$
существует правило вида $A \rightarrow \alpha XZ\beta$ такое, что $Z \Rightarrow^+ Y\gamma$, \Leftrightarrow
существует $Z \in \Gamma$ такой, что $X \doteq Z$ и $Y \in \text{FIRST}'(Z)$.
- $X \triangleright Y \Leftrightarrow$
 $Y \in \Sigma$ и существует правило вида $A \rightarrow \alpha Z_1 Z_2 \beta$ такое, что
 $Z_1 \Rightarrow^+ \gamma_1 X$ и $Z_2 \Rightarrow^* Y\gamma_2$, \Leftrightarrow
 $Y \in \Sigma$ и существуют $Z_1 \in \Gamma$ и $Z_2 \in \Sigma \cup \Gamma$ такие, что $Z_1 \doteq Z_2$,
 $X \in \text{LAST}'(Z_1)$ и либо $Z_2 = Y$, либо $Y \in \text{FIRST}'(Z_2)$.

Пример вычисления отношений простого предшествования

- Грамматика $G_1 = \{S \rightarrow aSSb|c\}$.
- $FIRST'(S) = \{a, c\}$, $LAST'(S) = \{b, c\}$.
- $a \doteq S$, $S \doteq S$, $S \doteq b$.
- $a \doteq S$ даёт $a \triangleleft Y$, где $Y \in FIRST'(S)$.
- $S \doteq S$ даёт $\triangleleft: S \triangleleft Y$, где $Y \in FIRST'(S)$.
- $S \doteq S$ даёт $\triangleright: X \triangleright Y$, где $X \in LAST'(S)$, $Y \in FIRST'(S)$.
- $S \doteq b$ даёт $X \triangleright b$, где $X \in LAST'(S)$.

	S	a	b	c
S	\doteq	\triangleleft	\doteq	\triangleleft
a	\doteq	\triangleleft		\triangleleft
b		\triangleright	\triangleright	\triangleright
c		\triangleright	\triangleright	\triangleright

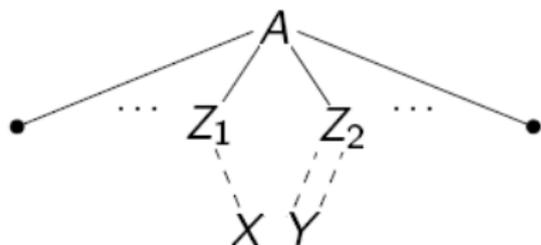
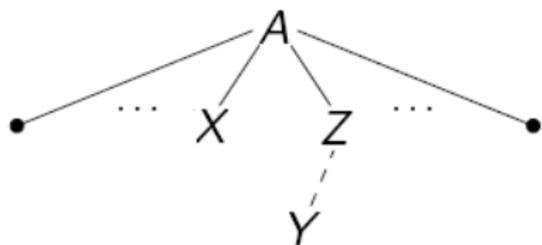
Предложение о выполнении отношений простого предшествования для соседних символов r -формы

Предложение

Пусть $\gamma = \alpha XY\beta$ — r -форма. Тогда $X \doteq Y$, $X \triangleleft Y$ или $X \triangleright Y$.

Доказательство.

- Пусть вывод формы γ представлен деревом T , в котором два соседних листа x и y помечены X и Y соответственно.
- Далее, пусть z_0 — ближайший общий предок x и y , z_0 помечен A .
- Если x и y — сыновья z_0 , то $X \doteq Y$.
- Если из x и y только x является сыном z_0 , то $X \triangleleft Y$.
- Иначе $X \triangleright Y$.



План

- 1 Отношения простого предшествования
- 2 Грамматики простого предшествования

Грамматика простого предшествования: определение

- Грамматика называется *обратимой*, если любые два различных правила вывода имеют различные правые части.
- КС-грамматика $G = (\Sigma, \Gamma, P, S)$ называется *грамматикой простого предшествования* (ПП-грамматикой), если она без циклов, обратима и для любой упорядоченной пары символов из $\Sigma \cup \Gamma$ выполняется не более одного отношения простого предшествования.
- Пример. Грамматика G_1 на слайде 12 является ПП-грамматикой.
- Для восходящего синтаксического анализа будем использовать маркер \vdash в начале входной строки и маркер \dashv в конце входной строки.
- Будем считать, что $\vdash \prec Y$ для любого символа Y , с которого может начинаться r -форма, и $X \succ \dashv$ для любого символа X , которым может заканчиваться r -форма.

Предложение о выделении основы для ПП-грамматики

Предложение

Пусть γ — r -форма ПП-грамматики, $\vdash \gamma \dashv = X_0 X_1 \dots X_n X_{n+1}$, $n \geq 1$. Тогда основой формы γ является такая цепочка $X_k \dots X_m$, что m — минимальный номер, для которого выполнено $X_m \succ X_{m+1}$, $k \leq m$ и

$$X_{k-1} \prec X_k, X_k \doteq X_{k+1}, \dots, X_{m-1} \doteq X_m, X_m \succ X_{m+1}.$$

Доказательство.

- $X_0 \prec X_1$ и $X_n \succ X_{n+1}$, поэтому такие индексы k и m существуют.
- Пусть T — дерево, представляющее вывод формы γ с листьями x_1, \dots, x_n , которые помечены X_1, \dots, X_n соответственно.
- Пусть основа формы γ есть $X_s \dots X_t$ ($s \leq t$).
- Тогда $X_t \succ X_{t+1}$, откуда $m \leq t$.
- Предположим, что $m < t$.

Предложение о выделении основы для ПП-грамматики: окончание доказательства

- Символы X_m и X_{m+1} не входят в основу $X_s \dots X_t$ одновременно.

$$\begin{array}{cc} X_s \dots X_t & X_s \dots X_t \\ X_k \dots X_m X_{m+1} & X_k \dots X_m \end{array}$$

Следовательно, $m < s$.

- Самый левый куст дерева T имеет листья x_s, \dots, x_t , так что лист x_m не принадлежит никакому кусту дерева T .
- Тогда у x_m есть брат — внутренний узел, значит, в T есть куст левее самого левого куста.
- Полученное противоречие показывает, что $m = t$.
- $X_{s-1} \triangleleft X_s$, $X_s \doteq X_{s+1}, \dots, X_{t-1} \doteq X_t$, следовательно, $k = s$.



Алгоритм восходящего анализа для ПП-грамматики

Пусть дана ПП-грамматика $G = (\Sigma, \Gamma, P, S)$;

$X \circ Y$ означает, что ни $X \doteq Y$, ни $X \leq Y$, ни $X \geq Y$.

Вход. Цепочка $w \in \Sigma^+$.

Выход. «ДА», если $w \in L(G)$, «НЕТ» иначе.

1. $\gamma := \vdash w \dashv$; // далее считаем, что $\gamma = X_0 \dots X_{n+1}$
2. **пока** ($\gamma \neq \vdash S \dashv$) **повторять**
3. $i := 0$; $k := 0$; $m := 0$;
4. **пока** ($i \leq n \wedge m = 0$) **повторять**
5. **если** ($X_i \circ X_{i+1}$) **возвратить** «НЕТ»;
6. **иначе если** ($X_i \leq X_{i+1}$) $k := i + 1$;
7. **иначе если** ($X_i \geq X_{i+1}$) $m := i$;
8. $i := i + 1$;
9. **если** ($m = 0$) **возвратить** «НЕТ»;
10. **иначе если** ($\nexists A : (A \rightarrow X_k \dots X_m) \in P$) **возвратить** «НЕТ»;
11. **иначе** $\gamma := X_0 \dots X_{k-1} A X_{m+1} \dots X_{n+1}$; **перейти на** 2;
12. **возвратить** «ДА»

Упражнение. Оптимизировать поиск основы.

Пример восходящего анализа для ПП-грамматики

Грамматика $G_1 = \{S \rightarrow aSSb|c\}$, цепочка $w = acaccbb$.

	S	a	b	c	\vdash
S	$\dot{=}$	\langle	$\dot{=}$	\langle	\rangle
a	$\dot{=}$	\langle		\langle	
b		\rangle	\rangle	\rangle	\rangle
c		\rangle	\rangle	\rangle	\rangle
\vdash	\langle	\langle		\langle	

$$\begin{aligned}
 & \vdash \langle a \langle \underline{c} \rangle a \langle c \rangle c \rangle b \rangle b \rangle \vdash \\
 & \vdash \langle a \dot{=} S \langle a \langle \underline{c} \rangle c \rangle b \rangle b \rangle \vdash \\
 & \vdash \langle a \dot{=} S \langle a \dot{=} S \langle \underline{c} \rangle b \rangle b \rangle \vdash \\
 & \vdash \langle a \dot{=} S \langle \underline{a \dot{=} S \dot{=} S \dot{=} b} \rangle b \rangle \vdash \\
 & \quad \vdash \langle \underline{a \dot{=} S \dot{=} S \dot{=} b} \rangle \vdash \\
 & \quad \vdash \langle S \rangle \vdash
 \end{aligned}$$

Предложение о корректности алгоритма восходящего анализа для ПП-грамматики

Предложение

Алгоритм на слайде 18 останавливается для любой входной цепочки $w \in \Sigma^+$, и если он возвращает «ДА», то $w \in L(G)$, а если «НЕТ», то $w \notin L(G)$.

Доказательство.

- Итерация внешнего цикла этого алгоритма заканчивается либо ответом «НЕТ», либо свёрткой.
- Длина цепочки γ может не уменьшаться не более чем $|G|$ свёрток подряд, поскольку грамматика G неукорачивающая и без циклов.
- Следовательно, после конечного числа свёрток γ будет иметь вид $\vdash A \dashv$. Тогда алгоритм завершается, сделав ещё не более одной свёртки.

Предложение о корректности алгоритма восходящего анализа для ПП-грамматики: окончание доказательства

- Если алгоритм возвращает «ДА», то цепочка w свёрнута в аксиому S , т. е. найден вывод цепочки w из S ; значит, $w \in L(G)$.
- Если алгоритм возвращает «НЕТ», то γ не является r -формой.
- В силу обратимости грамматики G все свёртки определяются однозначно. Тогда $w \notin L(G)$, поскольку γ не является r -формой.



Предложение об оценке числа шагов алгоритма типа «перенос-свёртка» восходящего анализа для ПП-грамматики

Алгоритм восходящего анализа для ПП-грамматики легко можно представить как алгоритм типа «перенос-свёртка».

Предложение

Число шагов (переносов и свёрток) такого алгоритма линейно зависит от длины входной цепочки.

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>).