Лекции по теории формальных языков

Лекция 11.
Грамматики слабого предшествования.
Отношения операторного 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

22 апреля 2011 г.

План

1. Грамматики слабого предшествования

2. Отношения операторного предшествования

План

Грамматики слабого предшествования: наводящие соображения и определение

- Расширим класс ПП-грамматик, допустив, что отношения $\doteq$ и $\lessdot$ могут пересекаться.
- Нахождение основы: $\alpha\beta\underline{\gamma} \gtrdot u$ или $\alpha\underline{\beta\gamma} \gtrdot u$, где $A \to \gamma$, $B \to \beta\gamma$ — некоторые правила вывода?
- Для каких грамматик правильно применять самое длинное правило вывода?

- КС-грамматика называется *грамматикой слабого предшествования* (*СП-грамматикой*), если
  - она без циклов, обратима,
  - отношение $\gtrdot$ не пересекается с объединением отношений $\doteq$ и $\lessdot$, и
  - в ней нет правил вида $A \to \alpha X\beta$ и $B \to \beta$ таких, что $X \doteq B$ или $X \lessdot B$.

Предложение о включении класса ПП-грамматик в класс СП-грамматик

Предложение

Любая ПП-грамматика является СП-грамматикой.

Доказательство.

- Пусть в ПП-грамматике G имеются правила $A \to \alpha X \beta$ и $B \to \beta$.
- Если $\beta = \varepsilon$, то очевидно, что ни $X \doteq B$, ни $X \lessdot B$.
- Далее $\beta \neq \varepsilon$. Следовательно, $\beta = Y\gamma$ для некоторых Y и γ.
- $A \to \alpha X \beta = \alpha X Y \gamma$, поэтому $X \doteq Y$.
- Предположим, что $X \doteq B$.
  - Существует правило $C \to \mu X B \nu$.
  - $B \Rightarrow^+ \beta = Y\gamma$.
  - Тогда $X \lessdot Y$ — противоречие.
- Предположим, что $X \lessdot B$.
  - Существует такое правило $C \to \mu X Z \nu$, что $Z \Rightarrow^+ B\eta$.
  - $Z \Rightarrow^+ B\eta \Rightarrow^+ Y\gamma\eta'$.
  - Тогда $X \lessdot Y$ — противоречие.

□

Вычисление отношений простого предшествования (напоминание)

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

$$\textsc{first}'(B) = \{X \in \Sigma \cup \Gamma \mid \exists\beta\ (B \Rightarrow^{+} X\beta)\},$$
$$\textsc{last}'(B) = \{X \in \Sigma \cup \Gamma \mid \exists\beta\ (B \Rightarrow^{+} \beta X)\}.$$

- $X \doteq Y \iff$ существует правило вида $A \to \alpha XY\beta$.
- $X \lessdot Y \iff$
  существует правило вида $A \to \alpha XZ\beta$ такое, что $Z \Rightarrow^{+} Y\gamma$, $\iff$
  существует $Z \in \Gamma$ такой, что $X \doteq Z$ и $Y \in \textsc{first}'(Z)$.
- $X \gtrdot Y \iff$
  $Y \in \Sigma$ и существует правило вида $A \to \alpha Z_1 Z_2 \beta$ такое, что
  $Z_1 \Rightarrow^{+} \gamma_1 X$ и $Z_2 \Rightarrow^{*} Y\gamma_2$, $\iff$
  $Y \in \Sigma$ и существуют $Z_1 \in \Gamma$ и $Z_2 \in \Sigma \cup \Gamma$ такие, что $Z_1 \doteq Z_2$,
  $X \in \textsc{last}'(Z_1)$ и либо $Z_2 = Y$, либо $Y \in \textsc{first}'(Z_2)$.

Пример СП-грамматики

$$GA_2 = \{E \to E + T \mid T, \quad T \to T * F \mid F, \quad F \to (E) \mid \mathtt{x}\}.$$

	E	T	F	$($	$)$	$\mathtt{x}$	$+$	$*$	$\dashv$
E					$\doteq$		$\doteq$		$\gtrdot$
T					$\gtrdot$		$\gtrdot$	$\doteq$	$\gtrdot$
F					$\gtrdot$		$\gtrdot$	$\gtrdot$	$\gtrdot$
$($	$\doteq, \lessdot$	$\lessdot$	$\lessdot$	$\lessdot$		$\lessdot$			
$)$					$\gtrdot$		$\gtrdot$	$\gtrdot$	$\gtrdot$
$\mathtt{x}$					$\gtrdot$		$\gtrdot$	$\gtrdot$	$\gtrdot$
$+$		$\doteq, \lessdot$	$\lessdot$	$\lessdot$		$\lessdot$			
$*$			$\doteq$	$\lessdot$		$\lessdot$			
$\vdash$	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$		$\lessdot$			

Есть ли правила вида $A \to \alpha X \beta$ и $B \to \beta$, где $X \doteq B$ или $X \lessdot B$?

- $E \to E + T$ и $E \to T$: ни $+ \doteq E$, ни $+ \lessdot E$;
- $T \to T * F$ и $T \to F$: ни $* \doteq T$, ни $* \lessdot T$.

Предложение о выборе правила СП-грамматики

Предложение

Пусть G — СП-грамматика с аксиомой S, $B \to \beta$ — одно из правил вывода этой грамматики,

$$S \Rightarrow^{+} \mu C w \Rightarrow \nu X \beta w.$$

Тогда правило $B \to \beta$ применяется последним в данном выводе, если и только если в G нет правила вида $A \to \alpha X \beta$.

Доказательство.

- Пусть правило $B \to \beta$ применяется последним в данном выводе.
- Тогда $\nu X B w$ является r-формой, следовательно, имеет место $X \doteq B$, $X \lessdot B$ или $X \gtrdot B$.
- Однако поскольку B — нетерминал, то неверно, что $X \gtrdot B$.
- Значит, $X \doteq B$ или $X \lessdot B$, поэтому по определению СП-грамматики в G нет правила вида $A \to \alpha X \beta$.

Предложение о выборе правила СП-грамматики: окончание доказательства

- Обратно, пусть в G нет правила вида $A \to \alpha X \beta$.
- Конец основы формы $\nu X \beta w$ находится непосредственно слева от выделенного вхождения w.
- Предположим, что эта основа является собственным суффиксом β' цепочки $\beta = \gamma X' \beta'$.
  - Произведя свёртку формы $\nu X \beta w = \nu X \gamma X' \beta' w$ по некоторому правилу $D \to \beta'$, получим форму $\nu X \gamma X' D w$.
  - Тогда $X' \doteq D$ или $X' \lessdot D$.
  - Но в G есть правило $B \to \beta = \gamma X' \beta'$. Противоречие.
- Основа формы $\nu X \beta w$ не может иметь суффикс $X \beta$, так как в G нет правила вида $A \to \alpha X \beta$.
- Таким образом, основа формы $\nu X \beta w$ есть β.

$\square$

Поиск основы при восходящем анализе для СП-грамматики

- Пусть γ — r-форма СП-грамматики, $\vdash \gamma \dashv = X_0 X_1 \ldots X_n X_{n+1}$, $n \geqslant 1$.
- Индекс m конца основы $X_k \ldots X_m$ такой, что m — минимальное число, для которого выполнено $X_m \gtrdot X_{m+1}$.
- Кандидаты на индекс начала основы: такие $k_1 < \ldots < k_s$, что

$$\ldots \lessdot X_{k_1} \doteq \ldots \doteq X_{k_2-1} \mathrel{\substack{\lessdot \\ \doteq}} X_{k_2} \doteq \ldots \doteq X_{k_s-1} \mathrel{\substack{\lessdot \\ \doteq}} X_{k_s} \doteq \ldots \doteq X_m \gtrdot \ldots$$

Пример восходящего анализа для СП-грамматики

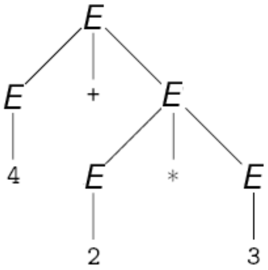
$$GA_2 = \{E \to E + T \mid T, \quad T \to T * F \mid F, \quad F \to (E) \mid \mathtt{x}\}, \qquad \mathtt{x} + \mathtt{x} * \mathtt{x}.$$

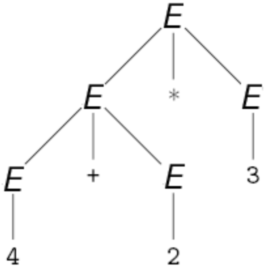
- $\vdash \lessdot \mathtt{x} \gtrdot + \lessdot \mathtt{x} \gtrdot * \lessdot \mathtt{x} \gtrdot \dashv$
- $\vdash \lessdot F \gtrdot + \lessdot \mathtt{x} \gtrdot * \lessdot \mathtt{x} \gtrdot \dashv$
- $\vdash \lessdot T \gtrdot + \lessdot \mathtt{x} \gtrdot * \lessdot \mathtt{x} \gtrdot \dashv$
- $\vdash \lessdot E \doteq + \lessdot \mathtt{x} \gtrdot * \lessdot \mathtt{x} \gtrdot \dashv$
- $\vdash \lessdot E \doteq + \lessdot F \gtrdot * \lessdot \mathtt{x} \gtrdot \dashv$
- $\vdash \lessdot E \doteq + \overset{\lessdot}{\doteq} T \doteq * \lessdot \mathtt{x} \gtrdot \dashv$
- $\vdash \lessdot E \doteq + \overset{\lessdot}{\doteq} T \doteq * \doteq F \gtrdot \dashv$
- $\vdash \lessdot E \doteq + \overset{\lessdot}{\doteq} T \gtrdot \dashv$
- $\vdash \lessdot E \gtrdot \dashv$

План

Идея основанного на отношениях операторного предшествования метода анализа выражений — 1

- Пусть *выражение* состоит из операндов, операторов и скобок. Считаем, что язык выражений задаётся грамматикой выражений (быть может, неоднозначной) с единственным нетерминалом E (операндом).
- Семантическое значение выражения определяется значениями атомарных операндов и порядком выполнения операторов.
- Порядок выполнения операторов можно описать с помощь дерева.





Идея основанного на отношениях операторного предшествования метода анализа выражений — 2

- Если порядок выполнения операторов не указан явно с помощью скобок, то он выявляется соглашениями о приоритете операторов и их ассоциативности.
- В соответствии с порядком выполнения операторов можно построить дерево снизу вверх.
- Выполнение оператора рассматривается как свёртка. Тогда как выделять основу для свёртки?
- По приоритету и ассоциативности операторов строятся *отношения операторного предшествования* (*отношения приоритета операторов*) $\doteq$, $\lessdot$ и $\gtrdot$, определённые на парах терминалов и маркеров $\vdash$, $\dashv$.
- Последний терминал основы будет ограничен самым левым отношением $\gtrdot$, первый терминал основы — ближайшим к концу основы отношением $\lessdot$ таким, что между этими $\lessdot$ и $\gtrdot$ выполняется только $\doteq$.

Замечания о форме и основе формы грамматики выражений

- В синтаксически правильном выражении два операнда не могут стоять рядом. Следовательно, никакая форма не содержит двух идущих подряд нетерминалов.
- Два терминала, входящие в форму, назовём *соседними*, если между ними в этой форме не стоят другие терминалы. (Что может стоять в форме между соседними терминалами?)
- Если $\mathtt{a}E\mathtt{b}$ входит в форму и одно из этих вхождений $\mathtt{a}$ или $\mathtt{b}$ входит в основу этой формы, то это вхождение E входит в основу.

Определение отношений приоритета

Пусть $\mathtt{a}$ и $\mathtt{b}$ — упорядоченная пара соседних терминалов. Тогда положим

- $\mathtt{a} \doteq \mathtt{b}$, если $\mathtt{a}$ и $\mathtt{b}$ должны быть свёрнуты на одном шаге;
- $\mathtt{a} \lessdot \mathtt{b}$, если $\mathtt{b}$ должен быть свёрнут раньше $\mathtt{a}$;
- $\mathtt{a} \gtrdot \mathtt{b}$, если $\mathtt{a}$ должен быть свёрнут раньше $\mathtt{b}$;
- $\vdash \lessdot \mathtt{b}$, если $\mathtt{b}$ может быть первым терминалом формы;
- $\mathtt{a} \gtrdot \dashv$, если $\mathtt{a}$ может быть последним терминалом формы.

Пример вычисления отношений приоритета

- Грамматика: $E \to E * E \mid E + E \mid -E \mid \mathrm{min}(E; E) \mid (E) \mid \mathtt{x}$.
- Операторы по убыванию приоритета: min, $-, *, +$.
- Ассоциативность операторов: $-$ — правая, $+$ и $*$ — левая.

	$+$	$*$	$-$	min	$\mathtt{x}$	(	)	;	$\dashv$
$+$	$\gtrdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\gtrdot$	$\gtrdot$	$\gtrdot$
$*$	$\gtrdot$	$\gtrdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\gtrdot$	$\gtrdot$	$\gtrdot$
$-$	$\gtrdot$	$\gtrdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\gtrdot$	$\gtrdot$	$\gtrdot$
min						$\doteq$			
$\mathtt{x}$	$\gtrdot$	$\gtrdot$	$\gtrdot$				$\gtrdot$	$\gtrdot$	$\gtrdot$
(	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\doteq$	$\doteq$	
)	$\gtrdot$	$\gtrdot$	$\gtrdot$				$\gtrdot$	$\gtrdot$	$\gtrdot$
;	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\doteq$		
$\vdash$	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$	$\lessdot$			

Пример восходящего анализа на основе отношений приоритета

Грамматика: $E \to E * E \mid E + E \mid -E \mid \mathsf{min}(E; E) \mid (E) \mid \mathrm{x}$.
Цепочка $\mathrm{x} * -\mathsf{min}(\mathrm{x}, \mathrm{x})$.

- $\vdash \lessdot \mathrm{x} \gtrdot * \lessdot - \lessdot \mathsf{min} \doteq (\lessdot \mathrm{x} \gtrdot ; \lessdot \mathrm{x} \gtrdot) \gtrdot \dashv$
- $\vdash_E \lessdot * \lessdot - \lessdot \mathsf{min} \doteq (\lessdot \mathrm{x} \gtrdot ; \lessdot \mathrm{x} \gtrdot) \gtrdot \dashv$
- $\vdash_E \lessdot * \lessdot - \lessdot \mathsf{min} \doteq (_E \doteq ; \lessdot \mathrm{x} \gtrdot) \gtrdot \dashv$
- $\vdash_E \lessdot * \lessdot - \lessdot \mathsf{min} \doteq (_E \doteq ;_E \doteq) \gtrdot \dashv$
- $\vdash_E \lessdot * \lessdot -_E \gtrdot \dashv$
- $\vdash_E \lessdot *_E \gtrdot \dashv$
- $\vdash_E \dashv$

Цепочка $\mathrm{x} - \mathsf{min}(\mathrm{x}, \mathrm{x})$.

- $\ldots$
- $\vdash_E \lessdot -_E \gtrdot \dashv$, основа: $E - E$, синтаксическая ошибка.

Другой подход к построению отношений операторного предшествования — 1

- Пусть дана грамматика $G = (\Sigma, \Gamma, P, S)$, в которой правые части правил не содержат подряд идущих нетерминалов (*операторная* грамматика).
- *Отношения операторного предшествования* $\doteq$, $\lessdot$ и $\gtrdot$ определяются на множестве $\Sigma \cup \{\vdash, \dashv\}$:
  - $\mathtt{a} \doteq \mathtt{b}$, если $A \to \alpha\mathtt{a}\gamma\mathtt{b}\beta \in P$ и $\gamma \in \Gamma \cup \{\varepsilon\}$;
  - $\mathtt{a} \lessdot \mathtt{b}$, если $A \to \alpha\mathtt{a}B\beta \in P$ и $B \Rightarrow^{+} \gamma\mathtt{b}\delta$, где $\gamma \in \Gamma \cup \{\varepsilon\}$;
  - $\mathtt{a} \gtrdot \mathtt{b}$, если $A \to \alpha B\mathtt{b}\beta \in P$ и $B \Rightarrow^{+} \delta\mathtt{a}\gamma$, где $\gamma \in \Gamma \cup \{\varepsilon\}$;
  - $\vdash \lessdot \mathtt{b}$, если $S \Rightarrow^{+} \gamma\mathtt{b}\alpha$ и $\gamma \in \Gamma \cup \{\varepsilon\}$;
  - $\mathtt{a} \gtrdot \dashv$, если $S \Rightarrow^{+} \alpha\mathtt{a}\gamma$ и $\gamma \in \Gamma \cup \{\varepsilon\}$.
- Операторная грамматика называется *грамматикой операторного предшествования*, если для любой упорядоченной пары терминалов выполняется не более одного отношения операторного предшествования.

Другой подход к построению отношений операторного предшествования — 2

Предложение

Пусть в операторной грамматике

$$\vdash S \dashv \Rightarrow^{*} \alpha A w \Rightarrow \alpha\beta w.$$

Тогда

- *отношение* $\doteq$ *или* $\lessdot$ *выполняется между соседними терминалами (и маркером* $\vdash$*) цепочки* α*;*
- *отношение* $\lessdot$ *выполняется между самым правым терминалом цепочки* α *и самым левым терминалом цепочки* β*;*
- *отношение* $\doteq$ *выполняется между соседними терминалами цепочки* β*;*
- *отношение* $\gtrdot$ *выполняется между самым правым терминалом цепочки* β *и первым символом цепочки* w*.*

Другой подход к построению отношений операторного предшествования — 3

- После построения отношений операторного предшествования все нетерминалы грамматики операторного предшествования отождествляются и удаляется правило вида $A \to A$.
- Теперь можно проводить восходящий анализ для полученной грамматики. (Но язык, порождаемый полученной грамматикой, может оказаться собственным надмножеством языка, порождаемого исходной грамматикой.)
- Не порождает ли исходная грамматика некоторый язык выражений? Тогда анализатор можно построить так, как описано в начале этого раздела.

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`).