Лекции по теории формальных языков

Лекция 12.
Общая схема LR-анализа.
Активные префиксы и LR(k)-пункты

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

29 апреля 2011 г.

План

1. Общая схема LR-анализа
2. Активные префиксы и LR(k)-пункты

План

1. Общая схема LR-анализа
2. Активные префиксы и LR(k)-пункты

Соглашения

- Восходящий анализ для класса LR(k)-грамматик.
- k — число символов анализируемой цепочки правее конца предполагаемой основы, которые просматриваются для принятия решения о свёртке.
- По умолчанию считаем, что
  - $k = 1$,
  - рассматриваемые грамматики приведённые и однозначные (но, быть может, с ε-правилами),
  - выводы — правые.

Схема LR-анализатора — 1

- Анализатор типа «перенос-свёртка».
- В стек записываются не символы грамматики, а *состояния*, но состояние однозначно определяет символ.
- Символу X может соответствовать несколько состояний, обозначаемых X^i (если одно, то оно обозначается самим символом X).
- Если $X_1^{i_1} \dots X_n^{i_n}$ — содержимое стека (вершина справа), v — необработанная часть входа, то $X_1 \dots X_n v$ — r-форма.
- Действия (шаги) анализатора:
  - перенос,
  - свёртка,
  - допуск.
- Решение об очередном действии анализатор принимает, исходя из
  - состояния на вершине стека (*текущего* состояния) и
  - k следующих символов входной цепочки.

Схема LR-анализатора — 2

- Действия анализатора указываются в *таблице анализа*, строки которой проиндексированы состояниями и которая состоит из подтаблиц ACTION и GOTO.
- Столбцы подтаблицы ACTION проиндексированы терминалами и маркером $\dashv$.
- Столбцы подтаблицы GOTO проиндексированы нетерминалами.
- В клетке подтаблицы ACTION указано, что нужно
  - перенести данное состояние в стек,
  - сделать свёртку по данному правилу или
  - допустить входную цепочку.
- В клетке таблицы GOTO указано состояние, которое кладётся в стек сразу после свёртки.

Таблица LR-анализа: пример

Грамматика GA_2: (1) $E \to E + T$, (2) $E \to T$, (3) $T \to T * F$,
(4) $T \to F$, (5) $F \to (E)$, (6) $F \to \mathtt{x}$.

	ACTION						GOTO		
	$+$	$*$	$\mathtt{x}$	$($	$)$	$\dashv$	E	T	F
E^1	$\leftarrow +$					✓			
T^1	$\otimes 2$	$\leftarrow *$			$\otimes 2$	$\otimes 2$			
F^1	$\otimes 4$	$\otimes 4$			$\otimes 4$	$\otimes 4$			
$($			$\leftarrow \mathtt{x}$	$\leftarrow ($			E^2	T^1	F^1
$\mathtt{x}$	$\otimes 6$	$\otimes 6$			$\otimes 6$	$\otimes 6$			
$+$			$\leftarrow \mathtt{x}$	$\leftarrow ($				T^2	F^1
$*$			$\leftarrow \mathtt{x}$	$\leftarrow ($					F^2
E^2	$\leftarrow +$				$\leftarrow)$				
T^2	$\otimes 1$	$\leftarrow *$			$\otimes 1$	$\otimes 1$			
F^2	$\otimes 3$	$\otimes 3$			$\otimes 3$	$\otimes 3$			
$)$	$\otimes 5$	$\otimes 5$			$\otimes 5$	$\otimes 5$			
∇			$\leftarrow \mathtt{x}$	$\leftarrow ($			E^1	T^1	F^1

Протокол работы LR-анализатора

Такт	Содержимое стека	Позиция указателя	Действие
1	∇	$\diamond \mathtt{x} + \mathtt{x} * (\mathtt{x} + \mathtt{x}) \dashv$	$\leftarrow \mathtt{x}$
2	$\nabla \mathtt{x}$	$\mathtt{x} \diamond + \mathtt{x} * (\mathtt{x} + \mathtt{x}) \dashv$	$\otimes 6\ F \longrightarrow \mathtt{x}$
3	∇F^1	$\mathtt{x} \diamond + \mathtt{x} * (\mathtt{x} + \mathtt{x}) \dashv$	$\otimes 4\ T \longrightarrow F$
4	∇T^1	$\mathtt{x} \diamond + \mathtt{x} * (\mathtt{x} + \mathtt{x}) \dashv$	$\otimes 2\ E \longrightarrow T$
5	∇E^1	$\mathtt{x} \diamond + \mathtt{x} * (\mathtt{x} + \mathtt{x}) \dashv$	$\leftarrow +$
6	$\nabla E^1 +$	$\mathtt{x} + \diamond \mathtt{x} * (\mathtt{x} + \mathtt{x}) \dashv$	$\leftarrow \mathtt{x}$
7	$\nabla E^1 + \mathtt{x}$	$\mathtt{x} + \mathtt{x} \diamond * (\mathtt{x} + \mathtt{x}) \dashv$	$\otimes 6\ F \longrightarrow \mathtt{x}$
8	$\nabla E^1 + F^1$	$\mathtt{x} + \mathtt{x} \diamond * (\mathtt{x} + \mathtt{x}) \dashv$	$\otimes 4\ T \longrightarrow F$
9	$\nabla E^1 + T^2$	$\mathtt{x} + \mathtt{x} \diamond * (\mathtt{x} + \mathtt{x}) \dashv$	$\leftarrow *$
10	$\nabla E^1 + T^2 *$	$\mathtt{x} + \mathtt{x} * \diamond (\mathtt{x} + \mathtt{x}) \dashv$	$\leftarrow ($
11	$\nabla E^1 + T^2 * ($	$\mathtt{x} + \mathtt{x} * (\diamond \mathtt{x} + \mathtt{x}) \dashv$	$\leftarrow \mathtt{x}$
12	$\nabla E^1 + T^2 * (\mathtt{x}$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} \diamond + \mathtt{x}) \dashv$	$\otimes 6\ F \longrightarrow \mathtt{x}$

Протокол работы LR-анализатора: окончание

Такт	Содержимое стека	Позиция указателя	Действие
12	$\nabla E^1 + T^2 * (\mathtt{x}$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} \diamond + \mathtt{x}) \dashv$	$\otimes 6\ F \to \mathtt{x}$
13	$\nabla E^1 + T^2 * (F^1$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} \diamond + \mathtt{x}) \dashv$	$\otimes 4\ T \to F$
14	$\nabla E^1 + T^2 * (T^1$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} \diamond + \mathtt{x}) \dashv$	$\otimes 2\ E \to T$
15	$\nabla E^1 + T^2 * (E^2$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} \diamond + \mathtt{x}) \dashv$	$\leftarrow +$
16	$\nabla E^1 + T^2 * (E^2 +$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} + \diamond \mathtt{x}) \dashv$	$\leftarrow \mathtt{x}$
17	$\nabla E^1 + T^2 * (E^2 + \mathtt{x}$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} + \mathtt{x} \diamond) \dashv$	$\otimes 6\ F \to \mathtt{x}$
18	$\nabla E^1 + T^2 * (E^2 + F^1$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} + \mathtt{x} \diamond) \dashv$	$\otimes 4\ T \to F$
19	$\nabla E^1 + T^2 * (E^2 + T^2$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} + \mathtt{x} \diamond) \dashv$	$\otimes 1\ E \to E + T$
20	$\nabla E^1 + T^2 * (E^2$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} + \mathtt{x} \diamond) \dashv$	$\leftarrow)$
21	$\nabla E^1 + T^2 * (E^2)$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} + \mathtt{x}) \diamond \dashv$	$\otimes 5\ F \to (E)$
22	$\nabla E^1 + T^2 * F^2$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} + \mathtt{x}) \diamond \dashv$	$\otimes 3\ T \to T * F$
23	$\nabla E^1 + T^2$	$\mathtt{x} + \mathtt{x} * (\mathtt{x} + \mathtt{x}) \diamond \dashv$	$\otimes 1\ E \to E + T$
24	∇E^1	$\mathtt{x} + \mathtt{x} * (\mathtt{x} + \mathtt{x}) \diamond \dashv$	$\checkmark$

Расширенная грамматика

- Пусть $G = (\Sigma, \Gamma, P, S)$ — КС-грамматика. Её *расширенной* (*пополненной*) грамматикой называется грамматика $G' = (\Sigma,\ \Gamma \cup \{S'\},\ P \cup \{S' \to S\},\ S')$, где S' — новый нетерминал.
- Далее предполагаем, что перед построением LR-анализатора грамматика расширяется.
- Свёртка по правилу $S' \to S$ сигнализирует о допуске входной цепочки.

План

1 Общая схема LR-анализа

2 Активные префиксы и LR(k)-пункты

Активные префиксы и LR(k)-пункты: определения

- Ключевым этапом построения LR-анализатора является построение ДКА, распознающего в точности все цепочки, которые могут оказаться в стеке при анализе правильных входных цепочек.
- Префикс r-формы грамматики, не выходящий за (правый) конец основы этой формы, называется *активным префиксом*.
- *LR(k)-пунктом* (*LR(k)-ситуацией*) грамматики называется $[A \to \beta_1 \cdot \beta_2, v]$, где
  - $A \to \beta_1\beta_2$ — правило вывода,
  - v — цепочка терминалов длины k или $v = u \dashv$ для некоторой цепочки терминалов u длины, меньшей k.

  Пустые цепочки в записи LR(k)-пункта опускаются.
- LR(k)-пункт $[A \to \beta_1 \cdot \beta_2, v]$ называется *допустимым для активного префикса* $\alpha\beta_1$, если существует (правый) вывод

$$S' \Rightarrow^* \alpha A w \Rightarrow \alpha\beta_1\beta_2 w \Rightarrow^* uw$$

и цепочка v является префиксом цепочки $w \dashv$.

Автомат пунктов грамматики. Формулировка основной теоремы LR-анализа

- В текущей лекции мы будем рассматривать только LR(0)-пункты и называть их просто *пунктами*.
- *Автоматом пунктов* (расширенной) грамматики $G = (\Sigma, \Gamma, P, S')$ называется ε-НКА $\mathcal{I}_G = (Q,\ \Sigma \cup \Gamma,\ \delta,\ \{\mathtt{i}_0\},\ Q)$, где
  - Q — множество всех пунктов грамматики G,
  - $\mathtt{i}_0 = [S' \to \cdot S]$,
  - отношение переходов состоит из всех переходов вида $([A \to \beta_1 \cdot X \beta_2],\ X,\ [A \to \beta_1 X \cdot \beta_2])$ и $([A \to \beta_1 \cdot B \beta_2],\ \varepsilon,\ [B \to \cdot \beta])$.

Теорема (основная теорема LR-анализа)

Пункт $\mathtt{i}$ *грамматики* G *является допустимым для активного префикса* γ *этой грамматики тогда и только тогда, когда в автомате* $\mathcal{I}_G$ *существует путь из* $\mathtt{i}_0$ *в* $\mathtt{i}$, *помеченный цепочкой* γ.

Базисные пункты и переходы. Лемма об активном префиксе и базисном пункте

- *Базисный пункт* — любой пункт, в котором точка стоит не в начале правой части правила, а также пункт $[S' \to \cdot S]$.
- *Базисный переход* — любой переход автомата $\mathcal{I}_G$ с меткой, отличной от ε.

Лемма

Пусть γ — активный префикс. Тогда найдётся базисный пункт, допустимый для γ.

Доказательство.

- Пусть $\gamma\alpha$ — первая (ближайшая к S) r-форма в выводе $S' \Rightarrow^* w$, для которой γ является активным префиксом.
- Если форма $\gamma\alpha$ появилась на первом шаге этого вывода, то на этом шаге применено правило $S' \to S$, и тогда
  - либо $\gamma = \varepsilon$ и допустимый для γ базисный пункт есть $[S' \to \cdot S]$,
  - либо $\gamma = S$ и допустимый для γ базисный пункт есть $[S' \to S\cdot]$.

Лемма: окончание доказательства

- Далее рассматриваем случай, когда форма $\gamma\alpha$ появилась не на первом шаге данного вывода.
- Основа формы $\gamma\alpha$ не входит в α, поскольку иначе γ был бы активным префиксом предыдущей формы.
- Тогда так как γ не выходит за конец основы формы $\gamma\alpha$, то $\gamma\alpha = \underbrace{\gamma'\beta_1}_{\gamma}\underbrace{\beta_2 u}_{\alpha}$, где $\beta_1\beta_2$ — основа формы $\gamma\alpha$, $\beta_1 \neq \varepsilon$.
- Следовательно, существует правило вывода $A \to \beta_1\beta_2$ и данный вывод $S' \Rightarrow^* w$ имеет вид

$$S' \Rightarrow^* \gamma' A u \Rightarrow \gamma'\beta_1\beta_2 u = \gamma\alpha \Rightarrow^* w.$$

- Таким образом, пункт $[A \to \beta_1 \cdot \beta_2]$ является базисным и допустимым для γ.

□

Замечание. Из доказательства этой леммы видно, что активный префикс γ впервые появляется в выводе сразу после применения правила, один из базисных пунктов которого допустим для γ.

Лемма о достижимости допустимого пункта из базисного пункта

Лемма

Пункт $[B \to \cdot\beta]$ допустим для активного префикса γ тогда и только тогда, когда он достижим по ε-переходам автомата $\mathcal{I}_G$ из некоторого базисного пункта, допустимого для γ.

Доказательство.

Достаточность.

- Пусть пункт $[A \to \beta_1 \cdot B\beta_2]$ допустим для $\gamma = \gamma'\beta_1$ и из этого пункта по ε-переходу $([A \to \beta_1 \cdot B\beta_2],\ \varepsilon,\ [B \to \cdot\beta])$ достижим пункт $[B \to \cdot\beta]$.
- Тогда

$$S' \Rightarrow^* \gamma' Au \Rightarrow \gamma'\beta_1 B\beta_2 u = \gamma B\beta_2 u \Rightarrow^* \gamma Bvu \Rightarrow \gamma\beta vu \Rightarrow^* w,$$

  поэтому пункт $[B \to \cdot\beta]$ допустим для γ.
- Следовательно, пункт $[B \to \cdot\beta]$ допустим для γ, если он достижим по ε-переходам из допустимого для γ пункта.

Лемма: продолжение доказательства

Необходимость.

- Пункт $[B \to \cdot\beta]$ допустим для активного префикса γ.
- Тогда существует вывод

$$S' \Rightarrow^* \gamma B v \Rightarrow \gamma\beta v \Rightarrow^* w. \qquad (\star)$$

- По предыдущему замечанию активный префикс γ впервые появляется в этом выводе сразу после применения правила, один из базисных пунктов которого допустим для γ. Пусть таким базисным пунктом является $[A \to \beta_1 \cdot \beta_2]$.
- Тогда вывод $(\star)$ имеет вид

$$S' \Rightarrow^* \gamma' A u \Rightarrow \gamma'\beta_1\beta_2 u = \gamma\beta_2 u \Rightarrow^* \gamma B v \Rightarrow \gamma\beta v \Rightarrow^* w,$$

где $\beta_2 = C\beta_2' \Rightarrow^* Bv'$.

- Если $C = B$, то из пункта $[A \to \beta_1 \cdot B\beta_2']$ по ε-переходу достижим пункт $[B \to \cdot\beta]$.

Лемма: окончание доказательства

- Иначе найдётся вывод

$$C \Rightarrow B_1\alpha_1 \Rightarrow^* B_1u_1 \Rightarrow B_2\alpha_2u_1 \Rightarrow^* B_2u_2 \Rightarrow \ldots \Rightarrow^* B_ku_k = Bu_k,$$

в котором используются (среди прочих) правила

$$C \to B_1\alpha,\ B_1 \to B_2\alpha_2,\ \ldots,\ B_{k-1} \to B\alpha_k.$$

- Значит, в автомате $\mathcal{I}_G$ имеются ε-переходы

$$\begin{gathered}([A \to \beta_1 \cdot C\beta_2'], \varepsilon,\ [C \to \cdot B_1\alpha_1]),\\ ([C \to \cdot B_1\alpha_1], \varepsilon,\ [B_1 \to \cdot B_2\alpha_2]),\\ \ldots,\\ ([B_{k-1} \to \cdot B\alpha_k], \varepsilon,\ [B \to \cdot\beta]).\end{gathered}$$

Теорема (основная теорема LR-анализа)

Пункт $\mathtt{i}$ *грамматики* G *является допустимым для активного префикса* γ *этой грамматики тогда и только тогда, когда в автомате* $\mathcal{I}_G$ *существует путь из* $\mathtt{i}_0$ *в* $\mathtt{i}$, *помеченный цепочкой* γ.

Доказательство. *Достаточность.* Индукция по $|\gamma|$.

- База индукции. $\gamma = \varepsilon$.
  - Базисный пункт $\mathtt{i} = \mathtt{i}_0 = [S' \to \cdot S]$ допустим для γ.
  - По предыдущей лемме любой пункт, достижимый из $\mathtt{i}_0$ по ε-переходам, допустим для γ.
- Индукционный переход. $\gamma = \gamma_0 X$.
  - Последний базисный переход в пути из $\mathtt{i}_0$ в $\mathtt{i}$, помеченном γ, имеет вид $([A \to \beta_1 \cdot X\beta_2],\ X,\ [A \to \beta_1 X \cdot \beta_2])$.
  - По индукционному предположению $[A \to \beta_1 \cdot X\beta_2]$ допустим для γ_0.
  - Следовательно, $\gamma_0 = \gamma'\beta_1$ и существует вывод

$$S' \Rightarrow^* \gamma' A u \Rightarrow \gamma'\beta_1 X \beta_2 u \Rightarrow^* w. \qquad (\star\star)$$

  - Тогда базисный пункт $\widetilde{\mathtt{i}} = [A \to \beta_1 X \cdot \beta_2]$ допустим для $\gamma = \gamma'\beta_1 X$.
  - Если $\widetilde{\mathtt{i}} \neq \mathtt{i}$, то из пункта $\widetilde{\mathtt{i}}$ по ε-переходам достижим пункт $\mathtt{i}$. Поэтому в силу предыдущей леммы пункт $\mathtt{i}$ допустим для γ.

Теорема: окончание доказательства

Необходимость. Индукция по $|\gamma|$.

- База индукции. $\gamma = \varepsilon$.
  - Пусть пункт $\mathtt{i} = [A \to \beta_1 \cdot \beta_2]$ допустим для γ.
  - Тогда $\beta_1 = \gamma = \varepsilon$, и потому единственный базисный пункт, допустимый для γ, есть $\mathtt{i}_0 = [S' \to \cdot S]$.
  - По предыдущей лемме пункт $\mathtt{i}$ достижим из $\mathtt{i}_0$ по ε-переходам.
- Индукционный переход. $\gamma = \gamma_0 X$.
  - По лемме на слайде 14 для активного префикса γ найдётся допустимый базисный пункт $\widetilde{\mathtt{i}} = [A \to \beta_1 X \cdot \beta_2]$.
  - Тогда $\gamma = \gamma' \beta_1 X$ и существует вывод $(\star\star)$.
  - Следовательно, пункт $[A \to \beta_1 \cdot X \beta_2]$ допустим для $\gamma_0 = \gamma' \beta_1$.
  - По индукционному предположению существует путь из $\mathtt{i}_0$ в $[A \to \beta_1 \cdot X \beta_2]$, помеченный цепочкой γ_0.
  - Добавим к этому пути переход $([A \to \beta_1 \cdot X \beta_2],\ X,\ [A \to \beta_1 X \cdot \beta_2])$ и получим путь из $\mathtt{i}_0$ в $\widetilde{\mathtt{i}}$, помеченный γ.
  - Проведённое рассуждение верно для любого базисного пункта $\widetilde{\mathtt{i}}$, допустимого для γ.
  - Если пункт $\mathtt{i}$ не базисный, то по предыдущей лемме он достижим из некоторого базисного допустимого для γ пункта по ε-переходам.

Следствия основной теоремы LR-анализа. LR(0)-автомат

Следствие

Язык, распознаваемый автоматом $\mathcal{I}_G$, совпадает с множеством всех активных префиксов грамматики G.

- ДКА (неполный) $\mathcal{A}_G$, построенный по ε-НКА $\mathcal{I}_G$ алгоритмом из лекции 2, будем называть *LR(0)-автоматом* грамматики G.
- Множество состояний $\mathcal{A}_G$ называют *канонической системой пунктов* грамматики G.
- Состояние ДКА есть множество всех состояний ε-НКА, которые достижимы из начального состояния по путям, помеченным фиксированной цепочкой.

Следствие

Состояние автомата $\mathcal{A}_G$, достижимое из начального состояния по пути, помеченному цепочкой γ, есть множество всех пунктов, допустимых для активного префикса γ.

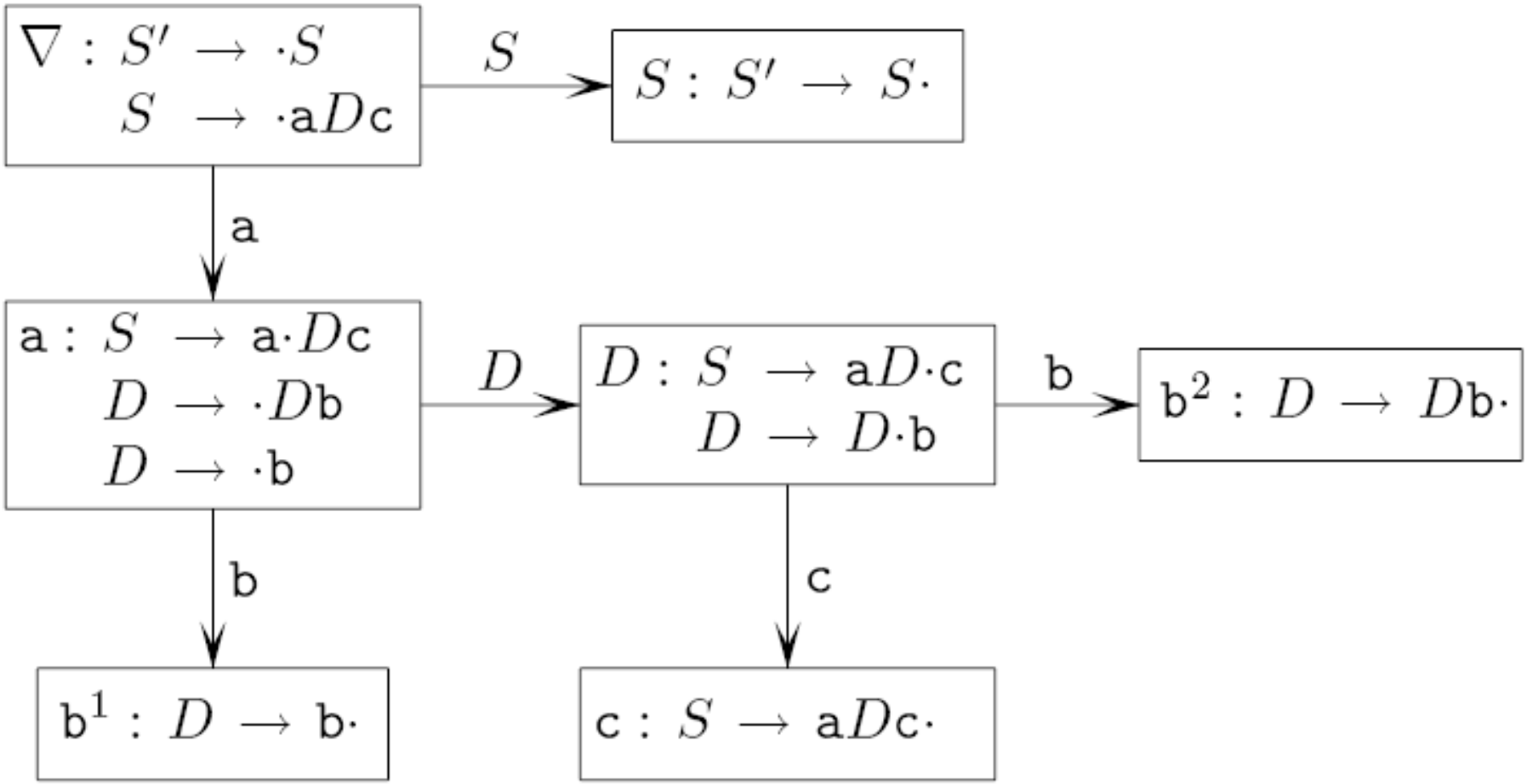
LR(0)-автомат: замечание об определении состоянием последнего прочитанного символа

- В автомате $\mathcal{I}_G$ в пункт $[A \to \alpha_1 X \cdot \alpha_2]$ ведёт только переход с меткой X.
- Если пункты $[A \to \alpha_1 X \cdot \alpha_2]$ и $[B \to \beta_1 Y \cdot \beta_2]$ принадлежат одному состоянию LR(0)-автомата, то $X = Y$.
- Состояние, в которое пришёл LR(0)-автомат при обработке некоторой цепочки, однозначно определяет последний прочитанный символ; причём если это состояние является начальным, то прочитанная часть цепочки пуста.
- Этот символ используем для обозначения состояния, при необходимости добавляя индекс.
- Для обозначения начального состояния используем символ ∇.

Пример построения LR(0)-автомата

Грамматика $G_1 = \{S' \to S,\ S \to \mathtt{a}D\mathtt{c},\ D \to D\mathtt{b},\ D \to \mathtt{b}\}$.

Переходы в $\mathcal{I}_{G_1}$ имеют вид $([A \to \beta_1 \cdot X\beta_2],\ X,\ [A \to \beta_1 X \cdot \beta_2])$ и $([A \to \beta_1 \cdot B\beta_2],\ \varepsilon,\ [B \to \cdot\beta])$.



Функции CLOSURE и GOTO

Пусть M — произвольное множество пунктов расширенной грамматики G.

- Положим, что $\textsc{closure}(M)$ есть минимальное по включению множество пунктов такое, что
  - $M \subseteq \textsc{closure}(M)$ и
  - $[A \to \alpha \cdot B\beta] \in \textsc{closure}(M)$ влечёт $[B \to \cdot\gamma] \in \textsc{closure}(M)$ для каждого правила вывода вида $B \to \gamma$.
- Пусть X — терминал или нетерминал грамматики G. Положим

$$\textsc{goto}(M, X) = \textsc{closure}(\{[A \to \alpha X \cdot \beta] \mid [A \to \alpha \cdot X\beta] \in M\}).$$

- GOTO есть функция переходов LR(0)-автомата $\mathcal{A}_G$.

Алгоритм построения LR(0)-автомата

Вход. Расширенная грамматика $G = (\Sigma, \Gamma, P, S')$.
Выход. ДКА $\mathcal{A}_G = (Q, \Sigma \cup \Gamma, \delta, \mathtt{I}_0, Q)$.

1. $\mathtt{I}_0 := \textsc{closure}(\{[S' \to \cdot S]\})$; $\mathit{label}(\mathtt{I}_0) := 0$;
2. $Q := \{\mathtt{I}_0\}$;
3. **пока** $(\exists \mathtt{I} \in Q : \mathit{label}(\mathtt{I}) = 0)$ **повторять**
4. **для каждого** $X \in \Sigma \cup \Gamma$
5. $\delta(\mathtt{I}, X) := \textsc{goto}(\mathtt{I}, X)$;
6. **если** $(\delta(\mathtt{I}, X) \notin Q)$
7. $Q := Q \cup \{\delta(\mathtt{I}, X)\}$;
8. $\mathit{label}(\delta(\mathtt{I}, X)) := 0$;
9. $\mathit{label}(\mathtt{I}) := 1$

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`).