§ 1. LR(0)-автомат. LR(0)-грамматика. Анализатор на основе LR(0)-автомата

Ключевые определения:

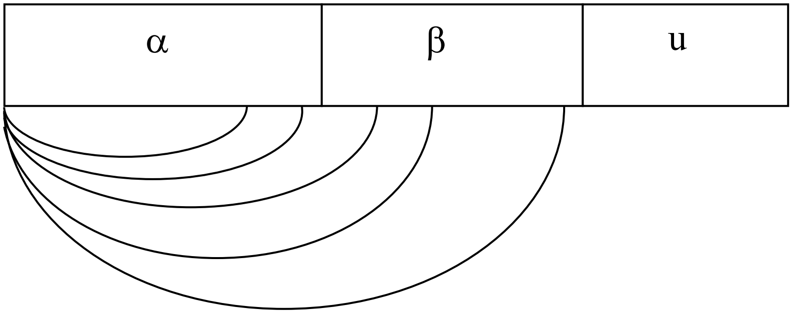
Опр. Пусть $S \Rightarrow \gamma_1 \Rightarrow \ldots \Rightarrow \gamma_m = \omega$ правосторонний вывод цепочки ω из аксиомы S.

r-формой (right-form) называется каждая γ_i .

Опр. Пусть в правостороннем выводе $\alpha A u \Rightarrow \alpha\beta u$ было использовано правило $A \to \beta$.
Основой r-формы $\alpha\beta u$ называется β.

Комментарий: при работе анализатора цепочка β на верхушке стека свернулась в A.

Опр. Активным префиксом r-формы $\alpha\beta u$ называется любой префикс цепочки $\alpha\beta$ (т.е. префикс, не выходящий за правый конец основы).



Опр. LR(0)-пунктом (пункт) называется набор (A, β_1, β_2), где $A \to \beta_1 \beta_2$ – правило грамматики.

Обозначение пункта: $[A \to \beta_1 \bullet \beta_2]$ («Правило с точкой»).

Все LR(0)-пункты – допустимые. Определение допустимости будет позже.

Комментарий: очевидно для любой грамматики количество всех возможных пунктов конечно.

Опр. Пусть $G = (\Sigma, \Gamma, P, S)$ – грамматика. Расширенной грамматикой называется $G' = (\Sigma, \Gamma\cup\{S'\}, P\cup\{S'\to S\}, S')$, где $S' \notin \Gamma$ (новый нетерминал = новая аксиома).

Опр. Автоматом пунктов расширенной грамматики G = (Σ, Γ, P, S′) называется ε-НКА (Q, Σ ∪ Γ, δ, q_0, Q), где

Q – множество пунктов (один пункт = одно состояние);

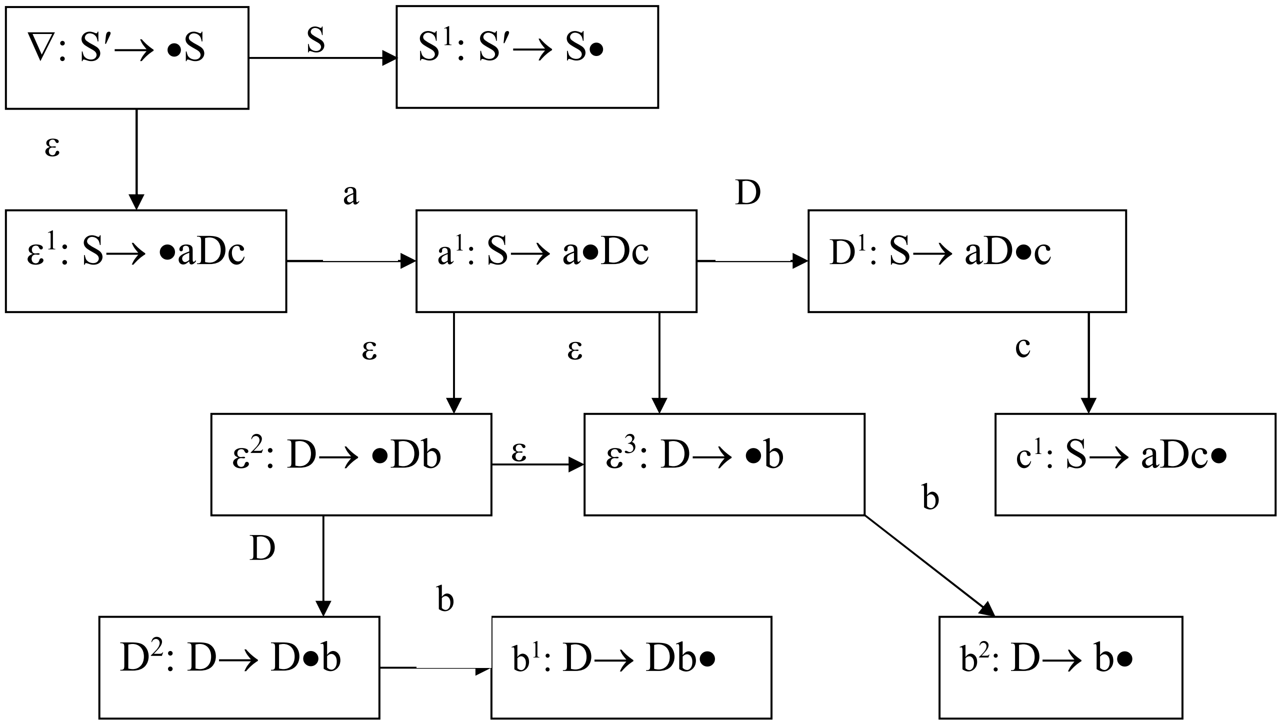
q_0 = [S′→ •S];

$$\delta = \bigcup_{\forall X \in \Sigma \cup \Gamma} \left(\delta\left(\left[A \to \beta_1 \bullet X\beta_3 \right], X \right) = \left[A \to \beta_1 X \bullet \beta_3 \right] \right) \cup$$

$$\cup \bigcup_{\forall B \in \Gamma} \left(\delta\left(\left[A \to \beta_1 \bullet B\beta_3 \right], \varepsilon \right) = \left[B \to \bullet\gamma \right] \right).$$

Пример. $G=\{S' \to S,\ S \to aDc,\ D \to Db \mid b\}$.
(стр. 161-162 книжки Шура и Замятина)

$G=\{S'\to S,\ S\to aDc,\ D\to Db \mid b\}$.



Опр. LR(0)-автоматом грамматики G называется ε-замыкание автомата пунктов.

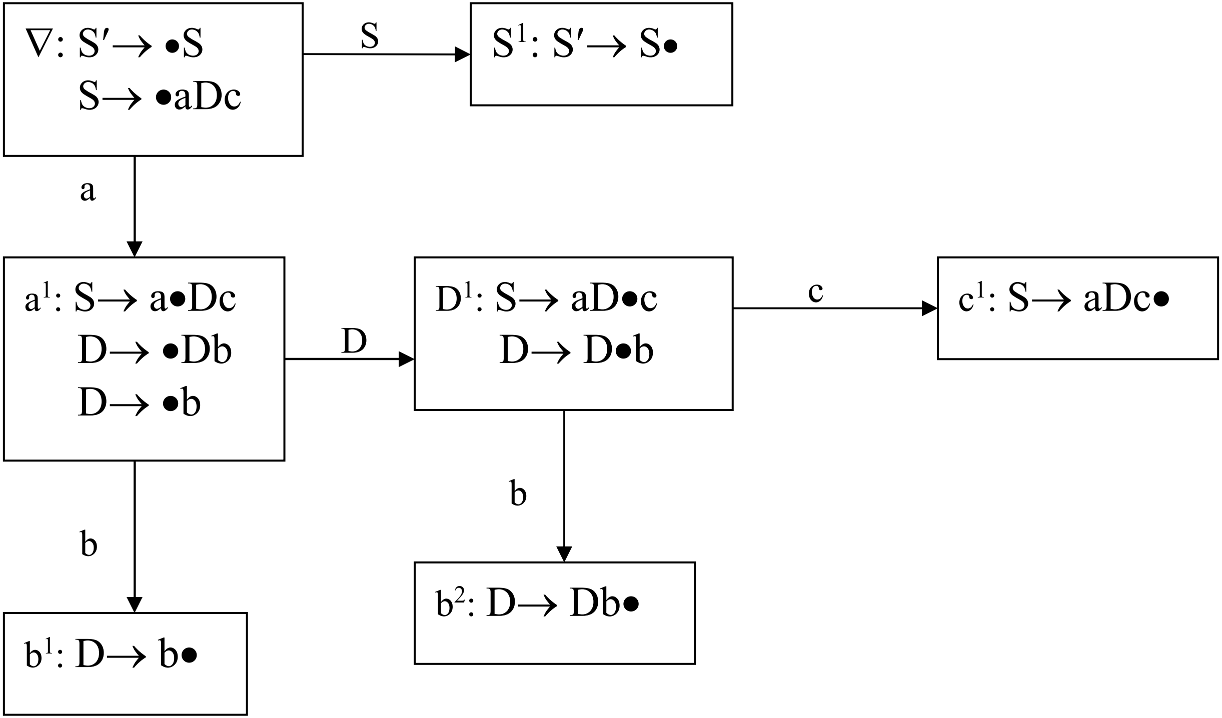
Комментарий: в задачах на практических занятиях будем строить ε-замыкание без поиска ε-НКА пунктов.

Пример. $G=\{S' \to S,\ S \to aDc,\ D \to Db \mid b\}$.

1) Каждое состояние LR(0)-автомата – набор пунктов, где $A \to \alpha \bullet B\beta$ влечет за собой $B \to \bullet\gamma$.

2) Один пункт может содержаться в нескольких состояниях.

3) Состояния равны, только если множества пунктов совпадают.


∇: S′→ •S
S→ •aDc
S
S¹: S′→ S•
a
a¹: S→ a•Dc
D→ •Db
D→ •b
D
D¹: S→ aD•c
D→ D•b
c
c¹: S→ aDc•
b
b¹: D→ b•
b
b²: D→ Db•

Опр. LR-анализатором называется таблица ACTION/GOTO, в которой строки соответствуют состояниям автомата, столбцы – символам $\Sigma \cup \Gamma \cup \{\dashv\}$:

	ACTION		GOTO
	Σ	$\dashv$	Γ
состояние			

В части ACTION применяются пометки вида:

$\surd$ = «допуск» (цепочка принадлежит языку);

$\leftarrow q$ = «перенос» (где q – название состояния в автомате, по которому строится анализатор);

$\otimes n$ = «свертка» по правилу с номером n.

Опр. Таблица анализатора бесконфликтная, если в каждой ячейке не более одного действия.

Существует два вида конфликтов:
перенос-свертка,
свертка-свертка.

Опр. (Один хороший случай бесконфликтного анализатора). Грамматика называется LR(0)-грамматикой, если в LR(0)-автомате каждое состояние, содержащее пункт с точкой на конце ($[A \to \alpha \bullet]$), состоит из единственного пункта.

Пример, рассмотренный ранее. $G = \{S' \to S,\ S \to aDc,\ D \to Db \mid b\}$.

Вопрос: является ли LR(0)-грамматикой?

Алгоритм построения LR(0)-анализатора
(бесконфликтный анализатор для LR(0)-грамматики).

Пусть $(Q, \Sigma \cup \Gamma, \delta, \nabla, Q)$ – LR(0)-автомат.
Заполнение ACTION:
1) В строке, соответствующей $[S' \to S\bullet]$ в столбце $\dashv$ заносим $\surd$ (допуск).
2) В строках, соотв. $[A \to \alpha\bullet]$, кроме заполненной в 1), в каждом столбце, включая $\dashv$ заносим $\otimes n$ (свертка).
3) В строках q, не заполненных в 1) и 2), для $a \in \Sigma \cup \{\dashv\}$ заносим $\leftarrow \delta(q,a)$, если $\delta(q,a) \neq \varnothing$ (перенос).

Заполнение GOTO:
В каждой строке q для $A \in \Gamma$ заносим $\delta(q,A)$, если $\delta(q,A) \neq \varnothing$.

Пример, рассмотренный ранее. $G = \{\underset{1}{S' \rightarrow S},\ \underset{2}{S \rightarrow aDc},\ \underset{3}{D \rightarrow Db} \mid \underset{4}{b}\}$.

	ACTION				GOTO	
	a	b	c	$\dashv$	S	D
∇						
a^1						
b^1						
b^2						
c^1						
S^1						
D^1						

	ACTION				GOTO	
	a	b	c	$\dashv$	S	D
∇	$\leftarrow a^1$				S^1	
a^1		$\leftarrow b^1$				D^1
b^1	$\otimes 4$	$\otimes 4$	$\otimes 4$	$\otimes 4$		
b^2	$\otimes 3$	$\otimes 3$	$\otimes 3$	$\otimes 3$		
c^1	$\otimes 2$	$\otimes 2$	$\otimes 2$	$\otimes 2$		
S^1				$\surd$		
D^1		$\leftarrow b^2$	$\leftarrow c^1$			

Комментарий: пустые ячейки в анализаторе означают, что входная цепочка не соответствует правилам языка.

В таблице ACTION можно добавить действия, называемые «обработчиком ошибок».

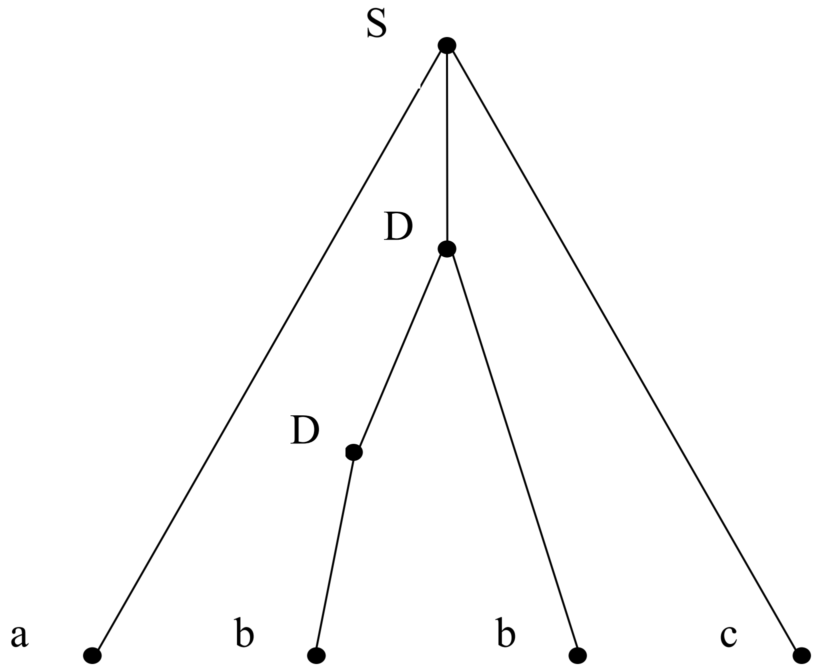
В таблице GOTO обращение к пустым ячейкам невозможно.

Пример. Рассмотрим анализ цепочки ω = *abbc*.

состояние	№ такта	стек	указатель	действие
∇	1	∇	$\lozenge abbc\dashv$	

состояние	№ такта	стек	указатель	действие
∇	1	∇	$\lozenge abbc\dashv$	
				перенос
a^1	2	$\nabla\ a^1$	$a\lozenge bbc\dashv$	
				перенос
b^1	3	$\nabla\ a^1\underline{b^1}$	$ab\lozenge bc\dashv$	
				свертка по 4
D^1=GOTO(a^1,D)	4	$\nabla\ a^1\mathrm{D}^1$	$ab\lozenge bc\dashv$	
				перенос
b^2	5	$\nabla\ a^1\underline{\mathrm{D}^1\mathrm{b}^2}$	$abb\lozenge c\dashv$	
				свертка по 3
D^1=GOTO(a^1,D)	6	$\nabla\ a^1\mathrm{D}^1$	$abb\lozenge c\dashv$	
				перенос
c^1	7	$\nabla\ \underline{a^1\mathrm{D}^1\mathrm{c}^1}$	$abbc\lozenge\dashv$	
				свертка по 2
S=GOTO(∇,S)	8	$\nabla\mathrm{S}$	$abbc\lozenge\dashv$	
				допуск

Восстановленное дерево вывода:



Восстановленный вывод:

$S \Rightarrow aDc \Rightarrow aDbc \Rightarrow abbc.$