§ 6. LR(k)-автомат. LR(k)-грамматики. LR-язык.

Опр. LR(k)-пунктом грамматики G называется набор $(A, \beta_1, \beta_2, \mathrm{v})$, где $A \to \beta_1\beta_2$ – правило грамматики, v – цепочка терминалов длины равной k, либо цепочка терминалов длины меньше k, дополненная в конце символом $\dashv$.

Обозначение: $\left[A \to \beta_1 \bullet \beta_2, \mathrm{v}\right]$

Опр. LR(k)-пункт $\left[A \to \beta_1 \bullet \beta_2, \mathrm{v}\right]$ называется допустимым для активного префикса $\alpha\beta_1$ некоторой r-формы, если существует вывод $S' \Rightarrow^* \alpha A\mathrm{w} \Rightarrow \alpha\beta_1\beta_2\mathrm{w} \Rightarrow^* \mathrm{uw}$, и цепочка v является префиксом цепочки $\mathrm{w}\dashv$.

Опр. LR(k)-автоматом грамматики G называется автомат, состояниями которого являются различные множества LR(k)-пунктов, полученные при ε-замыкании автомата LR(k)-пунктов.

Комментарий: LR(k)-автомат распознает множество всех активных префиксов грамматики G.

Опр. Грамматика G называется LR(k)-грамматикой, если из того, что при правостороннем выводе некоторой r-формы $\alpha\beta$u последним применялось правило $A \to \beta$, следует, что это же правило применялось последним при выводе любой r-формы $\alpha\beta$v, такой, что $\mathrm{FIRST}_k(v) = \mathrm{FIRST}_k(u)$.

(Напоминание: $u, v \in \Sigma^*$)

Обозначение $\mathrm{FIRST}_k(v)$ = первые k символов цепочки v, или все ее символы, если $|v| < k$.

Опр. Грамматика G называется LR-грамматикой, если она является LR(k)-грамматикой для $k \geq 0$.

Комментарий: $LR(0) \subseteq LR(1) \subseteq LR(2) \subseteq ... \subseteq LR(k) \subseteq \ldots \subseteq LR$.

Теорема. Для любого k выполняется строгое включение
$LR(k) \subset LR(k+1)$.

Исправленный комментарий:
$LR(0) \subset LR(1) \subset LR(2) \subset ... \subset LR(k) \subset \ldots \subset LR$.

Замечание: Существуют грамматики, не являющиеся LR(k)-грамматикой.

Пример. $G = \{S' \to S,\ S \to Bb \mid Cc,\ B \to Ba \mid \varepsilon,\ C \to Ca \mid \varepsilon,\ \}$.

Вопрос: $L(G) = ?$

Замечание: Существуют грамматики, не являющиеся LR(k)-грамматикой.

Пример. $G = \{S' \to S,\ S \to Bb \mid Cc,\ B \to Ba \mid \varepsilon,\ C \to Ca \mid \varepsilon,\ \}$.

$L(G) = \{a^*b,\ a^*c\}$.

Покажем, что ни для какого k G не является LR(k)-грамматикой.

(от противного) Предположим, G является LR(k)-грамматикой для некоторого k.

Рассмотрим выводы цепочек $a^k b$, $a^k c$.

$$S' \Rightarrow^* Ba^{k-1}b \Rightarrow Ba^k b \Rightarrow \underbrace{a^k b}_{\varepsilon u}, \quad (B\to\varepsilon \text{ – последнее правило})$$

$$S' \Rightarrow^* Ca^{k-1}c \Rightarrow Ca^k c \Rightarrow \underbrace{a^k c}_{\varepsilon v}, \quad (C\to\varepsilon \text{ – последнее правило})$$

Определение LR(k)-грамматики нарушено.

Опр. Язык называется LR(k)-языком, если существует LR(k)-грамматика, порождающая этот язык.

Список основных результатов без доказательства.

Теорема 1. Любой язык, порождаемый LR-грамматикой, порождается LR(1)-грамматикой.

Теорема 2. Любой язык, порождаемый LR-грамматикой, порождается SLR(1)-грамматикой.

Теорема 3. Класс LR-языков совпадает с классом языков, распознаваемых ДАМП.