

Дополнительное задание для лабораторной работы №5

1. Описание грамматики

```
<grammar> ::= <rule> <rule-sep> <grammar-tail>
<grammar-tail> ::= <rule> <rule-sep> <grammar-tail>
<grammar-tail> ::=

<rule> ::= <nonterm> <arrow> <rule-right>

<rule-right> ::= <production> <rule-tail>
<rule-tail> ::= <production-sep> <production> <rule-tail>
<rule-tail> ::=

<production> ::= <epsilon>
<production> ::= <term> <production-tail>
<production> ::= <nonterm> <production-tail>

<production-tail> ::= <term> <production-tail>
<production-tail> ::= <nonterm> <production-tail>
<production-tail> ::=

<term> ::= <term-regex>
<nonterm> ::= <nterm-start> <nterm-regex> <nterm-end>

// токены-параметры
<rule-sep> ::= ;
<arrow> ::= →
<production-sep> ::= |
<nterm-start> ::= [
<nterm-end> ::= ]
<term-regex> ::= (a-zA-Z0-9)+
<nterm-regex> ::= (a-zA-Z0-9)+
```

Пример грамматики в этом синтаксисе:

```
[A] → b[A] | b[C]a;
[C] → c[C] | ε;
```

2. Определим к какому классу LL(k), LR(k) и при каких k принадлежит грамматика

Для начала докажем детерминированность грамматики. При условии уникальности пользовательских токенов, в грамматике не возникнет неоднозначности в выборе правил во время переходов. Детерминированность была достигнута в том числе благодаря «развертке» правил в tail-правила. Поэтому для данной грамматики довольно просто построить DPDA. Так как грамматика детерминирована, можем перейти к анализу

свойств $LL(k)$, $LR(k)$.

Для начала проанализируем грамматику на $LL(k)$. Определим, сколько символов необходимо просмотреть для однозначного выбора правила грамматики. Будем считать, что все токены уникальны, начиная с первого символа, а также, что языки терминалов и нетерминалов не включают в себя специальные токены. Тогда в данной грамматике, посмотрев на первый символ вперед, мы можем точно понять, по какому правилу переходить. Следовательно, данная грамматика $LL(1)$. (другие случаи пользовательских токенов рассмотрены в п.3)

Теперь перейдем к анализу на $LR(k)$. Нам достаточно доказать, что грамматика не $LR(0)$. Для этого докажем, что префикс-свойство не выполняется. Рассмотрим грамматику:

$[S] \rightarrow a;$

$[S] \rightarrow b;$

Префикс $[S] \rightarrow a;$, который тоже является грамматикой, лежит в нашем языке. Следовательно, префикс свойство не выполняется и грамматика не $LR(0)$. Делаем вывод, что наша грамматика $LR(1)$.

3. Проанализируем, при каком выборе значений токенов грамматика теряет свои свойства.

Чтобы понять, когда грамматика теряет свои свойства, определим случаи, в которых может возникнуть неоднозначность выбора правила. Разберем такие случаи.

В первом случае неоднозначность может возникнуть при выборе между $\langle term \rangle$ и $\langle nonterm \rangle$ в правилах:

$\langle production \rangle ::= \langle term \rangle \langle production\text{-tail} \rangle$

$\langle production \rangle ::= \langle nonterm \rangle \langle production\text{-tail} \rangle$

$\langle production\text{-tail} \rangle ::= \langle term \rangle \langle production\text{-tail} \rangle$

$\langle production\text{-tail} \rangle ::= \langle nonterm \rangle \langle production\text{-tail} \rangle$

В общем случае множества значений токенов $\langle term \rangle$ и $\langle nonterm \rangle$ могут пересекаться. Для того, чтобы их различить в грамматике используются разделители $\langle nterm\text{-start} \rangle$ и $\langle nterm\text{-end} \rangle$. Грамматика будет недетерминированной, если:

1. $\langle nterm\text{-start} \rangle$, $\langle nterm\text{-end} \rangle$ будут пустыми, но множества значений $\langle term \rangle$ и $\langle nonterm \rangle$ при этом будут пересекаться:

$S \rightarrow ab \mid a;$

$a \rightarrow b;$

(здесь a - нетерминал)

2. $\langle term \rangle$ и $\langle nonterm \rangle$ будут включать пустое слово;
3. $\langle nterm\text{-start} \rangle$ и $\langle nterm\text{-end} \rangle$ входят в язык $\langle term \rangle$:

$[S] \rightarrow [a] \mid [A];$

$[A] \rightarrow a;$

(здесь $[a]$ – терминал)

4. $\langle nterm\text{-start} \rangle$ или $\langle nterm\text{-end} \rangle$ входят в язык $\langle nonterm \rangle$.

$[S] \rightarrow ab \mid [A][C];$

$[A][C] \rightarrow ac;$

(здесь A)[C – нетерминал)

5. Если $\langle nterm-start \rangle$ или $\langle nterm-end \rangle$ состоят из нескольких символов и при этом первые k из них входят в язык $\langle term \rangle$, но начиная с k-го не входит, то грамматика LL(k).

$_ [S] _ \rightarrow _ a _ b \mid _ [A] _;$

$_ [A] _ \rightarrow ac _;$

($_$ входит в множество терминалов, $\langle nterm-start \rangle := _ [$, $\langle nterm-end \rangle :=] _$)

Далее посмотрим на нетерминал $\langle production \rangle$:

$\langle production \rangle ::= \langle epsilon \rangle$

$\langle production \rangle ::= \langle term \rangle \langle production-tail \rangle$

$\langle production \rangle ::= \langle nonterm \rangle \langle production-tail \rangle$

В нем может появиться неоднозначность при выборе между $\langle term \rangle$ и $\langle nonterm \rangle$, которую мы уже разобрали, а также может добавиться неоднозначность для $\langle epsilon \rangle$. В этом случае мы можем дополнить наши рассуждения:

1. Если убрать разделители $\langle nterm-start \rangle$, $\langle nterm-end \rangle$, но с условием, что $\langle term \rangle$ и $\langle nonterm \rangle$ будут отличаться друг от друга, начиная с k-ого символа, и не включать в себя $\langle epsilon \rangle$. Тогда грамматика будет детерминированной, но уже не LL[1], а LL[k];

$S \rightarrow abc \mid abcd;$

$abcd \rightarrow a;$

(abc – терминал, abcd – нетерминал)

2. Если $\langle epsilon \rangle$ не уникален и встречается в языках $\langle term \rangle$, $\langle nonterm \rangle$, $\langle nterm-start \rangle$, то грамматика недетерминированная.

$[S] \rightarrow eps \mid epsa \mid [A];$

$[A] \rightarrow eps;$

(eps – значение для $\langle epsilon \rangle$)

3. Если первые k символов $\langle epsilon \rangle$ совпадают с $\langle nterm-start \rangle$, но отличаются с k-го, то грамматика LL(k).

$eps[S] \rightarrow eps \mid eps[A];$

$eps[A] \rightarrow eps;$

(eps – значение для $\langle epsilon \rangle$, $\langle nterm-start \rangle = eps[$)

Теперь проанализируем нетерминалы $\langle rule-tail \rangle$ и $\langle production-tail \rangle$.

$\langle rule-tail \rangle ::= \langle production-sep \rangle \langle production \rangle \langle rule-tail \rangle$

$\langle production-tail \rangle ::= \langle term \rangle \langle production-tail \rangle$

$\langle production-tail \rangle ::= \langle nonterm \rangle \langle production-tail \rangle$

$\langle production-tail \rangle ::=$

В $\langle production\text{-}tail \rangle$ возникает недетерминированность при выборе между $\langle production\text{-}sep \rangle$, $\langle term \rangle$ и $\langle nonterm \rangle$. Рассмотрим их:

1. Если $\langle production\text{-}sep \rangle$ будет входить во множество $\langle term \rangle$, $\langle nterm\text{-}start \rangle$, то грамматика будет недетерминированной:

$$\begin{aligned} [S] &\rightarrow a[b[[A]; \\ [A] &\rightarrow b; \end{aligned}$$

2. Если $\langle production\text{-}sep \rangle$ будет пустым, грамматика будет недетерминированной.

$$\begin{aligned} [S] &\rightarrow ab[A]; \\ [A] &\rightarrow b; \end{aligned}$$

Далее рассмотрим нетерминал $\langle rule \rangle$.

$$\langle rule \rangle ::= \langle nonterm \rangle \langle arrow \rangle \langle rule\text{-}right \rangle$$

Тут нет неоднозначности, однако все равно проанализируем как меняются свойства грамматики при разных значениях пользовательского токена $\langle arrow \rangle$:

1. Если токен пустой, то грамматика все равно детерминированная и LL(1);

$$\begin{aligned} [S] &ab[A]; \\ [A] &b; \end{aligned}$$

2. Если $\langle arrow \rangle$ совпадает с $\langle production\text{-}sep \rangle$, то грамматика остается детерминированной и LL(1);

$$\begin{aligned} [S] &| ab | [A]; \\ [A] &| b; \end{aligned}$$

3. Если $\langle arrow \rangle$ входит во множество $\langle term \rangle$, то грамматика остается детерминированной и LL(1);

$$\begin{aligned} [S] &\rightarrow \rightarrow\rightarrow | [A]; \\ [A] &\rightarrow b; \end{aligned}$$

4. Если $\langle arrow \rangle$ совпадает с $\langle nterm\text{-}start \rangle$ или $\langle nterm\text{-}end \rangle$, то грамматика остается детерминированной и LL(1);

$$\begin{aligned} &\rightarrow S] \rightarrow a | b \rightarrow A]; \\ &\rightarrow A] \rightarrow b; \end{aligned}$$

Проанализируем нетерминалы $\langle grammar\text{-}tail \rangle$ и $\langle grammar \rangle$.

$$\langle grammar\text{-}tail \rangle ::= \langle rule \rangle \langle rule\text{-}sep \rangle \langle grammar\text{-}tail \rangle$$
$$\langle grammar\text{-}tail \rangle ::=$$
$$\langle rule\text{-}right \rangle ::= \langle production \rangle \langle rule\text{-}tail \rangle$$
$$\langle rule\text{-}tail \rangle ::= \langle production\text{-}sep \rangle \langle production \rangle \langle rule\text{-}tail \rangle$$
$$\langle rule\text{-}tail \rangle ::=$$

Тут возникает неоднозначность на этапе, когда мы выходим из $\langle rule\text{-}tail \rangle$, то есть неоднозначность появляется в выборе между $\langle rule\text{-}sep \rangle$ и $\langle production\text{-}sep \rangle$. Рассмотрим

разделитель $\langle rule-sep \rangle$:

1. Если $\langle rule-sep \rangle$ пустой, то грамматика останется детерминированной, но LL свойство теряется. Однако, если все нетерминалы имеют максимальную длину k , то грамматика будет LL($k+3$), так как нам нужно дойти до $\langle arrow \rangle$ для однозначного определения перехода.

$$[S] \rightarrow ab \mid [A]$$
$$[A] \rightarrow b$$

2. Если $\langle production-sep \rangle$ и $\langle rule-sep \rangle$ совпадают, грамматика будет детерминированной, но LL свойство теряется. Однако, если все нетерминалы имеют максимальную длину k , то грамматика будет LL($k+3$), так как нам нужно дойти до $\langle arrow \rangle$ для однозначного определения перехода.

$$[S] \rightarrow ab \mid [A] \mid$$
$$[A] \rightarrow b \mid$$

3. Если $\langle rule-sep \rangle$ совпадает с $\langle arrow \rangle$, то грамматика останется детерминированной и LL(1);

$$[S]; ab[A]ab[B];$$
$$[A]; b;$$

4. Если $\langle rule-sep \rangle$ совпадает с $\langle nterm-start \rangle$ и $\langle nterm-end \rangle$, грамматика детерминированная и LL(2);

$$;S; \rightarrow ;b \mid ;A;;$$
$$;A; \rightarrow b;$$

5. Если $\langle rule-sep \rangle$ входит во множество $\langle term \rangle$, грамматика детерминированная, но не LL(1). Однако, если все нетерминалы имеют максимальную длину k , то грамматика будет LL($k+3$), так как нам нужно дойти до $\langle arrow \rangle$ для однозначного определения перехода.

$$[S] \rightarrow ;b \mid [A];$$
$$[A] \rightarrow b;$$

Если же говорить об LR(k) в случаях, когда грамматика детерминированная, то можно сделать вывод, что какие бы значения токенов мы ни задавали, грамматика всегда будет LR(1). Потому что одинарное правило всегда может быть отдельной грамматикой, и мы можем дописать после него еще одно правило. И это можно сделать при любых значениях пользовательских токенов. Например:

$$[S] \rightarrow a[S]b \mid ab;$$
$$[B] \rightarrow ab;$$

Получается, что префикс свойство никогда не выполняется, и язык всегда LR(1). Если же грамматика при каком-то выборе токенов становится недетерминированной, то она, соответственно, и не LR.