Ручное построение нисходящих синтаксических анализаторов.

Шайхутдинов Тимур

24 марта 2016 г.

1 Разработка грамматики

1.1 Задание

Регулярные выражения с операциями конкатенации (простая последовательная запись строк), выбора (вертикальная черта), замыкания Клини. Приоритет операций стандартный. Скобки могут использоваться для изменения приоритета. Для обозначения базовых языков используются маленькие буквы латинского алфавита. Используйте один терминал для всех символов. Пример: ((abc*b|a)*ab(aa|b*)b)*

1.2 Грамматика

Построим грамматику по описанию задания

$$
\begin{aligned}
S &\to C \\
S &\to \varepsilon \\
C &\to (C) \\
C &\to n \\
C &\to C|C \\
C &\to C* \\
C &\to CC
\end{aligned}
$$

Нетерминал	**Описание**
S	Правльное регулярное выражение
C	Непустое правильное регулярное выражение

В грамматике есть левая рекурсия, устраним её.

$$
\begin{aligned}
S &\to C \\
S &\to \varepsilon \\
C &\to (C)C' \\
C &\to nC' \\
C &\to (C) \\
C &\to n \\
C' &\to |CC' \\
C' &\to *C' \\
C' &\to CC' \\
C' &\to |C \\
C' &\to *C \\
C' &\to C
\end{aligned}
$$

Нетерминал	Описание
S	Правльное регулярное выражение
C	Непустое правильное регулярное выражение
C'	Непустое продолжение правильного регулярного выражения

Избавимся от правого ветвления.

$$
\begin{aligned}
S &\to C \\
S &\to \varepsilon \\
C &\to (C)S' \\
C &\to nS' \\
C' &\to |CS' \\
C' &\to *S' \\
C' &\to S' \\
S' &\to C' \\
S' &\to \varepsilon
\end{aligned}
$$

Нетерминал	Описание
S	Правльное регулярное выражение
C	Непустое правильное регулярное выражение
S'	Продолжение правильного регулярного выражения
C'	Непустое продолжение правильного регулярного выражения

2 Построение лексического анализатора

В грамматике пять терминалов: '|', '*****', '(', ')', '**n**' (универсальный терминал для обозначение всех символов).

Заведём класс **Token** для хранения терминалов.

```
public enum Token {
    OPEN_BRACKET, CLOSE_BRACKET,
     KLEENE_CLOSURE, CHOOSE, CHARACTER, END
}
```

Терминал	Токен
(	OPEN_BRACKET
)	CLOSE_BRACKET
*	KLEENE_CLOSURE
|	CHOOSE
n	CHARACTER
$	END

3 Построение синтаксического анализатора

Множества **FIRST** и **FOLLOW** для нетерминалов грамматики.

Нетерминал	FIRST	FOLLOW
S	(, n, ε	(, n, ε, *, |,), $
C	(, n	(, n, ε, *, |,), $
S'	(, n, |, *, ε	(, n, ε, *, |,), $
C'	(, n, |, *	(, n, ε, *, |,), $