

# Ненаправленные алгоритмы анализа

## Метод Ангера. СУК. НФХ.

В. С. Полозов

Кафедра системного программирования СПбГУ



Теория автоматов и формальных языков

## Альтернативные способы задания грамматик

Ещё способы задания грамматик:

- Форма Бэкуса-Наура (BNF) – ALGOL 60

$\langle \text{name} \rangle ::= \text{tom} \mid \text{dick} \mid \text{harry}$

$\langle \text{sentence} \rangle_s ::= \langle \text{name} \rangle \mid \langle \text{list} \rangle \text{ and } \langle \text{name} \rangle$

$\langle \text{list} \rangle ::= \langle \text{name} \rangle, \langle \text{list} \rangle \mid \langle \text{name} \rangle$

- Форма ван Вейнгаардена (van Wijngaarden) – ALGOL 68

name: tom symbol; dick symbol; harry symbol.

sentences<sub>s</sub>: name; list, and symbol, name.

list : name, comma symbol, list; name.

- Расширенные грамматики (ECF):

- $\text{Something}^+ \rightarrow \text{Something} \mid \text{Something } \text{SomethingSequence}$

- $\text{Something}^* \rightarrow \varepsilon \mid \text{Something } \text{SomethingSequence}$

- $\text{Something}^? \rightarrow \varepsilon \mid \text{Something}$

- $(\text{Something} ;)^?$

- $\text{Something} + 4$

## Альтернативные способы задания грамматик

Ещё способы задания грамматик:

- Форма Бэкуса-Наура (BNF) – ALGOL 60

$\langle \text{name} \rangle ::= \text{tom} \mid \text{dick} \mid \text{harry}$

$\langle \text{sentence} \rangle_S ::= \langle \text{name} \rangle \mid \langle \text{list} \rangle \text{ and } \langle \text{name} \rangle$

$\langle \text{list} \rangle ::= \langle \text{name} \rangle, \langle \text{list} \rangle \mid \langle \text{name} \rangle$

- Форма ван Вейнгаардена (van Wijngaarden) – ALGOL 68

name: tom symbol; dick symbol; harry symbol.

sentences<sub>S</sub>: name; list, and symbol, name.

list : name, comma symbol, list; name.

- Расширенные грамматики (ECF):

- $\text{Something}^+ \rightarrow \text{Something} \mid \text{Something } \text{SomethingSequence}$

- $\text{Something}^* \rightarrow \varepsilon \mid \text{Something } \text{SomethingSequence}$

- $\text{Something}^? \rightarrow \varepsilon \mid \text{Something}$

- $(\text{Something } ;)^?$

- $\text{Something} + 4$

## Альтернативные способы задания грамматик

Ещё способы задания грамматик:

- Форма Бэкуса-Наура (BNF) – ALGOL 60

$\langle \text{name} \rangle ::= \text{tom} \mid \text{dick} \mid \text{harry}$

$\langle \text{sentence} \rangle_S ::= \langle \text{name} \rangle \mid \langle \text{list} \rangle \text{ and } \langle \text{name} \rangle$

$\langle \text{list} \rangle ::= \langle \text{name} \rangle, \langle \text{list} \rangle \mid \langle \text{name} \rangle$

- Форма ван Вейнгаардена (van Wijngaarden) – ALGOL 68

name: tom symbol; dick symbol; harry symbol.

sentences<sub>S</sub>: name; list, and symbol, name.

list : name, comma symbol, list; name.

- Расширенные грамматики (ECF):

- $\text{Something}^+ \rightarrow \text{Something} \mid \text{Something SomethingSequence}$

- $\text{Something}^* \rightarrow \varepsilon \mid \text{Something SomethingSequence}$

- $\text{Something}^? \rightarrow \varepsilon \mid \text{Something}$

- $(\text{Something} ;)^?$

- $\text{Something} + 4$

## Ещё про грамматики

Аномалии в грамматиках:

- Неопределенные нетерминалы
- Неиспользуемые нетерминалы
- Непродуктивные нетерминалы  
Например:  $X \rightarrow \alpha X$

- Циклы

Например:  $A \rightarrow A$  или  $A \rightarrow B, B \rightarrow C, C \rightarrow A$

## Ещё про грамматики

Аномалии в грамматиках:

- Неопределенные нетерминалы
- Неиспользуемые нетерминалы
- Непродуктивные нетерминалы  
Например:  $X \rightarrow \alpha X$
- Циклы  
Например:  $A \rightarrow A$  или  $A \rightarrow B, B \rightarrow C, C \rightarrow A$

## Метод Ангера без $\epsilon$ -правил и циклов

Автор: S.H. Unger

Принцип. Возьмём правило:

$$S \rightarrow ABC|DE|F$$

разобьем вход на соответствующие корзины. Проверим.  
Можно использовать DFS или BFS. В авторской статье  
используется DFS.

## Метод Ангера без $\epsilon$ -правил и циклов

Автор: S.H. Unger

Принцип. Возьмём правило:

$$S \rightarrow ABC|DE|F$$

разобьем вход на соответствующие корзины. Проверим.  
Можно использовать DFS или BFS. В авторской статье  
используется DFS.



## Пример

Рассмотрим грамматику:

$$\begin{aligned} \textit{Expr} &\rightarrow \textit{Expr} + \textit{Term} \mid \textit{Term} \\ \textit{Term} &\rightarrow \textit{Term} \times \textit{Factor} \mid \textit{Factor} \\ \textit{Factor} &\rightarrow (\textit{Expr}) \mid i \end{aligned}$$

И вход

$$(i + i) \times i$$

## Метод Ангера с $\varepsilon$ -правилами

Рассмотрим правила:

$$\begin{aligned} S &\rightarrow A B C \\ B &\rightarrow S A \end{aligned}$$

и вход  $pqr$ .

## Пример

Рассмотрим правила:

$$\begin{aligned} S &\rightarrow L S D \mid \varepsilon \\ L &\rightarrow \varepsilon \\ D &\rightarrow d \end{aligned}$$

и пару входов:  $d$  и  $dd$ .

# Метод СҮК

Назван по именам создателей: J. Cocke, D.H. Younger, и T. Kasami.

Две фазы

- 1 Строится таблица соответствия выводам подстрок входного предложения нетерминалам
- 2 Используя построенную таблицу строим все возможные выводы

## Пример

$Number \rightarrow Integer | Real$

$Integer \rightarrow Digit | Integer Digit$

$Real \rightarrow Integer Fraction Scale$

$Fraction \rightarrow . Integer$

$Scale \rightarrow e Sign Integer | Empty$

$Digit \rightarrow 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9$

$Empty \rightarrow \epsilon$

$Sign \rightarrow + | -$

и вход с  $\epsilon$  - 43.1

## Обозначения

Введем обозначения:

- $R_\varepsilon = \{Empty, Scale\}$  - множество нетерминалов  $A$  таких, что  $A \rightarrow \varepsilon$
- $s_{i,l}$  - подстрока  $z_i z_{i+1} \dots z_{i+l-1}$  строки  $z = z_1 z_2 \dots z_n$  начиная с позиции  $i$  и длиной  $l$ .
- $R_{s_{i,l}}$  - множество нетерминалов из которых выводима строка  $s_{i,l}$ .
- Для строк длины 0 положим:  $s_{i,0} = \varepsilon$  и  $R_{s_{i,0}} = R_\varepsilon$

## СУК: НФХ

### Определение

*Нормальная форма Хомского (CNF):* все правила имеют вид  $A \rightarrow a$  или  $A \rightarrow BC$ .

Получаем:

- $R_\epsilon = \emptyset$
- $R_{s_{i,1}}$  вычисляется непосредственно из правил  $A \rightarrow a$ .
- Для правила  $A \rightarrow BC$   
из  $B$  должно выводиться  $s_{i,k}$ , т.е.  $B \in R_{s_{i,k}}$ ,  
из  $C$  должно выводиться  $s_{i+k,l-k}$ , т.е.  $C \in R_{s_{i+k,l-k}}$

# СУК: НФХ

Т.е. можно организовать *таблицу распознавания* (рис.).

$R_{s_{1,n}}$					
$R_{s_{1,n-1}}$	$R_{s_{2,n-1}}$				
...	...	...			
$R_{s_{1,2}}$	...	...	$R_{s_{n-1,2}}$		
$R_{s_{1,1}}$	$R_{s_{2,1}}$	...	$R_{s_{n-1,1}}$	$R_{s_{n,1}}$	
$s_1$	$s_2$	...	$s_{n-1}$	$s_n$	

$R_{i,j} = ?$

Соответственно, можно изменить порядок обхода:

- off-line order
- on-line order

Сложность?



# СУК: НФХ

Т.е. можно организовать *таблицу распознавания* (рис.).

$R_{s_{1,n}}$					
$R_{s_{1,n-1}}$	$R_{s_{2,n-1}}$				
...	...	...			
$R_{s_{1,2}}$	...	...	$R_{s_{n-1,2}}$		
$R_{s_{1,1}}$	$R_{s_{2,1}}$	...	$R_{s_{n-1,1}}$	$R_{s_{n,1}}$	
$s_1$	$s_2$	...	$s_{n-1}$	$s_n$	

$R_{i,j} = ?$

Соответственно, можно изменить порядок обхода:

- off-line order
- on-line order

Сложность?

# СУК: НФХ

Т.е. можно организовать *таблицу распознавания* (рис.).

$R_{s_{1,n}}$					
$R_{s_{1,n-1}}$	$R_{s_{2,n-1}}$				
...	...	...			
$R_{s_{1,2}}$	...	...	$R_{s_{n-1,2}}$		
$R_{s_{1,1}}$	$R_{s_{2,1}}$	...	$R_{s_{n-1,1}}$	$R_{s_{n,1}}$	
$s_1$	$s_2$	...	$s_{n-1}$	$s_n$	

$R_{i,j} = ?$

Соответственно, можно изменить порядок обхода:

- off-line order
- on-line order

Сложность?

# СУК: НФХ

Т.е. можно организовать *таблицу распознавания* (рис.).

$R_{s_{1,n}}$					
$R_{s_{1,n-1}}$	$R_{s_{2,n-1}}$				
...	...	...			
$R_{s_{1,2}}$	...	...	$R_{s_{n-1,2}}$		
$R_{s_{1,1}}$	$R_{s_{2,1}}$	...	$R_{s_{n-1,1}}$	$R_{s_{n,1}}$	
$s_1$	$s_2$	...	$s_{n-1}$	$s_n$	

$R_{i,j} = ?$

Соответственно, можно изменить порядок обхода:

- off-line order
- on-line order

Сложность?

## Преобразование к НФХ

Преобразование к НФХ:

- удаление  $\epsilon$ -правил
- удаление цепных правил ( $A \rightarrow B$ )
- удаление непродуктивных нетерминалов
- удаление недостижимых нетерминалов
- оставшаяся грамматика преобразуется добавлением правил, пока все правила не будут вида  $A \rightarrow a$  или  $A \rightarrow BC$ .

## Пример

$Numbers_S \rightarrow 0|1|2|3|4|5|6|7|8|9$

$Numbers_S \rightarrow Integer\ Digit$

$Numbers_S \rightarrow N1\ Scale' | Integer\ Fraction$

$N1 \rightarrow Integer\ Fraction$

$Integer \rightarrow 0|1|2|3|4|5|6|7|8|9$

$Integer \rightarrow Integer\ Digit$

$Fraction \rightarrow T1\ Integer$

$T1 \rightarrow .$

$Scale' \rightarrow N2\ Integer$

$N2 \rightarrow T2\ Sign$

$T2 \rightarrow e$

$Digit \rightarrow 0|1|2|3|4|5|6|7|8|9$

## Откат преобразования к CNF

При преобразовании к CNF могут потеряться нетерминалы:

- при удалении  $\epsilon$ -правил (В примере: Scale, Real, Empty)
- удаление unit-правил (В примере: Real)

Дополним грамматику правилом:

*Real*  $\rightarrow$  *N1 Scale'* | *Integer Fraction*

## Откат преобразования к CNF

При преобразовании к CNF могут потеряться нетерминалы:

- при удалении  $\epsilon$ -правил (В примере: Scale, Real, Empty)
- удаление unit-правил (В примере: Real)

Дополним грамматику правилом:

$$Real \rightarrow N1 \text{ Scale}' \mid Integer \text{ Fraction}$$

# Вопросы