

计算文法符号X的FIRST(X)

- \triangleright FIRST(X): 可以从X推导出的所有串首终结符构成的集合
- ▶如果 $X \Rightarrow^* \varepsilon$, 那么 $\varepsilon \in FIRST(X)$

①
$$E \rightarrow TE'$$
 $FIRST(E) = \{$ (id ?)
② $E' \rightarrow +TE' | \varepsilon$ $FIRST(E') = \{$ + ε }
③ $T \rightarrow FT'$ $FIRST(T) = \{$ (id ?)
④ $T' \rightarrow *FT' | \varepsilon$ $FIRST(T) = \{$ * ε }
⑤ $F \rightarrow (E) | id$ $FIRST(F) = \{$ (id }

算法

- ►不断应用下列规则,直到没有新的终结符或ε可以被加入到任何FIRST集合中为止
 - ▶如果X是一个终结符,那么 $FIRST(X) = \{X\}$
 - 》如果X是一个非终结符,且 $X \rightarrow Y_1 ... Y_k \in P(k \geq 1)$,那么如果对于某个i,a在FIRST (Y_i) 中且 ε 在所有的 FIRST (Y_1) , ... , FIRST (Y_{i-1}) 中 $(\operatorname{pr} Y_1 ... Y_{i-1} \Rightarrow^* \varepsilon)$, 就把a加入到FIRST(X)中。如果对于所有的 $j=1,2,\ldots,k$, ε 在 FIRST (Y_i) 中,那么将 ε 加入到FIRST(X)
 - → 如果 $X \rightarrow ε ∈ P$,那么将ε加入到FIRST(X)中

计算串 $X_1X_2...X_n$ 的FIRST集合

- \rightarrow 向 $FIRST(X_1X_2...X_n)$ 加入 $FIRST(X_1)$ 中所有的非 ϵ 符号
- ightharpoonup如果ε在 $FIRST(X_1)$ 中,再加入 $FIRST(X_2)$ 中的所有非ε符号;如果ε在 $FIRST(X_1)$ 和 $FIRST(X_2)$ 中,再加入 $FIRST(X_3)$ 中的所有非ε符号,以此类推
- \triangleright 最后,如果对所有的i, ε 都在 $FIRST(X_i)$ 中,那么将 ε 加入到 $FIRST(X_1X_2...X_n)$ 中

计算非终结符A的FOLLOW(A)

- ightarrow FOLLOW(A): 可能在某个句型中紧跟在A后边的终结符a的集合 FOLLOW(A)={a|S \Rightarrow * $\alpha A a \beta$, a $\in V_T$, $\alpha,\beta \in (V_T \cup V_N)^*$ }
- ▶如果A是某个句型的的最右符号,则将结束符"\$"添加到FOLLOW(A)中

```
(a) E \to TE' FIRST(E) = \{ (id) \} FOLLOW(E) = \{ \} \} \{ E' \to +TE' | \varepsilon \} FIRST(E') = \{ + \} FOLLOW(E') = \{ + \} \} \{ T \to FT' \} FIRST(T) = \{ (id) \} FOLLOW(T) = \{ + \} \} \{ T' \to *FT' | \varepsilon \} FIRST(T') = \{ * \} FOLLOW(T') = \{ + \} \} \{ T' \to *FT' | \varepsilon \} FIRST(T') = \{ * \} FOLLOW(T') = \{ + \} \} \{ T' \to *FT' | \varepsilon \} FIRST(T') = \{ * \} FOLLOW(T') = \{ + \} \}
```

需要计算多次,直到不再更新

算法

- 一不断应用下列规则,直到没有新的终结符可以被加入到任何FOLLOW集合中为止
 - ▶ 将\$放入FOLLOW(S)中,其中S是开始符号,\$是输入右端的结束标记
 - 如果存在一个产生式 $A \rightarrow \alpha B \beta$,那么 $FIRST(\beta)$ 中除 ϵ 之外的所有符号都在FOLLOW(B)中
 - →如果存在一个产生式 $A \rightarrow \alpha B$,或存在产生式 $A \rightarrow \alpha B \beta$ 且 $FIRST(\beta)$ 包含 ϵ ,那么 FOLLOW(A)中的所有符号都在 FOLLOW(B)中

例:表达式文法各产生式的SELECT集

X	FIRST(X)	FOLLOW(X)
E	(id	\$)
E '	3 +	8)
T	(id 🗸	+)\$
<i>T'</i>	3 *	+) \$
F	(id	*+)\$

表达式文法是 LL(1) 文法

(1)
$$E \to TE'$$
 $SELECT(1) = \{ (id) \}$
(2) $E' \to + TE'$ $SELECT(2) = \{ + \}$
(3) $E' \to \varepsilon$ $SELECT(3) = \{ \$) \}$
(4) $T \to FT'$ $SELECT(4) = \{ (id) \}$
(5) $T' \to *FT'$ $SELECT(5) = \{ * \}$
(6) $T' \to \varepsilon$ $SELECT(6) = \{ + \} \}$
(7) $F \to (E)$ $SELECT(8) = \{ id \}$

预测分析表

	产生式	SELECT
E	E→TE'	(id
E'	$E' \rightarrow +TE'$	+
	$E'\!\!\to\!\!arepsilon$	\$)
T	$T \rightarrow FT'$	(id
T'	<i>T'</i> →* <i>FT</i> '	*
F	$F \rightarrow (E)$	(
	<i>F</i> →id	id

非终结符	输入符号					
	id	+	*	()	\$
$\boldsymbol{\mathit{E}}$	$E \rightarrow TE'$			$E \rightarrow TE'$		
E'		$E' \rightarrow +TE'$			$E'\!\!\!\to\!\! arepsilon$	$E'\!\!\!\to\!\!\! arepsilon$
T	T→FT'			T→FT'		
T'		<i>T'</i> →ε	<i>T'</i> →* <i>FT</i> '		$T' \rightarrow \varepsilon$	<i>T'</i> →ε
$oldsymbol{F}$	$F \rightarrow id$			$F \rightarrow (E)$		

LL(1)文法的分析方法

- ▶递归的预测分析法
- ▶非递归的预测分析法





递归的预测分析法

- ▶ 递归的预测分析法是指:在递归下降分析中,根据预测分析表进行产生式的选择
 - ▶根据每个非终结符的产生式和LL(1)文法的预测分析表,为 每个非终结符编写对应的过程

- (1) <PROGRAM> → program <DECLIST> :<TYPE> ; <STLIST> end
- (2) $\langle DECLIST \rangle \rightarrow id \langle DECLISTN \rangle$
- (3) $\langle DECLISTN \rangle \rightarrow$, id $\langle DECLISTN \rangle$
- (4) $\langle DECLISTN \rangle \rightarrow \varepsilon$
- (5) $\langle STLIST \rangle \rightarrow s \langle STLISTN \rangle$
- (6) $\langle STLISTN \rangle \rightarrow ; s \langle STLISTN \rangle$
- (7) $\langle STLISTN \rangle \rightarrow \varepsilon$
- $(8) < TYPE > \rightarrow real$
- $(9) < TYPE > \rightarrow int$

SELECT(4)={:} SELECT(7)={end}

```
program DESCENT;
begin
GETNEXT(TOKEN);
PROGRAM(TOKEN);
GETNEXT(TOKEN);
if TOKEN≠'$' then ERROR;
end
```

- (1) <PROGRAM> → program <DECLIST> :<TYPE> ; <STLIST> en
- (2) $\langle DECLIST \rangle \rightarrow id \langle DECLISTN \rangle$
- (3) $\langle DECLISTN \rangle \rightarrow$, id $\langle DECLISTN \rangle$
- (4) $\langle DECLISTN \rangle \rightarrow \varepsilon$
- (5) $\langle STLIST \rangle \rightarrow s \langle STLISTN \rangle$
- (6) $\langle STLISTN \rangle \rightarrow ; s \langle STLISTN \rangle$
- (7) $\langle STLISTN \rangle \rightarrow \varepsilon$
- $(8) < TYPE > \rightarrow real$
- (9) < TYPE > \rightarrow int

SELECT(4)={:}
SELECT(7)={end}

```
procedure PROGRAM(TOKEN);
    begin
      if TOKEN≠'program' then ERROR;
      GETNEXT(TOKEN);
      DECLIST(TOKEN);
                  if TOKEN≠':' then ERROR;
      GETNEXT(TOKEN);
      TYPE(TOKEN);
      GETNEXT(TOKEN);
      if TOKEN≠';' then ERROR;
      GETNEXT(TOKEN);
      STLIST(TOKEN);
      if TOKEN≠'end' then ERROR;
    end
```

若下一个符号为终结符则进行对比,非终结符则调用对应的 过程

- (1) <PROGRAM> → program <DECLIST> :<TYPE> ; <STLIST> end
- (2) $\langle DECLIST \rangle \rightarrow id \langle DECLISTN \rangle$
- (3) $\langle DECLISTN \rangle \rightarrow$, id $\langle DECLISTN \rangle$
- (4) $\langle DECLISTN \rangle \rightarrow \varepsilon$
- (5) $\langle STLIST \rangle \rightarrow s \langle STLISTN \rangle$
- (6) $\langle STLISTN \rangle \rightarrow ; s \langle STLISTN \rangle$
- (7) $\langle STLISTN \rangle \rightarrow \varepsilon$
- $(8) < TYPE > \rightarrow real$
- $(9) < TYPE > \rightarrow int$

SELECT(4)={:}
SELECT(7)={end}

```
procedure DECLIST(TOKEN);
begin
if TOKEN≠'id' then ERROR;

GETNEXT(TOKEN);
DECLISTN(TOKEN);
end
```

- (1) <PROGRAM> → program <DECLIST> :<TYPE> ; <STLIST> end
- (2) $\langle DECLIST \rangle \rightarrow id \langle DECLISTN \rangle$
- (3) $\langle DECLISTN \rangle \rightarrow$, id $\langle DECLISTN \rangle$
- (4) $\langle DECLISTN \rangle \rightarrow \varepsilon$
- (5) $\langle STLIST \rangle \rightarrow s \langle STLISTN \rangle$
- (6) $\langle STLISTN \rangle \rightarrow ; s \langle STLISTN \rangle$
- (7) $\langle STLISTN \rangle \rightarrow \varepsilon$
- $(8) < TYPE > \rightarrow real$
- $(9) < TYPE > \rightarrow int$

```
SELECT(4)={:}
SELECT(7)={end}
```

```
procedure DECLISTN(TOKEN);
    begin
      if TOKEN =',' then
        begin
         GETNEXT(TOKEN);
         if TOKEN≠'id' then ERROR;
         GETNEXT(TOKEN);
         DECLISTN(TOKEN);
        end
       else if TOKEN≠':' then ERROR;
    end
```

- (1) <PROGRAM> → program <DECLIST> :<TYPE> ; <STLIST> end
- (2) $\langle DECLIST \rangle \rightarrow id \langle DECLISTN \rangle$
- (3) $\langle DECLISTN \rangle \rightarrow$, id $\langle DECLISTN \rangle$
- (4) $\langle DECLISTN \rangle \rightarrow \varepsilon$
- (5) $\langle STLIST \rangle \rightarrow s \langle STLISTN \rangle$
- (6) $\langle STLISTN \rangle \rightarrow ; s \langle STLISTN \rangle$
- (7) $\langle STLISTN \rangle \rightarrow \varepsilon$
- $(8) < TYPE > \rightarrow real$
- $(9) < TYPE > \rightarrow int$

SELECT(4)={:}
SELECT(7)={end}

```
procedure STLIST(TOKEN);
begin

if TOKEN≠'s' then ERROR;

GETNEXT(TOKEN);

STLISTN(TOKEN);

end
```

- (1) <PROGRAM> → program <DECLIST> :<TYPE> ; <STLIST> end
- (2) $\langle DECLIST \rangle \rightarrow id \langle DECLISTN \rangle$
- (3) $\langle DECLISTN \rangle \rightarrow$, id $\langle DECLISTN \rangle$
- (4) $\langle DECLISTN \rangle \rightarrow \varepsilon$
- (5) $\langle STLIST \rangle \rightarrow s \langle STLISTN \rangle$
- (6) $\langle STLISTN \rangle \rightarrow ; s \langle STLISTN \rangle$
- (7) $\langle STLISTN \rangle \rightarrow \varepsilon$
- $(8) < TYPE > \rightarrow real$
- $(9) < TYPE > \rightarrow int$

```
SELECT(4)={:} SELECT(7)={end}
```

```
procedure STLISTN(TOKEN);
     begin
       if TOKEN ="; then
         begin
          GETNEXT(TOKEN);
          if TOKEN≠'s' then ERROR;
         GETNEXT(TOKEN);
         STLISTN(TOKEN);
         end
       else if TOKEN≠'end' then ERROR;
     end
```

- (1) <PROGRAM> → program <DECLIST> :<TYPE> ; <STLIST> end
- (2) $\langle DECLIST \rangle \rightarrow id \langle DECLISTN \rangle$
- (3) $\langle DECLISTN \rangle \rightarrow$, id $\langle DECLISTN \rangle$
- (4) $\langle DECLISTN \rangle \rightarrow \varepsilon$
- (5) $\langle STLIST \rangle \rightarrow s \langle STLISTN \rangle$
- (6) $\langle STLISTN \rangle \rightarrow ; s \langle STLISTN \rangle$
- (7) $\langle STLISTN \rangle \rightarrow \varepsilon$
- $(8) < TYPE > \rightarrow real$
- $(9) < TYPE > \rightarrow int$

```
SELECT(4)={:}
SELECT(7)={end}
```

```
procedure TYPE(TOKEN);

begin

if TOKEN≠'real' or TOKEN≠'int'

then ERROR;

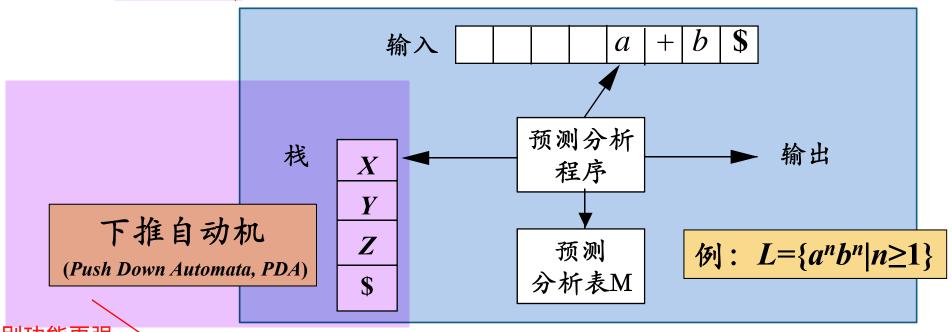
end
```





非递归的预测分析法

》非递归的预测分析不需要为每个非终结符编写递归下降过程, 而是根据预测分析表构造一个自动机, 也叫表驱动的预测分析



识别功能更强 因为有记忆功能

与有穷自动机相似

不同之处

a则n+1 b则n-1

有穷自动机没有记忆功能,只能通过状态记录n个数 但是n趋于无穷

非终	输入符号					
结符	id	+	*	()	\$
E	$E \rightarrow TE'$			$E \rightarrow TE'$		
E'		$E' \rightarrow +TE'$			$E'\!\!\!\to\!\!\! \varepsilon$	$E'\!\!\to\!\! \varepsilon$
T	T→FT'			T→FT'		
T'		<i>T'</i> →ε	<i>T'</i> →* <i>FT'</i>		<i>T'</i> →ε	<i>T'</i> →ε
F	$F \rightarrow id$			$F \rightarrow (E)$		

如果w是至今为止已经匹配完成的输入部分,那么栈中保存的文法符号序列 α 满足 $S \Rightarrow^*_{lm} w \alpha$

栈	剩余输入
E \$	id+id*id \$
<i>TE'</i> \$	id+id*id \$
<i>FT'E'</i> \$	id+id*id \$
id <i>T'E'</i> \$	id+id*id \$
<i>T'E'</i> \$	+id*id \$
E'\$	+id*id \$
+ <i>TE'</i> \$	+id*id \$
<i>TE'</i> \$	id*id \$
<i>FT'E'</i> \$	id*id \$
id <i>T'E'</i> \$	id*id \$
<i>T'E'</i> \$	*id \$
* <i>FT'E'</i> \$	*id \$
<i>FT'E'</i> \$	id \$
id <i>T'E'</i> \$	id \$
<i>T'E'</i> \$	\$
E'\$	\$
\$	\$

输出

$$E \rightarrow TE'$$
 $T \rightarrow FT'$
 $F \rightarrow id$
 $T' \rightarrow \varepsilon$
 $E' \rightarrow +TE'$
 $T \rightarrow FT'$
 $F \rightarrow id$
 $T' \rightarrow *FT'$
 $F \rightarrow id$
 $T' \rightarrow \varepsilon$

最左推导



表驱动的预测分析法

- ▶ 输入: 一个串w和文法G的分析表 M
- \triangleright 输出:如果w在L(G)中,输出w的最左推导;否则给出错误指示
- ▶ 方法:最初,语法分析器的格局如下:輸入缓冲区中是w\$, G的开始符号位于栈顶,其下面是\$。下面的程序使用预测分析表M生成了处理这个输入的预测分析过程

```
设置ip使它指向w的第一个符号,其中ip 是输入指针; 令X=栈顶符号; while (X \neq \$) { /* 栈非空 */ if (X 等于ip所指向的符号a) 执行栈的弹出操作,将ip向前移动一个位置; else if (M是一个终结符号)error (); else if (M[X, a]是一个报错条目)error (); X 非终结符 else if (M[X, a] = X \rightarrow Y_1 Y_2 \dots Y_k ) { 输出产生式X \rightarrow Y_1 Y_2 \dots Y_k ; 弹出栈顶符号; 将Y_k, Y_{k-1} …, Y_i 压入栈中,其中Y_i位于栈顶。 } 令X=栈顶符号
```

递归的预测分析法vs.非递归的预测分析法

	递归的预测分析法	非递归的预测分析法
42 克 坝 挡	程序规模较大,	主控程序规模较小, 🙂
程序规模	不需载入分析表	需载入分析表 (表较小)
直观性	较好 🙂	较差
效率	较低	分析时间大约正比于待分 析程序的长度 🙂
自动生成	较难	较易 🙂

预测分析法实现步骤

- 1) 构造文法
- 2) 改造文法: 消除二义性、消除左递归、消除回溯
- 3) 求每个变量的FIRST集和FOLLOW集,从而求得每个 候选式的SELECT集
- 4) 检查是不是 LL(1) 文法。若是,构造预测分析表
- 5) 对于递归的预测分析,根据预测分析表为每一个非终结 符编写一个过程;对于非递归的预测分析,实现表驱动 的预测分析算法





预测分析中的错误检测

- ▶两种情况下可以检测到错误
 - ▶ 栈顶的终结符和当前输入符号不匹配
 - ▶ 栈顶非终结符与当前输入符号在预测分析表对应项中的信息为空

预测分析中的错误恢复

- > 恐慌模式
 - ▶忽略输入中的一些符号,直到输入中出现由设计者选定的同步词法单元(synchronizing token)集合中的某个词法单元
 - ▶ 其效果依赖于同步集合的选取。集合的选取应该使得语法分析器 能从实际遇到的错误中快速恢复
 - ▶例如可以把FOLLOW(A)中的所有终结符放入非终结符A的同步记号集合 则放弃对A的识别,直接从A后面的字符开始分析
 - 》如果终结符在栈顶而不能匹配,一个简单的办法就是弹 出此终结符

非终		输入符号				
结符	id	+	*	()	*
E	$E \rightarrow TE'$			$E \rightarrow TE'$	synch	synch
E'		$E' \rightarrow +TE'$			$E' \rightarrow \varepsilon$	$E'\!\!\to\!\! arepsilon$
T	$T \rightarrow FT'$	synch		$T \rightarrow FT'$	synch	synch
T'		<i>T'</i> →ε	$T' \rightarrow *FT'$		<i>T'</i> →ε	$T' \rightarrow \varepsilon$
F	$F \rightarrow id$	synch	synch	$F \rightarrow (E)$	synch	synch

同步词法单元

X	FOLLOW(X)
E	\$)
E'	\$)
T	+)\$
T'	+)\$
F	*+)\$

Synch表示根据相应非终结符的FOLLOW集得到的同步词法单元

> 分析表的使用方法

- ▶ 如果M[A,a]是空,表示检测到错误,根据恐慌模式,忽略输入符号a
- \triangleright 如果M[A,a]是synch,则弹出栈顶的非终结符A,试图继续分析后面的语法成分
- > 如果栈顶的终结符和输入符号不匹配,则弹出栈顶的终结符

非终	输入符号					
结符	id	+	*	()	\$
E	E→TE'			E→TE'	synch	synch
E'		$E' \rightarrow +TE'$			Ε' →ε	$E'\!\!\!\to\!\!\! arepsilon$
T	$T \rightarrow FT'$	synch		$T \rightarrow FT'$	synch	synch
T'		<i>T'</i> →ε	$T' \rightarrow *FT'$		<i>T'</i> →ε	T' →ε
F	$F \rightarrow id$	synch	synch	$F \rightarrow (E)$	synch	synch

栈	剩余输入	
$m{E}$ \$	*id*+id \$	ignore +
E\$	id*+id\$	
<i>TE'</i> \$	id*+id\$	
<i>FT'E'</i> \$	id*+id \$	
id <i>T'E'</i> \$	id*+id\$	
<i>T'E'</i> \$	*+id \$	
* <i>FT'E'</i> \$	*+id \$	
<i>FT'E'</i> \$	+id \$	error
<i>T'E'</i> \$	+id \$	
<i>E'</i> \$	+id \$	
+ <i>TE'</i> \$	+id \$	
<i>TE'</i> \$	id \$	
<i>FT'E'</i> \$	id \$	
id <i>T'E'</i> \$	id \$	
<i>T'E'</i> \$		\$
E' \$	\$	
\$	\$	

