

词法分析

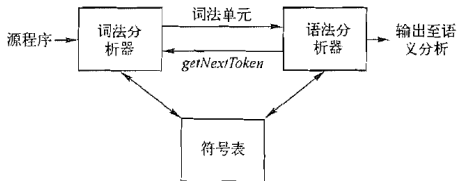
魏恒峰

hfwei@nju.edu.cn

2020 年 11 月 2 日



输入: 程序文本/字符串 s & **词法单元 (token) 的规约**



输出: 词法单元流

token : $\langle \text{token-class}, \text{attribute-value} \rangle$

词法单元	非正式描述	词素示例
if	字符 i, f	if
else	字符 e, l, s, e	else
comparison	< 或 > 或 <= 或 >= 或 == 或 !=	<=, !=
id	字母开头的字母 / 数字串	pi, score, D2
number	任何数字常量	3.14159, 0, 6.02e23
literal	在两个 " 之间, 除 " 以外的任何字符	"core dumped"

token : \langle token-class, attribute-value \rangle

词法单元	非正式描述	词素示例
if	字符 i, f	if
else	字符 e, l, s, e	else
comparison	< 或 > 或 <= 或 >= 或 == 或 !=	<=, !=
id	字母开头的字母 / 数字串	pi, score, D2
number	任何数字常量	3.14159, 0, 6.02e23
literal	在两个 " 之间, 除 " 以外的任何字符	"core dumped"

int/if 关键词

ws 空格、制表符、换行符

comment “//” 开头的一行注释或者 “/* */” 包围的多行注释

```
int main(void)
{
    printf("hello, world\n");
}
```

```
int main(void)
{
    printf("hello, world\n");
}
```

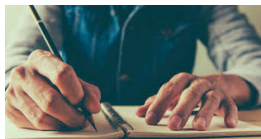
int ws **main/id** LP void RP ws
 LB ws
 ws id LP literal RP SC_{ws}
 RB

```
int main(void)
{
    printf("hello, world\n");
}
```

int ws main/id LP void RP ws
 LB ws
 ws id LP literal RP SCws
 RB

本质上, 这就是一个**字符串 (匹配/识别) 算法**

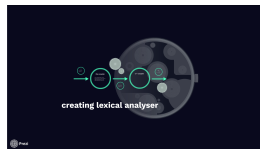
词法分析器的三种设计方法



手写词法分析器



词法分析器的生成器



自动化词法分析器

生产环境下的编译器 (如 gcc) 通常选择**手写词法分析器**

识别字符串 s 中符合某种词法单元模式的所有词素

```
if ab42>=42
    xyz =3.14
else xyz = 2.718
```

ws if else id integer real relop

识别字符串 s 中符合某种词法单元模式的所有词素

```
if ab42>=42
    xyz =3.14
else xyz = 2.718
```

ws if else id integer real relop

识别字符串 s 中符合某种词法单元模式的**第一个词素**

识别字符串 s 中符合某种词法单元模式的所有词素

```
if ab42>=42
    xyz =3.14
else xyz = 2.718
```

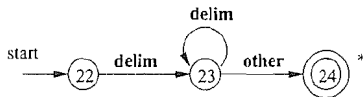
ws if else id integer real relop

识别字符串 s 中符合某种词法单元模式的**第一个词素**

识别字符串 s 中符合**特定词法单元模式**的第一个词素

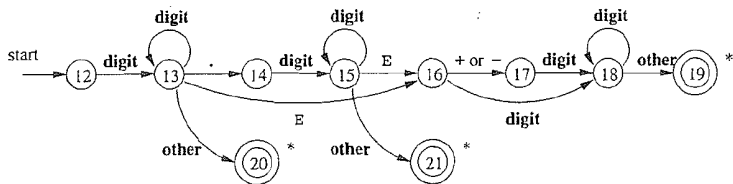
识别字符串 s 中符合**特定词法单元模式**的第一个词素

ws: blank tab newline



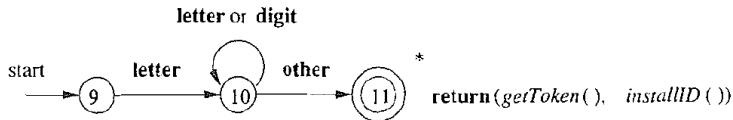
识别字符串 s 中符合**特定词法单元模式**的第一个词素

num: 整数 (允许以 0 开头)



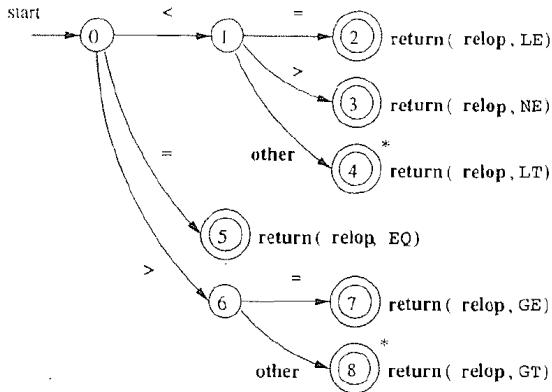
识别字符串 s 中符合**特定词法单元模式**的第一个词素

id: 字母开头的字母/数字串

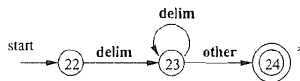
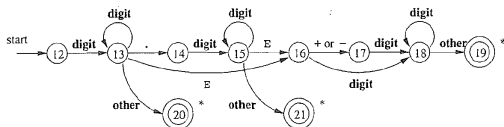
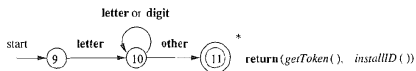
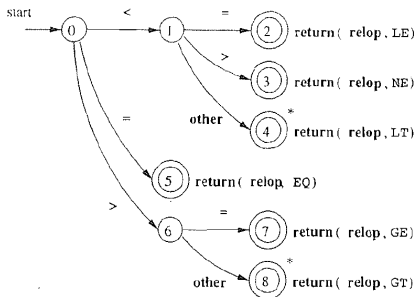


识别字符串 s 中符合**特定词法单元模式**的第一个词素

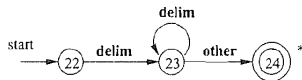
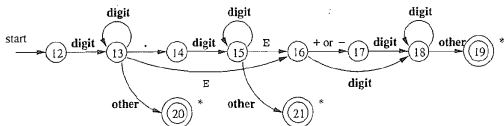
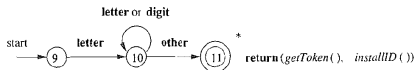
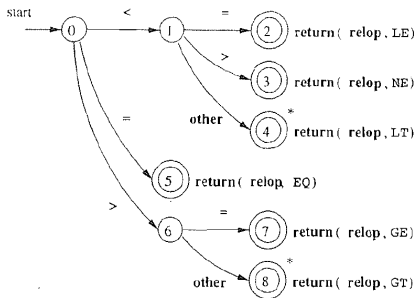
relop: < > <= >= == <>



识别字符串 s 中符合某种词法单元模式的第一个词素 (SCAN())

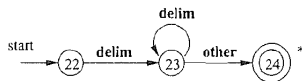
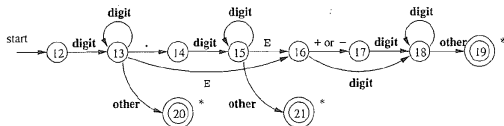
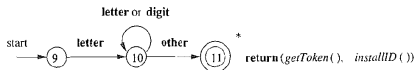
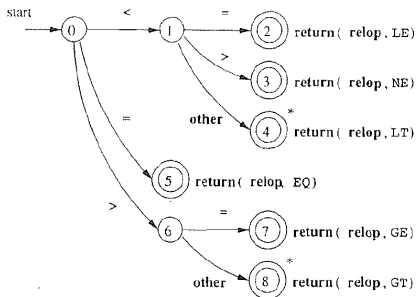


识别字符串 s 中符合**某种词法单元模式**的第一个词素 (SCAN())



根据**下一个字符**即可判定词法单元的类型

识别字符串 s 中符合**某种词法单元模式**的第一个词素 (SCAN())



根据**下一个字符**即可判定词法单元的类型

否则, 报告**该字符有误**, 并忽略该字符

识别字符串 s 中符合某种词法单元模式的所有词素

最外层循环调用 `SCAN()`

或者, 由语法分析器按需调用 `SCAN()`

Thank
You!



Office 926

hfwei@nju.edu.cn