

词法分析 (1)

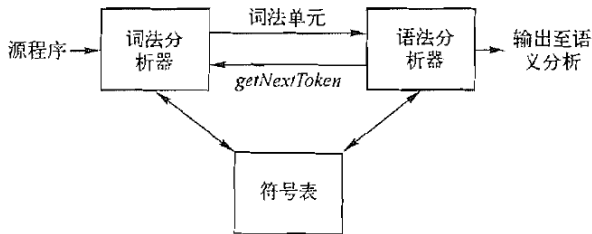
魏恒峰

hfwei@nju.edu.cn

2021 年 11 月 05 日 (周五)



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

```
CharStream input = CharStreams.fromStream(is);
SimpleExprLexer lexer = new SimpleExprLexer(input);

CommonTokenStream tokens = new CommonTokenStream(lexer);

SimpleExprParser parser = new SimpleExprParser(tokens);
ParseTree tree = parser.prog();
```

SimpleExpr.g4

```
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");
}
```

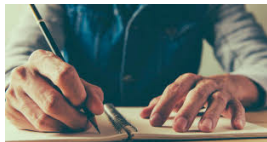
int WS main/id LP void RP WS
 LB WS
 WS id LP literal RP SC WS
 RB

本质上, 这就是一个**字符串 (匹配/识别) 算法**

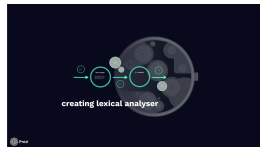
词法分析器的三种设计方法



词法分析器生成器



手写词法分析器



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**



master ▾

[gcc / gcc / c-family / c-lex.c](#)

iains Objective-C/C++ : Improve '@' keyword locations. ...

19 contributors +7

1435 lines (1278 sloc) | 38.8 KB

master ▾

[gcc / libcpp / lex.c](#)

urnathan cpplib: EOF in pragmas ...

25 contributors +13

4364 lines (3825 sloc) | 119 KB

Thank
You!



Office 926

hfwei@nju.edu.cn