

# 编译小组作业说明文档

苏行健 史志远 易郁杰 孙元旭

## 作业介绍

本次作业希望实现一个C语言到LLVM中间码的翻译器，作业以C11标准为参考，实现了部分的语法支持，并使用 ANTLR4 工具进行词法和语法分析。

## 语法支持

以下是作业实现的 ANTLR4 语法文件支持的精简语法特性（满足基本作业要求之外实现的特性已经加粗标出）：

### 1. 程序结构

- 编译单元和翻译单元
- 函数定义与声明

### 2. 类型系统

- 基本类型（`void`，`int`，`char` 等）
- `struct` 和 `typedef`
- 类型限定符（`const`）与存储类（`typedef`，`extern`，`static`）

### 3. 声明与初始化

- 变量声明、初始化
- 指针、多级指针
- 数组和函数声明

### 4. 表达式

- 赋值、条件、逻辑、算术
- 一元与后缀表达式

### 5. 语句

- 复合语句 `{ }`
- 条件语句（`if`，`else`）
- 循环语句（`while`，`do-while`，`for`）
- 跳转语句（`goto`，`continue`，`break`，`return`）

### 6. 运算符

- 算术（`+`，`-`，`*`，`/`，`%`）
- 关系（`<`，`>`，`<=`，`>=`，`=`，`!=`）
- 逻辑（`&&`，`||`，`!`）
- 其他（`&`，`→`，`.`，`++`，`--`，`?:`）

### 7. 字面量与标识符

- 标识符命名
- 数值、字符与字符串常量

### 8. 辅助

- 单行和多行注释
- 忽略空白字符

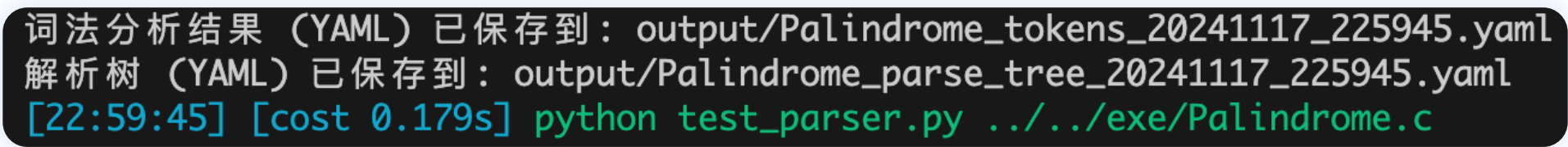
基本涵盖了 C 语言的核心语法。

词法分析与语法分析

对于下面判断回文字符串的C语言程序

```
1  #include <stdio.h>
2  #include <string.h>
3
4  void isPalindrome(char *s) {
5      int left = 0;
6      int right = strlen(s) - 1;
7      while (left < right) {
8          if (s[left] != s[right]) {
9              printf("False\n");
10             return;
11         }
12         left++;
13         right--;
14     }
15     printf("True\n");
16 }
17
18 int main() {
19     isPalindrome("abbcdbcba");
20     isPalindrome("abc");
21     return 0;
22 }
```

使用test\_parser.py进行词法和语法分析， 命令行输出截图如下



词法分析结果前三个token结果如下， 每一个token提供了类型、文本、所在行、所在列。

```
1  - type: VOID
2    text: void
3    line: 2
4    column: 0
5  - type: Identifier
6    text: isPalindrome
7    line: 2
8    column: 5
9  - type: LEFT_PAREN
10   text: (
11   line: 2
12   column: 17
```

语法分析树的前15行如下

```
1  rule: compilationUnit
2  children:
3  - rule: translationUnit
4    children:
5    - rule: functionDefinition
6      children:
```

```
7      - rule: declarationSpecifiers
8      children:
9      - rule: declarationSpecifier
10     children:
11     - rule: typeSpecifier
12     children:
13     - terminal: void
14       type: VOID
15     - rule: declarator
```

## 难点与创新点

1. LLVM中间代码生成（待第二次提交实现）
2. 指针、结构体等基础要求之外的C语法支持（暂时词法和语法分析部分）
3. 四则运算源文件翻译的实现（暂时词法和语法分析部分）

## 小组分工

- 词法分析和语法分析、符号表：苏行健、史志远
- LLVM中间代码生成：易郁杰、孙元旭

## 参考资料

- [C11标准](#)
- [ANTLR4工具使用方法](#)
- [ANTLR4提供的C语言g4文件](#)