

# 编译原理

课程Project说明

COMP130014.01

2021.09

# 简介

- 本学期共有两个Project:
- Project 1: 词法分析, 占比**40%**
- Project 2: 语法分析, 占比**60%**
- 每组1-2人, **2021年10月7日23:59**之前确定组员名单发送至负责Project的TA邮箱
- TA联系方式:
  - 孙玉齐 [20210240018@fudan.edu.cn](mailto:20210240018@fudan.edu.cn) (负责Project)
  - 白玉 [20210240008@fudan.edu.cn](mailto:20210240008@fudan.edu.cn) (负责课程作业)
- 面对面建群, 发布Project/答疑/预约汇报时间

# 实验环境

- OS: linux
- 依赖: gcc/g++, flex, bison
- flex与bison安装(以Ubuntu为例):
  - `sudo apt-get install flex`
  - `sudo apt-get install bison`
- 实验环境也可自行在MAC OS以及WINDOWS下配置
- 简单来说, 就是C/C++配合flex与bison两个工具完成实验

# 实验目的

- 通过flex与bison，分析目标PCAT语言，并生成目标语言的语法树
- PCAT语言可看作一种简化版的PASCAL语言：

```
PROGRAM IS
  VAR i, j : INTEGER := 1;
  VAR x : REAL := 2.0;
  VAR y : REAL := 3.0;
BEGIN
  WRITE ("i = ", i, ", j = ", j);
  WRITE ("x = ", x, ", y = ", y);
END;
```

# Project 1 词法分析 (40%)

- 任务：使用flex工具对于给定的PCAT语言样例做词法分析
- 左侧：PCAT语言代码
- 右侧：词法分析的结果

```
PROGRAM IS
  VAR i, j : INTEGER := 1;
  VAR x : REAL := 2.0;
  VAR y : REAL := 3.0;
BEGIN
  WRITE ("i = ", i, ", j = ", j);
  WRITE ("x = ", x, ", y = ", y);
END;
```

ROW	COL	TYPE	TOKEN/ERROR MESSAGE
1	1	reserved keyword	PROGRAM
1	9	reserved keyword	IS
2	5	reserved keyword	VAR
2	9	identifier	i
2	10	delimiter	,
2	12	identifier	j
2	14	delimiter	:
2	16	identifier	INTEGER
2	24	operator	:=
2	27	integer	1
2	28	delimiter	;
3	5	reserved keyword	VAR
3	9	identifier	x
3	11	delimiter	:
3	13	identifier	REAL
3	18	operator	:=
3	21	real	2.0
3	24	delimiter	;
4	5	reserved keyword	VAR
4	9	identifier	y
4	11	delimiter	:
4	13	identifier	REAL
4	18	operator	:=
4	21	real	3.0
4	24	delimiter	;
5	1	reserved keyword	BEGIN
6	5	reserved keyword	WRITE
6	11	delimiter	(
6	12	string	"i = "
6	18	delimiter	,
6	20	identifier	i
6	21	delimiter	,
6	23	string	", j = "
6	31	delimiter	,

# Flex简介

- 一种可以使用正则表达式完成文本词法分析的工具
- 举例：提取出只有加法和减法的表达式的token

# Flex简介

lexer.lex

```
1 %{
2   #include "lexer.h"
3   %}
4   %option      nounput
5   %option      noyywrap
6
7   DIGIT        [0-9]
8   INTEGER      {DIGIT}+
9   REAL         {DIGIT}+"."{DIGIT}*
10  WS           [ \t]+
11
12  %%
13  {WS}          /* skip blanks and tabs */
14  <<EOF>>       return T_EOF;
15  "+"           return ADD;
16  "-"           return SUB;
17  {INTEGER}|{REAL} return NUMBER;
18  %%
19
```

定义区

规则区

用户代码区

lexer.h

```
1  #ifndef _LEXER_H_
2  #define _LEXER_H_
3
4  #define T_EOF      0
5  #define ADD        1
6  #define SUB        2
7  #define NUMBER     3
8
9  #endif
10
```

# Flex简介

编译:

flex -o lexer.c lexer.lex

g++ -c lexer.c -o lexer.o

g++ main.cpp lexer.o -o lexer

```
(base) user2@DIVPInspur250:~/Compile/flex_demo$ ./lexer
1+2-3.3+2.2
1
+
2
-
3.3
+
2.2
```

main.cpp

```
1  #include <iostream>
2  #include <cstdio>
3  #include "lexer.h"
4  using namespace std;
5
6  int yylex();
7  extern "C" FILE *yyin;
8  extern "C" char *yytext;
9
10 int main(int argc, char **argv)
11 {
12     if (argc > 1) {
13         yyin = fopen(argv[1], "r");
14     } else {
15         yyin = stdin;
16     }
17
18     while (true) {
19         int n = yylex();
20         if (n == T_EOF) {
21             break;
22         }
23         cout << yytext << endl;
24     }
25
26     return 0;
27 }
```



# Project 1 评分细则

## 项目完成度及正确性：（共计70分）

1. 正确分析case 1-10中出现的所有tokens并统计tokens的总数，将每个case的词法分析结果存储成txt格式 **（必做，每个样例完全正确得5分，共计50分）**
2. 正确输出每一个token的起始行号、列号与类型（选做，每个样例完全正确得1分，共计10分）
3. 正确分析case 11中出现的各种无需语法分析的词法错误，提供相应报错信息（选做，25个测试点，其中存在10个错误，共计10分）

## 项目报告及展示：（共计30分）

1. 撰写项目报告，说明flex的用法，识别不同token所使用的正则表达式及其原理，如何判断token的行列号及类型，如何实现报错功能等等，在结尾标明分工及贡献百分比 **（必做，20分）**
2. 项目报告完成后，与TA预约，在上机课时间向TA展示样例的词法分析结果，TA会就项目相关内容进行简单的提问 **（必做，10分）**

# Project 1 提交方式

- 项目代码及项目报告PDF请打包并发送至TA邮箱 [20210240018@fudan.edu.cn](mailto:20210240018@fudan.edu.cn)
- **项目报告DDL: 2021年11月1日 23:59**
- 提交报告后与助教预约展示时间
- **展示DDL: 2021年11月12日上机课**
- 如果文件太大，可先上传至百度云或者复旦云，再将网盘分享地址发送到TA邮箱
- 若对Project有疑问，或想在上机课外时间展示，可与TA联系
- TA办公地址：江湾校区交叉学科2号楼A4008室
- **严禁抄袭，一经发现Project作0分处理**
- **只实现必做功能也一定可以顺利通过，不要铤而走险**

# 参考资料

- Flex manual:

[http://ranger.uta.edu/~fegaras/cse5317/flex/flex\\_toc.html](http://ranger.uta.edu/~fegaras/cse5317/flex/flex_toc.html)

- Bison manual:

[http://ranger.uta.edu/~fegaras/cse5317/bison/bison\\_toc.html](http://ranger.uta.edu/~fegaras/cse5317/bison/bison_toc.html)

- PCAT语言参考中有相应的LL1文法参考