

LAB3 ROSE

ex2

21307099 李英骏

目录

1	Oberson-0 语言词汇表	2
2	Oberon-0 语言词法规则	3
2.1	词法规则的正则定义式描述	3
2.2	词法规则异同	3
3	lex 族工具的输入差异	4
4	实验记录: 词法分析器的生成	5

1 Oberson-0 语言词汇表

保留字列表		运算符列表	
类型	说明	类型	说明
MODULE	程序模块声明（保留字）	+	加法（取正）
BEGIN	程序段开始声明（保留字）	-	减法（取负）
END	程序段结束声明（保留字）	*	乘法运算
CONST	常量声明（保留字）	>	大于
TYPE	自定义类型定义（保留字）	>=	大于等于
VAR	定义变量（保留字）	<	小于
PROCEDURE	定义子过程（保留字）	<=	小于等于
RECORD	定义结构体（保留字）	#	不等于
ARRAY	定义数组（保留字）	=	等于
OF	语句关键字（保留字）	&	与运算
OR	或运算（保留字）)	右括号
DIV	整数除法运算（保留字）	:	类型声明标志
MOD	取模运算（保留字）	;	语句结束标志
WHILE	循环语句开始声明（关键字）	~	逻辑非操作
DO	循环语句（关键字）	:=	赋值运算
IF	条件语句开始声明（关键字）	.	域操作运算符
THEN	条件语句（关键字）	[数组下标左括号
ELSIF	条件语句（关键字）]	数组下标右括号
ELSE	条件语句（关键字）	(左括号
INTEGER	整数类型定义（关键字）	,	变量间隔标志
BOOLEAN	布尔类型定义（关键字）	EOF	源程序结束标志

常量	标识符
[1-9][0-9]* 0[0-7]*	letter(letter digit)*

2 Oberon-0 语言词法规则

2.1 词法规则的正则定义式描述

```

Number      -> [1-9][0-9]* | 0[0-7]*
Identifier  -> [a-zA-Z][0-9a-zA-Z]*
WhiteSpace  -> " " | \r | \n | \r\n | \t | \f | \b
Comment     -> \(\*([^\*] | (\*[^*\])))* (\(\*))\* \* \))

```

2.2 词法规则异同

相同点

- 标识符不能以数字开头.
- 八进制数以 0 开头.

不同点

- 标识符规则:
 - C/C++ 和 Java: 标识符可以包含下划线, 并且长度不限.
 - Oberon-0: 标识符不能包含下划线, 并且长度限制为 24 个字符.
- 注释风格:
 - C/C++: 支持 /* */ 块注释和 // 行注释.
 - Oberon-0: 使用 (* *) 作为注释.
- 整数表示:
 - C/C++ 和 Java: 整数表示不受进制影响, 使用二进制表示法.
 - Oberon-0: 整数表示受十进制和八进制的区别影响.
- 运算符:
 - C/C++ 和 Java: 使用
 - Oberon-0: 使用 MOD 进行取模, 使用 DIV 进行除法, 使用 = 进行等于比较, 使用 # 进行不等于比较, 使用 OR 进行或操作, 使用 进 进行非操作.

3 lex 族工具的输入差异

JFlex

- **正则表达式**: JFlex 支持扩展的正则表达式, 包括字符类、组、量词、断言等.
- **状态机**: JFlex 允许定义多个词法分析状态, 并在不同状态之间切换. 这使得处理复杂的词法规则变得更加容易.
- **Unicode 支持**: JFlex 完全支持 Unicode, 可以处理多种语言的输入.
- **用户代码**: JFlex 允许在词法规则中嵌入 Java 代码, 以便在匹配到特定模式时执行.

JLex

- **正则表达式**: JLex 支持基本的正则表达式, 但不如 JFlex. 它主要提供了字符类、组、量词等基本功能.
- **状态机**: JLex 也支持词法分析状态的定义, 但功能和灵活性不及 JFlex.
- **Unicode 支持**: JLex 的 Unicode 支持较为有限, 主要用于处理 ASCII 和基本的 UTF-8 字符集.
- **用户代码**: JLex 允许在词法规则中嵌入 Java 代码, 但嵌入方式较为简单, 缺乏 JFlex 的灵活性.

GNU Flex

- **正则表达式**: GNU Flex 支持广泛的正则表达式, 包括字符类、组、量词、断言等. 正则表达式功能非常强大, 与 JFlex 相当.
- **状态机**: GNU Flex 允许定义多个词法分析状态, 并在不同状态之间切换. 处理复杂的词法规则时更方便.
- **Unicode 支持**: GNU Flex 的 Unicode 支持较为有限, 主要用于处理 ASCII 和基本的 UTF-8 字符集.
- **用户代码**: GNU Flex 允许在词法规则中嵌入 C 代码.

JFlex 的基本框架

- 用户代码
- %% 选项与声明
- 词法规则

JFlex 的选项与声明

- %class 中定义了生成的 Java 文件的类名以及文件名.
- %eofval 中定义了当识别到 EOF 的时候执行的操作.
- %yylexthrow 中定义了当前词法分析程序中要抛出的异常.
- %cup 的声明是为 java_cup 提供接口, 因此生成了 next_token() 函数.
- %line 以及 %column 是声明了可以使用 yyline 以及 yycolumn 获得当前的位置.
- %..... 用户代码.....% 中填写的是用户代码, 一般在这里实现 main 函数.
- <YYINITIAL> 中定义了对于匹配每一个正则表达式后要执行的操作.

比较

- **正则表达式功能**: JFlex 和 GNU Flex 支持更强大的正则表达式功能.
- **Unicode 支持**: 只有 JFlex 提供了更全面的 Unicode 支持.
- **用户代码嵌入**: JFlex 和 GNU Flex 的嵌入方式更为灵活.

4 实验记录: 词法分析器的生成

软装置中似乎并没有文档提及的 Exceptions 和测例. Exception 自行编写, 有几个是复制了上次 LAB2 计算器的, 测例使用 ex1 中构建的词法错误测例.