

MIDL 的词法规则和编译原理课程实验要求

西工大软件学院

1. MIDL 的词法规则

1) 语言的关键字 (16 个) :

struct	float	boolean	short	long
double	int8	int16	int32	int64
unit8	unit16	unit32	unit64	char
unsigned				

注：关键字是保留字，并且不区分大小写。

2) 语言的专用符号 17 个:

{	}	;	[]
*	+	-	~	/
%	>>	<<	&	^
	,			

3) 标示符 **ID**、整数 **INTEGER**、字符串 **STRING** 和 **BOOLEAN** 等 4 个词法规则（标黑的）通过下列正则表达式定义：

ID = LETTER (UNDERLINE?(LETTER | DIGIT))*

LETTER = [a-z] | [A- Z]

DIGIT = [0-9]

UNDERLINE= _

INTEGER = (0 | [1-9] [0-9]*) **INTEGER_TYPE_SUFFIX**?

INTEGER_TYPE_SUFFIX = 1 | L

STRING = " (~\ | ~")* "

BOOLEAN = TRUE | FALSE

4) 补充说明的注意事项

其中，正则符号标识含义如下表：

~\ ~"	表示匹配除了 \， "外所有字符
[a-z]	表示 a 字符到 z 字符，与课件 ppt 中含义一致
?	表示匹配 0 或 1 次，与课件 ppt 中含义一致
+	表示匹配 1 或 多次，与课件 ppt 中含义一致
*	表示匹配 0 或 多次，与课件 ppt 中含义一致
a b	表示匹配 a 字符或 b 字符，与课件 ppt 中含义一致

()	表示一个组，比如()*表示*对括号内组整体起作用，与课件 ppt 中含义一致
-----	--

空白、换行符和制表符在词法分析时可以忽略掉，但是字符串本身的空白、换行符和制表符不能忽略。

2. 词法分析程序课程设计要求

词法分析程序的设计与实现需要按照以下要求完成。

1) 基于词法规则设计词法分析器（20 分）

画出确定的有穷自动机（确定化），并提供必要的文字说明。提交状态转换图.doc

2) 词法分析程序的编程实现与测试（80 分）

(1) 编程实现词法分析器，提交可执行词法分析程序的源程序（语言不限，60 分）

编程实现词法分析程序，该词法分析程序能够读取测试输入文件中的源程序，并将其词法分析的结果即 token 序列或词法分析的所有错误信息输出到 tokenOut.txt 中，以便于检查你的词法分析程序的正确性。输出的错误信息要包括错误所在行数，错误类型等信息（同目录下有测试文件夹，共五大类测试输入）。

(2) 提交所生成的词法分析程序的测试方案（20 分）.doc

注：将状态转换图.doc、源程序、以及词法分析程序测试方案.doc 文件放在一个文件夹下，按照“班级_学号_姓名”的命名方式打包提交。如果还有其它需要说明的问题须写在 readme.doc 中。

3. 有关附加分的说明

关于字符串，如果词法分析程序考虑转义字符（正规表达式的定义如下），则作为加分项目，可以选做。即如果词法分析程序考虑了对转义字符的支持，额外加分 10 分，但总分不得超过 100 分。例如，根据词法分析程序的完成情况，某同学取得 92 分的成绩，但由于该同学的词法分析程序正确处理了转义字符（必须在状态转换图、词法分析源程序和 `readme.doc` 中分别进行备注），该同学的总分合计为 100 分。

STRING = " (ESCAPE_SEQUENCE | (~\|~"))* "

ESCAPE_SEQUENCE = \ (b | t | n | f | r | " | \)