



编译原理

实验一：词法分析

规格严格，功夫到家

授课教师：李旭涛
实验中心：汪花梅

实验安排

✓ 12个学时，共计4次实验课，完成4个实验；

✓ 上课时间：【10】周三1-2节，【11】周四9-12节，【12】周四1-4节，【13】周三1-2节；

实验一	第10周三1-2节	2学时		词法分析
实验二	第11周四9-12节	4学时	提交实验一	自底向上的语法分析（LR(1)）
实验三	第12周四1-4节	4学时	提交实验二	典型语句的语义分析及中间代码生成
实验四	第13周三1-2节	2学时	提交实验三	目标代码生成

✓ 实验四作业的截止日为：第14周，周五17:00前

实验安排

✓作业所需提交：

1. 纸质版实验报告（下次实验课上统一收齐）
2. 电子版代码压缩包（截止时间同报告）

电子版提交内容：工程文件（含源代码），输入输出文件
压缩包&实验报告命名：学号_姓名_实验X.rar

例：18xxxx_小明_实验一.rar

✓实验总分数：20分。实验报告占40%，代码占60%



编译原理实验报告
-模板

实验一	词法分析	16分(百分制)
实验二	自底向上的语法分析 (LR(1))	34分(百分制)
实验三	典型语句的语义分析及中间代码生成	34分(百分制)
实验四	代码生成	16分(百分制)



实验提交

实验提交网址: <http://grader.tery.top/>

 登录

用户名

 180110103

密码

 密码同用户名 

登录

← 后退

2020秋_编译原理_1&2&3&4&5班作业

实验一_词法分析器 文件作业 开放中

开始时间	截止时间
2020-11-03 00:00	2020-11-12 00:00
答案公布时间:	2021-01-09 00:00
上次提交	得分
无	无

开始作答

2020秋_编译原理_1&2&3&4&5班

课程负责人

汪花梅

近期作业

实验一_词法分析器 8天后截止

无

无

进入课程

题目1

分数: 100

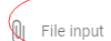
实验内容:

- 1、读取文件, 对文件内自定义的类C语言程序段进行词法分析;
- 2、输出以文件形式存放的TOKEN串和简单符号表;
- 3、要求所输入的C语言程序段包含常见的关键字, 标识符, 常数, 运算符和分界符等。

实验步骤:

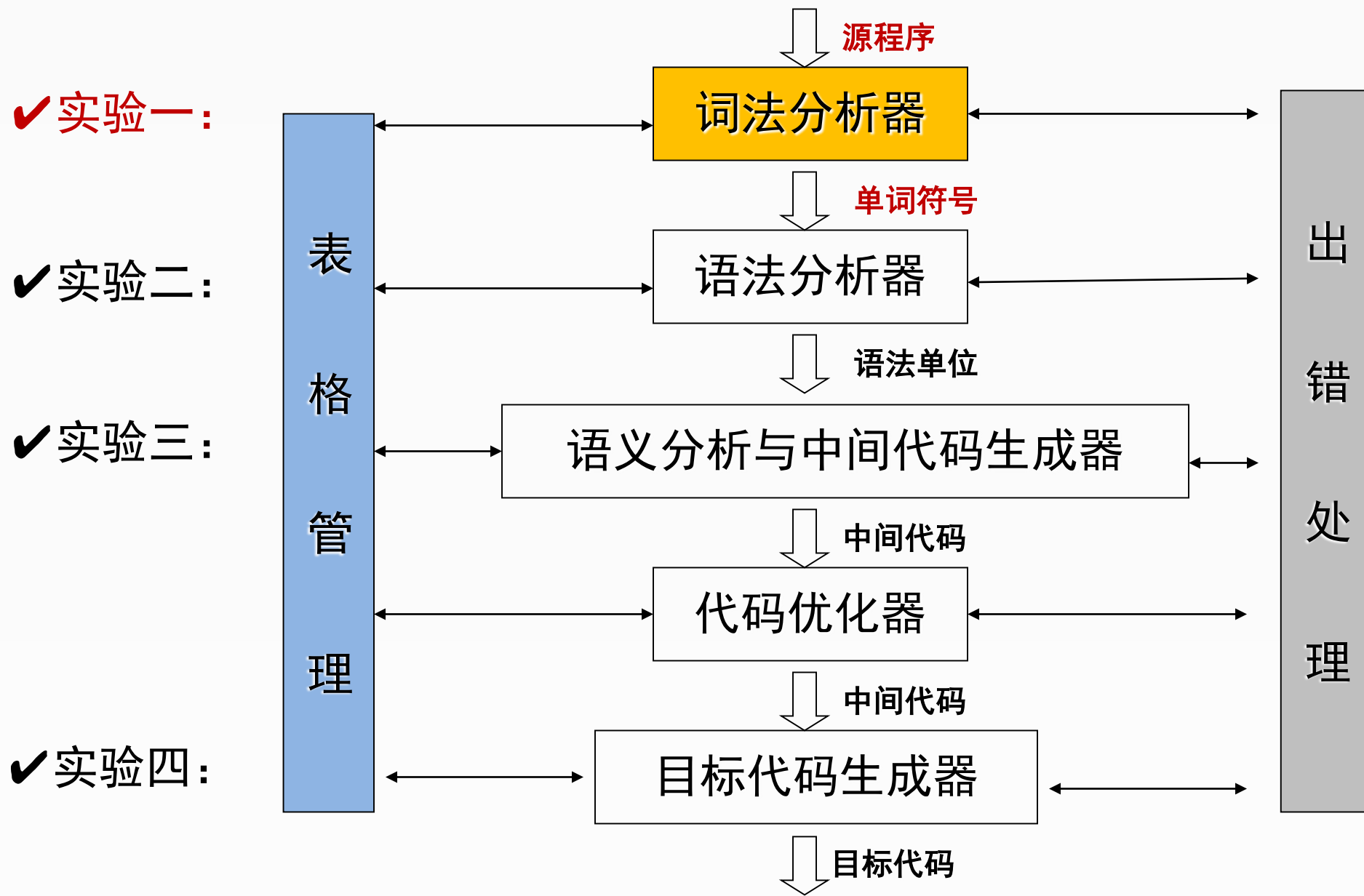
创建编码表;
创建类C语言文法, 画出有穷自动机, 状态转换图;
根据有穷自动机的状态转移编写代码;
定义数据结构: TOKEN串, 符号表等;
完成代码编写, 实现实验内容;

允许的文件后缀: zip



点击上传压缩文件

编译程序的总体结构



实验目的

1. 加深对词法分析程序的功能及实现方法的理解。
2. 对类C语言的文法描述有更深入的认识，理解有穷自动机、编码表和符号表在编译的整个过程中的应用。
3. 设计并编程实现一个词法分析程序，对类C语言源程序段进行词法分析，加深对高级语言的认识。

实验学时数：2学时

实验内容

编写一个词法分析程序，读取文件，对文件内自定义的类C语言程序段进行词法分析。

1. **输入**：以文件形式存放自定义的类C语言程序段；
2. **输出**：以文件形式存放的**TOKEN串**和简单**符号表**；
3. **要求**：输入的C语言程序段包含常见的**关键字**，**标识符**，**常数**，**运算符**和**分界符**等。

实验总体步骤

1. 创建**编码表**；
2. 定义数据结构：**TOKEN串**，**符号表**等；
3. 创建属于自己的**类C语言文法**；
4. 根据所建文法画出**有穷自动机**，**状态转换图**；
5. 根据有穷自动机的状态转移**编写代码**；
6. **输入**自定义的类C语言程序段文件；
7. **输出**以文件形式存放的TOKEN串和符号表；
8. **完成实验报告**。

编码表

单词名称	类别编码	单词值
标识符	1	内部字符串
无符号常数(整)	2	整数值
无符号浮点数	3	数值
布尔常数	4	0 或 1
字符串常数	5	内部字符串
BEGIN	6	—
END	7	—
FOR	8	—
DO	9	—
.....
:	20	—
+	21	—
*	22	—
,	23	—
(.....	—

注意:

- 标识符一个编码
- 关键字、运算符、分界符一符一码
- 常量一类一码

词法分析器的输出——TOKEN串（单词符号）

单词的种类

- (1) **标识符**：由用户定义，表示各种名字；
- (2) **关键字**：也称基本字，do、while、int、char、sizeof...
- (3) **常数**：整常数、实常数、字符常量、字符串常量、符号常量；
- (4) **运算符**：算术运算符+、-、*、/等；
逻辑运算符not、or与and等；
关系运算符=、<>、>=、<=、>和<等；
- (5) **分界符**：，、；、（、）...

词法分析器输出实例

举例：程序语句 `if count>7 then result := 3.14` 的单词符号序列

(IF, 0)
(ID, 指向count的符号表入口)
(GT, 0)
(INT, 7)
(THEN, 0)
(ID, 指向result的符号表入口)
(ASSIGN, 0)
(REAL, 3.14)

表示单词的种类，可用整数
编码或记忆符表示

种别码

属性值

不同的单词不同的值

标识符和常数的属性用两种：

- 1、如果还未建符号表，可以用标识符的单词本身表示。
- 2、标识符存入符号表中，用指向该标识符所在符号表入口的指针作为其属性值。

词法分析器的输出——符号表

符号表：以名字为关键字来记录其信息的**数据结构**；

支持的基本操作包括**插入**、**查找**和**删除**。

符号表的数据结构类型：线性表（优点：简单直观，缺点：插入时间复杂度高）

散列表（查找、插入效率高）

注意点：

名字字段的长度是否需要限制？如何限制？

	名字	属性
符号表表项1	abc	...
符号表表项2	i	...
⋮
符号表表项 n

图：符号表

文法设计

➤ 正则文法表示

$G=(V,T,P,S)$, 其中 $V=\{S,A,B,C,\text{digit},\text{no_0_digit},\text{char}\}$, $T=\{\text{任意符号}\}$, P 定义如下

约定: 用 digit 表示数字: $0,1,2,\dots,9$; no_0_digit 表示数字: $1,2,\dots,9$;

用 letter 表示字母: $A,B,\dots,Z,a,b,\dots,z,_$

标识符: $S \rightarrow \text{letter } A \quad A \rightarrow \text{letter } A | \text{digit } A | \varepsilon$

整常数: $S \rightarrow \text{no_0_digit } B \quad B \rightarrow \text{digit } B | \varepsilon$

运算符: $S \rightarrow C \quad C \rightarrow = | * | + | - | /$

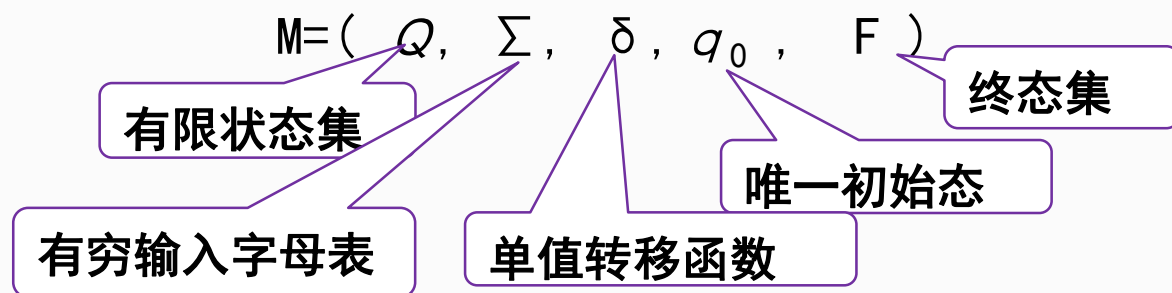
➤ 正则表达式表示法

标识符: $\text{id} \rightarrow \text{letter} (\text{letter} | \text{digit})^*$

整常数: $\text{id} \rightarrow \text{no_0_digit} (\text{digit})^*$

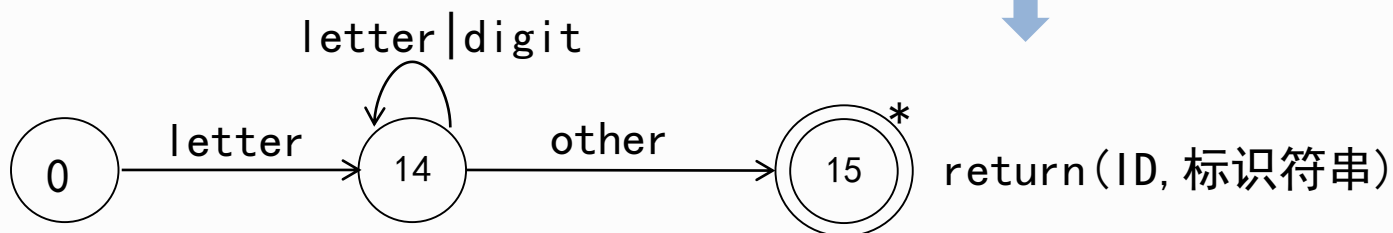
有穷自动机

确定的有穷自动机 (DFA):



标识符正则表达式: $\text{id} \rightarrow \text{letter} (\text{letter} | \text{digit})^*$

识别标识符和关键字的状态转换图:

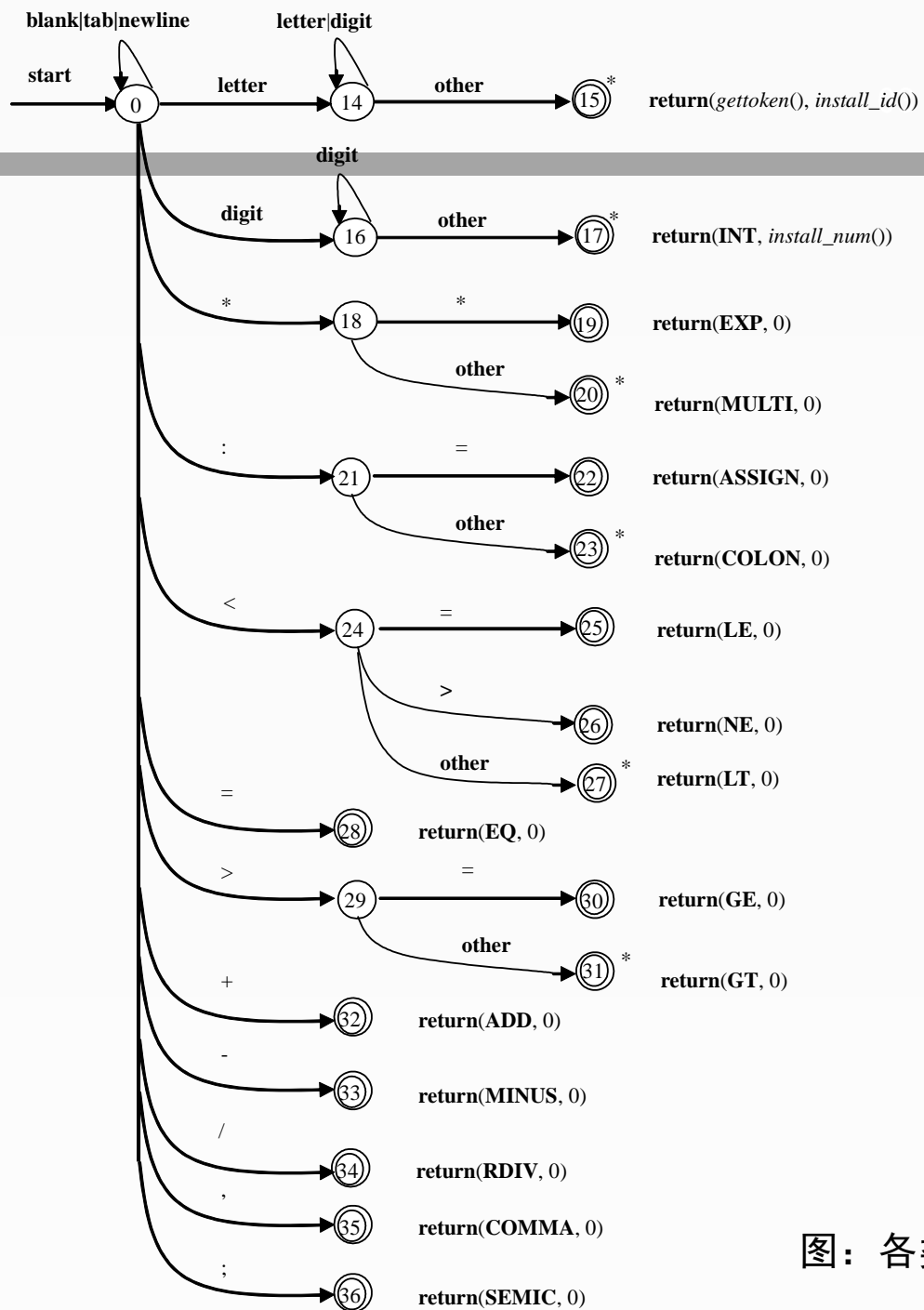


DFA的
三种表示

转换函数

转移矩阵

状态转换图



实验步骤



图：各类状态转化图合并

附加功能

非必要完成项，如完成，请在实验报告中另行标注实现方法并画出完整的有穷自动机，酌情加分但不超出实验部分的总分。

- 实现词法分析程序的拓展功能，比如：
对八进制数、十六进制数（包括但不限于）等进行词法分析。

实验报告内容

1. 写出实验中所用的C语言程序段的**文法描述**；
2. 画出完整的**有穷自动机**（参考教材P93(chapter3. ppt第55页)）；
3. 画出实验中用到的单词的**词类编码表**（参考教材P66(chapter3. ppt第7页)）；
4. 设计并描述**符号表**的逻辑结构及存贮结构；
5. 描述词法分析程序中主要模块的算法；
 - a) 分类函数的算法；
 - b) 识别标识符的算法；
 - c) 查填符号表的算法；
6. 词法分析后生成的**TOKEN串文件**和**符号表文件**；
7. 实验中遇到的问题及解决方案。

实验思考

- 思考词法分析程序能否能够使用空格来区分单词？
- 思考词法分析程序作为一个独立子程序具有哪些优点？

词法分析器输入输出举例

输入代码:

```
s = 123;  
mul_x = s * 2;  
while(mul_x > 100)  
{  
    mul_x = mul_x - 1;  
}  
test = 321.0 / 0x1234;
```

词法分析器输出单词:

```
(1,s)  
(9,=)  
(2,123)  
(16,;)  
(1,mul_x)  
(9,=)  
(1,s)  
(3,*)  
(2,2)  
(16,;)  
(1,while)  
(21,()  
(1,mul_x)  
(11,>)  
(2,100)  
(22,))  
(19,{)  
(1,mul_x)  
(9,=)  
(1,mul_x)  
(13,-)  
(2,1)  
(16,;)  
(20,})  
(1,test)  
(9,=)  
(18,321.0)  
(14,/  
(26,0x1234)  
(16,.)
```

词法分析器输出符号表:

```
(s,-1,?)  
(mul_x,-1,?)  
(test,-1,?)  
(a,-1,?)  
(b,-1,?)  
(printf,-1,?)
```

同学们，
独立开始实验