**《软件构造》实验讲义**

黑龙江大学计算机科学技术学院

黑龙江大学软件学院

2023年3月

**一、开设本实验的目的**

本实验课程的开设是为配合《软件构造》课程的教学，加深学生对程序设计语言编译器基本概念、基本理论与实现方法的理解，使学生能够通过计算机语言处理技术的学习掌握软件系统设计与构造的基本方法。通过本实验课程，使学生系统地掌握编译系统的结构、工作流程及编译程序核心组成部分的设计原理和实现技术，从而具备初步开发系统软件和应用软件的能力。

**二、实验内容安排**

实验要求完成两部分内容：

**第一部分：词法分析程序的构造。**

构造识别各类单词的自动机，掌握符号表的建立与单词的分类方法，掌握词法分析器的设计方法。

**第二部分：基于预测分析法的语法分析程序的构造。**

构造识别指定表达式的语法分析程序，掌握自顶向下语法分析的基本思想，掌握利用预测分析法进行语法分析的基本原理与过程。

**三、实验要求**

1、每个实验项目都有规定的完成时间，需要在规定的时间完成相应的任务。一个实验项目结束方可进行下一项。

2、每名同学按照自己完成的各个实验内容根据给定的实验报告模板撰写相应的实验报告。

3、全部实验完成后每名同学需要提交实验报告及完整的程序代码。

**四、成绩评定依据**

学生要对各个实验环节进行理论学习、实验分析与设计、编码实现、最后进行分析与总结，每个实验完成后要提交实验报告。

本实验要求每个学生完成词法分析程序和语法分析程序的构造。实验环节占20分。完成选做内容的同学可加分。

整个实验的评分依据包括实验报告、每个实验项目的完成时间、程序代码三个方面。

实验课程的具体内容、进度安排、考核标准及学时安排如下表：

表1 实验进度安排及成绩构成表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 实验项目名称 | 内容提要及考核标准 | 时间安排 | 分数 |
| 1 | 词法分析程序的构造 | 将输入的源程序从左到右进行扫描和分解，提取出单词并依次输出各个单词的类别码及自身值；如遇到错误显示“ERROR”;同时进行标识符登记符号表的管理。按时保质完成全部实验要求可得10分。 | 4学时 | 10分 |
| 2 | 基于预测分析法的语法分析程序的构造 | 根据给定的表达式文法（满足LL(1)文法的条件）构造预测分析表，并利用预测分析表实现对给定表达式的语法分析过程。如遇到错误显示“ERROR”。按时保质完成全部实验要求可得10分。 | 4学时 | 10分 |

**五、具体内容**

**实验一 词法分析程序的构造**

**（一）实验目的**

1、了解词法分析的任务。

2、掌握构造识别各类单词自动机的方法及过程。

3、掌握符号表的建立及单词的分类方法。

4、掌握词法分析程序的基本原理与构造方法。

**（二）实验内容**

对于给定的符合C语言文法所定义的一段程序，使用高级语言（c、c++、java、python）编写程序完成下列内容：

1、根据C语言所有单词的正规式（或正规文法）构造识别正规式的NFA。

2、将1构造出的NFA确定化。

3、将2构造出的DFA最小化。

4、根据DFA识别所有的标识符。

5、根据DFA识别所有的常数（至少包含整数）。

6、根据DFA识别C语言中定义的所有关键字（按照32个来做）

7、根据DFA识别C语言中的所有界符。

8、根据DFA识别C语言中的所有运算符。

其中1-3为选作，4-8为必作。选作内容有加分。

（三）实验指导

1、单词的正规文法及正规式示例。

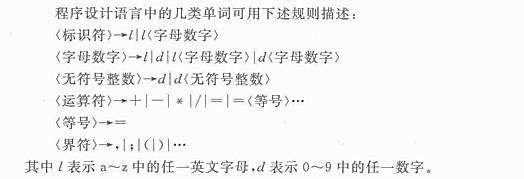
 正规文法如图1-1所示：

图1-1 单词正规文法示例

正规式：

标识符：l(l|d)\*； 无符号整数：dd\*

2、编写函数：构造识别正规式的NFA，存储构造出的NFA的状态矩阵。

根据构造出的NFA状态图，选用适当的存储结构存储该NFA的状态矩阵。

对于状态图中的每一条边，用户使用一个三元组（A,a,B）的形式进行输入，其

中A、B是状态（A、B可以是同一个状态），a是一个输入符，输入一个三元组

后，状态矩阵就相应改变，当把所有的三元组都输入后，NFA的状态矩阵存储

完毕。

3、编写函数：利用2构造出的NFA状态矩阵构造其对应的DFA的状态矩

阵。

根据子集法构造DFA的过程将1构造出来的NFA状态矩阵转化成它对应的DFA的状态矩阵。（具体过程自己设计并实现）

4、编写函数：将3构造出来的DFA进行最小化。

根据DFA最小化的方法，编写程序进行若干次划分，得到最小化的DFA的

状态矩阵。（具体过程自己设计并实现）

5、编写函数:利用最小化的DFA的状态矩阵，设计并编写程序识别输入的

源程序（source.txt）。

从最小化的DFA的开始状态出发，依次扫描输入符号串的每个字符，每扫描到一个字符，根据DFA状态矩阵跳转到对应的状态，若输入符号串的所有符号被识别后到达终态，以二元式（类别码，符号串）【常数】或三元式（类别码，符号串，地址）【关键字、标识符、运算符、常数、界符】的形式写入名为result.txt的文件中。

（1）C语言单词正规式的状态转换图示例如图1-2所示。

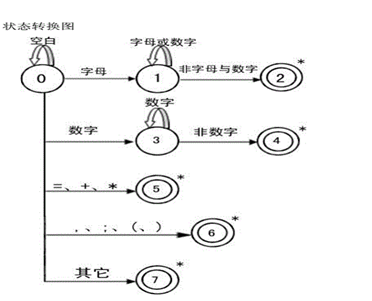


图1-2 状态转换图示例

（2）将一段（不少于5行）的c语言编写的程序存入source.txt中。

（3）使用适当的存储结构存储关键字表、界符表、运算符表、标识符表。

1. 关键字表

|  |
| --- |
| auto break case char const continue default do double else  enum extern float for goto if int long register return short signed sizeof static struct switch typedef union unsigned void volatile while |

1. 界符表：" ' [ ] { } ( ) ; , : | &
2. 运算符表：+ - \* / > < =
3. 标识符表【程序创建，动态增加】

（4）从名为source.txt的文件中读取字符，识别出符号串及其类型。【关键字、标识符、运算符、常数、界符】。

（5）对于识别出的标识符，将其写入名为id.txt的文件中，对于每个识别出

的符号串，以二元式（类别码，符号串）【常数】或三元式（类别码，符号串，地址）【关键字、标识符、运算符、常数、界符】的形式写入名为result.txt的文件中。

【说明】 我们假定类型 1-关键字；2-标识符；3-运算符；4-常数；5-界符

1. 对于识别出来的关键字，将其三元式（1，符号串，该关键字在关键字表

中的位置）写入result.txt中，例如如果识别出来的关键字是int，则其三元式为（1，int，17）。

1. 对于识别出来的标识符，将其写入标识符表中，例如如果识别出来的标

识符是a，则将其写入标识符表中；并将其三元式（2，符号串，该标识符在标识符表中的位置）写入result.txt中。

1. 对于识别出来的运算符，将其三元式（3，符号，该运算符在运算符表中

的位置）写入result.txt中，例如如果识别出来的运算符是\*，则其三元式为（3，\*，3）。

1. 对于识别出来的常数，将其二元式（4，符号串）写入result.txt中，例如

如果识别出来的常数是3.14，则其三元式为（4，3.14）。

1. 对于识别出来的界符，将其三元式（5，符号，该界符在界符表中的位置）

写入result.txt中，例如如果识别出来的运算符是{，则其三元式为（5，{，5）。

1. 将标识符表的内容写入id.txt中。

【说明】在result.txt中每写一个二元式或三元式就换行，在id.txt中每写一

个标识符就换行。

（6）对于不符合给定C语言文法规则的符号串给出错误提示。

【示例演示】

例如：若source.txt中的内容为

int a=0;

int i;

float b=1.55;

则id.txt的内容为：

a

i

b

result.txt的内容为：

(1 , int , 17)

(2 , a ， 1)

(3 , = , 7)

(4 , 0)

(5 , ; , 9)

(1 , int , 17)

(2 , i ，2)

(5 , ; , 9)

(1 , float , 13)

(2 , b ，3)

(3 , = , 7)

(4 , 1.55)

(5 , ; , 9)

**实验二 基于预测分析法的语法分析程序的构造**

**（一）实验目的**

1、掌握LL（1）分析法的基本原理。

2、掌握FIRST集、FOLLOW集、SELECT集的计算方法和过程。

3、掌握预测分析表的构造方法。

4、掌握使用预测分析表进行语法分析的过程。

**（二）实验内容**

对于某个给定的表达式文法（满足LL（1）文法的条件），使用高级语言（c、c++、java、python）编写程序，完成下列内容：

1、计算文法中每个文法符号以及每个产生式右部符号串的FIRST集。

2、计算文法中每个非终结符的FOLLOW集。

3、计算文法中每个产生式的SELECT集。

4、构造预测分析表设计算法编程实现判断输入的符号串是否满足文法定义。

（三）实验指导

1、已知某个消除左递归后的文法G[E]如下：

（文法可以保存在名为grammer.txt的文件中或使用适当的存储结构存储）

（1）E->TE’ （2）E’->+TE’

（3）E’->ε （4）T->FT’

（5）T’->\*FT’ （6）T’->ε

（7）F->i （8）F->(E)

2、编写函数：分别求出FIRST集、FOLLOW集和SELECT集，可参考按照图2-1的形式输出。（这里未列出每个产生式右部符号串的FIRST集，请同学们自行加入）。

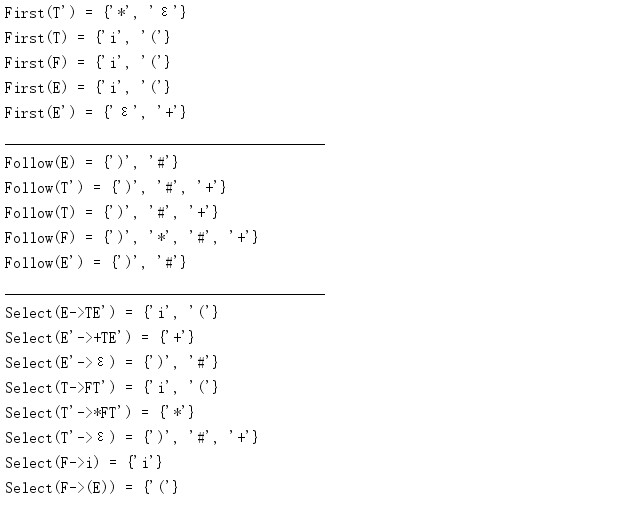


图2-1 FIRST集、FOLLOW集、SELECT集的结果示例

3、根据2中计算出来的SELECT集编写函数构造出预测分析表，存到名为

pat.txt的文件或适当的存储结构中。示例文法的预测分析表的内容应该如图2-2所示。

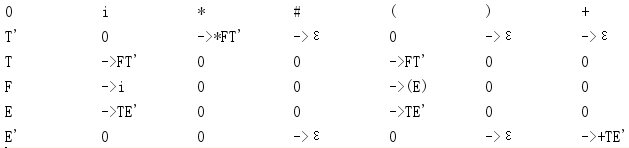
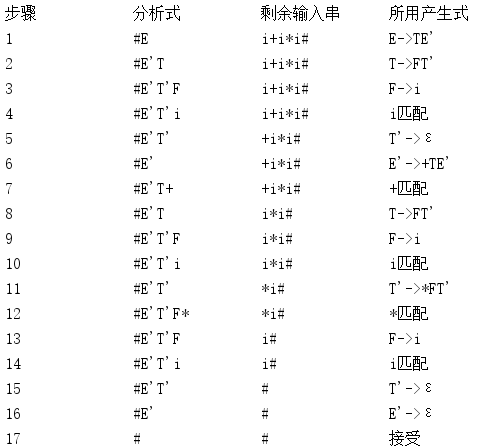


图2-2 示例文法的预测分析表

4、如果输入表达式i+i\*i，根据3构造出的预测分析表设计算法编程实现判别该符号串是否满足文法定义，如果满足，则输出“Yes”，否则输出“No”。可以将分析过程保存到名为result.txt的文件或适当的存储结构中。result.txt文件中的内容应该如图2-3所示。

图2-3 对表达式i+i\*i的分析过程

**六、参考教材**

1、王生原等. 编译原理（第三版）[M].北京：清华大学出版社，2015.

2、蒋宗礼、姜守旭. 编译原理（第2版）[M].北京：高等教育出版社，2017.