**深 圳 大 学 实 验 报 告**

**课程名称： 编译原理**

**实验项目名称： 词法分析程序设计**

**学院： 计算机与软件学院**

**专业： 软件工程**

**指导教师： 尹剑飞**

**报告人： 钟靖扬 学号：2020111051 班级： 01**

**实验时间： 2023年3月1号-2023年3月20号**

**实验报告提交时间： 2023.3.25**

**教务处制**

|  |
| --- |
| **实验目的与要求：**  **目的：**针对状态矩阵、NFA、DFA、正规式、确定化算法、最小化算法、Thompson算法等词法分析应用问题，设计并实现相应的解决方案，通过设计FA相关类族，以及实现前述几个重要的算法，既加深对抽象的词法、自动机、正规式等形式语言理论基础概念的理解与掌握，也加强对面向对象程序编写能力和计算思维的培养。  **要求:**  **第一部分： 无符号数的识别及DFA的应用**  这部分又分为2个小实验   1. 输入：一个文本文件（源代码文件）   输出：将源代码中的无符号数识别出来并输出到另一个文件中  示例：如果输入是“123\*abc+def/99.2+9.9E+c”，那么输出是：（数字, 123），（其它，\*abc+def/），（数字，99.2），（其它，+），（异常，9.9E+c）  说明：其它是非数字打头的字符串；异常是数字打头，但最后却是不符合定义的无符号数。   1. 假设：用字符“ABCDEFGHIJK”（大写）分别表示数字0..9和.、E、+、-，那么，字符串“BCLD”表示数字“12E3”=12000；   输入：一个文本文件  输出：将隐藏在文本文件中的有效无符号数识别出来。  示例：如果输入是“BCD\*abc+def/JJKC+JKJL+c”，那么输出是：（数字, 123）、（数字，99.2），无效（异常）的无符号数不输出  **第二部分：DFA/NFA的读写及确定化、最小化算法 （选做）**  这部分又分为3个小实验   1. 输入：一个文本文件（格式自定义）   输出：读入文件中的DFA/NFA，创建对应的DFA/NFA对象，再写回另一文件中。  示例：如果输入是“f(S,a)=A, f(A,b)=B, {B}” //假设默认开始状态是S，  那么输出是：K={S，A，B}；Σ={a,b}；f(S,a)=A, f(A,b)=B；S；Z={B}   1. 输入：上一实验输出的NFA文件   输出：将读入的NFA进行确定化后，输出结果，需要考虑有/无ε的情况。   1. 输入：上一实验输出的DFA文件   输出：将读入的DFA进行最小化后，输出结果。  **第三部分：正规式及其应用**  这部分又分为3个小实验   1. Thompson方法的实现   输入：一个正规式  处理：创建与正规式对应的NFA对象，再将NFA对象写回文件，格式与前述实验相同，使得能够读回；  示例：如果输入是“a\*b”，那么输出是：K={A , B , C , D , E , F}；Σ={a, b}；f(A,ε)=B, f(B,a)=C, f(C, ε)=B, f(C, ε)=D, f(A, ε)=D, f(D, ε)=E, f(E, b)=F；A；Z={F}  说明：优先级处理可用栈或递归，思路见算符优先分析法。概要来说，判断当前运算符与下一个运算符间的优先级，当前≥下一个则计算（产生NFA），否则入栈或递归。   1. 从正规式到DFA   将上述实验与前述实验连接起来，使得读入一个正规式文件，能输出一个DFA。正规式文件格式自定，通常每行一个正规式。  示例：如果正规式文件包括2行，每行分别是“a”和“b”，那么，输出的DFA实际上是“a | b”对应的DFA。   1. 字符串的识别   输入：一个正规式文件和一个字符串文件  输出：判断字符串文件中的每个字符串，能否被正规式对应的DFA所识别  其次，再给每个正规式增加一个类别，识别到给定字符串符合某个特定正规式时，输出该类别。  示例：如果输入  a\* 类A  b 类B  那么对字符串aaa输出：aaa，类A |
| **方法、步骤：**  要完成本实验，依据实验要求进行分解，需要从数据结构设计和面向对象设计角度，思考并回答下述问题：   1. 如何识别无符号数？   如何判断一个输入的字符串中，包括无符号数字？如何判断无符号数的开始？中间字符及结束？  根据状态矩阵画出相应的DFA和状态转换表来获取数据。    而我们需要的结果分别有三种，数字，其他和异常，因此根据我们的需要画出对应的状态表   1. 如何实现简单的字符映射加解密算法？   请在这里，补充完善你的设计思路，完成后，请删去这句话。  例如ABCD在ASCII表中是按顺序排列的，因此我们可以比较简单的直接用A～J来表示0～9用加法就可以完成一个简单的映射。   1. 如何定义DFA/NFA对象，特别的，对自动机里面的映射函数f如何定义、存储？（注意这里的“对象”，指用面向对象的程序设计思维方式（内部数据结构、消息）来定义一个DFA类、一个NFA类。）   请在这里，补充完善你的设计思路，完成后，请删去这句话。   1. 如何实现确定化算法？（选做）   确定化算法需要把一个旧的状态集合作为一个新的状态，这种情况如何处理？又如何求某个状态的ε闭包？  请在这里，补充完善你的设计思路，完成后，请删去这句话。   1. 如何实现最小化算法？（选做）   最小化算法需要对原状态集进行等价类划分，这个划分要如何处理？  请在这里，补充完善你的设计思路，完成后，请删去这句话。   1. 汤普森方法需要处理算符的优先级，该如何处理？如何进行递归调用？   请在这里，补充完善你的设计思路，完成后，请删去这句话。   1. 如何根据给定DFA，判断并输出给定输入串的类别？   请在这里，补充完善你的设计思路，完成后，请删去这句话。 |
| **实验过程及内容：**  实验过程及内容，处理代码设计说明、代码及其注释外，特别关注编程过程。 |
| **实验结论：**   1. 测试用例的设计与说明   对每部分的实验，你分别设计了什么测试数据（测试用例）进行测试，你设计测试数据的出发点是什么，有什么目的？   1. 测试结果的说明与分析   对测试（实验）结果进行分析说明 |
| **心得体会：**  除了实验的心得体会外，  还要额外增加说明你对文法、直接推导、推导、句型、句子、语言这些概念的理解。 |
| 指导教师批阅意见：  成绩评定：  指导教师签字：尹剑飞  年 月 日 |
| 备注： |