

2022-2023 年度《编译原理上机实习》指导书

一、上机实习目的

理解编译程序的构造原理，掌握编译程序的构造方法与技术。通过实习，使学生既加深对编译原理基础理论的理解，又提高动手能力，特别是提高软件设计能力。

二、上机实习要求

在理解编译原理基本思想的基础上，选择一个自己熟悉的程序设计语言，完成编译程序的设计和实现过程。

编译程序的设计可以采用自顶向下和自底向上两种不同的方法。由于许多高级语言（如 PASCAL，C）中的语法成分都是递归定义的，所以**本实习要求学生采用递归下降分析技术**，这是一种自顶向下的编译方法，其基本思想是对语言的每个（或若干个）语法成分编制一个处理子程序，从处理<程序>这个语法成分的子程序开始，在分析过程中调用一系列过程或函数，对源程序进行语法和语义分析，直到整个源程序处理完毕为止。

本上机实习是为一种架空语言设计一个基于**一次扫描**的编译程序，完成词法分析、语法分析、语义分析等功能，并生成中间代码（四元式）或某种机器上的目标代码（汇编语言）。

三、上机实习步骤

1. 阅读《上机实习指导书》，了解实习内容和考核要求（**考核要求见附录 2**）
2. 浏览并打印上机实习报告（**模板见群文件**）
2. 浏览并分析文法（**文法见附录 1**）
3. 根据文法，填写上机实习报告中过程部分
4. 根据文法，编写编译程序（**教学视频见学习通**）
5. 输入随机生成的源代码并上机测试结果，必要时对程序进行调试（**源代码生成程序和编译程序样例见群文件**）
6. 完成上机实习报告剩余内容
7. 预约并准备答辩（**答辩流程见附录 3**）
8. 答辩后提交相关材料

四、上机实习相关资源

- 课程群文件中【上机资料】文件夹：《上机实习指导书》、上机实习报告模板、源代码生成程序和编译程序样例（**样例程序仅作为对比答案的参照，根据考核要求编程**）
- 课程群文件中【群在线文件】文件夹：答辩预约登记表（包括拓广功能登记）
- 学习通：教学视频

五、学生应提交的上机实习材料

完成上机实习后，每个学生至少应提交如下两种材料：

5.1 上机实习报告

- 要求：正文部分必须用钢笔或签字笔等**手写**。**不可以用印刷体打印稿**；可以用ipad**手写后打印pdf**，但是**必须事先和任课教师说明**。若需附程序运行结果，可截图打印后夹订在报告书中。
- 提交方式：答辩时提交纸质报告。

5.2 程序数据

【重要！提交物是否齐全采用程序自动分析判断，因此文件名需严格按照下述格式命名·摆放，不符合要求一律判定为未提交】

- 要求：

1. 创建一个名为**【学号姓名】**（学号和姓名不要加空格或其他符号，确保学号正确）的文件夹，放入以下文件：

文件名	内容
source.txt (全小写字母，下同)	一个包括自己编写的所有源代码的文本文件。若源代码为单个文件，直接复制粘贴内容；若为多个文件，则合并在一个文件里（主程序的代码放在开头）。不要加入第三方函数库代码
源程序文件	自己编写的所有源代码文件，文件名无限制，文件个数不限，文件夹层数不限；直接拷贝到文件夹内。不需要配置类文件， 不要打包压缩
in.txt	一个输入用的文本文件
debug.txt	一个输出用的文本文件
out.txt	一个输出用的文本文件
compiler.编译后缀名	一个编译好的可以执行的文件，如 compiler.exe

其中，源程序语言和编译后文件的对应关系如下：

选择的编程语言	至少需要提交的源程序文件	对应的编译后文件
c语言	.c, .h	.exe
c++	.cpp	.exe
c#	.cs, .form	.exe
java	.java	.class
python	.py	无需重复提交
其他	事先和任课教师个人沟通	

2. 返回上一层文件夹，在**【学号姓名】**文件夹上右击鼠标压缩文件，**允许格式为zip或rar**

3. 自行检查，文件名为**【学号姓名.zip】**（或.rar），双击打开压缩包后，**目录里只有一个名为【学号姓名】的文件夹**，该文件夹里面包含了所有要求的文件。

- 提交方式：**根据通知提交至指定平台**。

六、截止日期

上机报告书的提交截止日期：最后一节上机课的下课时间。

程序数据的提交截止：最后一节上机课后24小时后。

注意：上机作业是课程考核的一部分（等同于考试），对作弊行为按照《南京工业大学学生纪律处分规定》，给予记过以上处分。

- 全员提交结束后，按年级为单位统一查重。抄袭和提供抄袭均给予处分。
- 未提交齐程序代码或上机实习报告的，上机部分最高成绩为及格分。
- 未完成答辩的，上机部分为零分。

附录1 架空文法（2020级版）

<程序> → <变量说明部分>;<语句部分> ←注意 “<程序>” 为开始符号
<变量说明部分> → <变量说明><标识符列表>
<变量说明> → string
<标识符列表> → <标识符列表>,<标识符>|<标识符>
<标识符> → <字母>|<标识符><字母>|<标识符><数字>
<语句部分> → <语句部分><语句>;|<语句>;
<语句> → <赋值语句>|<条件语句>|<循环语句>
<赋值语句> → <标识符>=<表达式>
<条件语句> → if (<条件>) <嵌套语句> else <嵌套语句>
<循环语句> → do <嵌套语句> while (<条件>)
<表达式> → <项>|<表达式><连接运算符><项>
<项> → <因子>|<项><重复运算符><数字>
<因子> → <标识符>|<字符串>|(<表达式>)
<字符串> → “<字母序列>” ←注意 “引号” 为英文半角
<字母序列> → <字母序列><字母>|ε
<连接运算符> → +
<重复运算符> → *
<关系运算符> → <|>|<> |>|=|<= |== ←注意 “<>” 为不等于
<条件> → <表达式><关系运算符><表达式>
<复合语句> → start <语句部分> end
<嵌套语句> → <语句>|<复合语句>
<字母> → a|b|c|d|e|f|g|h|i|j|k|l|m|n|o|p|q|r|s|t|u|v|w|x|y|z
<数字> → 0|1|2|3|4|5|6|7|8|9

约定

- 字符串的连接运算： “abc” + “def” = “abcdef”
- 字符串的重复运算： “abc” * 3 = “abcabcabc”
- 字符串的逻辑运算（<> | ==） 对字符串的符号进行对比计算：
“abc” == “abc” 为 true “abc” <> “def” 为 true
- 字符串的逻辑运算（< | > | >= | <=） 假设为对字符串的长度进行对比计算：
“ab” < “abc” 为 true “abc” <= “def” 为 true
- 标识符不能以关键字开头
- 标识符和关键字之间不一定由空格隔开，如 startabc=1 可以分解为 start,abc,=,1; 即，空格应该作为多余符号做忽略处理

附录 2 评价基准

上机作业占总成绩的 30%，由两部分组成：基本 90% + 拓广 10%。其中，基本部分包括词法分析、语法分析、语义分析、上机报告和答辩，总分值为 110%，超出 90% 部分舍去。

分值分配		要求
词法分析	基本（20%，每点 5%）	1. 从同目录下的名为 in.txt 的文件读入源程序， 2. 根据词法规则识别单词， 3. 程序内创建并填写相应的表（不要求输出）， 4. 往同目录下名为 debug.txt 的文件中写入每次识别出的二元式。
	附加（5%）	如果存在词法错误，则显示词法错误信息、出错位置。
语法分析	基本（20%，每点 10%）	1. 对源程序作语法分析，确定是否符合该文法， 2. 往同目录下名为 debug.txt 的文件中写入每次推导使用的一条产生式。
	附加（5%）	若存在语法错误，显示语法错误信息和位置。
语义分析与中间代码生成	基本（20%，每点 5%）	1. 在语法分析的基础上进行语义分析， 2. 往同目录下名为 debug.txt 的文件中写入每次产生的四元式， 3. 分析结束后往同目录下名为 debug.txt 的文件中写入标识符表、临时变量表和四元式表 4. 往同目录下名为 out.txt 的文件输出源程序的四元式表。
	附加（5%）	若存在语义错误，显示语义错误信息和位置。
上机报告	基本（20%）	内容完整，书写字迹工整。
	附加（5%）	第一节上机课完成提交 5%，第二节 4%，第三节 3%，第四节 2%，第五节 1%。
答辩	10%	回答正确。
拓广	10%	实现一种不与他人重复的拓广功能。

要求：**执行**程序后，程序**自动开始翻译**源代码，向对应文件写出数据后**自动关闭**。

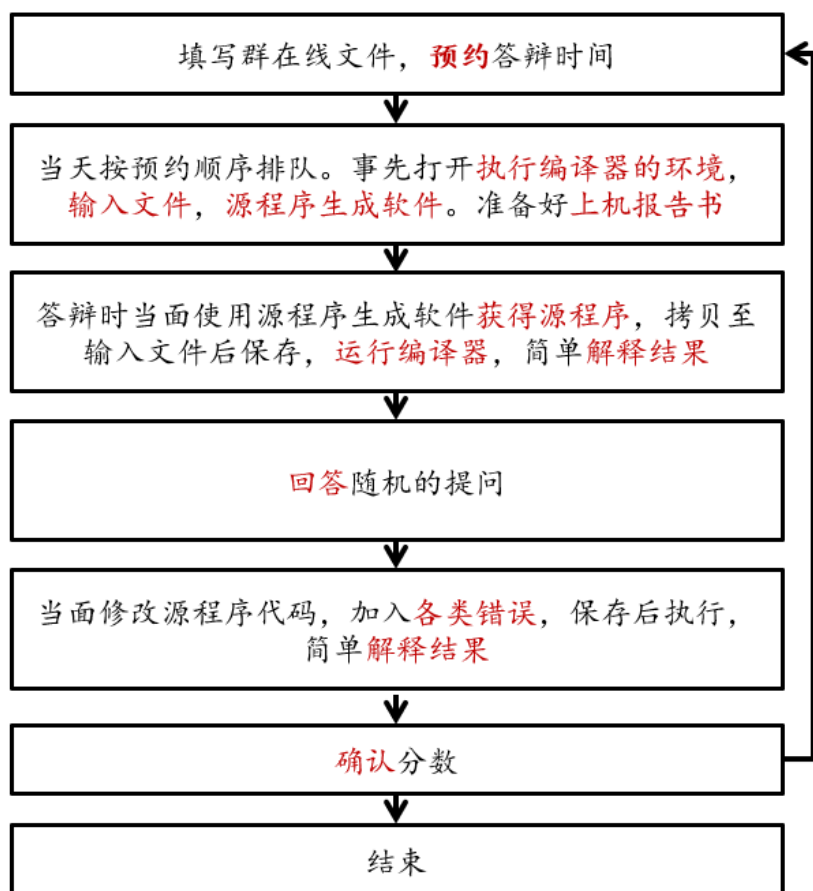
注意

- 基本部分的工作量对应上机 12 课时，预计最多 12 小时；拓广功能预计另外需要最多 12-24 小时。**若花费时间超过目测量**，及时调整学习方法或询问任课教师。
- 拓广功能采用群在线文件登记，动手实践前务必**先确认登记表中不与他人重复**，具体内容可以与任课教师沟通。

■ 拓广功能例（包括且不限于下述内容）：

- 针对某种特定的词法、语法或语义错误，出错时试图从错误中恢复（处理错误不同，算作不同拓广功能）
- 输出每个临时变量经过第一次计算后的值（限 1 人）；
- 获得正确计算结果的标识符表（限 1 人）；
- 判断死循环（限 1 人）；
- 对随机生成的源代码进行缩进调整等格式化（限 1 人）；
- 在翻译结果正确的前提下去除多余的临时变量（限 1 人）；
- 设计添加相关文法，实现对数组的定义、参照、赋值的翻译（若能正确实现，在拓广 10%的基础上，对最终成绩加 5 分，限 1 人）
 - 约定数组格式为 <标识符>[<数字>]，如 abc[9]
- 设计添加相关文法，实现对函数的定义、调用的翻译（若能正确实现，在拓广 10%的基础上，对最终成绩加 5 分，限 1 人）
 - 约定函数格式为 function <标识符>(string <标识符>) <复合语句>，其中传递一个字符串类型的参数，无返回值
- 生成所学的汇编语言生成目标代码（若能正确实现，在拓广 10%的基础上，对最终成绩加 5 分，限 1 人）
 - 约定可用寄存器为 R0, R1, R2, R3
- 其他对编译程序的拓展或优化

附录3 答辩流程



注意

- **上机课时用于检查结果，编程和撰写报告要利用课余自习时间。**预计每节课90分钟检查20人左右。
- 留学生无法到校、电脑不带windows系统等**特殊情况主动联系任课老师。**
- 提交的**报告书不退回。**