

编译原理与技术课程大作业

- 指导教师：陈雨亭老师，chenyt@cs.sjtu.edu.cn/34206975，电信群楼 3-511 办公室
- 编译课程助教（符易阳、胡宸章）
 - 符易阳，fuyiyang@sjtu.edu.cn，电信群楼东 311 实验室
 - 胡宸章，hcz1221@sjtu.edu.cn
- QQ 群：SJTU-SE-Compiler，群号：464964419
- 提交方式：
 - 每个时间节点前一天，请把文件提交到 elearning 上
 - 检查时间节点（具体时间安排将另行通知）
 - ◆ 第 2 周的周四前：提交组员信息（包括组员学号，姓名，联系 email，联系电话等）
 - ◆ 第 6 周的周一或者周二：检查任务 1
 - 提交内容：一份 word 文档，描述 MyLang 语言、词法、语法规则，及样例程序
 - ◆ 第 11 周的周一或者周二：检查任务 2 和 3 的部分
 - 提交内容：一份 word 文档，描述从 MyLang 程序转换为汇编代码的解决方案；QuickSort 汇编代码；MyLang 编译器（支持词法分析和语法分析功能）的可执行文件和源代码
 - 检查中请使用 PPT，描述编译器的特点、框架、关键算法等

◆ 第 16 周的周一或者周二：检查任务 3 和额外要求

- 提交内容：MyLang 编译器（集成词法分析、语法分析、语义分析、代码生成功能）的可执行文件和源代码；编译器说明、语法规则、测试用例（含 QuickSort 程序）等
- 检查中请使用 PPT，描述编译器的特点、框架、关键算法等

■ 检查地点：具体检查时间、地点由助教通知

■ 检查注意事项：

■ 注意点

- ◆ 每次检查结束后，我们将邀请部分较有特色的小组为大家介绍自己的工作。
- ◆ 允许组内合作，禁止组间合作，严禁任何抄袭、代码拷贝行为。
- ◆ 指导小组不协调分组、各开发小组内分工、工作量计算等问题，每小组各成员将获得一致的分数。

编译原理课程实验（2015 年）

在编译原理与技术课程的大作业中，每个小组将独立完成一个语言的设计及相关编译器的开发。你们将建立一个 1~3 人的项目小组，合作完成大作业。

主要任务 1（15%，截止日期：第 5 周）

1.1 设计一个语言（这是一个独一无二、较为轻便灵巧的语言，假设其名为 **MyLang**，你们可以借鉴其他语言将 **MyLang** 设计为过程式语言、面向对象语言、函数式语言、脚本语言），该语言能够描述快速排序算法 **QuickSort**。要求：

1.1.1 支持用户的输入输出、常见计算表达式（例如 +, -, *, /, (), 优先级处理等）；

1.1.2 你需要将 **MyLang** 定义为一种强类型语言。关于程序语言的类型，存在这样一些概念：

1.1.2.1 静态类型定义语言：一种在编译时，数据类型是固定的语言。

大多数静态类型定义语言强制这一点，它要求你在使用所有变量之前要声明它们的数据类型。**Java** 和 **C** 是静态类型定义语言。

1.1.2.2 动态类型定义语言：一种在执行期间才去发现数据类型的语言，与静态类型定义相反。**VBScript** 和 **Python** 是动态类型定义的，因为它们是在第一次给一个变量赋值的时候找出它的类型的。

1.1.2.3 强类型定义语言：一种总是强制类型定义的语言。**Java** 和

Python 是强制类型定义的。如果你有一个整数，如果不显式地进行转换，你不能将其视为一个字符串。

1.1.2.4 弱类型定义语言： 一种类型可以被忽略的语言，与强类型定义相反。VBScript 是弱类型定义的。在 VBScript 中，可以将字符串 '12' 和整数 3 进行连接得到字符串 '123'，然后可以把它看成整数 123，而不需要显示转换。

1.1.3 MyLang 需要支持数组；支持复合类型或者类（及继承），以保存复合型的数据结构；

1.1.4 支持至少复合语句、if-else、while、for 语句等常见语句或者结构。另外，MyLang 需要支持 foreach 语句，该语句为数组或集合中的每个元素重复一个嵌入语句组。典型的 foreach 语法为 foreach (type identifier in expression) statement，例如：

```
int[] nArray = new int[100];  
  
foreach( int i in nArray )  
  
    Debug.WriteLine( i.ToString() );
```

你可以参考 Java 或者 C#语言关于 foreach 语句的具体用法；

1.1.5 支持函数或者方法等相似的程序调用机制；

1.1.6 支持注释。

1.2 使用 MyLang，书写 QuickSort 程序；

1.3 采用 BNF 制定 MyLang 的词法、语法规则。

1.3.1 这是一个轻量级的语言，考虑到开发量的因素，语言规则不宜庞大。

1.3.2 例如，Java Syntax Specification（<http://www.daimi.au.dk/dRegAut/JavaBNF.html>）提供了 Java 语言的词法、语法规则。一些已有语言的词法、语法规则可以参考：
<http://java.net/projects/javacc/downloads/directory/contrib/grammars>。

主要任务 2（15%，截止日期：第 10 周）

2.1 写一段 C++ 程序 QuickSort，采用 Visual Studio 查看相应的汇编代码；

2.2 书写 QuickSort 的汇编代码，并采用 MASM、[FASM](#) 或者 TASM、AT&T 等汇编器执行。请尽早确定小组所选用的汇编器。考虑到网络资源的因素，推荐使用 MASM 或者 FASM。注意：

2.2.1 如何书写汇编代码，请参照《Intel 汇编语言程序设计》；

2.2.2 MASM 的使用，请参照
<http://kipirvine.com/asm/gettingStartedVS2010/index.htm> ([MASM withVisualStudio.pdf](#)) 和《[Microsoft MASM 6.1 Programmer's Guide](#)》；

2.2.3 FASM 的使用，请参照 <http://flatassembler.net> 和《Flat assembler 1.70 Programmer's Manual》（FASM 帮助文档）；

2.2.4 AT&T 是 GCC 生成的目标汇编语言。

2.3 针对 MyLang，提供一个程序转换为汇编代码的初步方案。

2.4 注意：主要任务 3 中，第 10 周需要检查词法分析和语法分析功能。

主要任务 3（60%,截止日期：第 10 周(3.1-3.3)、第 15 周(3.4-3.5)）

3.1 为你的编译器提供一个图形界面，以很清楚地显示四个主要阶段的输入和输出（例如，提供四个菜单按钮，分别对应词法分析、语法分析、语义检查、代码生成）。注意：

3.1.1 四个阶段的输入输出均需要能够独立显示。如果不能独立显示，则在相应的阶段分数基础上扣除 20%。

3.1.2 缺少图形界面的扣除 10%（我们既希望程序员方便使用编译器，又不想让程序员手工输入太多的命令）。

3.1.3 你可以基于 JavaCC 或者其他类似的编译器生成器开发，也可以自己手工开发。

3.2（10%）为 MyLang 所书写的 QuickSort 程序进行词法分析，生成 Token Stream。如果程序违反词法规则，则报出合适的错误信息；

3.3（20%）为 MyLang 所书写的 QuickSort 程序进行语法分析，生成一个可展示的抽象语法树结构。如果程序违反语法规则，则报出合适的错误信息；

3.4（15%）为 MyLang 所书写的 QuickSort 程序进行语义检查，显

示所有变量的类型，及检查其正确性。如果类型不正确，则予以报错，注意类型的兼容性，及函数的形参与实参兼容；

3.5（15%）为 MyLang 所书写的 QuickSort 程序进行代码生成，**生成汇编代码**。

3.5.1 试运行汇编代码，如果自动生成后需要手动修改汇编代码才能使其运行的话，修改量不能超过生成代码量的10%。

额外要求（10%，第 15 周(3.4-3.5)）

5.1为 MyLang 提供更强大或者更有特色、更精巧的语言机制及编译支持，在每次检查中告诉我们这些特点和如何完成相应的编译任务。

5.2用 MyLang 语言书写**另一个**稍复杂的程序（如八皇后问题，<http://baike.baidu.com/view/698719.htm>），并生成相应的汇编代码予以执行。

5.3项目最终提交的时候需要提供附录文件，包含相应编译器说明、语法规则、测试用例等。