

## 四、《编译原理》实验

### Oberon-0 逆向工程工具 ROSE<sup>1</sup>

#### 目 录

1 简介 .....	1
1.1 实验目的 .....	1
1.2 实验环境 .....	2
1.3 实验计划 .....	2
1.4 诚信声明 .....	2
2 实验内容 .....	3
2.1 实验一、熟悉 OBERON-0 语言定义 .....	3
2.1.1 实验步骤 1.1 编写一个正确的 OBERON-0 源程序 .....	3
2.1.2 实验步骤 1.2、编写上述 OBERON-0 源程序的变异程序 .....	3
2.1.3 实验步骤 1.3、讨论 OBERON-0 语言的特点 .....	4
2.1.4 实验步骤 1.4、讨论 OBERON-0 文法定义的二义性 .....	4
2.1.5 提交结果 .....	4
2.1.6 评价标准 .....	5
2.2 实验二、自动生成词法分析程序 (JFLEX) .....	5
2.2.1 实验步骤 2.1、总结 OBERON-0 语言的词汇表 .....	5
2.2.2 实验步骤 2.2、抽取 OBERON-0 语言的词法规则 .....	5
2.2.3 实验步骤 2.3、下载词法分析程序自动生成工具 JFLEX .....	5
2.2.4 实验步骤 2.4、生成 OBERON-0 语言的词法分析程序 .....	6
2.2.5 实验步骤 2.5、讨论不同词法分析程序生成工具的差异 .....	6
2.2.6 提交结果 .....	6
2.2.7 评价标准 .....	8
2.3 实验三、自动生成语法分析程序 (JAVACUP) .....	8
2.3.1 实验步骤 3.1、下载自动生成工具 JAVACUP .....	8
2.3.2 实验步骤 3.2、配置和试用 JAVACUP .....	8
2.3.3 实验步骤 3.3、生成 OBERON-0 语法分析和语法制导翻译程序 .....	8
2.3.4 实验步骤 3.4、讨论不同生成工具的差异 .....	9
2.3.5 提交结果 .....	9
2.3.6 评价标准 .....	10
2.4 实验四、手工编写递归下降预测分析程序 .....	10
2.4.1 实验步骤 4.1、设计 OBERON-0 语言的翻译模式 .....	10

2.4.2 实验步骤 4.2、编写递归下降预测分析程序 .....	11
2.4.3 实验步骤 4.3、语法分析讨论：自顶向下 vs. 自底向上 .....	11
2.4.4 提交结果 .....	11
2.4.5 评价标准 .....	12
 <b>3 OBERON-0 语言 .....</b>	<b>13</b>
 <b>3.1 简介 .....</b>	<b>13</b>
<b>3.2 词法定义 .....</b>	<b>14</b>
3.2.1 单词 .....	14
3.2.2 基本数据类型 .....	14
<b>3.3 语法定义 .....</b>	<b>14</b>
3.3.1 扩展 BNF (EBNF) .....	14
3.3.2 OBERON-0 语言的 EBNF 定义 .....	14
<b>3.4 语言描述 .....</b>	<b>16</b>
3.4.1 模块声明 .....	16
3.4.2 运算符与表达式 .....	16
3.4.3 类型声明 .....	17
3.4.4 选择符 .....	17
3.4.5 作用域规则 .....	17
3.4.6 过程声明与参数传递 .....	18
3.4.7 预定义函数 .....	18
<b>3.5 程序示例 .....</b>	<b>18</b>
 <b>4 逆向工程工具 .....</b>	<b>20</b>
 <b>4.1 调用图 .....</b>	<b>20</b>
4.1.1 调用图定义 .....	20
4.1.2 调用图例子 .....	20
4.1.3 参考文献 .....	21
<b>4.2 流程图 .....</b>	<b>22</b>
4.2.1 流程图定义 .....	22
4.2.2 流程图例子 .....	22
对于一个 ACCOUNT 模块中的存款和取款过程，由 ROSE 生成的两个流程图例子如图 4.2 所示。 .....	22
 <b>5 实验软装置 .....</b>	<b>24</b>
 <b>5.1 目录和文件组织 .....</b>	<b>24</b>
<b>5.2 异常类设计 .....</b>	<b>24</b>
5.2.1 异常类程序包 .....	24
5.2.2 通用异常类 .....	24
<b>5.3 词法分析实验软装置 .....</b>	<b>25</b>

---

5.3.1 词法错误的异常类 .....	25
5.3.2 词法分析主程序 .....	25
5.3.3 词法分析程序测试用例 .....	26
<b>5.4 语法分析和语法制导翻译实验软装置 .....</b>	<b>26</b>
5.4.1 语法错误的异常类 .....	26
5.4.2 语义错误的异常类 .....	26
<b>5.5 预定义设计图 API .....</b>	<b>27</b>
5.5.1 调用图绘制 API .....	27
5.5.2 流程图绘制 API .....	28

# 1 简介

---

在开始实验之前, 请务必花足够时间阅读完本文档关于实验的描述与约定!

## 1.1 实验目的

本实验是编译原理实验环节的一个综合型、应用型实验, 其中还包含部分研究与探索型实验内容。

本实验处理的程序设计语言是 Oberon-0, 它是著名的 Pascal 和 Modula-2 语言的后继者 Oberon 语言的一个精简子集。学生在本实验中需开发一个面向 Oberon-0 的逆向工程工具, 根据一个输入的 Oberon-0 源程序自动绘制对应的设计图, 这些设计图包括子程序调用关系图和程序控制流设计图 (例如程序流程图、PAD 图、N\_S 图等), 本实验项目将该工具命名为 **ROSE**, 意为 **Reverse Oberon Software Engineering**。

本实验借助于 **ROSE** 工具的设计与实现, 帮助学生通过实践深入理解和牢固掌握编译技术中的词法分析、语法分析、语法制导翻译、自动生成工具等重要环节。

本实验的主要目标包括:

- 掌握词法分析程序的工作原理与构造方法, 并能够推广到对文本的串匹配搜索等其他同类型应用。
- 掌握词法分析程序自动生成工具 **lex** 或类似工具的工作原理与使用方法, 学习如何编写一个 **lex** 源文件以解决词法分析或模式匹配问题, 在实践中进一步体会软件自动化的基本思路。
- 掌握递归下降的预测分析方法以及语法制导的翻译技术, 学习如何根据 BNF 语法定义和应用需求设计一个翻译模式 (Translation Scheme), 并利用高级程序设计语言的递归机制实现一个翻译模式。
- 掌握语法分析程序自动生成工具 **yacc** 或类似工具的工作原理与使用方法, 学习如何编写一个 **yacc** 源文件以解决语法分析及语法制导翻译问题, 进一步加深体会软件自动化的基本思路。
- 通过加强设计空间 (Design Sapce) 的深入探讨以及 Java 编程风格的实践, 提高对面向对象设计的认识, 养成良好的编程习惯与规范, 并学会多个工程文档的组织与交付。

## 1.2 实验环境

**编程语言:** Java JDK 1.5 或以上版本

**开发工具:** 可自由选择 Eclipse、JBuilder 等 IDE 环境, 也可直接采用 UltraEdit、EditPlus 等编辑器在命令行工作。但提交的实验结果必须独立于特定的 IDE, 可直接运行在 JDK 上。

**编码规范:** 要采用面向对象风格来设计实验中的所有类, 并遵循一种良好的程序设计习惯。例如, 如果你将一个程序的很多功能全部放在一个长长的 `main()` 方法中实现, 这样的设计与编码风格会被扣分。

在实验过程中应注意培养规范的编码风格。本实验要求所有源代码严格遵循 Sun 公司 (现为 Oracle 公司) 关于 Java 程序设计语言的编码规范 (Code Conventions for the Java Programming Language, Revised April 1999), 详细内容请参见: [Code Conventions for the Java Programming Language: Contents \(oracle.com\)](http://codeconventions.oracle.com)。

完成项目代码后, 应使用 JDK 附带的文档工具 javadoc, 根据源程序中的文档化注释, 自动生成相应的说明性文档。

## 1.3 实验计划

**ROSE** 实验项目是编译原理课程的一个综合型、应用型实验, 为高质量地完成这一课程实践环节, 你需要在开展实验之前精心规划自己的时间与精力投入, 因而推荐你参考以下建议及时开展实验活动。

根据任课教师对往届学生完成编译原理类似实验项目的经验估计, 你在语言预备实验 (实验一) 的投入时间不足 1 天; 在词法分析阶段的实验 (实验二) 约为 4 天; 在语法分析阶段及语法制导翻译阶段的实验 (实验三和实验四) 约为 10 天。另, 不要忘记撰写、整理和提交实验文档还需要一定的时间!

## 1.4 诚信声明

你必须独立完成编译原理课程实践环节的所有实验步骤, 并独立撰写实验报告和编写、调试程序。在实验过程中, 你可以参考各种各样的资源, 并与其他同学讨论问题 (并且鼓励在实验过程中与其他同学讨论问题), 但复制其他同学实验结果中的任何代码片段或任何文档片段都会被视作抄袭行为。一旦发现有任何抄袭行为, 你在本课程实践环节的所有成绩将被取消, 同时也意味着你本门课程的不及格。情节严重者, 整门课程的总评成绩将作为“作弊 0 分”呈报, 这也将导致你丧失申请中山大学学士学位的机会。望诸位同学慎之、戒之。切勿心存侥幸, 切勿挑战老师的智慧!

## 2 实验内容

---

对于一个像 **ROSE** 这样大规模、复杂实验项目而言, 精心分解实验的过程对实验的执行效果会产生重要影响。本实验分解为 4 个子实验项目, 每一子实验项目都规定了明确的实验步骤、实验要求以及需要提交的实验结果。

### 2.1 实验一、熟悉 Oberon-0 语言定义

本文档第 3 部分给出了 Oberon-0 语言的完整 BNF 定义, 即 Oberon-0 语言在语法方面的形式化规格说明 (Specification); 此外, 第 3 部分还以自然语言的方式提供了 Oberon-0 语言的语义描述。

请仔细阅读 Oberon-0 语言的定义, 并完成以下实验内容。

#### 2.1.1 实验步骤 1.1 编写一个正确的 Oberon-0 源程序

遵循 Oberon-0 语言的 BNF 定义, 编写一个正确的 Oberon-0 源程序。要求在这个源程序中, 用到 Oberon-0 语言的所有语法构造, 即你编写的源程序覆盖了 Oberon-0 语言提供的模块、声明 (类型、常量、变量等)、过程声明与调用、语句、表达式等各种构造。

如果有可能, 你编写的 Oberon-0 源程序最好是有其实际意义的, 譬如一个求阶乘的程序、一个求最大公因子的程序、一个用加法实现乘法的程序等等, 但这只是一个任选的要求。

注意, 这里仅要求你编写一个词法、语法和语义符合 Oberon-0 语言定义的源程序, 并未强制要求该源程序在逻辑上是正确的。

#### 2.1.2 实验步骤 1.2、编写上述 Oberon-0 源程序的变异程序

根据上述正确的源程序, 写出若干含有词法、语法或语义错误的 Oberon-0 源程序。其做法是对原有的正确源程序做出最小的改动, 使之成为一个包含至少一个错误的源程序; 为每一种可能存在的词法、语法和语义错误构造出一个变异程序。例如, 有一个变异程序中含有不合法的标识符, 有另一个变异程序中含有不合法的常量, 还有一个变异程序中丢失了左括号等。在每一变异程序的第一行用注释说明该变异产生的错误类型。

基于极限编程 XP (eXtreme Programming) 中的测试驱动编程 (Test-Driven Programming) 思想, 在进入正式编码前先设计好测试用例将有助于把握正确的设计。而上述正确的 Oberon-0 源程序及其一批含有错误的变异程序将在后述实验步骤中作为测试用例。



上述“程序变异”的思路源于软件测试技术中的变异测试 (Mutation Testing)。变异测试的目标是筛选更有效的测试数据，其思路是通过变异 (Mutation) 操作在程序中植入一个错误，变异后的程序称为变异程序 (Mutant)；通过检查测试数据能否发现变异程序中的这些错误，可判断测试数据的有效性。有兴趣的同学不妨查阅相关资料（例如：[http://en.wikipedia.org/wiki/Mutation\\_testing](http://en.wikipedia.org/wiki/Mutation_testing)）进一步学习变异测试技术。

### 2.1.3 实验步骤 1.3、讨论 Oberon-0 语言的特点

在 Oberon-0 语言中，与其前辈 Pascal 语言一样区别了保留字 (Reserved Word) 与关键字 (Keyword) 两个概念。例如，**IF**、**THEN**、**ELSIF** 等是保留字，而 **INTEGER**、**WRITE**、**WRITELN** 则是关键字。试解释这两个概念的区别。

根据 Oberon-0 语言的 BNF 定义，Oberon-0 程序中的表达式语法规则与 Java、C/C++ 等常见语言的表达式有何不同之处？试简要写出它们的差别。

### 2.1.4 实验步骤 1.4、讨论 Oberon-0 文法定义的二义性

根据 Oberon-0 语言的 BNF 定义，讨论 Oberon-0 程序的二义性问题，即讨论根据上述 BNF 定义的上下文无关文法是否存在二义性。

如果你认为该文法存在二义性，则举例说明在什么地方会出现二义性，并探讨如何改造文法以消除二义性。

如果你认为该文法没有二义性，则请解释：为何在其他高级程序设计语言中常见的那些二义性问题在 Oberon-0 语言中并未出现？

### 2.1.5 提交结果

将实验一的所有结果存放在子目录 **xxxxxxxxNNN\ex1** 中，其中 **xxxxxxxx** 是你的学号，**NNN** 是你的中文姓名。例如，学号为“05372001”、姓名为“马婷”的同学，应将其完成的实验一全部结果存放在“**05372001 马婷\ex1**”子目录中；注意学号与姓名之间不要加空格、姓与名之间亦不要加空格。

实验一最终提交的实验结果应包括：

- 在 **ex1** 子目录中存放自述文件 **readme.txt**，其中给出你的姓名、学号、电子邮件、联系电话、完成日期、以及其他补充说明。
- 在 **ex1** 子目录中存放你的实验报告 **Oberon-0.pdf**（允许、但不建议采用 **.doc**、**.docx**、**.txt**、**.wps** 等格式；下同），其中包括 Oberon-0 语言特点介绍、文法二义性讨论、实验心得体会等内容。
- 在 **ex1\testcases** 子目录中存放一个词法、语法、语义均正确的 Oberon-0 程序源代码，该文本文件的文件名由你自己根据程序的功能自由命名，但必须以 **.obr** 为文件扩展名。

- 在 `ex1\testcases` 子目录中存放一批含有词法、语法或语义错误的变异程序，文件名均同语法正确的 Oberon-0 源程序，但分别以 `.xxx` 为文件扩展名（`xxx` 表示序号，从 001 开始）。

### 2.1.6 评价标准

为了更好地分配时间与精力完成实验一，你可参考以下评分标准：

项 目	权重 (%)
1、一个正确的 Oberon-0 源程序	20
2、一批含有词法、语法或语义错误的变异程序	40
3、Oberon-0 语言特点的讨论	10
4、Oberon-0 文法二义性的讨论	20
5、文档组织及 readme 文件等	10

## 2.2 实验二、自动生成词法分析程序（JFlex）

实验二要求你下载一个词法分析程序自动生成工具 JFlex，并利用该工具自动产生 Oberon-0 语言的词法分析程序，该词法分析程序的源代码是用 Java 语言编写的。

### 2.2.1 实验步骤 2.1、总结 Oberon-0 语言的词汇表

根据 Oberon-0 语言的 BNF 定义和语言描述，抽取 Oberon-0 语言的词汇表以供词法分析程序设计与实现之用。你需要将 Oberon-0 的所有单词分类，并以表格形式列出各类词汇的预定义单词；譬如，在保留字表中列出所有的保留字，在运算符表中列出所有的运算符，等等。

请在实验报告中说明你的单词分类的理由，并解释如何处理 Oberon-0 语言中保留字和关键字两类略有不同的单词。

### 2.2.2 实验步骤 2.2、抽取 Oberon-0 语言的词法规则

在 Oberon-0 语言的 BNF 定义中，既包括 Oberon-0 语言的语法定义部分，也包括 Oberon-0 语言的词法定义部分。请将词法定义从 Oberon-0 语言的 BNF 中分离出来，并写成正则定义式（Regular Expression）的形式。

然后，在实验报告中讨论与 Pascal、C/C++、Java 等常见高级程序设计语言的词法规则相比，Oberon-0 语言的词法规则有何异同。

### 2.2.3 实验步骤 2.3、下载词法分析程序自动生成工具 JFlex

实验二指定的词法分析程序自动生成工具选用由 Gerwin Klein 开发的 JFlex。这是一个类似 Unix 平台上 lex 程序的开源（Open Source）软件工具，遵循 GNU



General Public License (GPL)。JFlex 本身采用 Java 语言编写, 并且生成 Java 语言的词法分析程序源代码。该软件工具的前身是由美国普林斯顿大学计算机科学系 Elliot Berk 开发、C. Scott Ananian 负责维护的 JLex。

从 <http://www.jflex.de/> 可下载该工具, 当前的最新稳定版本是 2020 年 5 月发布的 JFlex 1.8.2。该网站提供的压缩文件中已包含了你在本实验中所需的各类资源, 包括该工具的 Java 源代码、支持运行的库文件与脚本文件、用户文档、输入源文件例子等。

根据你自己的安装配置, 修改 JFlex 安装目录下脚本文件 `bin\jflex.bat` 中的两个环境变量 `JFLEX_HOME` 和 `JAVA_HOME` 的设置。然后运行 JFlex 附带的输入源文件例子, 以验证你是否正确安装并配置了 JFlex。

如果你觉得 JFlex 附带的用户手册仍不足以帮助你掌握 JFlex 的原理或用法, 自己动手在网上查找其他关于 JLex、GNU Flex、lex 等类似工具的大量电子资源。

## 2.2.4 实验步骤 2.4、生成 Oberon-0 语言的词法分析程序

仔细阅读 JFlex 的使用手册, 根据你在实验步骤 2.1 给出的 Oberon-0 语言词法规则的正则定义式, 编写一个 JFlex 输入源文件。

本实验提供的实验软装置为实验二预定义了代表 Oberon-0 语言各种词法错误的异常类, 并为实验二设计了主程序 `main()`。你在实验二中不要改变 `main()` 的内容, 并且在你的词法分析程序中应根据词法错误的类别抛出相应的异常类。实验软装置还提供了一批测试用例, 用于检测你的程序是否能够识别出词法错误并正确地对词法错误进行分类 (通过抛出的异常类标识错误类别)。关于 **ROSE** 实验软装置的详细描述请参阅本文档第 5 部分。

以你编写的源文件作为输入运行 JFlex, 得到一个 Oberon-0 语言词法分析程序的 Java 源代码; 编译该源程序生成 Java 字节码。通过实验软装置提供的测试用例检查后, 利用你在实验一编写的所有 Oberon-0 源程序测试你的词法分析程序。

## 2.2.5 实验步骤 2.5、讨论不同词法分析程序生成工具的差异

比较以下 3 种流行的词法分析程序自动生成工具之间的差异: JFlex、JLex 和 GNU Flex。主要讨论这些软件工具接收输入源文件时, 在词法规则定义方面存在的差异。

在网站 <http://www.cs.princeton.edu/~appel/modern/java/JLex/> 可找到关于 JLex 工具的权威资料; 关于 GNU Flex 的官方资料则位于 <https://github.com/westes/flex>。

## 2.2.6 提交结果

将实验二的所有结果存放在子目录 `xxxxxxxxNNN\ex2` 中, 其中 `xxxxxxxx` 是你的学号, `NNN` 是你的中文姓名, 命名规则同实验一。

实验二最终提交的实验结果应包括：

- 在 **ex2** 子目录中存放自述文件 **readme.txt**，其中给出你的姓名、学号、电子邮件、联系电话、完成日期、以及其他补充说明。
- 在 **ex2** 子目录中存放你在实验二撰写的实验报告 **lexgen.pdf**，其中的内容包括：
  - 以表格形式列出 Oberon-0 语言的词汇表。
  - 用正则定义式描述 Oberon-0 语言词法规则；若你使用纯文本书写正则定义式，其中的元符号“定义为”使用 “->” 表示，空串使用 “epsilon” 表示。
  - 紧随 Oberon-0 语言词法规则之后，给出一段关于 Oberon-0 语言与其他高级语言的词法规则之异同比较。
  - 讨论 3 种不同 lex 族软件工具的输入文件中，词法规则定义的差异或特点。
- 自动生成的 Oberon-0 语言词法分析程序及其测试用例，包括：
  - 在 **ex2** 子目录中存放各种脚本文件，包括：运行 JFlex 根据输入文件生成词法分析程序的脚本 **gen.bat**；编译词法分析程序的脚本 **build.bat**；运行词法分析程序扫描你编写的正确 Oberon-0 例子程序的脚本 **run.bat**；运行所有由实验软装置提供的测试用例的脚本 **test.bat**。
  - 在 **ex2\src** 子目录中存放面向 Oberon-0 语言的 JFlex 输入源文件 **oberon.flex** 以及你所需的其他相应文件。
  - 在 **ex2\src** 子目录中存放由 JFlex 生成的 Oberon-0 语言词法分析程序的 Java 源程序 **OberonScanner.java**（注意我们不使用默认的生成名字 **Yylex**）。
  - 在 **ex2\bin** 子目录中存放根据 Oberon-0 词法分析程序源代码编译得到的字节码文件 **OberonScanner.class** 以及其他相关的字节码文件。
  - 在 **ex2\doc** 子目录中存放根据你的源程序中文档化注释自动生成的 javadoc 文档。
  - 在 **ex2\jflex** 子目录中存放你所使用的 JFlex 工具。

注意，请不要提交那些由 IDE 自动生成的配置文件或备份文件（下同）。

### 2.2.7 评价标准

为了更好地分配时间与精力完成本次实验, 你可参考以下评分标准:

项 目	权重 (%)
1、Oberon-0 语言词法规则的正则定义式	20
2、JFlex 输入源文件的正确性	40
3、生成的词法分析程序的运行与测试	20
4、不同 lex 族工具差异比较	10
5、文档组织及 readme 文件等	10

## 2.3 实验三、自动生成语法分析程序 (JavaCUP)

实验三要求你下载一个语法分析程序自动生成工具 JavaCUP, 利用该工具自动产生一个 Oberon-0 语言的语法分析和语法制导翻译程序; 生成的程序源代码是以 Java 语言编写的。

### 2.3.1 实验步骤 3.1、下载自动生成工具 JavaCUP

实验三选用最早由美国卡内基·梅隆大学的 Scott E. Hudson 开发的一个语法分析程序自动生成工具 JavaCUP, 它是一个 LALR Parser Generator。JavaCUP 是一个类似 Unix 平台上 yacc 程序的开源 (Open Source) 软件工具, 遵循 GNU General Public License (GPL)。JavaCUP 本身采用 Java 编写, 并且生成 Java 语言的分析程序源代码。该软件工具经由美国普林斯顿大学计算机科学系 Andrew W. Appel 教授指导 Frank Flannery 等人改进, 目前由 C. Scott Ananian 负责维护。

从 <http://www2.cs.tum.edu/projects/cup/> 可下载该软件工具的最新版本 CUP 0.11b。该网站已包含了你在实验中所需的各类资源, 包括该工具的 Java 源代码、已编译生成的字节码、简明的用户手册、以及一个简单的命令行计算器例子等。

### 2.3.2 实验步骤 3.2、配置和试用 JavaCUP

成功下载并配置后, 试运行 JavaCUP 附带的输入源文件例子 (一个基于命令行的简单计算器应用), 以保证你正确安装并配置了 JavaCUP。

如果你觉得 JavaCUP 附带的用户手册仍不足以帮助你掌握 JavaCUP 的原理与用法, 自己动手在网上查找其他关于 GNU Bison、yacc 等类似工具的大量电子资源。

### 2.3.3 实验步骤 3.3、生成 Oberon-0 语法分析和语法制导翻译程序

仔细阅读 JavaCUP 使用手册, 根据 Oberon-0 语言的 BNF 定义编写一个 JavaCUP 输入源文件。

根据你的 JavaCUP 输入源文件生成的语法分析程序须完成以下功能:

1、对于一个存在词法、语法或语义错误的 Oberon-0 源程序，必须至少指出一处错误，并判断错误的类别及产生错误的位置（错误产生的位置定位允许有偏差），并以相应的异常对象向客户程序报告找出的错误的类别。是否支持其他功能取决于你的时间、精力与能力，譬如你可尝试从错误中恢复并继续执行语法分析，也可生成设计图时立即中止程序的执行。在错误检查与错误恢复（指找到错误后继续执行分析过程的能力）方面做得优秀的实验可获得更高的评分。

2、对于一个词法、语法和语义完全正确的 Oberon-0 源程序，自动绘制出不同模块中多个过程之间的调用图（Call Graph）。本文档第 4 部分详细介绍了 ROSE 实现的调用图的定义。

你在生成 Oberon-0 语言的语法分析程序时，可以直接使用实验二由 JFlex 生成的 Oberon-0 语言的词法分析程序。

以你编写的源文件为输入运行 JavaCUP，得到 Oberon-0 语言语法分析程序的若干 Java 源代码；编译这些 Java 源程序生成相应的字节码，再分别利用自己在实验一编写的 Oberon-0 源程序及其变异程序测试生成的结果，看看你生成的语法分析和语法制导翻译程序能否正确地识别出 Oberon-0 源程序中的各类错误，并且能否针对正确的 Oberon-0 源程序输出期望的调用图。

### 2.3.4 实验步骤 3.4、讨论不同生成工具的差异

比较两种流行的语法分析程序自动生成工具之间的差异：JavaCUP 和 GNU Bison，主要讨论这两种软件工具接收输入源文件时，在语法规则定义方面存在的差异。关于 GNU Bison 工具的官方资料可在网站 <http://www.gnu.org/software/bison/bison.html> 找到。

同样基于 Java 语言的分析器生成工具 (Parser Generator, 即 Compiler Compiler)，还有一个名为 JavaCC 的工具。在网上搜索并浏览关于 JavaCC 的相关信息，用最扼要的一两句话指出 JavaCC 与 JavaCUP 的最核心区别。

### 2.3.5 提交结果

将实验三的所有结果存放在子目录 `xxxxxxxxNNN\ex3` 中，其中 `xxxxxxxx` 是你的学号，`NNN` 是你的中文姓名，命名规则同实验一。

实验三最终提交的实验结果应包括：

- 在 `ex3` 子目录中存放自述文件 `readme.txt`，其中给出你的姓名、学号、电子邮件、联系电话、完成日期、以及其他补充说明。
- 在 `ex3` 子目录中存放讨论两个不同 yacc 族工具语法规则定义差异的文档 `yaccgen.pdf`。
- 自动生成的 Oberon-0 语言语法分析程序及其测试用例，包括：

- 在 **ex3** 子目录中存放各种脚本文件，包括：运行 JavaCUP 根据输入文件生成语法分析程序的脚本 **gen.bat**；编译语法分析程序的脚本 **build.bat**；运行语法分析程序处理你编写的正确 Oberon-0 例子程序的脚本 **run.bat**；运行所有由实验软装置提供的测试用例的脚本 **test.bat**。
- 在 **ex3\src** 子目录中存放面向 Oberon-0 语言的 JavaCUP 输入文件 **oberon.cup** 以及你所需的其他相应文件。
- 在 **ex3\src** 子目录中存放由 JavaCUP 生成的 Oberon-0 语法分析程序源代码 **Parser.java**、**Symbol.java** 和其他相应的 Java 源代码（注意我们不使用默认的生成名字 **parser** 和 **sym** 等）。
- 在 **ex3\bin** 子目录中存放根据 Oberon-0 语法分析程序源代码编译得到的字节码文件 **Parser.class** 以及其他相关的字节码文件。
- 在 **ex3\doc** 子目录中存放根据你的源程序中文档化注释自动生成的 javadoc 文档。
- 在 **ex3\javacup** 子目录中存放你所使用的 JavaCUP 工具。

### 2.3.6 评价标准

为了更好地分配时间与精力完成本次实验，你可参考以下评分标准：

项 目	权重 (%)
1、JavaCUP 输入源文件的正确性	50
2、生成的语法分析程序的运行与测试	30
3、两种 yacc 族工具差异比较	10
4、文档组织及 readme 文件等	10

## 2.4 实验四、手工编写递归下降预测分析程序

实验四要求你利用 Java 语言手工编写一个 Oberon-0 语言的语法分析程序，该语法分析程序执行与实验三自动生成的语法分析程序类似的功能，但实验三要求逆向工程工具生成的是调用图，而实验四要求生成的是流程图 (Flowchart)。

该语法分析程序采用递归下降预测分析技术，要求你遵循语法制导翻译思想，先设计 Oberon-0 语言的翻译模式，再据此编写语法分析程序。

### 2.4.1 实验步骤 4.1、设计 Oberon-0 语言的翻译模式

根据 Oberon-0 语言 BNF 定义中的语法规则，以及你完成 Oberon-0 源程序处理的应用需求，为 Oberon-0 语言设计一个合适的翻译模式。



注意，由于实验四规定使用递归下降的预测分析技术，你需要改造文法以适用于这种自顶向下的分析方法。例如，上下文无关文法中的左递归必须消除，以避免递归下降的预测分析程序进入死循环。

#### 2.4.2 实验步骤 4.2、编写递归下降预测分析程序

根据上一步骤获得的翻译模式，利用 Java 语言设计并实现一个 Oberon-0 语言的递归下降预测分析程序。

结合编译原理理论课所学知识，从一个翻译模式设计一个递归下降预测分析程序已有比较成熟的启发式规则，你应遵循这些规则设计你的语法分析程序。

例如：文法的每一非终结符号应对应着一个递归子程序，开始符号则对应着其中的主程序；由向前看符号（Lookahead）决定分支动作；每一个继承属性对应一个形式参数，所有综合属性对应返回值，子结点的每一属性对应一个局部变量；翻译模式中产生式右部的终结符号、非终结符号与语义动作分别执行匹配、递归调用和嵌入代码等动作。

#### 2.4.3 实验步骤 4.3、语法分析讨论：自顶向下 vs. 自底向上

通过你自己在实验三和实验四的实际体会，对递归下降预测分析技术和自底向上的 LR 分析技术这两种不同的分析策略进行比较。

建议你在比较两种技术的各自优点和不足时，考虑（但不必局限于）以下方面：

- 分析技术的简单性，包括分析程序是否易于调试。
- 分析技术的通用性，即能处理的语言范围。
- 是否便于表达语义动作以完成语法制导翻译。
- 是否易于实现出错恢复。
- 若以表格驱动方式取代递归程序实现，则分析表大小的优劣如何？
- 分析速度。

#### 2.4.4 提交结果

将实验四的所有结果存放在子目录 **xxxxxxxxNNN\ex4** 中，其中 **xxxxxxxx** 是你的学号，**NNN** 是你的中文姓名，命名规则同实验一。

实验四最终提交的实验结果应包括：

- 在 **ex4** 子目录中存放自述文件 **readme.txt**，其中给出你的姓名、学号、电子邮件、联系电话、完成日期、以及其他补充说明。



- 在 **ex4** 子目录中存放描述 Oberon-0 语言翻译模式的文本文件 **scheme.pdf**。
- 你手工编写的 Oberon-0 语言语法分析和语法制导翻译程序及其测试用例，包括：
  - 在 **ex4** 子目录中存放各种脚本文件，包括：编译语法分析程序的脚本 **build.bat**；运行语法分析程序扫描你编写的正确 Oberon-0 例子程序的脚本 **run.bat**；运行所有由实验软装置提供的测试用例的脚本 **test.bat**。
  - 在 **ex4\src** 子目录中存放 Oberon-0 语法分析程序的主程序 **OberonParser.java**，以及你在考虑面向对象设计时引入的其他 Java 类的源代码；你用程序包组织的各相关类亦存放在 **ex4\src** 之中的相应子目录中。
  - 在 **ex4\bin** 子目录中存放根据词法分析程序源代码编译得到的 Java 字节码文件 **OberonParser.class** 以及其它相关的字节码文件。
  - 在 **ex4\doc** 子目录中存放根据你的源程序中文档化注释自动生成的 javadoc 文档。

### 2.4.5 评价标准

为了更好地分配时间与精力完成本次实验，你可参考以下评分标准：

项 目	权重 (%)
1、Oberon-0 语言的翻译模式	30
2、语法分析和语法制导翻译程序的源代码	40
3、语法分析和语法制导翻译程序的运行与测试	15
4、面向对象设计与编码风格	10
5、文档组织及 readme 文件等	5

## 3 Oberon-0 语言

---

本实验的处理对象是 Oberon-0 语言, 该语言中包含了高级程序设计语言的表达式, 以及结构化程序设计中的结构化控制结构、子程序、参数传递等机制的抽象。

### 3.1 简介

用于编译原理实验的计算机语言应足够简单, 但又不失其代表性。据此, 本实验项目选择了 Oberon-0 语言为处理对象。

Oberon-0 的来历可追溯到近 50 年前, 在此期间的程序设计语言发展约每 10 年就有一标志性成果。1960 年, P. Naur 等人设计了 Algol 60 语言; 约 10 年后, 随着结构化程序设计思想的成熟, N. Wirth 设计出远比 Algol 68 语言成功 (特别是在教育界) 的 Pascal 语言; 又一个 10 年过去, N. Wirth 根据程序设计和软件工程的最新进展, 在 Pascal 基础上设计了 Modula-2 语言; 又过了约 10 年后, M. Reiser 与 N. Wirth 一起, 将 Pascal 语言和 Modula-2 语言的程序设计本质精华浓缩为 Oberon 语言。

Oberon-0 语言是 Oberon 语言的一个子集, 为程序员提供了良好的程序结构。在 Oberon-0 程序中, 最基本的语句是赋值语句; 复合语句支持顺序、条件 (if 语句) 和迭代 (while 语句) 执行。Oberon-0 中还支持子程序这一重要概念, 包括过程声明和过程调用两个范畴, 并且提供了两种不同的参数传递方式: 按值调用 (值参数) 和按引用调用 (可变参数)。

然而 Oberon-0 的类型系统却十分简洁, 仅有的基本数据类型是整数类型 (INTEGER) 和布尔类型 (BOOLEAN), 因而可声明整数类型的常量和变量, 也允许用算术运算符构造表达式; 而表达式的比较运算则产生 BOOLEAN 类型的值, 并可用于逻辑运算。Oberon-0 的复合数据类型包括数组和记录, 且允许任意嵌套; 但最基本的指针类型或引用类型就被省略了。

一个过程代表了由语句组成的功能单元, 因而在一个过程的写法中自然会关系到名字的局部性 (Locality) 问题。Oberon-0 语言支持将标识符声明为局部于某一过程, 即仅在该过程本身范围内标识符才是可见的或合法的。

由 N. Wirth 本人编著的 Theory and Techniques of Compiler Construction: An Introduction (Addison-Wesley, 1996, ISBN 0-201-40353-6, 本课程教学网站提供了该教材的电子版) 中, 第 6 章描述了 Oberon-0 程序设计语言的语法并给出一个样板程序。本文档的语法定义和例子程序即源于该教材, 但作了少数改动, 因而本实验处理的 Oberon-0 语言应以本文档的定义为准。

## 3.2 词法定义

Oberon-0 语言定义了非常简单的语法规则。

### 3.2.1 单词

与我们熟悉的 C、C++、Java 等语言不同, Oberon-0 语言是大小写无关的。譬如, 保留字 WHILE、While 和 while 三种写法是等价的; 而标识符 BALANCE、Balance 和 balance 是相同的标识符。

Oberon-0 的标识符长度不允许超过 24 个字符 (允许 24 个字符)。

在 Oberon-0 中还支持括号风格的注释, 在 “(” 和 “)” 之间的内容全部为注释; 注意, Oberon-0 注释不允许嵌套。

### 3.2.2 基本数据类型

在 Oberon-0 程序中仅支持 INTEGER 和 BOOLEAN 两种基本数据类型, 可以利用 VAR 声明这两种类型的变量。

INTEGER 类型的常量书写形式只允许 Pascal 语言的无符号整数; 这些常量可以由 0 开头, 但解释为八进制常量 (此时常量中不允许出现 8 和 9 两个数字), 非 0 开头则解释为十进制常量。无论十进制还是八进制整数常量, 每一常量中包含数字 (包括 0) 的个数不可超过 12 个 (从而限制了整数常量允许表达范围的最大值)。

INTEGER 类型常量与标识符之前必须以空白符号分隔; 例如, 扫描 25id 时应作为一个非法整数常量处理, 而不是理解为常量 25 和标识符 id 两个单词。注意, 不支持书写 BOOLEAN 类型的常量 TRUE 和 FALSE。

## 3.3 语法定义

本小节以 EBNF 定义了 Oberon-0 语言的形式语法。

### 3.3.1 扩展 BNF (EBNF)

EBNF 意即扩展的 BNF (Extended BNF), 是我们在实际应用中定义一门计算机语言的形式语法的国际标准, 参见 ISO/IEC 14977: 1996(E). The Standard Metalanguage Extended BNF (本课程教学网站提供了该标准的电子版文档)。

### 3.3.2 Oberon-0 语言的 EBNF 定义

```
module = "MODULE" identifier ";"  
        declarations  
        ["BEGIN"  
         statement_sequence]
```

	"END" identifier "." ;
declarations	= ["CONST" {identifier "=" expression ";" }] ["TYPE" {identifier "=" type ";" }] ["VAR" {identifier_list ":" type ";" }] {procedure_declaration ";" } ;
procedure_declaration	= procedure_heading ";" procedure_body ;
procedure_body	= declarations ["BEGIN" statement_sequence] "END" identifier ;
procedure_heading [formal_parameters] ;	= "PROCEDURE" identifier [formal_parameters] ;
formal_parameters	= "(" [fp_section {";" fp_section}] ")" ;
fp_section	= ["VAR"] identifier_list ":" type ;
type	= identifier   array_type   record_type ;
record_type	= "RECORD" field_list {";" field_list} "END" ;
field_list	= [identifier_list ":" type] ;
array_type	= "ARRAY" expression "OF" type ;
identifier_list	= identifier {" ," identifier} ;
statement_sequence	= statement {";" statement} ;
statement	= [assignment   procedure_call   if_statement   while_statement] ;
while_statement	= "WHILE" expression "DO" statement_sequence "END" ;
if_statement	= "IF" expression "THEN" statement_sequence {"ELSIF" expression "THEN" statement_sequence} ["ELSE" statement_sequence] "END" ;
procedure_call	= identifier [actual_parameters] ;
actual_parameters	= "(" [expression {" ," expression}] ")" ;

assignment	=	identifier selector ":" expression ;
expression	=	simple_expression [ ("="   "#"   "<"   "<="   ">"   ">=") simple_expression ] ;
simple_expression	=	["+ "   "- "] term { ("+"   "-"   "OR") term } ;
term factor	=	factor { ("*"   "DIV"   "MOD"   "&") factor } ;
factor	=	identifier selector   number   "(" expression ")"   "~" factor ;
number	=	integer ;
selector	=	"." identifier   "[" expression "]" ;
integer	=	digit {digit} ;
identifier	=	letter {letter   digit} ;

## 3.4 语言描述

在 Theory and Techniques of Compiler Construction: An Introduction 中并未详细描述 Oberon-0 语言的语义特性，因为该语言的语义可以按直观的方式参照其他程序设计语言（特别是沿 Algol 60、Pascal、Modula-2、Oberon 这一家族的语言）来理解。

本小节特别强调了 Oberon-0 可能引起理解上二义性的几个方面，使得在本实验项目中老师和所有学生对该语言的理解是无二义的。

### 3.4.1 模块声明

一个模块中可声明类型、常量、变量、过程等，也可声明该模块的主程序（程序体包括在 BEGIN 和 END 之间）。但一个模块也允许没有主程序，仅声明一些类型、常量、变量、过程等供其他模块使用（尽管本实验项目中尚未更深入地考虑这一问题）。

一个模块声明的 END 之后的标识符必须与该模块的名字相同。

### 3.4.2 运算符与表达式

算术表达式支持一元运算 “+” 和 “-”，分别表示取正和取负运算；支持二元运算 “+”、“-”、“\*”、“DIV” 和 “MOD”，分别表示加法、减法、乘法、整除、取模运算。参与算术表达式的运算量必须是 INTEGER 类型，否则会产生类型不兼容错误；由于 Oberon-0 语言不支持实数除法运算（仅支持整除运算），因而算术表达式的运算结果总是 INTEGER 类型。

算术表达式可参与“=”、“#”(不等于)、“<”、“<=”、“>”和“>=”等关系运算,运算结果为 BOOLEAN 类型的值。BOOLEAN 类型的值不允许参与任何关系运算,否则会产生类型不兼容错误。

但是 BOOLEAN 类型的值可能参与逻辑运算“&”、“OR”和“~”,分别表示“与”、“或”和“非”。INTEGER 类型的值不允许参加任何逻辑运算,否则会产生类型不兼容错误。

### 3.4.3 类型声明

Oberon-0 允许类型声明,即以一个标识符重命名一个基本数据类型或复合数据类型。例如,

---

```
TYPE
  UserId = INTEGER;
  VisitRecord = RECORD
    user: UserId;
    visits: INTEGER
  END;
```

---

此后,即可用这些标识符作为类型,以声明其他变量。例如,

---

```
VAR
  id1, id2: UserId;
  rec: VisitRecord;
```

---

### 3.4.4 选择符

Oberon-0 语言支持两种选择符:“.”用于访问记录中的一个域;“[]”用于以下标表达式访问一个数组的元素。这两种选择符均可以嵌套使用,但必须保证:“.”左边的标识符指称的是一个记录,右边的标识符是该记录中的一个域;“[]”左边的标识符指称的是一个数组,“[]”中间是一个求值结果为 INTEGER 的表达式。

例如,以下选择符的使用是合法的:

---

```
VAR
  accountList: ARRAY 100 OF RECORD
    account: INTEGER;
    balance: INTEGER
  END;
BEGIN
  accountList[1].account := 101;
  accountList[1].balance := 8500;
  ...
```

---

### 3.4.5 作用域规则

Oberon-0 语言是一个块(Block)结构语言;凡在一个块中声明的所有标识符,其作用域仅局限在该块之中。

例如,在一个 MODULE 中声明的所有类型、常量和变量(参见语法单位 declarations)在该模块中是可见的,包括该模块定义的所有块(过程声明)中都是可见的;而在一个 PROCEDURE 中声明的所有类型、常量和变量(参见语法



单位 declarations 和 FormalParameters) 则仅在该过程中是可见的, 其他过程中不可见之。

### 3.4.6 过程声明与参数传递

如果一个过程声明的某个形式参数之前声明有保留字 VAR, 则该参数称为一个可变参数, 将采用按引用传递 (Call by Reference) 的参数传递方式, 因而该过程可能有副作用; 否则, 该参数称为一个值参数, 将采用按值传递 (Call by Value) 的参数传递方式, 因而该过程不会产生副作用。严格的语义检查应保证可变参数必须是一个左值 (L-value)。

一个过程声明的 END 之后的标识符必须与该过程的名字相同。

### 3.4.7 预定义函数

Oberon-0 提供了 3 种预定义函数: read()、write()和 writeln, 分别表示控制台输入和输出, 其含义与 Pascal 语言中的相同。

## 3.5 程序示例

以下提供了以 Oberon-0 书写的一个模块 (Module) 的源代码清单, 可能会有助于你理解 Oberon-0 语言的特点。该模块中包含了几个简单的过程, 这些过程的名字表达了其功能。

---

```
(* A sample Oberon-0 source program. *)
MODULE Sample;
  PROCEDURE Multiply;
    VAR x, y, z: INTEGER;
  BEGIN
    Read(x); Read(y);
    z := 0;
    WHILE x > 0 DO
      IF x MOD 2 = 1 THEN z := z + y END;
      y := 2 * y;
      x := x DIV 2
    END ;
    Write(x); Write(y); Write(z); WriteLn
  END Multiply;

  PROCEDURE Divide;
    VAR x, y, r, q, w: INTEGER;
  BEGIN
    Read(x); Read(y);
    r := x; q := 0; w := y;
    WHILE w <= r DO
      w := 2 * w
    END;
    WHILE w > y DO
      q := 2 * q;
      w := w DIV 2;
      IF w <= r THEN
```

---



```
        r := r - w;  
        q := q + 1  
    END  
END;  
Write(x); Write(y); Write(q); Write(r); WriteLn  
END Divide;  
  
PROCEDURE BinSearch;  
    VAR i, j, k, n, x: INTEGER;  
        a: ARRAY 32 OF INTEGER;  
BEGIN  
    Read(n);  
    k := 0;  
    WHILE k < n DO  
        Read(a[k]);  
        k := k + 1  
    END;  
    Read(x);  
    i := 0; j := n;  
    WHILE i < j DO  
        k := (i + j) DIV 2;  
        IF x < a[k] THEN  
            j := k  
        ELSE  
            i := k + 1  
        END  
    END;  
    Write(i); Write(j); Write(a[j]); WriteLn  
END BinSearch;  
END Sample.
```

---

## 4 逆向工程工具

本实验项目要求开发的软件工具 **ROSE** 可根据一个 Oberon-0 语言源程序自动绘制出一个模块中所有过程的调用图 (实验三), 并为每一过程自动绘制出表达控制流设计的流程图 (实验四)。

### 4.1 调用图

在一个编译器执行过程间分析 (Interprocedural Analysis) 以实现全局优化时, 调用图 (Call Graph) 是开展分析的一个重要基础。

#### 4.1.1 调用图定义

在 A. Aho 等人合著的 *Compilers: Principles, Techniques, and Tools*, 2nd Ed (Addison-Wesley, 2006, ISBN 0-321-48681-1) 第 12 章 12.1 小节中, 调用图被定义为一个由结点和有向边组成的集合, 其中:

- 程序中的每一过程用一个结点表示; 在 **ROSE** 中采用黄色填充的矩形表示这类结点, 矩形中标注该过程的名字。
- 程序中每一个调用过程的地方称为调用点 (Call Site), 这些调用点也用一个结点表示; 在 **ROSE** 中用青色填充的矩形表示这类结点, 其中标注调用点的 Id、所处的过程名字、以及发出过程调用的那条语句。
- 在某一调用点调用某一过程这种调用关系在图中用一条有向边表示。

为一个面向对象程序构造调用图并不容易, 因为动态绑定 (Dynamic Binding; 亦称动态分派, Dynamic Dispatching) 使得调用关系更加复杂; 像 C 程序中允许将函数指针作为参数传递, 也给生成调用图造成麻烦。幸运的是我们在 **ROSE** 中不必考虑这一问题。

#### 4.1.2 调用图例子

给定如下所示的 Oberon-0 源程序:

---

```
(* An Oberon-0 demonstrating the usage of a call graph. *)  
MODULE CallGraph;  
  VAR x, y: INTEGER;  
  
  PROCEDURE a;  
  BEGIN  
    b(x, 1);  
    c(y, 2);  
  END  
END
```

---

```

END a;

PROCEDURE b(i: INTEGER; j: INTEGER);
BEGIN
  x := i + j;
END b;

PROCEDURE c(i: INTEGER; j: INTEGER);
BEGIN
  b(x + i, 2 * j);
  d(2 * i, y + j);
END c;

PROCEDURE d(i: INTEGER; j: INTEGER);
BEGIN
  y := i * j;
END d;
END CallGraph.
  
```

根据上述 Oberon-0 源程序，由 **ROSE** 工具自动生成的一个调用图如图 1 所示。

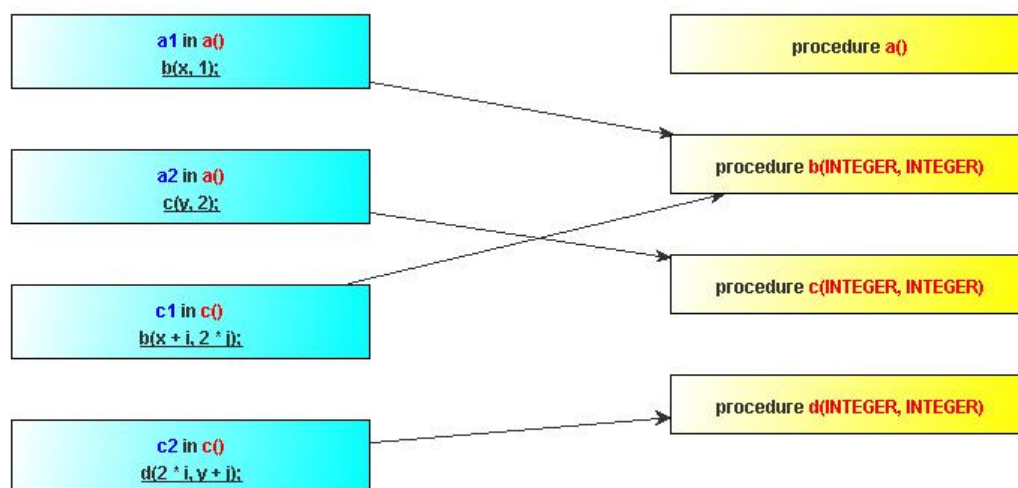


图 4.1 根据 Oberon-0 源程序生成的一个调用图例子

### 4.1.3 参考文献

对于调用图自动生成有兴趣的同学，可进一步阅读以下参考文献：

- [1] B. Ryder. **Constructing the Call Graph of a Program**. *IEEE Transactions on Software Engineering*, 5(3), May 1979, pp.216- 226
- [2] D. Grove, G. DeFouw, J. Dean, and C. Chambers. **Call Graph Construction in Object-Oriented Languages**. *SIGPLAN Notice*, 32(10), October 1997, pp.108-124
- [3] D. Callahan, A. Carle, M. Hall, and K. Kennedy. **Constructing the Procedure Call Multigraph**. *IEEE Transactions on Software Engineering*, 16(4), April 1990, pp.483-487

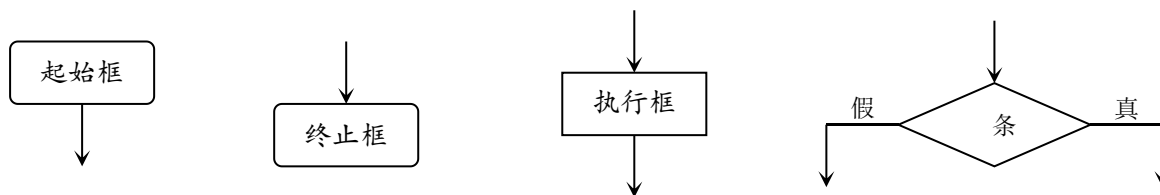
## 4.2 流程图

在结构化程序设计中，有许多表示技术可用于表达控制流设计，譬如控制流图 (Control Flow Graph, 简称 CFG)、N\_S 图 (又称盒图)、日立公司的 PAD 图、JSP 方法中的 Jackson 图等。ROSE 选择了流程图作为控制流的可视化工具。

### 4.2.1 流程图定义

流程图 (Flowchart) 又称程序流程图、程序框图，是一个子程序控制流设计的图形表示，其中展示了程序执行过程中可能遍历的所有路径。

流程图中具有以下基本图形：



关于流程图的更详细资料可参阅：<http://en.wikipedia.org/wiki/Flowchart>。

### 4.2.2 流程图例子

对于一个 Account 模块中的存款和取款过程，由 ROSE 生成的两个流程图例子如图 4.2 所示。

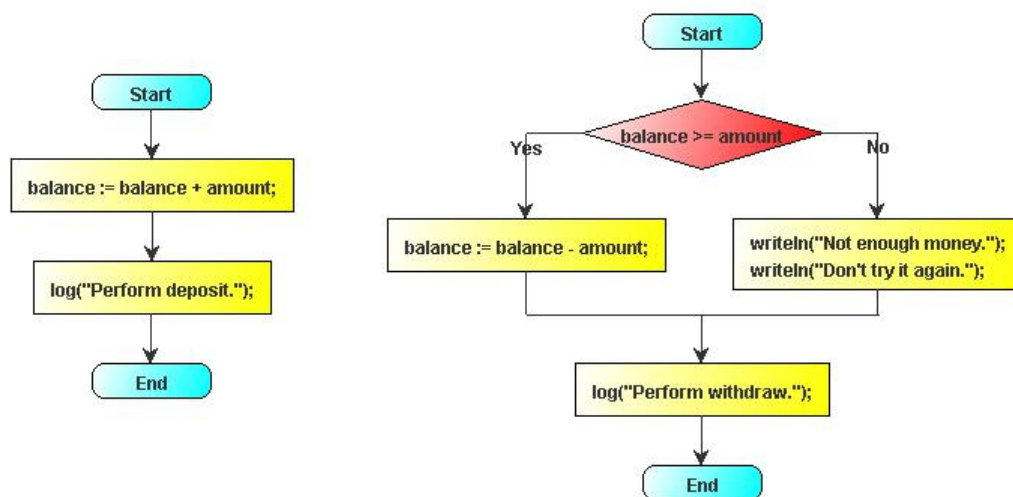


图 4.2 两个流程图的例子

对于本文档第三章 3.5 小节所示 Sample 模块中的 Multiply 过程, 由 **ROSE** 自动生成的一个流程图例子如图 4.3 所示。

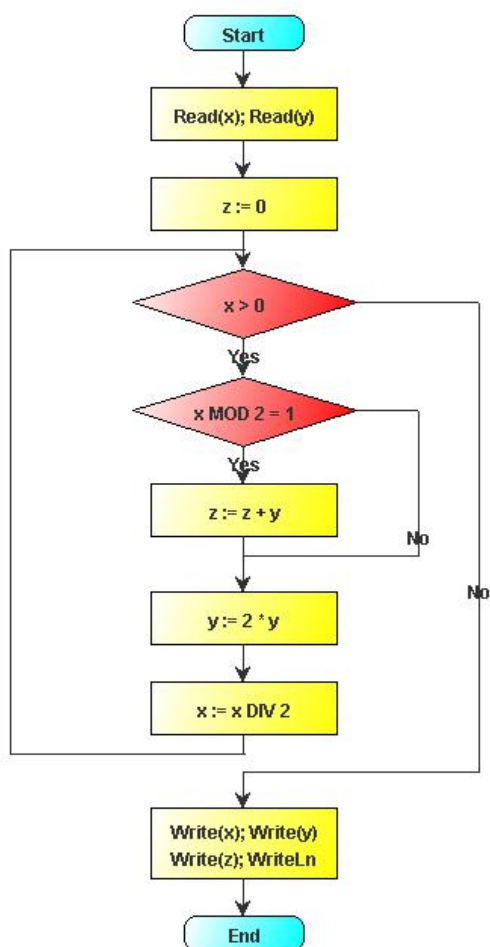


图 4.3 第三章 Oberon 例子程序中 Multiply 过程对应的流程图



## 5 实验软装置

为便于同学们开展实验，本实验提供了实验软装置。每位同学应下载该实验软装置，并基于它开展实验过程。

### 5.1 目录和文件组织

本实验项目提供的实验软装置包括以下目录和文件:

- 子目录 `exceptions` 中存放了所有预定义的异常类。
- 子目录 `callgraph` 中存放了为生成调用图而设计的可复用类。
- [PENDING]

### 5.2 异常类设计

本实验要求在词法分析、语法分析、语义分析等阶段发现源程序中的错误时，抛出相应的异常对象。实验软装置中预定义了所有的异常类型，你的分析程序中应根据相应的错误类别抛出最准确的异常对象；如果你认为某类异常还可以细分，则允许利用继承关系派生出更明细的异常类。

#### 5.2.1 异常类程序包

所有异常类均组织在一个名为 `exceptions` 的程序包中，相应的类文件存放在 `exceptions` 子目录中。

#### 5.2.2 通用异常类

为更好地对异常进行分类和组织，本实验项目的实验软装置中设计了几个通用的异常类。通常你的分析程序不应抛出这些通用异常类的对象实例，而应抛出它们的派生类异常对象，以便更准确地反映找到的错误的类别。

##### **OberonException**

所有找到的错误对应的异常对象的根 (Root) 类。

##### **LexicalException**

表示找到的错误属于词法错误，是 `OberonException` 的派生类。本文档 5.3.1 小节详细介绍了该类的所有派生异常类。

### **SyntacticException**

表示找到的错误属于语法错误，是 `OberonException` 的派生类。本文档 5.4.1 小节详细介绍了该类的所有派生异常类。

### **SemanticException**

表示找到的错误属于语义错误，是 `OberonException` 的派生类。本文档 0 小节详细介绍了该类的所有派生异常类。

## **5.3 词法分析实验软装置**

本小节介绍的词法分析实验软装置主要供本实验项目中实验二和实验三使用。

### **5.3.1 词法错误的异常类**

本小节所有异常均为 `LexicalException` 的派生类。

#### **IllegalSymbolException**

当识别一个单词时遇到不合法的输入符号（譬如@、\$等符号）则抛出该异常。

#### **IllegalIntegerException**

当整数常量（无论是十进制还是八进制）与其后的标识符之间无空白分隔时抛出该异常。

#### **IllegalIntegerRangeException**

当识别出的整数常量（无论是十进制还是八进制）大于本文档约定的整数常量值最大限制时抛出此异常。

#### **IllegalOctalException**

当 0 开头的整数常量中含有 0~7 之外的符号（包括 8 和 9）时抛出该异常。

#### **IllegalIdentifierLengthException**

当识别出的一个标识符长度超过最大限制时抛出该异常。

#### **MismatchedCommentException**

当 “(” 开头的注释直至扫描到最后一个符号都找不到配对的 “)” 时抛出该异常。

### **5.3.2 词法分析主程序**

本实验软装置提供了语法分析的主程序 `main()` 源代码，在你的实验 2 和实验 3 中均以此作为主程序；请不要修改主程序中的任何代码片段。

主程序 `main()` 简单地根据命令行参数读入测试用例, 并调用你的词法分析入口程序完成单词扫描过程, 再根据返回结果判断是否与预期输出一致。其源代码如下:

---

```
/**  
 * Main entry to the scanner of Oberon-0.  
 ... [PENDING]
```

---

### 5.3.3 词法分析程序测试用例

[PENDING]

## 5.4 语法分析和语法制导翻译实验软装置

本小节介绍的语法分析和语法制导翻译实验软装置主要供本实验项目中实验三和实验四使用。

### 5.4.1 语法错误的异常类

本小节所有异常均为 `SyntacticException` 的派生类。

#### **MissingLeftParenthesisException**

当分析到左右圆括号不匹配、且缺少左括号时则抛出该异常。

#### **MissingRightParenthesisException**

当分析到左右圆括号不匹配、且缺少右括号时则抛出该异常。

#### **MissingOperatorException** 异常

当分析到程序中缺少运算符时, 或调用预定义函数缺少相应的参数时则抛出该异常。

#### **MissingOperandException** 异常

当分析到程序中缺少操作数时则抛出该异常。

### 5.4.2 语义错误的异常类

本小节所有异常均为 `SemanticException` 的派生类。

#### **TypeMismatchedException**

当分析到表达式、赋值语句、或参数传递等构造中出现类型不兼容错误时则抛出该异常。

## ParameterMismatchedException

当分析到过程调用的实际参数数目与过程声明的形式参数数目不一致时则抛出该异常。注意, 当实际参数的数目与形式参数一致, 只是存在某一参数的类型不兼容时, 应抛出 `TypeMismatchedException` 异常而不是抛出本异常。

## 5.5 预定义设计图 API

**ROSE** 的实验软装置提供了预定义的 API 支持你绘制各类设计图, 强烈建议你在本实验项目中调用这些现成的 API 绘制各类设计图。学有余力的同学可以自己开发新的控制流设计图 API (譬如 PAD 图或 Jackson 图), 但请注意这并不是本实验项目的关注点。

实验软装置预定义的设计图 API 均基于开源项目 JGraph, 其中的类 JGraph 等价于 Java Swing 中的 JTable、JTree 等控件, 弥补了 Swing 未提供强有力图形支持的不足。但你在实验中不必理会 JGraph, 直接使用由实验软装置提供的 API 即可。

### 5.5.1 调用图绘制 API

要绘制类似 DBv2 第 12 章 pp.904-906 的例子, 可按如下方式调用预定义的 API (意即你的翻译模式应按如下次序执行其中的语义运作; 某些次序是任意的, 但 `addEdge()` 的次序受一定限制):

```
import callgraph.*;

public class Demo1 {
    public static void main(String[] args) throws Exception {
        // Create a call graph
        CallGraph graph = new CallGraph();

        // Add procedure nodes
        graph.addProcedure("f1", "fun1");
        graph.addProcedure("f2", "fun2");
        graph.addProcedure("main", "main");

        // Add call sites
        graph.addCallSite("c1", "fun1", "return (*pf)(x+1)");
        graph.addCallSite("c2", "fun2", "return (*pf)(y)");
        graph.addCallSite("c3", "main", "(*pf)(5)");

        // Add procedure calls
        graph.addEdge("c1", "f1");
        graph.addEdge("c1", "f2");
        graph.addEdge("c2", "f1");
        graph.addEdge("c2", "f2");
        graph.addEdge("c3", "f1");
        graph.addEdge("c3", "f2");
        // Show the graph
        graph.show();
    }
}
```

运行上述程序，可得与 DBv2 第 12 章 pp.906 调用图类似的结果，如图 5.1 所示：

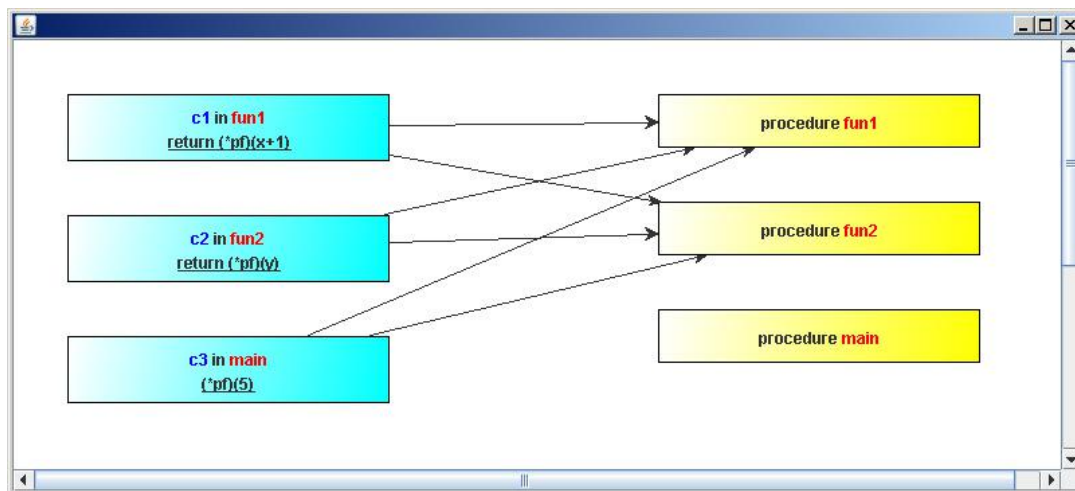


图 5.1 使用预定义 API 生成的调用图

### 5.5.2 流程图绘制 API

要绘制第三章 3.5 小节的 Oberon 源程序定义 Sample 模块中 3 个过程对应的流程图，可按如下方式调用预定义的 API（意即你的翻译模式应按如下次序执行其中的语义运作）：

```

import flowchart.*;

public class FlowchartDemoOberon {
  public static void main(String[] args) throws Exception {
    Module sampleModule = new Module("Sample");
    {
      Procedure proc = sampleModule.add("Multiply");
      proc.add(new PrimitiveStatement("Read(x); Read(y)"));
      proc.add(new PrimitiveStatement("z := 0"));
      WhileStatement wstmt = new WhileStatement("x > 0");
      proc.add(wstmt);
      {
        IfStatement istmt = new IfStatement("x MOD 2 = 1");
        wstmt.getLoopBody().add(istmt);
        {
          istmt.getTrueBody().add(new PrimitiveStatement("z := z +
y"));
        }
        wstmt.getLoopBody().add(new PrimitiveStatement("y := 2 * y"));
        wstmt.getLoopBody().add(new PrimitiveStatement("x := x DIV
2"));
      }
      proc.add(new PrimitiveStatement(
        "Write(x); Write(y)<br>Write(z); WriteLn"));
    }
    Procedure proc = sampleModule.add("Divide");
    proc.add(new PrimitiveStatement("Read(x); Read(y)"));
    proc.add(new PrimitiveStatement("r := x; q := 0; w := y"));
  }
}
  
```

```

WhileStatement wstmt = new WhileStatement("w <= r");
proc.add(wstmt);
{
    wstmt.getLoopBody().add(new PrimitiveStatement("w := 2 * w"));
}
wstmt = new WhileStatement("w > y");
proc.add(wstmt);
{
    wstmt.getLoopBody().add(new PrimitiveStatement("q := 2 * q"));
    wstmt.getLoopBody().add(new PrimitiveStatement("w := w DIV
2"));
    IfStatement istmt = new IfStatement("w <= r");
    wstmt.getLoopBody().add(istmt);
    {
        istmt.getTrueBody().add(new PrimitiveStatement("r := r -
w"));
        istmt.getTrueBody().add(new PrimitiveStatement("q := q +
1"));
    }
}
proc.add(new PrimitiveStatement(
    "Write(x); Write(y)<br>Write(q); Write(r); WriteLn"));
} {
    Procedure proc = sampleModule.add("BinSearch");
    proc.add(new PrimitiveStatement("Read(n)"));
    proc.add(new PrimitiveStatement("k := 0"));
    WhileStatement wstmt = new WhileStatement("k <= n");
    proc.add(wstmt);
    {
        wstmt.getLoopBody().add(new
PrimitiveStatement("Read(a[k])"));
    }
}

```

上述关于流程图的 API 由程序包 flowchart 提供, 该程序包的类结构设计如图 5.2 所示。关于这些 API 设计的设计空间 (Design Space) 及其折衷结论 (Trade-off and Consequence) 有许多值得深究之处, 留待本课程的实验交流环节进一步讨论。



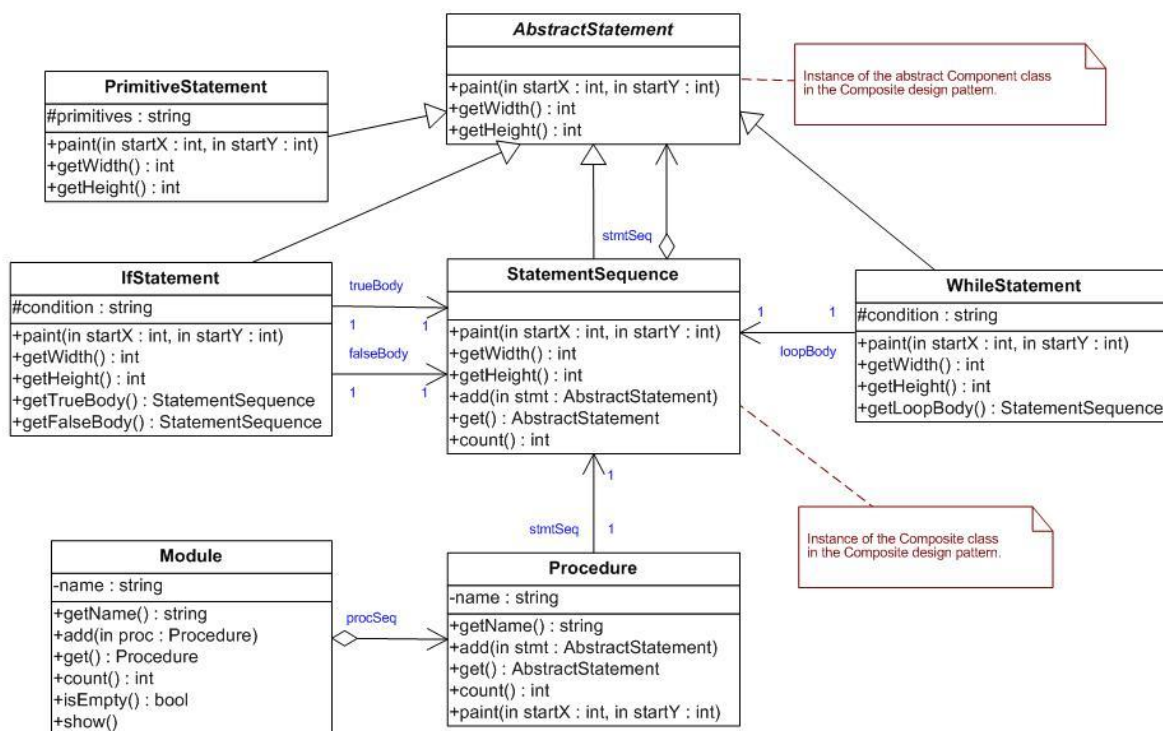


图 5.2 程序包 flowchart 的 UML 类图

□