

# 数据结构课程设计 指导书

合肥工业大学  
计算机与信息学院  
2021 年 6 月

## 一、概述

数据结构课程设计是对学生学习数据结构课程后,编程能力的一种全面综合训练,是与课堂听讲、自学和练习相辅相成的、必不可少的一个教学环节。通常,课程设计中的问题比平时的习题复杂的多,也更接近实际。课程设计着眼于原理与应用的结合点,使学生学会如何把书上学到的知识用于解决实际问题,培养软件工作所需要的动手能力;另一方面,能使书上的知识变“活”,起到深化理解和灵活掌握教学内容的目的。平时的习题较偏重于如何编写功能单一的“小”算法,局限于一个或两个知识点,而课程设计是软件设计的综合训练,包括问题分析,总体结构设计,用户界面设计、程序设计基本技能和技巧,多人合作,以至一整套软件工作规范的训练和科学作风的培养。此外,还有很重要的一点是:计算机是比任何教师更严厉的检查者。

为达到上述目的,使学生更好地掌握程序设计的基本方法和 C++语言的应用,本课程安排了课程设计环节,提供了各类题目供学生选择。每个设计题采取了统一的格式,由问题描述、基本要求、测试数据、实现提示和选做内容等五个部分组成。问题描述旨在为学生建立问题提出的背景,指明问题“是什么”。基本要求则对问题进一步求精,划出问题的边界,指出具体的参量或前提条件,并规定该题的最低限度要求。测试数据部分旨在为检查学生上机作业提供方便。在实现提示部分,对实现中的难点及其解法思路等问题作了简要提示,提示的实现方法未必是最好的,学生不应拘泥与此,而应努力设计和开发更好的方法和结构。选做部分向那些尚有余力的读者提出了更高的要求,同时也能开拓其它读者的思路,在完成基本要求时就力求避免就事论事的不良思想方法,尽可能寻求具有普遍意义的解法,使得程序结构合理,容易修改、扩充和重用。

## 二、总体要求

### 1. 系统分析与系统设计

“分析就是在采取行动之前,对问题的研究”(Demarco, 1978)。系统分析在软件开发过程中是非常重要的,其任务就是通过对问题本身的研究,产生一个系统需要做什么的规范的、一致的和可行的需求说明。

在此基础上,根据任务确定数据结构及其相应的存储方式,并利用或设计算法实现问题描述。

在此之上进一步考虑系统中所需设计的类(对象)以及类之间的关系、本系统中各个类所涉及的属性及针对这些属性的操作。类及类之间的关系可用类图来表示,对象之间的消息传递可用箭头表示,另外一些重要的操作应给出规格说明。

### 2. 详细设计与编码

对类中的属性和操作从实现的角度(如可扩充、在派生类中能否直接使用或

只需少量修改、访问的效率和方便性等)进一步考察;对类中的操作(即方法)进一步求精:用 if、while、for 和赋值语句加上自然语言写出算法框架;同时考虑能否使用已有类库(包括直接使用或通过派生)以减少编程的工作量和提高程序的可靠性。

编码,即程序设计,是对详细设计的结果的进一步求精,用面向对象语言(如 C++)表达出来。在充分理解和把握语言运行机制的基础上,编写出正确的、清晰的、易读易改和高效率的程序。另外,在标识符的命名、代码的长度(一个方法长度一般不超过 40 行,否则应划分为两个或多个方法)、程序书写的风格(如缩进格式、空格(空行)的应用、注释等)方面也应注意,遵循统一的规范。

### 3. 上机调试和测试

上机时要带一本面向对象语言的教材,若有开发环境的用户指南(手册)及类库(库函数)手册则更好。应仔细阅读程序编译和连接时的错误信息(通常是英文的),弄清其确切含义,提高调试效率。要学习并掌握开发环境所提供的调试工具。

经过调试,能够运行的程序并非就是一个正确的程序。实际上,在上机之前,就应根据系统的需求设计相应的测试数据集,特别是一些异常情况的处理(如用户输入数据未按指定格式、数据极大或极小时程序如何处理等一些极端的情况)。

### 4. 课程设计报告

课程设计报告的内容及要求:

#### (1) 需求和规格说明

描述问题,简述题目要解决的问题是什么?规定软件做什么。原题条件不足时补全

#### (2) 设计

1) 设计思想:程序结构(如类图),重要的数据结构。主要算法思想(文字描述,不要画框图)

2) 设计表示:类名及其作用,类中数据成员名称及作用,类中成员函数原型及其功能,可以用表格形式表达。

3) 实现注释:各项要求的实现程度、在完成基本要求的基础上还实现了什么功能?

4) 详细设计表示:主要算法的框架及实现此算法的成员函数接口。

#### (3) 用户手册

即使用说明(包括数据输入时的格式要求)。

#### (4) 调试及测试

调试过程中遇到的主要问题是如何解决的；对设计和编码的回顾讨论和分析；程序运行的时空效率分析；测试数据集；运行实例；改进设想；经验和体会等。

### (5) 附录

1) 源程序清单：打印文本和磁盘文件，磁盘文件是必须的。源程序要加注释，除原有注释外再用钢笔加一些必要的注释和断言。

2) 测试数据：即列出测试数据集

3) 运行结果：上面测试数据输入后程序运行的结果

### 注意事项：

以上要求为一般的要求，针对具体问题和具体的开发过程，某些方面可以做适当的增减。

各种文档资料要在程序开发过程中逐渐形成, 而不是最后补写（但不排斥最后誉清）。

各种文档要以统一格式的稿纸用钢笔书写，也可录入计算机用 Word 及其它文字编辑软件排版后打印输出。

## 三、示例

封面：

数据结构课程设计 报告	
设计题目：	*****
学生姓名：	
专    业：	
班    级：	
学    号：	
指导教师：	
完成日期：	
合肥工业大学计算机与信息学院	

### (1) 需求和规格说明 任务功能描述

某小型公司，主要有四类人员：经理、技术人员、销售经理和推销员。要求存储这些人员的姓名、编号、级别、当月薪水，计算月薪总额并显示全部信息。

人员编号基数为 1000，每输入一个人员的信息，编号顺序加 1。

程序要对所有人员有提升级别的功能。为简单起见，所有人员的初始级别均为 1 级，然后进行升级，经理升为 4 级，技术人员和销售经理升为 3 级，推销员仍为 1 级。

月薪计算办法是：经理拿固定月薪 8000 元；技术人员按每小时 100 元领取月薪；推销员的月薪按该推销员当月销售额的 4%提成；销售经理既拿固定月薪也领取销售提成，固定月薪为 5000 元，销售提成为所管辖部门当月销售总额的 5%。

### (2) 设计

据上述需求，设计一个基类\*\*，然后派生出\*\*类、\*\*类。子类与基类间的关系。

每个类中涉及哪些主要函数：

在基类中，除了定义构造函数和析构函数以外，还应统一定义对各类人员信息都应有的操作，这样可以规范各派生类的基本行为。但是各类人员的月薪计算方法不同，不能在基类 employee 中统一定义计算方法。各类人员信息的显示内容也不同，同样不能在基类中统一定义显示方法。因此，在 employee 类中用纯虚函数的方式定义了计算月薪函数 pay() 和显示信息函数 displayStatus()，然后在派生类中再根据各自的同名函数实现具体的功能。

由于 salesmanager 的两个基类又有公共基类 employee，为避免二义性，这里将 employee 类设计为虚基类。

#### 功能实现的主要流程图

### (3) 用户手册

程序运行时，首先提示输入雇员姓名。

对于经理直接输出其工资及其它信息；

对于技术人员，程序提示输入其本月工作时数，然后输出其工资及其它信息；

对于推销员，程序提示输入其本月销售额，然后输出其工资及其它信息；

对于销售经理，程序提示输入其管辖部门本月销售总额，然后输出其工资及其它信息。

### (4) 调试及测试

测试数据 1:

运行结果 1:

测试数据 2:

运行结果 2:

测试数据 3:

运行结果 3:

(5) 存在的问题

(6) 进一步改进及其思路

附录——源程序

## 四、题目

### 题 1：求字符串之间距离

**问题描述：**设有字符串  $X$ ，称在  $X$  的头尾及中间插入任意多个空格后构成的新字符串为  $X$  的扩展串，如字符串  $X$  为“abcbcd”，则字符串“abcb□cd”，“□a□bcbcd□”和“abcb□cd□”都是  $X$  的扩展串，这里“□”代表空格字符。如果  $A1$  是字符串  $A$  的扩展串， $B1$  是字符串  $B$  的扩展串， $A1$  与  $B1$  具有相同的长度，那么定义字符串  $A1$  与  $B1$  的距离为相应位置上的字符的距离总和，而两个非空格字符的距离定义为它们的 ASCII 码的差的绝对值，而空格字符与其它任意字符之间的距离为已知的定值  $K$ ，空格字符与空格字符的距离为 0。在字符串  $A$ 、 $B$  的所有扩展串中，必定存在两个等长的扩展串  $A1$ 、 $B1$ ，使得  $A1$  与  $B1$  之间的距离达到最小，将这一距离定义为字符串  $A$ 、 $B$  的距离。请编写程序，求出字符串  $A$ 、 $B$  的距离。

#### 实现要求：

- 1) 基于静态存储方式和动态存储方式，自行设计并实现 `myString` 类，并实现相应的函数以完成上述功能，建议考虑时空开销；
- 2) 建议结合实际应用数据，进一步扩展字符串距离的定义和计算方法。

### 题 2：后缀表达式计算器（实验课计算器扩展功能已做过的建议不选该题）

**问题描述：**后缀表达式不包含括号，运算符放在两个运算对象的后面，所有的计算按运算符出现的顺序，严格从左向右进行（不再考虑运算符的优先规则，如：2 1 + 3 \*，即  $(2 + 1) * 3$ 。

实现一个简单的后缀表达式计算器。假定表达式里的基本数值为实数，可用的运算符包括 +, -, \*, /, ^，其中的 ^ 表示求幂运算。

#### 实现要求：

- 1) 假定输入表达式里的数和运算符之间都有空格，这样可以简化输入的处理；
- 2) 输入的算术表达式以分号为结束符。计算器应该能输入并计算一系列表达式，遇到一行的第一个字符就是分号时程序结束；
- 3) 上题的计算器增加一元函数功能，允许表达式里写 `sin`, `cos`, `tan`, `log`（自然对数）等函数，还可以考虑加入自定义的其他数学函数；
- 4) 增加窗口界面。

### 题 3：二叉树结点染色问题

**问题描述：**一棵二叉树可以按照如下规则表示成一个由 0、1、2 组成的字符

序列，我们称之为“二叉树序列 S”：

$$S = \begin{cases} 0 & \text{表示该树没有子节点} \\ 1S_1 & \text{表示该树有一个子节点，} S_1 \text{为其子树的二叉树序列} \\ 2S_1S_2 & \text{表示该树有两个子节点，} S_1 \text{和} S_2 \text{分别表示其两个子树的二叉树序列} \end{cases}$$

例如，图 1 所表示的二叉树可以用二叉树序列  $S=21200110$  来表示。

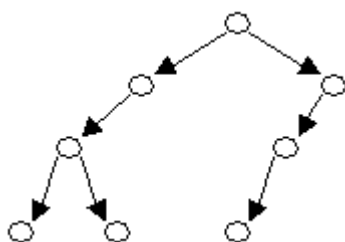


图 1 二叉树序列示意图

任务是要对一棵二叉树的节点进行染色。每个节点可以被染成红色、绿色或蓝色。并且，一个节点与其子节点的颜色必须不同，如果该节点有两个子节点，那么这两个子节点的颜色也必须不相同。给定一棵二叉树的二叉树序列，请求出这棵树中最多和最少有多少个点能够被染成绿色。

**实现要求：**

- 1) 用文件读入方式来构建二叉树结构，至少 5 个二叉树测试用例；
- 2) 用窗口界面显示出所有二叉树的染色结果；
- 3) 给出这棵树中最多和最少有多少个点能够被染成绿色。
- 4)

#### 题 4：模式匹配问题（实验课上已做过的建议不选该题）

**问题描述：**给定一个模式 P 和一个文本文件 T，设计并实现算法在 T 中查找 P。查阅相关资料，熟悉并设计模式匹配方法 index 和 KMP 算法(D.E. Knuth, V. R. Pratt, J. H. Morris)

**实现要求：**

- 1) 实现对 P 的匹配，返回匹配的位置以及次数；
- 2) 对比 KMP 和 index 算法的比较次数；
- 3)（选作）自行查阅资料带通配符的 P 扩展模式，并进行柔性匹配实现；
- 4)（选作）界面演示算法的匹配过程。

**参考：**

**Index 算法描述：**

```
int index(p,S)
{ i=1;
```



```

While (i<=length(S)-length(p)+1)
    if (substr(S,i, length(p))==p)
        return i;
    else i++;
}

```

KMP 算法思想：

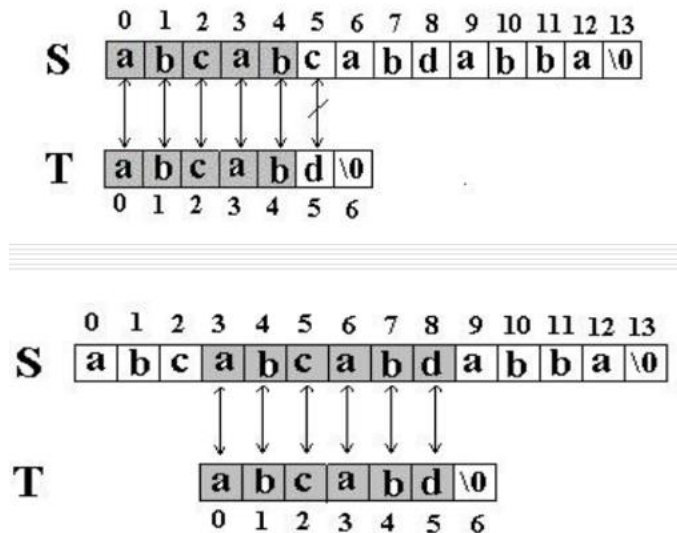


图 2 算法过程示意图

## 题 5：谣言传播问题

**问题描述：**股票经纪人往往对谣言很敏感，你老板希望你能找到一个好方法向他们散布谣言，从而使他在股市占有战术优势。为了达到最有效果，需要谣言传播的尽量快。不幸的是，股票经纪人只相信从他们信任的信源传播来的消息，因此，在你散布谣言之前，需要对他们的联系网进行详细考察。对于一个股票经纪人，他需要一定时间才能将信息传送给他的联系人，给你这些信息，你的任务是，决定选谁作为第一个传递谣言的人，以使谣言传遍所有人的时间最短，当然，如果谣言不能传遍所有人的话，你也要给出说明。

例如，假设共有 3 个联系人，联系人 1 传递信息给联系人 2 和 3 所用的时间分别为 4 和 5；联系人 2 传递信息给联系人 1 和 3 所用的时间分别为 2 和 6；联系人 3 传递信息给联系人 1 和 2 所用的时间均为 2，则选择联系人 3 作为第一个传递谣言的人，可以使谣言传遍所有的人时间最短，为 2。

（选择有向图中的一个源点，使它到其余各顶点的最短路径中最长的一条路径最短）

**实现要求：**

- 1) 通过文件读取方式来构建图结构，至少 5 个测试用例；

- 2) 界面展示出求解的传播路径以及长度;
- 3) (选作) 动态展示出谣言传播过程。

## 题 6: 网络布线

**问题描述:** 计算机网络要求网络中的计算机被连接起来, 本问题考虑一个“线性”的网络, 在这一网络中计算机被连接到一起, 并且除了首尾的两台计算机只分别连接着一台计算机外, 其它任意一台计算机恰连接着两台计算机。图 3 中用圆点表示计算机, 它们的位置用直角坐标表示。网络连接的计算机之间的距离单位为英尺。

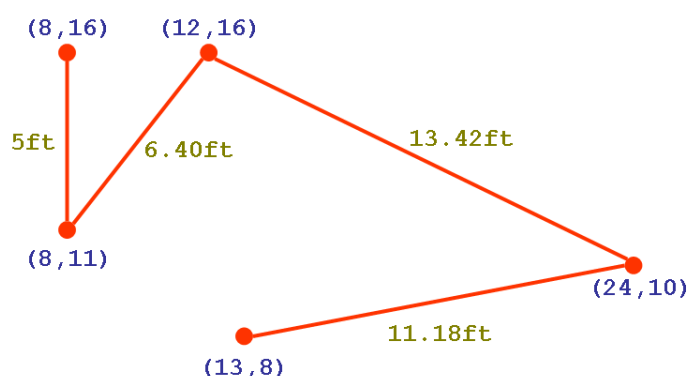


图 3 网络布线示意图

由于很多原因, 我们希望使用的电缆长度应可能地短。你的问题是去决定计算机应如何被连接以使你所使用的电缆长度最短。在设计方案施工时, 电缆将埋在地下, 因此连接两台计算机所要用的电缆总长度等于计算机之间的距离加上额外的 16 英尺电缆, 以从地下连接到计算机, 并为施工留一些余量。

图 4 是计算机的最优连接方案, 这样一个方案用电缆的总长度是

$$(4 + 16) + (5 + 16) + (5.38 + 16) + (11.18 + 16) = 90.01 \text{ 英尺}$$

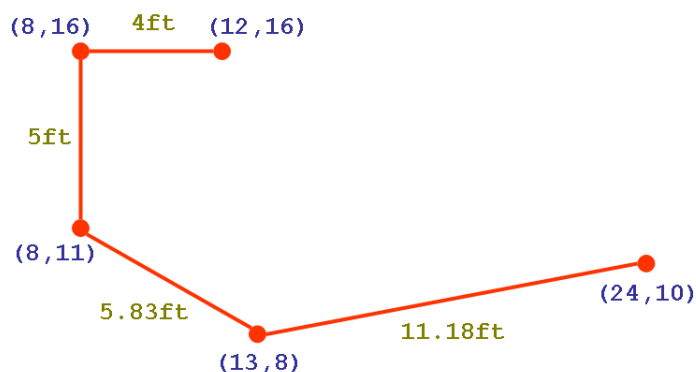


图 4 最优连接方案示意图

### 实现要求:

- 1) 输入网络中的计算机总数和每台计算机的坐标。
- 2) 输出使电缆长度最短的连接方案。给出最优连接方案中每两台相邻计算机之间的距离, 以及总的电缆长度。
- 3) (选作) 参考图 3、图 4, 用图形化的方式显示结果, 包括点的坐标、最优路径、相邻计算机之间的距离。

### 题 7: 中国邮路问题

**问题描述：**邮递员的工作是每天在邮局里选出邮件，然后送到他所管辖的客户中，再返回邮局。自然地，若他要完成当天的投递任务，则他必须要走过他所投递邮件的每一条街道至少一次。问怎样的走法使他的投递总行程为最短？这个问题就称为中国邮路问题。

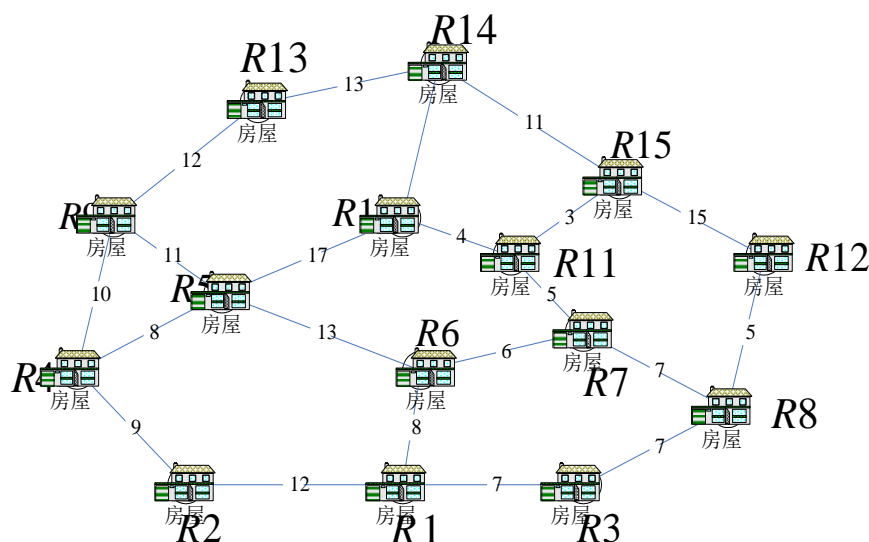


图 5 中国邮路问题示意图

实现要求:

- 1) 只求解用户输入的图形的中国邮路问题：要求用户输入图形，求解输入的图形的中国邮路问题，要求能显示图形和最终结果。
- 2) 加入图形编辑器：系统自动生成图形，系统求解生成的图形的中国邮路问题，要求能显示图形和最终结果。
- 3) (选作) 能够图形显示求解过程。

### 题 8：数独游戏

**问题描述：**在一个  $9 \times 9$  的大正方形中，包含 9 个  $3 \times 3$  的小正方形。如图 6 所示。可以看到，其每行、每列、每个小正方形，都有 9 个空格。

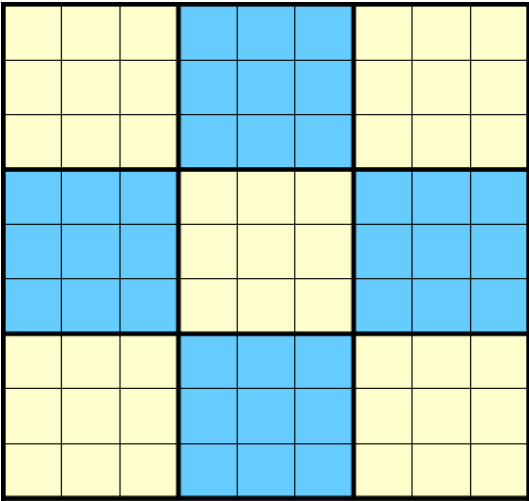


图 6 数独游戏棋盘示意图

要求只用 1 到 9 这些数字，填满大正方形中所有的 81 个空格，同时满足：

- (1) 在每列的 9 个空格中分别填入 1 到 9，且每个数字在此列中只能出现一次；
- (2) 在每行的 9 个空格中分别填入 1 到 9，且每个数字在此行中只能出现一次；
- (3) 在每个小正方形的 9 个空格中分别填入 1 到 9，且每个数字在此正方形中只能出现一次；

游戏一开始会给定了某些空格的值。参加游戏的人根据这些已知的值以及上面的约束条件，推理出剩余的空格的值。

		3			8	1	5	
2					7			
	8					3		
			3	2	4			7
	3	1			9			6
				8	1			
4	2							1
			5			6		
3			1		6			

图 7 数独题目示例

9	7	3	4	6	8	1	5	2
2	1	5	9	3	7	4	6	8
6	8	4	2	1	5	3	7	9
5	6	9	3	2	4	8	1	7
8	3	1	7	5	9	2	4	6
7	4	2	6	8	1	9	3	5
4	2	6	8	7	3	5	9	1
1	9	7	5	4	2	6	8	3
3	5	8	1	9	6	7	2	4

图 8 解题结果

### 实现要求:

1) 编写“数独计算器”: 显示一个空白的  $9 \times 9$  大正方形, 请玩家自己输入要求解的题目, 然后系统帮助玩家解答。

2) 加入“数独题目生成器”: 系统自动生成数独题目, 玩家进行解答, 系统可判定玩家答案的正确性。玩家也可以查看解答。

3) 让玩家选择题目难度, 生成不同难度级别的数独题目; 可以设置提示功能, 在玩家解题过程中帮他提示错误或给出若干空格的解答; 可以根据题目难度和解题时间, 对玩家的水平进行打分;

4) (选作) 窗口界面交互式游戏设计。

### 题 9: 最大匹配问题

**问题描述:** 写出求一个二分图的最大匹配的算法, 并用于解决下面的问题。

第二次世界大战时期, 英国皇家空军从沦陷国征募了大量外籍飞行员。由皇家空军派出的每一架飞机都需要配备在航行技能和语言上能互相配合的 2 名飞行员, 其中 1 名是英国飞行员, 另 1 名是外籍飞行员。在众多的飞行员中, 每一名外籍飞行员都可以与其他若干名英国飞行员很好地配合。如何选择配对飞行的飞行员才能使一次派出最多的飞机。对于给定的外籍飞行员与英国飞行员的配合情况, 试设计一个算法找出最佳飞行员配对方案, 使皇家空军一次能派出最多的飞机。

对于给定的外籍飞行员与英国飞行员的配合情况, 编程找出一个最佳飞行员配对方案, 使皇家空军一次能派出最多的飞机。

### 实现要求:

1) 数据输入: 由文件 input.txt 提供输入数据。文件第 1 行有 2 个正整数  $m$  和  $n$ 。 $n$  是皇家空军的飞行员总数 ( $n < 100$ );  $m$  是外籍飞行员数。外籍飞行员编号为  $1 \sim m$ ; 英国飞行员编号为  $m+1 \sim n$ 。接下来每行有 2 个正整数  $i$  和  $j$ , 表示外籍飞行员  $i$  可以和英国飞行员  $j$  配合。文件最后以 2 个 -1 结束。

2) 结果输出: 程序运行结束时, 将最佳飞行员配对方案输出到文件 output.txt 中。第 1 行是最佳飞行员配对方案一次能派出的最多的飞机数  $M$ 。接下来  $M$  行是最佳飞行员配对方案。每行有 2 个正整数  $i$  和  $j$ , 表示在最佳飞行员配对方案中, 飞行员  $i$  和飞行员  $j$  配对。如果所求的最佳飞行员配对方案不存在, 则输出 'No Solution!'。

3) 界面展示二分图, 并根据求解算法动态展示匹配方案, 即动态显示图中被选中的点。

输入文件示例:	输出文件示例:
input.txt 5 10 1 7 1 8 2 6 2 9 2 10 3 7 3 8 4 7 4 8 5 10 -1 -1	output.txt 4 1 7 2 9 3 8 5 10

#### 题 10：最佳匹配问题

**问题描述：**羽毛球队有男女运动员各  $n$  人。给定 2 个  $n \times n$  矩阵  $P$  和  $Q$ 。 $P[i][j]$  是男运动员  $i$  和女运动员  $j$  配对组成混合双打的男运动员竞赛优势； $Q[i][j]$  是女运动员  $i$  和男运动员  $j$  配合的女运动员竞赛优势。由于技术配合和心理状态等各种因素影响， $P[i][j]$  不一定等于  $Q[j][i]$ 。男运动员  $i$  和女运动员  $j$  配对组成混合双打的男女双方竞赛优势为  $P[i][j] * Q[j][i]$ 。设计一个算法，计算男女运动员最佳配对法，使各组男女双方竞赛优势的总和达到最大。

设计一个优先队列式分支限界法，对于给定的男女运动员竞赛优势，计算男女运动员最佳配对法，使各组男女双方竞赛优势的总和达到最大。

#### 实现要求：

1) 数据输入：第一行有 1 个正整数  $n$  ( $1 \leq n \leq 20$ )。接下来的  $2n$  行，每行  $n$  个数。前  $n$  行是  $p$ ，后  $n$  行是  $q$ ；

2) 结果输出：将计算出的男女双方竞赛优势的总和的最大值输出；

样例输入：

```

3
10 2 3
2 3 4
3 4 5
2 2 2

```

3 5 3

4 5 1

样例输出：52

#### 题 11：构造哈夫曼树

**问题描述：**设计程序以实现构造哈夫曼树的哈夫曼算法

**实现要求：**

- 1) 文件输入至少 5 个测试用例；
- 2) 求解出所有可能的哈夫曼树及其对应的带权路径长度。
- 3) (选作) 动态演示构造过程。

#### 题 12：压缩软件

**问题描述：**采用哈夫曼编码思想实现文件的压缩和恢复功能，并提供压缩前后的占用空间之比。要求如下：

**实现要求：**

- 1) 描述压缩基本符号的选择方法。
- 2) 运行时的压缩原文件的规模应不小于 5K。
- 3) 提供恢复文件与原文件的相同性对比功能。

#### 题 13：排序算法实验分析

**问题描述：**给出一组实验来比较下列排序算法的时间性能：插入排序、冒泡排序、选择排序、快速排序、堆排序、希尔排序、归并排序、基数排序（其它排序也可以作为比较的对象）。

**实现要求：**

- 1) 文件给出至少 10 组测试用例，每组测试用例建议长度在[100-10000]之间；
- 2) 给出排序结果，并计算出元素比较和移动次数以及算法所用机器时间；
- 3) 分析性能包括平均时间性能、最好情况下的时间性能、最差情况下的时间性能等。

#### 题 14：小型文本编辑器

**问题描述：**设计一个文本编辑器，使其具有通常编辑器所应具备的基本功能。

**实现要求：**

- 1) 要求该编辑器在串基本抽象数据类型上构建。

- 2) 编辑器应具备如字符串查找、剪切、粘贴、替换、字数统计、行数统计等基本功能。
- 3) 具备图形化界面。

### 题 15：电梯调度模拟系统

**问题描述：**设计一个模拟电梯工作过程的图形演示系统。要求所设计的电梯能符合市场上大多数系统的要求。

**实现要求：**

- 1) 两部以上的电梯；
- 2) 设计并实现两种以上的调度策略；
- 3) 通过大量的测试用例对比两种策略的性能，包括平均响应时间和电梯能耗。

### 题 16：随机决策树构造

**问题描述：**决策树是通过一系列规则对数据进行分类的过程。它提供一种在什么条件下会得到什么值的类似规则的方法。它是一个从上到下、分而治之的归纳过程，是决策树的一个经典的构造算法。应用于很多预测的领域，如通过对信用卡客户数据构建分类模型，可预测下一个客户他是否属于优质客户。

**分类过程：**分类是数据挖掘、机器学习和模式识别中一个重要的研究领域。数据分类是一个两步过程。第一步，使用已知类别标记的训练数据集建立一个分类模型。例如：图 9 是一个决策树模型。第二步，对未知标记的数据使用模型进行分类。例如，根据图 9 的决策树模型，运用自顶而下的属性测试过程，将表 2 中的样例 1-6 分别分类为“Y”、“Y”、“Y”、“Y”、“N”、“N”。

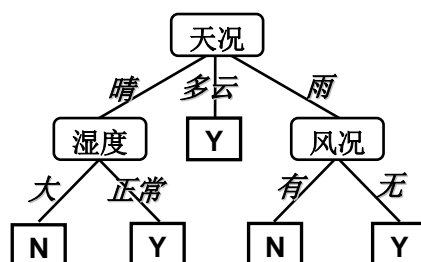


图 9 一个决策树模型的例子

决策树算法（ID3 算法）描述：

输入：训练样例集 S，未标记的节点 T，属性集 A

输出：以 T 为根的决策树

- ① 如果 S 中所有样例都是正例，则标记节点 T 为“Y”，并结束；



- ② 如果 S 中所有样例都是反例，则标记节点 T 为 “N”，并结束；
  - ③ 否则，从 A 中选择一个属性 X，（可随机选）标记节点 T 为 X；
  - ④ 设 X 的所有取值为 V1, V2, ..., Vn, 依据这些取值将 S 划分为 n 个子集 S1, S2, ..., Sn, 建 T 的 n 个孩子节点 Ti, 并分别以 Vi 作为从 T 到 Ti 的分支标号；
  - ⑤ 对每对 (Si, Ti, A-{X}), 递归调用 ID3 算法建立一棵以 Ti 为根的子树；
- END

举例：对表 1 运用算法构建决策树

表 1. 一个训练数据集

编号	天况	温度	湿度	风况	分类
1	晴	热	大	无	N
2	晴	热	大	有	N
3	多云	热	大	无	Y
4	雨	中	大	无	Y
5	雨	冷	正常	无	Y
6	雨	冷	正常	有	N
7	多云	冷	正常	有	Y
8	晴	中	大	无	N
9	晴	冷	正常	无	Y
10	雨	中	正常	无	Y
11	晴	中	正常	有	Y
12	多云	中	大	有	Y
13	多云	热	正常	无	Y
14	雨	中	大	有	N

对下列样例输入使用构建的决策树模型预测其分类属性：

表 2. 一个待分类的数据集

编号	天况	温度	湿度	风况	分类
1	晴	热	正常	无	?
2	晴	热	正常	有	?
3	雨	热	正常	无	?
4	晴	中	正常	无	?
5	晴	冷	大	有	?
6	晴	冷	大	无	?

**实现要求：**

- 1) 设计合理的数据结构，编程实现决策树构造算法；

2) 给定训练数据集，运用构建的决策树模型，设计合理的文件格式，保存于外存之中；

3) 设计决策树分类算法，根据保存在外存的决策树模型，实现决策树的分类过程，完成对未知类别属性数据样例的分类。

### 题 17：关联规则求解

**问题描述：**关联分析是数据挖掘中的一个重要任务，Apriori 算法是一种典型的关联分析算法。多用于超市的销售决策，如通过统计一段时间内，用户买商品 A 和 B 同时发生的概率，得出了顾客买 A 则很可能会买 B 的一条规则。

本课题要求对给定的训练数据，实现 Apriori 算法，构件关联规则集合。

Apriori 算法描述如下：

输入：训练样例集 S，属性集 A，支持度阈值 minsup，置信度阈值 minconf

输出：关联规则集

(1) 令  $k=1$

(2) 生成长度为 1 的频繁项集

(3) 重复以下操作直到不在生成新的频繁项集

(a) 剪除一些候选项集，它们包含非频繁的长度为  $K$  子集

(b) 从长度为  $k$  的频繁项集中生成长度为  $k+1$  的候选项集

(c) 通过扫描数据库，计算每个候选项集的支持度

(d) 除去那些非频繁的候选项集，只留下那些频繁的候选项集

END

示例：

表 3 基本概念表：关联规则的简单例子

TID	网球拍	网 球	运动鞋	羽毛球
1	1	1	1	0
2	1	1	0	0
3	1	0	0	0
4	1	0	1	0
5	0	1	1	1
6	1	1	0	0

用一个简单的例子说明。表 3 是顾客购买记录的数据库 D，包含 6 个事务。项集  $I=\{\text{网球拍, 网球, 运动鞋, 羽毛球}\}$ 。考虑关联规则（频繁二项集）：网球拍与网球，事务 1, 2, 3, 4, 6 包含网球拍，事务 1, 2, 6 同时包含网球拍和网球，支持

度  $(X \wedge Y)/D=0.5$ ，置信度  $(X \wedge Y)/X=0.6$ 。若给定最小支持度  $\alpha = 0.5$ ，最小置信度  $\beta = 0.6$ ，认为购买网球拍和购买网球之间存在关联。

**实现要求：**

- 1) 设计合理的数据结构，编程实现算法；
- 2) 给定训练数据集，设计合理的文件格式，保存于外存中；
- 3) 设计 Apriori 算法，保存关联规则在外存中。

**题 18：老鼠走迷宫**

**问题描述：**给定一个迷宫地图，迷宫中央有一只老鼠，迷宫的右下方有一个粮仓。游戏的任务是使用键盘上的方向键操纵老鼠在规定的时间内走到粮仓处。

**实现要求：**

- 1) 老鼠形象可辨认，可用键盘操纵老鼠上下左右移动；
- 2) 迷宫的墙足够结实，老鼠不能穿墙而过；
- 3) 正确检测结果，若老鼠在规定时间内走到粮仓处，提示成功，否则提示失败；
- 4) 添加编辑迷宫功能，可修改当前迷宫，修改内容：墙变路、路变墙；
- 5) 找出走出迷宫的所有路径，以及最短路径；
- 6) 利用序列化功能实现迷宫地图文件的存盘和读出等功能。

**题 19：图的简单路径**

**问题描述：**给定无向图或有向图，求解其简单路径

**实现要求：**

- 1) 用文件输入方式构建图结构，至少 5 个测试用例，要求包括有向图和无向图；
- 2) 求出图中从起点到终点的所有简单路径。其中起点和终点可以由用户自行设定；
- 3) 求出图中从起点到终点的指定长度（如 K）的所有简单路径。其中起点和终点可以由用户自行设定。

**题 20：散列表的设计与实现**

**问题描述：**设计散列表实现电话号码查找系统。

**实现要求：**

- 1) 设每个记录有下列数据项：用户名、电话号码、地址；
- 2) 从键盘输入各记录，以用户名（汉语拼音形式）为关键字建立散列表；

- 3) 采用一定的方法解决冲突;
- 4) 查找并显示给定电话号码的记录;
- 5) (选作) 设计不同的散列函数, 比较冲突率;
- 6) (选作) 在散列函数确定的前提下, 尝试各种不同类型处理冲突的方法, 考察平均查找长度的变化。

## 题 21: 英文文本统计及处理

**问题描述:** 给出一篇英文文章, 文件不小于 5M 的大小。统计其中的每个不同英文单词和总单词的数量, (含标点符号), 并实现对停用词(给定)和低频率词(出现次数小于某个参数, 如小于 3)的过滤。

**实现要求:**

- 1) 分别用链表和哈希表来实现, 注意要给出不同大小文件耗费的时间, 对时间性能进行进一步分析;
- 2) 关于英文文章, 可以自动生成文本文件, 也可以从网络上下载几篇英文文章。

## 题 22: 本科生导师制问题

**问题描述:** 在高校的教学改革中, 有很多学校实行了本科生导师制。一个班级的学生被分给几个老师, 每个老师带领  $n$  个学生, 如果老师还带研究生, 那么研究生也可直接负责本科生。

本科生导师制问题中的数据元素具有如下形式:

- (1) 导师带研究生: (老师, ((研究生 1, (本科生 1, ..., 本科生  $m$ )), ...))
- (2) 导师不带研究生: (老师, (本科生 1, ..., 本科生  $m$ ))

导师的自然情况只包括姓名、职称;

研究生的自然情况只包括姓名、班级;

本科生的自然情况只包括姓名、班级。

**实现要求:**

- 1) 插入: 将某位本科生或研究生插入到广义表的相应位置;
- 2) 删除: 将某本科生或研究生从广义表中删除;
- 3) 查询: 查询导师、本科生(研究生)的情况;
- 4) 统计: 某导师带了多少个研究生和本科生;
- 5) 输出: 将某导师所带学生情况输出。

### 题 23: 矩阵运算工具包

**问题描述:** 矩阵运算工具包

**实现要求:**

- 1) 设计一个矩阵类, 实现矩阵的基本运算, 包括加、减、乘、点乘、连乘、逆置、秩、90 度、180 度、270 度的旋转及其他运算;
- 2) (选作) 给定一个小矩阵, 用小矩阵在大矩阵上进行卷积运算, 求出大矩阵的卷积结果。

### 题 24: 特殊矩阵工具包

**问题描述:** 设有两个矩阵  $A=(a_{ij})_{m \times n}$ ,  $B=(b_{ij})_{p \times q}$ 。

**实现要求:**

- 1) 设计一个矩阵类, 矩阵包含稀疏矩阵和三角矩阵;
- 2) 实现矩阵的基本运算, 包括加、减、乘、点乘、连乘、逆置、秩、90 度、180 度、270 度的旋转及其他运算;
- 3) (选作) 给定一个小矩阵, 用小矩阵在大矩阵上进行卷积运算, 求出大矩阵的卷积结果。

**参考:**

- 1) 求矩阵的转置, 矩阵的转置  $A'=(a_{ji})_{n \times m}$ , 转置前输出原矩阵, 转置后输出转置矩阵;

- 2) 求矩阵 A、B 的和。矩阵 A 和 B 能够相加的条件是:  $m=p$ ,  $n=q$ ; 矩阵 A 和 B 如果不能相加, 请给出提示信息; 若能够相加, 则求和矩阵 C 并输出 C;

$$C=A+B=(c_{ij})_{m \times n}, \text{ 其中 } c_{ij}=a_{ij}+b_{ij}$$

- 3) 求矩阵 A、B 的差。矩阵 A 和 B 能够相减的条件是:  $m=p$ ,  $n=q$ ; 矩阵 A 和 B 如果不能相减, 请给出提示信息; 若能够相减, 则求差矩阵 C 并输出 C;

$$C=A-B=(c_{ij})_{m \times n}, \text{ 其中 } c_{ij}=a_{ij}-b_{ij}$$

- 4) 求矩阵 A、B 的积。矩阵 A 和 B 能够相乘的条件是:  $p=n$ ; 矩阵 A 和 B 如果不能相乘, 请给出提示信息; 若能够相乘, 则求积矩阵 D 并输出 D;

$$D=A \times B=(d_{ij})_{m \times q}, \text{ 其中 } d_{ij}=\sum a_{ik} \times b_{kj}, k=1, 2, \dots, n$$

### 题 25: 城市管理

**问题描述:** 自定义一个城市数据库, 每条数据库记录包括城市名 (任意长的字符串) 和城市的坐标 (用整数 x 和 y 表示)。实现数据的插入、删除、查询功能, 并实现指定距离内的所有城市。设计算法实现指定一定数目的具体城市, 寻找遍历这些城市并回到出发点的最佳路径, 观察随着城市数目的增加, 算法执行

效率的变化。

### 编程任务：

- 1) 用列表对城市进行记录和管理，实现城市的增加、删除和查询功能，并实现文件保存和读取
- 2) 计算城市之间距离，统计输出距离某城市一定范围内的所有城市。
- 3) 实现一定规模城市的遍历最佳路径选择。
- 4) 分析随着城市数目增加时，算法执行效果的改变，深刻理解旅行商问题。

## 题 26：数字图像处理

**问题描述：**一幅图像是包含位置集和颜色集的数据。考虑二维灰度图像，位置集就是一个矩阵的行列，矩阵的内容为颜色值，颜色为  $0 \sim 255$  间的整数，表示该位置的灰度等级，0 为黑色，255 为白色。

图像处理就是与该矩阵相关的计算，一种常见的计算就是通过一点和周围 8 个点的信息，共同决定该点的新值：如一点的新值为该点和周围 8 点颜色之和的平均，这一操作可用图 10 表示。

$1/9$	$1/9$	$1/9$
$1/9$	$1/9$	$1/9$
$1/9$	$1/9$	$1/9$

图 10 平滑操作示意图

这样处理后图像会变得平滑，因此，称为平滑操作。

如果将上述操作变为图 11：

-1	-1	-1
-1	9	-1
-1	-1	-1

图 11 锐化操作后示意图

操作后图像的边缘变得更加突出，被称为锐化操作。

实现上述图像的平滑和锐化操作。

### 实现任务：

- 1) 常见格式图像的读写(灰度图)；
- 2) 设计并实现上述平滑算子和锐化算子；
- 3) (选作) 查阅其他的平滑和锐化算子并实现；
- 4) 可视化展示处理前后图像的变化，并尝试进行分析。

## 题 27：电子字典

**问题描述：**设计一个中英文字典查询系统。如给出中文“苹果”，能查询出其对应的英文单词“apple”。

**实现要求：**

- 1) 自建系统的词库和存储结构，以方便更好的进行查询；
- 2) 该系统在一定的权限控制下具有基本的查询、增加、删除、修改功能，并尽可能的保证良好的时间开销；
  - (1) 查询操作时应考虑精确查询和模糊查询；
  - (2) 增加操作时：该程序输入一个英语单词，若该单词不存在，可向词库添加该词和释义（应考虑一个单词可以有多个释义）；
- 3) 实现输入输出交互界面。

## 题 28：人群划分问题

**问题描述：**构建一个人际关系网，并根据人与人之间的关系紧密度进行图的划分，将人按照其紧密度划分成几个群。（数据集下载地址：<https://blog.csdn.net/wzgang123/article/details/51089521>）

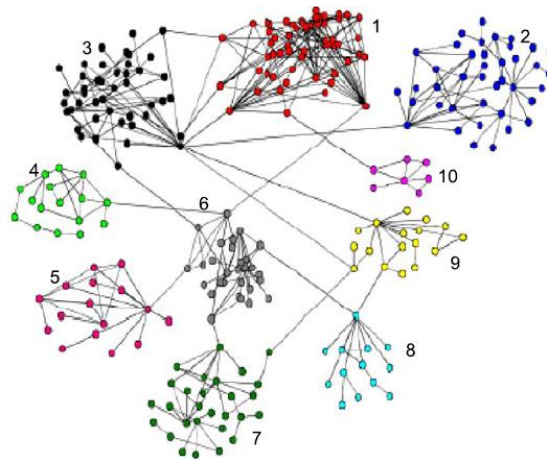


图 12 人群关系划分示意

**实现要求：**

- 1) 关系网中的人数不少于 50 人；
- 2) 关系紧密度可用[0-1]间的数值进行表示，也可通过两种方式进行计算：
  - a) 通过人的相关信息（如性别，年龄，籍贯、教育程度、职业等）进行计算；
  - b) 通过两个节点的共同邻居个数来衡量两个节点间的关系紧密度；
- 3) 可指定群的数目，也可算法自动计算；
- 4)（选作）查阅图划分或社团划分相关资料，并尝试对其中某一算法进行实

现。

### 题 29：图结构的教学演示

**问题描述：**对图模型进行教学演示，包括多种存储结构，图结构的基本算法（遍历算法，最小生成树，最短路径，拓扑排序等），并给出其演示过程。

**实现要求：**

- 1) 动态演示算法求解过程，便于学习；
- 2) (选作) 选择其中一种算法进行改进。

### 题 30：排序算法的教学演示

**问题描述：**多种排序算法的对比

**实现要求：**

- 1) 实现多种（不少于 6 种）排序算法；
- 2) 对比分析其时间空间性能；
- 3) 动态演示其排序的过程，便于学习；
- 4) (选作) 选择其中一种算法进行改进。

### 题 31：中文文本的实体统计。

**问题描述：**给定一本电子书（如三国演义），统计该书中人物的出场次数，考虑电子书中的人物及这些人物的别号或绰号为同一人，如曹操，曹孟德。

**实现要求：**

- 1) 输出：一共有多少人物，各个人物的出场次数；
- 2) (选作) 对指代词如“公”，“汝”等所指代的对应任务进行计数。

### 题 32：校园导航

**问题描述：**给定校园内 10 个以上的地点及路径图，给出实时导航推荐路径。需给出自行车和步行两种方案。

**实现要求：**

- 1) 输入出发地址和目的地址，求解出其最短路径并推荐，并估计其大概需要花费的时间；
- 2) 对校园道路考虑步行道和车行道，因为道路推荐考虑步行和车行两种路径；
- 3) 地图展示其导航路径。



### 题 33：饭店推荐

**问题描述：**学校周围有一些饭店提供外卖，综合考虑多种因素如口味、距离、类型、价格等，设计并实现一个查询推荐系统。

**实现要求：**

- 1) 饭店数量在 40 家以上；
- 2) 涉及的考虑因素包括饭店的地理位置坐标（用于计算距离），饭店主营类型（用于筛选），供餐速度，口味评价等；
- 3) 用户输入类型，根据距离，供餐速度及口味等因素对饭店进行排序推荐。

### 题 34：平衡二叉树的构建与查找算法教学演示

**问题描述：**设计实现基于平衡二叉树的查找算法与演示系统；

**实现要求：**

- 1) 平衡二叉树的建立：要求用户输入数据（可界面输入或者文件输入），边建立二叉树边调平衡，可通过手动/自动演示两种方法图形化展示算法过程；
- 2) 用户输入查找数据，界面动态演示基于平衡二叉树的查找过程。

### 题 35：B-树的构建与查找算法教学演示

**问题描述：**设计实现基于 B-树的查找算法与演示系统；

**实现要求：**

- 1) B-树的构建：要求用户输入数据（可界面输入或者文件输入），可通过手动/自动演示两种方法图形化展示算法过程；
- 2) 用户输入查找数据，界面动态演示基于 B-树的查找过程。

### 题 36：查找算法的教学演示

**问题描述：**设计实现若干查找算法及其演示系统。

**实现要求：**

- 1) 实现多种查找算法（顺序表、树表、散列表三种数据类型不小于 2 类），并进行对比分析；
- 2) 用户输入查找数据，界面动态演示基于查找过程的图形化显示。

### 题 37：线索二叉树的教学演示

**问题描述：**设计实现线索二叉树的相关运算及其演示系统。

**实现要求：**

- 1)实现动态存储方式的线索二叉树类;
- 2)要求能够从文件读取数据构建线索二叉树, 进行遍历、节点的插入和删除, 并保存到文件中;
- 3)任意构建二叉树, 实现 3 种线索化操作, 及其前驱与后继的查找;
- 4)实现线索二叉树的图形用户界面操作。

### 题 38: 森林的教学演示

**问题描述:** 设计实现基于森林的应用及其演示系统。

**实现要求:**

- 1) 实现静态与动态存储方式的森林类;
- 2) 要求能够从文件读取数据构建森林, 进行森林到二叉树的转化, 并保存到文件中;
- 3) 要求根据所建森林查找父子对并输出, 同时输出森林中每个节点的层次和森林的高度;
- 4) 实现森林的图形用户界面操作。

### 题 39: 教学计划排课系统

**问题描述:** 假设任何专业都有固定的学习年限, 每学年含两学期, 每学期的时间长度和学分上限均相等。每个专业开设的课程都是确定的, 且课程在开设时间的安排必须满足先修关系。每门课程的先修课程是确定的(可自行设计), 可有任意多门, 也可以没有。每门课恰好占一个学期。试根据此问题描述, 设计一个教学计划编制程序, 完成如下功能:

**实现要求:**

- 1) 输入参数包括: 学期总数, 一学期的学分上限, 每门课的课程号(固定占 8 位的字母数字串)、学分和直接先修课程的课程号; (数据课程教师提供)
- 2) 自行设计编排策略, 每个学期的学分尽量均衡;
- 3) 利用拓扑排序求出所有的拓扑序列, 查看课程之间的拓扑序列; 若无解, 可反馈教学计划编制出错信息;
- 4) 输出任意一个学期的课程信息, 包括课程名称、课程与学分;
- 5) 按照课程查询任一门课程所属的学期;
- 6) 采用图形用户界面实现。

#### 题 40：最长公共子序列

**问题描述：**问题求解 2 个字符串的最长公共子序列。输入的 2 个字符串可以从键盘读入，也可以从两个文本文件中读入。比如字符串 1：BDCABA；字符串 2：ABCBADAB，则这两个字符串的最长公共子序列长度为 4，最长公共子序列是：BCBA。

**实现要求：**

- 1) 采用至少两种算法设计实现；
- 2) 通过设定多个测试用例，分析算法的时间复杂度和空间复杂度；