**数据结构课程设计任务书**

**适用专业： 计算机科学与技术、物联网**

**适用年级：**  **2014级**

**华中科技大学计算机科学与技术学院**

**数据结构课程组**

**2015年12月**

**数据结构课程设计任务书**

数据结构是计算机科学技术、信息安全与物联网等专业的一门重要专业基础课，牢固掌握数据结构的基础知识，熟练地运用数据结构的思想与技术方法解决实际应用问题是是本课程学习的基本任务与目标。而课程设计是实现这一学习目标的重要环节和组成部分。通过课程设计的训练，使学生加深对数据结构知识的理解，牢固掌握其应用方法，并合理灵活地解决一定实际问题，增强和提高综合分析问题与解决问题的能力。

本学期课程设计提供五个实际问题或综合实验专题，同学们可根据个人的兴趣偏好，选择其中一个作为自己数据结构课程设计的题目。

**题目一 哈夫曼编码与文件压缩**

* **设计目的：**掌握二叉树、哈夫曼树的概念，性质与存储结构，能够利用哈夫曼算法实现哈夫曼编码，并应用于文件压缩，从而提高学生综合运用知识的技能与实践能力。
* **设计内容：**分析与设计哈夫曼树的存储结构，实现哈夫曼算法以及编码与译码基本功能，并对任意文本文件利用哈夫曼编码进行压缩得到压缩文件，然后进行解压缩得到解压文件。有兴趣的同学可以查阅资料实现Lempel-Ziv sliding window压缩方法，并与之比较。
* **设计要求：**

（1）要求界面友好，输入文本文件可带路径（如：D:\doc\original.txt），哈夫曼算法所得到的压缩文件名为\*.cod，哈夫曼树也以文件形式保存，文件名为\*.hfm。

（2）显示压缩比、压缩时间、解压时间与对应的编码表。

* **设计提示：**统计文本文件中各字符的频度并作为权值生成哈夫曼树，并利用哈夫曼树进行二进制编码。
* **参考文献：**

[1] 严蔚敏, 吴伟民. 数据结构（C语言版）. 北京: 清华大学出版社,1997

[2] 王晓东. 计算机算法设计与分析. 北京: 电子工业出版社, 2007

[3] 严蔚敏, 吴伟民, 米宁. 数据结构题集（C语言版）. 北京: 清华大学出版社,1999

**题目二 基于堆的优先级队列ADT实现及其应用**

* **设计目的：**优先级队列在计算机系统与技术中应用广泛而深入。本设计使学生掌握堆作为优先级队列的数据表示方法，物理存储结构与相关算法，从而为后续学习，研究与实践奠定良好的数据结构基础，提高数据结构方法与技术的应用实践能力。
* **设计内容：**传统队列是一种符合先插入的元素必须先删除（FIFO）的处理逻辑，这不总是满足应用要求；很多时候需要优先级高的任务先处理（即后插入的可能先删除）。**（1）**基于堆的概念设计优先级队列(Priority Queue)抽象数据类型，至少包含Init\_PriorityQue, Destroy\_PriorityQue, Clear\_PriorityQue，PriorityQue\_Insert, PriorityQue\_DeletMin, PriorityQue\_Empty, PriorityQue\_Full等操作；**（2）**选择适当的物理存储结构实现优先级队列ADT; **(3)**应用优先级队列ADT设计与实现一个医院门诊医师与病人看诊服务事件仿真程序，使医师服务效率尽量高。
* **设计要求：**

（1）仿真事件（如病人到达，病情复杂度/就诊时间，病人离开等）可根据某种概率分布或随机模型生成。

（2）要求对各种算法进行理论分析，同时也对实测结果进行统计分析。测试数据要求有一定规模。

（3）要求界面整洁、美观，操作方便。

* **参考文献：**

[1] 严蔚敏, 吴伟民. 数据结构（C语言版）. 北京: 清华大学出版社,1997

[2] 严蔚敏, 吴伟民, 米宁. 数据结构题集（C语言版）. 北京: 清华大学出版社,1999

[3] Mark Allen Weiss.Data Structures and Algorithm Analysis in C, 机械工业出版社，2010, 177-192

**题目三 武昌地区公交查询与换乘推荐**

* **设计目的：**掌握图、查找、排序等数据结构的物理存储结构与基本算法，通过解决较复杂的基于图模型的实际问题，提高学生对数据结构知识综合运用的技能与实践能力。
* **设计内容：**以经停武汉市武昌地区的公交线路及地理信息为素材，基于需求分析，设计有效的数据结构及其存储结构表示武昌地区公交线路网，其中含各条线路信息、站点信息；设计有效的算法对线路按某关键字排序、对线路进行典型查询、以及基于已知起点与终点进行换乘推荐。
* **设计要求：**

（1）从互联网或相关资料获取可靠的武汉公交线路及其地理数据，通过线性结构与图模型对其进行表示，且以文件保存。

（2）图形方式显示上述图模型与求解结果。

（3）界面友好，实现的功能包括：录入与修改公交线路信息；查询所有线路信息（线路名/号、起点、终点、首末车时间、票价规则），按线路名或起点站名排序；查询指定线路的详情（沿途站点、首末车时间、票价规则、站间距离等）；查询某一位置途经的所有公交线路、指定起点与终点，推荐乘车方案（如要求换乘次数最少、路线最短或无要求条件等）。

* **设计提示：**以图作为主要数据结构，其中顶点表示站点、边表示两个站点有公交直达，邻接矩阵作为主要存储结构，边的权值与信息要考虑两站点间的距离及全部运营线路；求解最短路径的算法可以引入解决换乘推荐问题。一条公交线路可表示为由站点构成的线性表。设计与实现时可对线路信息与软件功能进行自由扩展。有兴趣的同学还可以用直观图形方式显示公交线路网，图形要素包括站点（stop point）与形点(shape point)，形点序列勾画出线路的地理形式。
* **参考文献：**

[1] 严蔚敏, 吴伟民. 数据结构（C语言版）. 北京: 清华大学出版社,1997

[2] 王晓东. 计算机算法设计与分析. 北京: 电子工业出版社, 2007

[3] 严蔚敏, 吴伟民, 米宁. 数据结构题集（C语言版）. 北京: 清华大学出版社,1999

[4] URL: <http://wuhan.8684.cn>

[5] URL: http://wh.bus.58.com

**题目四C语言代码复制/相似度检测软件**

* **设计目的：**巩固数据结构基本知识，特别是串的基本操作与模式匹配算法等，利用串的处理解决较复杂的实际问题，提高学生对数据结构知识综合运用的技能与实践能力。
* **设计内容：**设计有效的存储结构与处理算法对两个C语言代码源文件进行分析、比较，判定代码的重复性、涉嫌复制的程度或相似度，给出判定的定性或量化分析结果。
* **设计要求：**

（1）C语言代码以\*.c文本文件形式存储与提供，在分析处理时采用串的形式表示。

（2）主要算法处理要利用串的基本操作，包括串的模式匹配操作。

（3）根据所用检测技术的不同，在对串的处理基础上可以使用数据结构中其他如树、Hash表等内容。

（4）界面友好，便于指定输入文件、能够显示结果与一定的重要分析细节。

* **设计提示：**在读取C源文件之后，基于串的操作对代码串进行分割、删除多余的非特征子串、转换等。代码相似度检测技术已有一定的研究，有一些不同的方法。查阅资料，了解典型解决方法。最后设计主要以串匹配为基础的解决方案。
* **参考文献：**

[1] 严蔚敏, 吴伟民. 数据结构（C语言版）. 北京: 清华大学出版社,1997

[2] 王晓东. 计算机算法设计与分析. 北京: 电子工业出版社, 2007

[3] 邓爱萍,徐国梁,肖奔.程序源代码剽窃检测串匹配算法的研究.计算机工程与科学，2008,30(3):62~68

[4] 邓爱萍,徐国梁,肖奔.基于串匹配方法的源代码复制检测技术研究.科学技术与工程，2007,7(10):2251~2254

[5] 程金宏.程序代码相似度度量研究：[硕士学位论文].内蒙古师范大学,2007

[6] 侯敏，刘东升.基于串匹配的源代码抄袭检测技术研究.电脑编程技巧与维护，2011,(24):6~9

**题目五B树的维护**、**查询及图形仿真**

* **设计目的：**牢固掌握B-树结构及其基本算法，通过实现图形方式对B-树操作的动态模拟，提高学生实践能力。
* **设计内容：**设计与实现B-树存储结构，以及B-树构造、插入、删除、查找等操作，并以图形方式模拟这一动态过程。
* **设计要求：**

（1）采用图形用户界面，且要求用图形方式显示B-树操作的动态过程。

（2）数据元素的值可以从终端输入，操作结果以文件形式保存。

（3）测量操作执行的时间性能。

* **设计提示：**B-树的阶设置为参数，可自行确定。B-树对范围查询支持不够，可设计与实现基于B-树的范围查询扩展功能。
* **参考文献：**

[1] 严蔚敏, 吴伟民. 数据结构（C语言版）. 北京: 清华大学出版社,1997

[2] 王晓东. 计算机算法设计与分析. 北京: 电子工业出版社, 2007

[3] 严蔚敏, 吴伟民, 米宁. 数据结构题集（C语言版）. 北京: 清华大学出版社,1999

**课程设计要求**

1. 五个题目任选其一，计算机科学与技术专业，假期独立完成，下学期检查程序。(注意：所用的数据结构与算法应与课程紧密相关；抄袭他人程序者记零分。)
2. 各题目的难度系数

（1）哈夫曼编码与文件压缩 0.85 (注：综合成绩不超过85分)

（2）基于堆的优先级队列ADT实现及其应用 1.05

（3）武昌地区公交查询与换乘推荐 1.0

（4）C语言代码复制/相似度检测软件 0.95

（5）B树的维护、查询及图形仿真 1.0

3. 按华中科技大学计算机科学与技术学院制定的课程设计报告文本结构及写作规范撰写课程设计报告，报告主要内容应至少涵盖如下方面(以下非报告目录)。

一、问题描述与技术现状分析

二、系统总体设计(含模块结构图)

三、数据结构和算法详细设计

四、C语言程序实现的简要说明

(开发环境、支持包、函数原型与功能及调用关系；全部源程序以电子版提供)

五、程序测试及结果分析

六、复杂度分析

七、软件的用户使用说明

八、特色与不足

九、主要参考文献

4. 报告提交方式

（1）打印《课程设计报告》(A4纸双面打印)；

（2）以班为单位提交光盘(至少含课程设计报告、源程序和脱离开发环境独立运行的目标程序)；

（3）计算机科学与技术专业按课表上课程设计结束2-3周后提交。

**附：**

**数据结构课程设计指导用书**

——数据结构题集（C语言版），严蔚敏等编著，清华大学出版社

**数据结构课程设计参考书**

——数据结构（C语言版），严蔚敏、吴伟民编著，清华大学出版社