课 程 设 计 任 务 书

**《数据结构》大作业**

目的任务

数据结构大作业是对软件设计的综合训练，包括问题分析、总体设计、用户界面设计、程序设计基本技能和技巧，多人合作，以至一套软件工作规范的训练和科学作风的培养。在数据结构实验中，完成的只是单一而"小"的算法，而本课程设计是对学生的整体编程能力的锻炼。数据结构大作业的目的是训练学生对问题的抽象能力和算法的运用能力。

设计内容

大作业由2~4**名**同学组队完成，可从备选题目中选择一个题目来完成或者自选题目。原则上鼓励自选题目。

**【备选题目】**

从下面的备选题目中选择一个或者自拟题目。请发挥你的想象力，选用合适的数据结构与算法。需要理解的是，数据结构大作业重点考核的是数据结构和算法的运用；好的界面呈现可以锦上添花，但并不是决定因素。

**1. 航空客运订票系统（难度★★）**

航线信息：终点站名、航班号、飞机号、飞行日（星期几）、乘员定额、余票量、已订票的客户名单（包括姓名、订票量、舱位等级1，2或3）以及等候替补的客户名单（包括姓名、所需票量）。

系统实现如下功能：

查询：根据旅客提出的终点站名输出信息（例如航班号、飞机号、星期几飞行，最近一天航班的日期和余票额等）；

订票：根据客户提出的要求（航班号、订票数额）查询该航班票额情况，若尚有余票，则为客户办理订票手续；若已满员或剩余票额少于定票额，则需重新询问客户要求。若无票，可登记排队候补或提供相关可选择航班；

退票：根据客户提供的情况（日期、航班），为客户办理退票手续，然后查询该航班是否有人排队候补，首先询问排在第一的客户，若所退票额能满足他的要求，则为他办理订票手续，否则依次询问其他排队候补的客户；

修改航班信息：当航班信息改变可以修改航班数据文件。

测试要求：

构造大量的测试数据，充分展示查询性能。

**简易电子表格（难度★★）**

设计一个支持基本计算与统计功能和其他一些表格管理/处理功能的计算机软件，使用户可在该软件的支持下，用交互方式进行表格建立、数据输入、数据编辑、统计、计算及其他一些表格操作。

实现功能：

按表格形式显示表格，并支持用户使用简单的功能键（按键选择式的简单菜单）进行操作；

建立空白表格，同时在屏幕上显示，使其处于可输入数据状态；

通过键盘将数据输入在屏幕上的电子表格上，同时要支持基本的数据输入编辑；

将表格中任意数据块复制到另一块中。复制到目标块时，对目标块中原内容的处理，可选择的方式有：代替、相加、相减、按条件替换；

支持鼠标操作各项功能；

支持汉字显示、输入；

单元格内可输入公式（表达式），使对应单元格的最终内容为公式的计算结果；

统计计算：包括合计、求平均、求最大/小；

统计计算方式：表格按行/列统计计算，表格按块统计计算；

排序：使任一行/列中的数据按大小（升或降）排列，对字符串型数据，还要可选大小写敏感；

表格保存：使电子表格存储在磁盘上（磁盘文件），并可随时读入，供继续处理。

**教务管理系统 （难度★★）**

根据学校的需求，设计一个教务管理系统，提供给师生使用。该系统有以下要求：

**目标用户**：学生，教师，教务管理员。

**软件配置**：要求生成3个可执行文件，分别提供给学生、教师和教务管理员使用，以下分别称为学生端、教师端和教务员端软件（三个软件都在同一台计算机上运行，暂不考虑网络，以后学到网络，可对该软件进行扩展）。采用文件来保存学生、教师、课程等记录，学生端、教师端和教务员端共同访问这些记录文件。

**功能要求**：下面列举的功能有些是必须实现，有些则是根据自己的能力实现的扩展功能。

F1. 教务员端的功能

F1.1（必须实现）管理学生登录帐号和教师登录帐号。

F1.2（必须实现）增加/删除/修改学生信息、班级信息、教师信息、课程信息。

F1.3（必须实现）支持对学生、课程信息、教师信息的快速查找。数据量较大，要**建立索引**以支持对文件的高效查找。可以根据学生名、学号、课程名、教师名等进行**多关键字组合查找**。

F1.4（扩展功能）支持课程信息的部分匹配查找，比如输入"数据"能找到"数据结构"这个课程的信息。

F1.5（扩展功能）实现课程成绩的统计、排序，按课程或班级或学生生成成绩表。

F1.6（必须实现）退出保存功能。当该软件被关闭，再打开时，数据没有丢失。

F2. 教师端的功能

F2.1（必须实现）根据自己的帐号和密码进行登录。登录出错应提示。

F2.2（扩展功能）发布作业。

F2.3（扩展功能）检查作业。统计交作业的情况，提交学生作业的成绩。

F2.4（扩展功能）自动检查抄袭。比较作业目录下的作业文件，如果有任何两个文件的相似度超过80%，则自动判定为抄袭，向教师提示。

F2.5（必须实现）提交课程成绩。成绩由平时分和期末分的平均得出，其中平时分和期末分由教师输入，计算总评时的比例由老师设置（比如平时分占30％，期末分占70％）。在教师输入了课程成绩之后，教务员端应该能查找到，学生端也能查找到自己的成绩。

实现提示：在输入成绩后，该成绩被记录到文件中，以供查找。

F2.6（必须实现）退出保存功能。当该软件被关闭，再打开时，数据没有丢失。

F3. 学生端的功能：

F3.1（必须实现）根据自己的学号和密码进行登录。登录出错应提示。

F3.2（扩展功能）显示当前的作业。当教师发布作业时，要求界面上马上提示有作业的任务。学生可以打开查看。

实现提示：可以考虑将教师发布的作业信息记录在文件中，学生端周期性的去检查文件的变化，如果有变化，则在界面上提示。

F3.3（扩展功能）提交作业。学生可以选择某个作业文件进行提交。提交之后，该文件被拷贝到某个作业目录下，以供教师端进行作业检查。

F3.4（必须实现）查看自己的课程成绩。

F3.5（必须实现）退出保存功能。当该软件被关闭，再打开时，数据没有丢失。

**测试要求：**

C1: 构造20000个学生，2000个教师，200门课程测试数据，用于测试系统的功能。提示：可以写一个小程序来产生一些随机的数据.

C2: 测试环境：在同一台机器上，同时运行教务员端、教师端和学生端。

**家谱管理系统（难度★★）**

实现功能：

家庭成员信息存储：基本信息至少应包括姓名，出生地，出生日期，死亡日期，性别，身高，学历，职业，最高职务/职称等；

家族关系存储：将各家庭成员之间的关系，存储在计算机中（可永久保存）；

家谱数据的更新：修改、删除、加入；

将家谱以较友好的格式输出（显示）；

按基本信息查询成员，按亲戚关系查询；

统计：平均寿命、平均身高、男女比例、家庭平均人口、平均（最高/低）学历等；

屏幕显示家谱树形结构（类似Windows 目录）；

支持鼠标操作。

**人机对弈游戏（难度★★）**

开发一款人机对弈的游戏，如井字棋、五子棋、象棋等。

基本要求：

实现基本的人机交互界面，建议采用图形化界面显示棋盘和棋子。

电脑能够自动根据人下的棋，判断下一步较优的选择。

实现提示：

每下一步棋，都会形成新的棋局，不同的下法，会衍生出不同的棋局。从而在下棋的过程中，当玩家下了一步棋后，电脑对当前的棋局的推断过程如下：电脑会判断有哪些地方电脑是可以下的，如果下了某个棋后，将会形成新的棋局，在这个新的棋局中，人又有可能在哪里下棋，如果人下了某部棋后，又会形成新的棋局，然后接着电脑又可能在那里下…。我们可以构建以棋局为结点的树，棋局之间的衍生关系对应为树上的连边关系，借助棋局树，寻找最优的下一步棋。

**导航系统（难度★★★）**

车载导航系统是在城市中驾车的好帮手，不仅能够计算出到达目的地的最优路线，而且可以显示当前位置附近的一些附加信息，比如附近的加油站，餐馆等等。请设计一款导航软件，实现导航系统的核心功能。

**目标用户**：汽车司机。

**数据配置**：导航系统中的地图可以抽象为一个图，地点信息和路径信息可以抽象为图中的顶点和边。请设计算法，来产生模拟的地图数据：

随机产生N个二维平面上的顶点（N>=10000），每个顶点对应地图中的一个地点

对于每个地点x，随机建立若干个连边到x附近的地点。每条连边代表一条路径，路径的长度等于边的两个顶点之间的二维坐标距离。

模拟数据必须保证，**产生的图是一个连通图**。

**功能要求**：

F1. （必须实现）地图显示功能。输入一个坐标，显示距离该坐标最近的100个顶点以及相关联的边。关于如何用vc++画图，或java画图，请在baidu或google上查找相关方法，或者借阅相关的参考书。

F2. （必须实现）地图缩放功能。提示：地图缩的越小，屏幕上显示的点数就越多，但是太多的点，会看不清楚。所以可以考虑只选择一个单元区域内只显示一个代表的点。

F2. （必须实现）任意指定两个地点A和B，能计算出A到B的最短路径。并将该最短路径经过的顶点以及连边显示出来。

F3. (扩展实现) 模拟车流。请为每条连边增加两个属性：车容量v（饱和状态下这条路所能容纳的汽车的数量）、当前在这条路上的车辆数目n（n>v时为超负荷运作）。假设该路的长度为L，则该路的通行时间可模拟为cLf(n/v)，其中c是常数；f(x)是一个分段函数, x小于等于某个常数时，f(x) = 1，当x大于该常数时，f(x) = 1+ex。每条道路的车容量v和道路的长度L为预先指定的固定参数。请模拟产生汽车在地图中行驶，为简化模型假设在同一条路上每架汽车穿越该路的时间均等于cLf(n/v)。要求实现模拟车流的动态变化，任意时刻，给定一个坐标，能在界面上显示该坐标附近的所有路径，并动态显示各个路径上的车流量的大小（可用不同颜色或其他方法区分车流量的大小级别）。

F4. (扩展实现)综合考虑路况的最短路径。任意时刻，指定两个地点A和B，能根据当前的路况，计算出从A到B最短行车时间，以及相应的最佳路径，并在界面上，将该最短路径经过的顶点以及连边显示出来。

**科学文献管理系统（难度★★★★）**

科研工作者的日常工作离不开查阅科学文献，并对其中的信息进行分析、筛选、挖掘和管理。请你为科研工作者设计一个管理系统，提高科学文献的管理效率。

**目标用户**：科研工作者。

**数据配置**：请通过以下方法下载数据文件dblp.xml.gz.

<http://dblp.uni-trier.de/xml/dblp.xml.gz>

将该数据文件解压后，其中包含一个dblp.xml文件。该文件由科学文献的记录序列组成，记录的格式如下所示：

<article mdate="2002-01-03" key="persons/Codd71a">

<author>E. F. Codd</author>

<title>Further Normalization of the Data Base Relational Model.</title>

<journal>IBM Research Report, San Jose, California</journal>

<volume>RJ909</volume>

<month>August</month>

<year>1971</year>

<cdrom>ibmTR/rj909.pdf</cdrom>

<ee>db/labs/ibm/RJ909.html</ee>

</article>

每个记录对应一篇文章，其中包含对作者，题名，发表杂志，卷号，出版时间等的详细说明。请基于该数据，设计能满足后述功能的文献管理系统。

**注：1）dblp.xml的大小超过1G，所以不要直接点击打开该文件。需要通过命令行命令’more’ 或者自行编程查看。**

**2） 编程过程中，不允许使用数据库系统。需要自己建立管理文献数据的数据结构。**

**功能要求**：

F1. （必须实现） 基本搜索功能。输入作者名，能展示该作者发表的所有论文信息。输入完整的论文的题目，能展示该论文的其他相关信息

F2. （必须实现） 相关搜索。输入作者名，能展示于该作者有合作关系的其他所以作者。

F2. （必须实现） 作者统计功能。输出写文章最多的前100名作者。

F3. (扩展实现) 热点分析功能。分析每一年发表的文章中，题目所包含的单词中，出现频率排名前10的关键词。

F4. (扩展实现) 部分匹配搜索功能。给定若干个关键字，能快速搜索到题目中包含该关键字的文章信息

F5. (扩展实现) 聚团分析。作者之间的合作关系可以看成是一个图，每个作者对应一个顶点，任两个作者之间如果存在合作关系，则在两个顶点之间建立连边。这个图中的每一个完全子图我们称为一个聚团（所谓完全子图指的是该子图的任意顶点都和该子图的其他顶点有连边，完全子图的顶点个数称为该完全子图的阶数），请统计整个图中各阶完全子图的个数。

F6. (扩展实现) 可视化显示。通过图形化界面，展示作者之间合作关系图及其相关文章信息。

**出租车轨迹分析（难度★★★★★）**

出租车是城市中主要出行交通工具之一。通过分析城市中的出租车的轨迹，可以获得大量有意义的城市动态信息，包括城市中人群的移动规律、城市交通状况、城市区域的繁华程度、城市区域之间的关联关系、旅游路线推荐等等。

目前，由于车载GPS设备的广泛使用，以及具有海量信息存储能力的数据中心的推行，采集、存储并实时处理出租车的轨迹已经成为现实。如何对这些轨迹进行有效的管理、分析与挖掘，是一个不仅富有意义而且富有挑战的任务。请根据下面给定的出租车轨迹数据，进行相应的分析以及可视化的展示。

**数据配置**：

请通过以下方法下载数据文件01.zip – 014.zip

<http://research.microsoft.com/apps/pubs/?id=152883>

将该数据文件解压后，得到10375个txt文件。文件记录了北京市10,000辆出租车1星期的GPS轨迹数据。记录的格式如下所示：

1,2008-02-02 15:36:08,116.51172,39.92123

1,2008-02-02 15:46:08,116.51135,39.93883

1,2008-02-02 15:46:08,116.51135,39.93883

…

每个记录对应一个时空点，其中包含用户标识Id,时间Time，经度longitude，纬度latitude。请基于该数据，进行出租车轨迹的分析。

**功能要求**：

F1. （必须实现）出租车轨迹可视化。显示所有或者某部出租车的轨迹。显示的方式可以是在编程绘制的窗口中画出地图并显示出租车的轨迹点；也可以通过调用google地图或者baidu地图的API，在google地图或者baidu地图中显示。请在baidu或google上查找相关方法，或者借阅相关的参考书。

F2. （扩展实现）地图缩放功能。可以对地图进行放大或者缩小，并相应的调整出租车轨迹的展示。

F3. （必须实现）区域范围查找。统计在某个特定的时间段，用户指定的矩形区域内的出租车的数目。这里的矩形区域可以通过给出矩形的左上角和右下角的经纬度坐标来确定。

F4. (扩展实现) 区域车流密度分析。给定距离参数r,将整个地图划分成网格，每个格子的大小是r\*r。统计分析在不同的时间段，经过所有网格区域内的车流密度的变化。

F5. (必须实现) 区域关联分析1。用户指定两个矩形区域，统计在不同的时间段，往来这两个区域之间的车流量的变化。

F6. (必须实现) 区域关联分析2。用户指定一个矩形区域，统计往来该矩形区域和其他区域的车流量随时间的变化。

F7. (扩展实现) 频繁路径分析1。一条路径的频繁度可以定义为这条路径上通行的汽车的总数。根据用户给定的参数k，距离参数x，统计在整个城市中，长度超过x的最为频繁的前k个路径。

F8. (扩展实现)频繁路径分析2。给定两个矩形区域A和B，分析从A到B的最为频繁的前k个通行路径。

F9. (扩展实现)通信时间分析。给定两个矩形区域A和B，分析在不同的时间段，出租车从A到B的通行时间最短的路径，以及相应的通行时间。

**P2P网络电视模拟器（难度★★★★★）**

网络电视是通过互联网观看实时播放的视频节目的一类应用。常见的网络电视客户端可运行于普通的PC机、移动终端或电视机顶盒。目前比较比较流行的软件包括PPTV、腾讯视频、乐视网络电视等。

传统的网络电视采用的是客户端-服务器（C/S）的网络架构，所有的网络电视客户端都和服务器建立连接，并向服务器请求视频数据。在这种架构下，服务器的负载和客户端的数目呈线性比例关系，随着用户的增长运营成本急剧增长。为了解决服务器瓶颈问题，一种称为Peer-to-Peer (P2P) 的网络架构被提出并获得广泛推广使用。在基于P2P架构的网络电视系统中，客户端之间建立网络连接，并相互分享下载的视频数据。每个客户端都可以从其他客户端处下载视频数据，并且缓存一定数量的视频数据，同时也可以向其他客户端提供缓存的视频数据。通过这种人人为我，我为人人的方式，发挥客户端之间的网络传输能力，极大降低服务器的负载。

如何优化P2P网络从而最大程度提升电视客户端的观看质量是一个重要的研究课题。在研究P2P网络优化策略的时候，经常需要模拟P2P网络，并在该网络中观察各种策略对用户体验的影响。请你根据下面的要求，实现一个P2P网络电视模拟器。

**数据配置**：P2P网络可以抽象为一个图，服务器和每个网络电视客户端都可以抽象为图中的1个顶点，网络连接可以抽象为图中的连边。请设计算法，来产生模拟的数据：

网络中包含N个客户端顶点（N >= 100），1个服务器顶点。每个客户端/服务器随机分配一个二维平面中的坐标(x, y) 用于描述该客户端/服务器的地理位置。

每个客户端或服务器随机挑选网络中的t个客户端或服务器，建立到它们的网络连接

任何两个顶点之间的网络连接的传输速率和节点之间距离成反比，传输速率的取值范围是[20Kbyte/s, 100Kbyte/s]。

服务器每秒产生的视频数据的大小是30Kbyte。视频数据被切分为等大的数据块，每个数据块的大小为1Kbyte。每个数据块都对应一个全局唯一的整数序号，该序号反映数据块产生的时间先后。例如，第一个数据块序号为1，第二个数据块序号为2，依此类推。

每个客户端会定期检测它的邻居节点拥有的数据块序号，并向邻居节点请求自己没有的数据块。邻居节点收到请求后，会将对应的数据块传给该客户端。（本次模拟实验中，可忽略发起请求所需要消耗的时间）

客户端可以连续播放的前提是拥有连续的完整的M个分块 (M>= 5)。也即是客户端能播放数据块i的前提是，该客户端已经下载了数据块i, i+1, i+2,…, i+M-1.

每个客户端可缓存N个数据分块，（N可自由配置）。在客户端上维护一个长度为N的缓存队列，新到达的数据分块会将最老的数据块进行覆盖。

**功能要求**：

F1. （必须实现）网络可视化功能。显示所有客户端的位置和服务器的位置，以及彼此之间的网络连接。关于如何用vc++画图，或java画图，请在baidu或google上查找相关方法，或者借阅相关的参考书。

F2. （扩展实现）网络缩放功能。可以对网络显示区域进行缩放。

F3. （必须实现）模拟视频数据流。按上述数据配置要求，模拟服务器产生视频块，每个客户端向邻居节点请求数据块。测试客户端播放的流畅度（流畅度可定义为该客户端的连续播放时间占总时间的比例）。（提示：可基于队列实现"离散事件模拟"，可通过百度、google、电子图书馆等方式获取网络模拟相关知识）

F4. (必须实现) 测试客户端的播放延迟。（客户端的播放延迟 = 数据块在客户端上的播放时间 – 数据块在服务器上的产生时间）。

F5. （扩展实现）上述数据配置中构建的网络本质上是随机网络。请综合考虑节点之间的传输速率等因素，设计更佳的网络构建方案，使得各个客户端的播放延迟较小，播放流畅度较好。

F6. （扩展实现）模拟节点退出对网络的影响。让网络中R个客户端节点（R< N）停止工作，测试其他客户端的播放延迟和流畅度所受的影响。

F7. （扩展实现）设计邻居更新策略，以降低节点退出对网络的影响。每个节点在检测到邻居节点退出时，按某种既定的策略建立新的网络连接。请给出你的具体策略，并测试该策略的效果。

**开放式探索题目**

可通过下面大数据比赛相关平台，选择自己感兴趣的题目，并下载数据进行实验。最终大作业项目可在这些数据比赛题目的基础上进行扩展，增加用户交互界面，补充一些有趣而使用的功能。请充分发挥你的想象力。

[https://www.kaggle.com/](https://www.baidu.com/link?url=F3ne8F-IqU9ORi-Dadg2VGowhK0f9QrBokG_fDB06PO3ZX1GO8Ct5JuleJ_TDWeT&wd=&eqid=8f27baac000552cc000000035df054aa%22%20\t%20%22https://www.baidu.com/_blank)

<https://tianchi.aliyun.com/home/>

<https://www.datafountain.cn/>

**自选题目**

根据自己兴趣选定题目，要求：

题目合理，难度适合，能实现对数据结构课程内容应用的目的。

自选题目在寒假前要发给老师审核确定，再开始做。

时间安排

大作业时间和进度安排建议如下：

|  |  |
| --- | --- |
| 项 目 内 容 | 时间 |
| 选定题目和任务分配 | 2019.1.1 ～ 2019.1.15 |
| 需求分析和概要设计 | 2019. 1.21 ～ 2019.3.1 |
| 算法设计和编码调试 | 2019. 3.2 ～ 2019.4.15 |
| 系统整合和撰写文档 | 2019. 4.16 ～ 2019.4.20 |
| 答辩（包括结果演示、  提交小组和个人文档） | 2019年4月下旬 |

**4月20日前**提交到教学在线，每组同学把要提交的所有内容打包，并以"题目+姓名1+姓名2+。。。"作为压缩文件名。答辩时间另行通知。

设计工作要求

根据问题的工作量和难易程度，由2~4**名**同学组队完成。

**下学期将安排大作业答辩。答辩以小组为单位进行答辩，论述实现的软件的主要功能、所采用的数据结构和核心算法，以及测试的结果。**

所实现的程序，须满足以下要求：

界面友好：图形界面，支持鼠标操作。有合理的中文提示，每个功能可以设立菜单，根据提示，可以完成相关的功能。出现非法输入，会给出异常提示。

物理存储：相关数据要求存储在数据文件中，在程序中完成文件的读写操作。

逻辑结构：根据问题的要求，采用线性或非线性结构。需要考虑大数据量问题。

**每位同学**还需要针对自己所承担的部分工作，提交报告文档**。**该文档是评分的重要依据之一，请认真对待。该文档须包括如下内容：

需求分析： (陈述要解决的问题，要实现的功能),

详细设计：包括设计算法流程图、算法分析、使用的数据结构 **（要求详细论证）**；

软件测试：包括测试数据和测试结果记录

总结： 设计过程中遇到的问题及解决方法；尚未解决的问题及考虑应对的策略；收获和心得；

参考文献：在提交的设计报告中**必须要有参考文献！必须按学校规定的模板来写！**可以是课本以及与课程设计相关的工具书、论文、网页等。要求在报告中按规定格式详细列举所用参考文献。

**具体格式请参照《课程设计报告书》模板。**

**在提交的设计报告中，要能体现出合理的分工和充足的工作量。要求每人都要交设计报告，系统整体设计部分可以相同或相似，但自己做的部分内容要详细。**

**以小组作为单元，提交如下内容**：

源代码。要求**注释清晰**，编写规范，模块化。编程语言: C++/Java/phython/…

可运行的文件。

答辩PPT。

成绩评定

1. 考核办法：在学生完成设计、调试后，组织验收。以小组为单位进行答辩。

2. 成绩评定：

小组表现分：（50分）

包括程序是否可以正确运行，能否明确回答问题，小组整体完成情况。

个人表现分：（50分）

包括能否按时完成自己负责的部分，个人完成情况。

参考文献

Clifford A.Shaffer.数据结构与算法分析（C++版） （第三版）. 电子工业出版社. 2013.

严蔚敏，吴伟民. 数据结构(C语言版)（第二版）. 人民邮电出版社.2016.

殷人昆等. 数据结构（用面向对象方法与C++语言描述）第二版. 清华大学出版社. 2007.