

北京林业大学 20 22 --20 23 学年第 1 学期课程设计任务书

课程名称: 数据结构 A

开课学院: 信息学院

考试班级: 计算机类 21 级 1-7, 信息 21 级 1-2 命题人: 李冬梅 付慧 孟伟

设计题目(范围): 植物百科数据的管理与分析

请详细说明该设计的方案、内容、要求、进度等

【课程设计学时】32 学时

【课程设计目的】

1. “三知”知识目标: 熟练掌握线性表、栈、队列、串、数组、树和图等基本数据结构的逻辑特性和存储表示方法; 熟练掌握各种基本数据结构的基本算法和其应用; 熟练掌握问题分析、总体结构设计、程序设计的基本技能和技术。
2. “三能”能力目标: 能够综合运用数据结构与算法和相关的数学等理论知识对复杂工程中的算法问题进行抽象、分析和建模; 能够依据工程实际需求合理组织数据、并在计算机中有效地存储数据; 能够针对复杂工程中的算法问题, 设计出比较合理的解决方案, 利用具体的编程语言实现解决方案, 并具有一定的创新思维能力。
3. “三有”素养目标: 具有良好的工程素养和职业素养, 诚信守法, 能够坚持职业操守和道德规范; 具有精益求精的工匠精神、创新精神和探索未知终身学习的意识; 具有科技报国的社会责任感、使命感和爱国主义情操。

【课程设计内容】



党的二十大报告中指出“推动绿色发展，促进人与自然和谐共生”，为林草事业发展提出了更高的要求，指明了前进的方向。林草系统要牢固树立和践行绿水青山就是金山银山的理念，坚持山水林田湖草沙一体化保护和系统治理，加快提升现代化建设步伐，使林草事业发展成果更多、更好地惠及群众，奋力谱写新时代林草事业高质量发展新篇章。在党的二十大背景下，新时代林草人应积极投身于我国林草事业，将专业知识服务于我国的生态文明建设，将个人理想与国家的发展需求结合起来，为建设生态文明和美丽中国做出新的更大贡献。

“原本山川，极命草木”，植物既是重要的生物资源，又是环境的重要组成部分。在信息化的社会发展背景下，为促进经济发展与自然发展的和谐统一，将植物信息整合，对植物数据进行管理和分析显得尤为必要。

本次课程设计要求基于给定的植物百科数据，实现植物数据的管理与分析。通过对植物信息进行分析，可以进一步开展各地区植物及其多样性研究，对于合理开发利用和保护植物资源、因地制宜地制定林业生产和旅游业发展规划等具有重要意义。利用专业知识为生态文明建设提供解决方案和决策支持，增强专业自信。

另外，结合有关植物分布和植物从属关系划分等数据，借助数据结构课程所学的相关知识，通过对数据的处理和分析，可以全面了解植物分类体系以及我国在生态环境建设与植物多样性发展中做出的贡献，认识到植物多样性对建设美丽中国的重要性。

植物相关数据存储存储在 8 个文本文件中，相应的文件信息说明如表 1 所示。其中，各个文件中不同的数据项之间均使用“#”分隔，如文件 plant.txt 中每一植物的植物名称、学名、分布地、详情描述之间使用“#”分隔，而分布地可能包括多个省份，各省份之间使用“@”分隔。图 1 给出了文件 plant.txt 中植物“独叶草”对应的数据示例。

表 1 数据文件相关信息

文件名	存储信息	具体信息
plant.txt	植物基本信息	植物名称、学名、分布地、详情描述
distance.txt	省份间距离信息	省份 1 名称、省份 2 名称、距离
relation1.txt	种-属从属信息	种名、属名
relation2.txt	属-科从属信息	属名、科名
relation3.txt	科-目从属信息	科名、目名
relation4.txt	目-纲从属信息	目名、纲名
relation5.txt	纲-门从属信息	纲名、门名
relation6.txt	门-界从属信息	门名、界名

独叶草#Kingdonia uniflora#云南@四川@甘肃@陕西#多年生小草本，无毛。根状茎细长，自顶端芽中生出1叶和1条花草；芽鳞约3个，膜质，卵形，长4-7毫米。叶基生，有长柄，叶片心脏圆形，宽3.5-7厘米，五全裂，中、侧全裂片三浅裂，最下面的全裂片不等二深裂，顶部边缘有小牙齿，背面粉绿色。叶柄长5-11厘米。花草高7-12厘米。花直径约8毫米。萼片(4-5-6(-7)，淡绿色，卵形，长5-7.5毫米，顶端渐尖，退化雄蕊长1.6-2.2毫米；雄蕊长2-3毫米，花药长约0.3毫米；心皮长约1.4毫米，花柱与子房近等长。瘦果扁，狭倒披针形，长1-1.3厘米，宽约2.2毫米，宿存花柱长3.5-4毫米，向下反曲，种子狭椭圆球形，长约3毫米。5月至6月开花。

图1 植物数据示例—独叶草

【植物小科普】独叶草(Kingdonia uniflora)被人们称为“世界上最孤独的植物”，为什么说它最孤独呢？因为独叶草一生只长一片叶子，只开一朵淡绿色的小花，民间有一种说法，“独花独叶一根草”，说的就是这种草。独叶草 1999 年被列为国家一级濒危珍稀保护植物，是我国特有的植物，以无性繁殖为主，天然更新能力差，加之森林的破坏采挖，植株数量逐渐减少，自然分布日益缩减，近年已很难见到。国外学者 Foster A. S. 博士曾愿意以一辆轿车换取一株独叶草的活体，我国知名植物学专家胡正海教授和任毅教授坚持认为活体植物不可出境，保护我国植物资源，他们潜心研究并发表了一系列重要的学术论文，我国科研人员在植物学研究领域已经取得了显著的国际地位，我国生物多样性数量也位于世界前列。



“植物百科数据的管理与分析”实践项目由植物基本信息管理模块、植物分布地分析模块和植物从属关系检索模块三个子模块组成。

1. 植物基本信息管理模块

文件 plant.txt 中存储了 6490 种植物信息，每种植物的基本信息包括植物名称、

学名、分布地和详情描述。通过对 plant.txt 进行读写等操作，实现对植物基本信息的增加、删除和修改，并利用多种不同的策略完成植物信息的查找和排序，借助相关评价指标比较不同策略的优劣。

【综合素养拓展】通过实现此模块的各项功能，可以意识到在解决实际问题时，策略选择对实现效率具有决定性的作用，培养自己精益求精的工匠精神和探索未知终身学习的意识。

1.1 植物基本信息的增加、删除与修改

从 plant.txt 中读取植物的基本信息，采用后插法创建一个植物信息的链表，基于该链表，实现植物基本信息的增加、删除与修改功能。

(1) 增加植物信息

用户可以增加植物信息，输入植物名称、学名、分布地和详情描述信息，将该植物的基本信息添加到 plant.txt 中的最后。新增成功后，输出 plant.txt 的最后一行即新增植物信息。（注：由于植物名称是唯一的，因此在增加时需要判断该植物名称是否存在于 plant.txt 中，若存在，则输出“增加失败”。）

(2) 删除植物信息

用户可以删除植物信息，输入要删除的植物名称，将该植物的全部信息从 plant.txt 中删除，同时要求输出已删除植物对应的全部信息。（注：需要判断该植物名称是否存在于 plant.txt 中，若存在，则执行删除操作，否则，输出“删除失败”。）

(3) 修改植物信息

用户可以对植物的详情描述信息进行修改，输入植物名称和修改后的详情描述信息，将修改后的信息存储到 plant.txt 中，并输出已修改植物对应的全部信息。（注：需要判断该植物名称是否存在于 plant.txt 中，若存在，则执行修改操作，否则，输出“修改失败”。）

1.2 基于多种查找策略的植物信息查找

分别基于线性表、二叉排序树和散列表 3 种不同的数据结构，实现 6 种不同的植物信息的查找策略。对于任何一种查找策略，当用户输入植物学名后，若查找成功，则输出该植物对应的基本信息（植物名称、分布地、详情描述），并输出查找成功时的平均查找长度 ASL (Average Search Length)，否则，输出“查找失败”。

通过比较不同查找方法对应的 ASL，分析各种查找方法的优劣。

具体的查找策略包括以下 6 种：

- (1) 基于顺序表的顺序查找
- (2) 基于链表的顺序查找
- (3) 基于顺序表的折半查找
- (4) 基于二叉排序树的查找
- (5) 基于开放地址法的散列查找
- (6) 基于链地址法的散列查找

注：为尽可能减少冲突，现给出散列函数的构造规则及样例参考。如表 2 所示，假设散列表表长 m 为 6600，散列函数的除数取小于 m 的最大质数，即 6599。

表 2 散列函数的构造规则及样例参考

构造规则	样例 1	样例 2
$key = "c_0 c_1 \dots c_{n-1}"$ (其中, n 为字符串 key 的长度)	$key = "abc"$	$key = "Gentiana omeiensis"$
$Hash(key) = (\sum_{i=0}^{n-1} int(c_i) i^2) \% 6490$ (其中, $int(c_i)$ 得到字符 c_i 的 ASCII 码)	$Hash(key) = (97*0^2 + 98*1^2 + 99*2^2) \% 6599 = 494$	$Hash(key) = (71*0^2 + 101*1^2 + 110*2^2 + \dots + 115*17^2) \% 6599 = 3530$

1.3 基于字符串模式匹配算法的植物关键信息查询

在植物的详情描述信息中存在很多用户可能感兴趣的关键信息。当用户输入感兴趣的关键字后，通过字符串模式匹配算法可以将输入的关键字与详情描述信息进行匹配，输出所有匹配到的植物名称。例如，用户输入“根状茎细长”进行查询，而“独叶草”对应的详情描述信息中含有“根状茎细长”（如图 2），则输出“独叶草”。（同时输出所有详情描述信息包括“根状茎细长”的其他植物名称。）

独叶草#Kingdonia uniflora#云南@四川@甘肃@陕西#多年生小草本，无毛。
根状茎细长 自顶端芽中生出1叶和1条花草；芽鳞约3个，膜质，卵形，长4-7毫米。叶基生，有长柄，叶片心状圆形，宽3.5-7厘米，五全裂，中、侧全裂片三浅裂，最下面的全裂片不等二深裂，顶部边缘有小牙齿，背面粉绿色。叶柄长5-11厘米。花草高7-12厘米。花直径约8毫米。萼片(4-5-6(-7)，淡绿色，卵形，长5-7.5毫米，顶端渐尖，退化雄蕊长1.6-2.1毫米；雄蕊长2-3毫米，花药长约0.3毫米；心皮长约1.4毫米，花柱与子房近等长。瘦果扁，狭倒披针形，长1-1.3厘米，宽约2.2毫米，宿存花柱长3.5-4毫米，向下反曲，种子狭椭圆球形，长约3毫米。5月至6月开花。

图 2 植物关键信息查询示例

模式匹配算法要求分别基于以下 2 种策略来实现:

- (1) 基于 BF 算法的关键信息查询
- (2) 基于 KMP 算法的关键信息查询

1.4 基于多种排序策略的植物信息排序分析

运用不同的排序策略,将植物信息按植物学名的字典序进行排列,从而能够根据学名快速找到相应植物。对于任何一种排序策略,在排序的同时,要求输出完成排序时总的关键字比较次数 KCN(Key Comparing Number)和记录移动次数 RMN(Record Moving Number)。

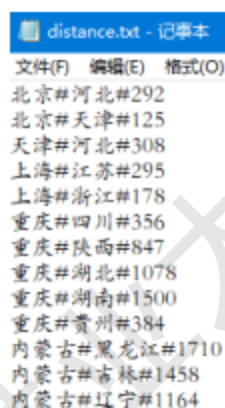
通过比较不同排序方法对应的 KCN 和 RMN,分析各种排序方法的优劣及适用情况。

具体的排序策略包括以下 5 种:

- (1) 直接插入排序
- (2) 折半插入排序
- (3) 简单选择排序
- (4) 冒泡排序
- (5) 快速排序

2. 植物分布地分析模块

植物基本信息文件 plant.txt 中存有植物的省份分布信息。在文件 distance.txt 中,存储两两省份间的距离信息,这里只存储相邻省份之间的距离(假设距离为两省份省会城市之间的直线距离,单位为公里),如图 3 所示,北京到河北(石家庄)的直线距离为 292 公里,内蒙古(呼和浩特)到黑龙江(哈尔滨)的直线距离为 1710 公里。



```
distance.txt - 记事本
文件(F) 编辑(E) 格式(O)
北京#河北#292
北京#天津#125
天津#河北#308
上海#江苏#295
上海#浙江#178
重庆#四川#356
重庆#陕西#847
重庆#湖北#1078
重庆#湖南#1500
重庆#贵州#384
内蒙古#黑龙江#1710
内蒙古#吉林#1458
内蒙古#辽宁#1164
```

图 3 相邻省份间的距离信息示例

2.1 植物移植最短路径分析

当需要对植物进行跨省移植时，花费的成本与运载植物的路程长度息息相关（这里暂不考虑气候等环境因素对植物生长的影响）。用户可以通过输入植物名称和待移植的目的地，得到运输该植物需要花费的最短路径。其中，若移植的目的省份中已有该植物的分布，则输出“该省份已存在，无需移植”；否则，输出移植的出发地和抵达目的省份的最短路径长度。

注：若植物有多个分布地，则需要分别算出每个分布地到目的省份的最短路径，再取其中的最小值作为输出结果。

2.2 可移植路径分析

当在某地发现一株新植物时，需要及时对其进行易地保护。易地移植的过程中，若路程过长，植物容易在运载途中死亡，因此需要对植物的可移植路径进行分析。用户输入新植物发现地、移植目的地及该植物能接受的最大路程，通过对省份图进行遍历，输出所有可行的运输路径（按路程总长度从升序输出）。

【综合素养拓展】通过对植物在易地保护中可移植路径的分析与实现，增强当代大学生对植物的保护意识和危机意识，一起助力生物多样性保护，进一步认识到植物保护对推进生态文明建设的重要意义，坚信“绿水青山就是金山银山”的重要发展理念。

3. 植物从属关系检索模块（选做）

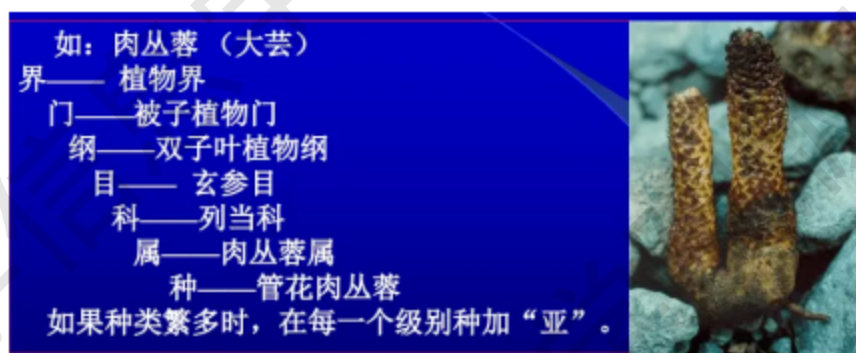
3.1 植物分类树构建

根据生物分类学的相关知识，生物的分类系统通常包含七个级别：界、门、纲、目、科、属、种。对于植物领域而言，可以通过一个分类树存储所有的植物及其分类信息。其中，“植物界”即为树的根结点，每一个植物名称可视为一个“种”，并存储在树的叶子结点中，中间的门、纲、目、科、属可以存储在树的非叶子结点中。（这里叶子结点和非叶子结点是针对树的逻辑结构而言的，实际存储时不一定如此。）

【综合素养拓展】植物分类学是植物学其他分支学科的基础，也是植物资源开发利用和生物多样性保护的基础。利用多学科研究手段，对重要植物类群进行深入、细致的分类学和系统研究，澄清物种的范围和物种间的界限，确定物种的数目、特征和分布格局，提出更科学的分类框架，仍是当前和今后一段时间内植物分类学研究的重要

任务。此外，随着学科的发展、计算和分析技术的提升，以及国家、社会和公众对植物学知识需求的增加，对植物数据进行整合分析、建立智能化“植物百科全书”和植物图像的快速识别系统等，也成为植物分类学的重要内容。

通过利用所学的数据结构专业知识构建植物分类树，可以实现对植物信息的整合和分析。利用科学手段分析某一地区的植物是否具有多样性特征，在植物栽培种植时可以根据植物的亲缘关系进行选择和优化，为我国自然保护区建设提供科学支撑，对资源合理利用和可持续发展具有重要意义。



根据图 4 所示的 6 个从属关系文件 relationi.txt (i=1...6)，采用树的二叉链表表示法（即孩子兄弟法），构建出该植物分类树。随后，输入植物名称，根据构建出的植物分类树，依次输出该植物的属、科、目、纲、门、界信息。



图 4 植物 6 个从属关系文件信息示例

3.2 同属植物信息检索

输入植物名称，根据构建出的植物分类树，输出与该植物同一个“属”的其他植物，即该叶子节点的全部兄弟节点，不包含自身。

3.3 下属植物信息检索

输入植物类别（门、纲、目、科、属任何一个），根据构建出的植物分类树，输出隶属于该类别的所有植物，即该植物类别对应的分支节点下的全部叶子节点。

【课程设计进度安排】

表 3 课程设计进度安排

序号	时间	完成内容
1	16 周周一	教师讲解课程设计任务书中的要求，介绍“推动绿色发展，促进人与自然和谐共生”的二十大报告精神，布置思政调研工作，要求学生明确课程设计的基本任务和具体要求（知识、能力、素养），熟悉所给数据文件的格式，学习文件读写方法。
2	16 周周二	完成 1.1、1.2
3	16 周周三	完成 1.3、1.4
4	16 周周五	完成 2.1
5	17 周周一	完成 2.2（当天进行中期检查，要求至少完成 1.1—1.4）
6	17 周周二	根据自身情况完成选做、代码测试
7	17 周周三	根据自身情况完成选做、代码测试
8	17 周周五	一对一答辩验收，根据模板撰写课程设计报告

【课程设计考核要求】

1. 总评成绩组成说明

（1）中期检查占 10%，程序验收占 60%，课程设计报告占 30%（严格按照模板要求撰写）。

（2）选做内容没有涉及者成绩最高 85 分。

2. 程序验收要求

（1）程序至少实现任务书中的必做内容，无编译错误，可以正确运行。

（2）程序编写时注意格式和命名要规范，适当添加注释，程序可读性强。能够解释程序中所涉及的每个问题求解的具体过程、每个功能算法的思想和具体的执行过程、每个算法的时空复杂度、每一行代码的含义，程序中关键的语句在去掉后能够当场复

原。

(3) 程序验收时至少提问三个问题，如果两个答不上，则不通过，下学期补考。

(4) 可以借鉴别人的程序，但不能完全相同，否则视为 copy，下学期补考。

3. 文档提交要求

(1) 提交时间：20 周周三（2023 年 1 月 12 日）24 点前

(2) 提交内容：严格按模板撰写实习报告，将实习报告上传到 FTP 对应班级的“课程设计”文件夹，报告命名为“学号姓名-实习报告.docx”。

教研室主任意见：

签字：李冬梅 2022 年 12 月 10 日

学院负责人意见：

签字：崔晓晖 2022 年 12 月 10 日

注：此表签字后，扫描成 PDF 于考前上传至教务系统“课程考核材料上传管理”，纸质版随学生课程设计材料上交学院。