

華中科技大學

程序設計 綜合課程設計

(2022 級網絡空間安全、信息安全、密碼技術專業)

網絡空間安全學院

程序設計綜合課程設計課程組

2024 年 1 月

目录

1 课程概述	1
1.1 课程背景.....	1
1.2 课程目标.....	1
1.3 课程任务	2
2 设计问题：磁盘目录文件扫描和查询.....	3
2.1 问题概述.....	3
2.2 设计建议	4
2.3 主要功能要求	5
2.4 扩展功能要求	6
2.5 升级功能要求	7
2.6 约定说明	7
2.7 数据文件格式说明.....	7
2.8 设计要求说明	9
2.9 评测说明	9
3 总体要求	10
3.1 坚守学术诚信	10
3.2 程序规范	10
3.3 报告规范	10
3.4 课堂与考勤要求	11
3.5 检查与验收.....	11
指导参考书目.....	12
附录 A 评价指标.....	13

1 课程概述

1.1 课程背景

对于网络空间安全、信息安全、密码技术专业大二学生，在前三个学期已经学习了 C 语言程序设计、数据结构两门面向编程知识与技术的基础理论课和数据库系统课程，以及 C 语言程序设计实验、数据结构实验两门编程实践课程，不仅具有较为系统性的 C 语言、常用数据结构基本知识和基本的数据库应用技术基础，而且具有初步的程序设计、数据抽象与建模、问题求解与算法设计的能力，奠定了进行复杂程序设计的知识基础。但两门实验课仍属于对基本编程模型与技术的验证性训练，而“程序设计综合课程设计”正是使大家从简单验证到综合应用，甚至在编程中实现智慧与风格升华的重要实践环节，为后续学习高级算法分析等课程与进行网络安全系统编程打下坚实的基础，让综合编程技能成为大家的固有能力和通向未来专业之门的钥匙。

1.2 课程目标

基于“程序设计综合”课程设计实践课程规划原则及其在计算机相关专业人才培养中的地位，其应该体现与达到如下目标：

(1)综合性训练目标：在该课程中涉及 C 语言的主要编程要素，如典型的数据类型与控制结构；覆盖多种典型的数据结构如线性结构、二叉树与树结构、图结构、查找表结构，数据文件的组织运用；以及数据库关系模式的设计与编程应用等。从先前实验课的单要素或单一结构训练向多要素，多结构综合应用训练转变。

(2)培养应用问题的求解能力：程序设计是为问题求解服务的，提高对应用问题进行分析、数据抽象与建模、问题定义与功能划分、基础数据收集与测试用例构造等综合分析与表示能力。

(3)程序编写向程序设计转化：在实验课程中，老师基本描述了相关数据结构，程序框架及主要算法，基于此进行程序编写训练，其属于验证与复现性编程实践。“程序设计综合”要求同学们基于对应用问题的分析，建立求解模型，设计数据结构与主要算法，从而进行程序设计，更多地体现“设计”的内涵与份量。

(4)进一步培养编程规范性与工程化素养:通过“程序设计综合课程设计”的实践进一步培养良好的规范性编程习惯,以及一定的程序设计与软件开发的工程化素养,按照问题定义、必要的需求分析、系统设计、编程实现、程序测试分析及编制程序设计综合课程设计报告的流程,来组织本实践课程的开展与进行,形成初步的工程化程序设计素养。

1.3 课程任务

在选择与确定了“程序设计综合课程设计”的设计问题之后,按工程化的基本流程分别完成如下任务:

(1)阅读“程序设计综合课程设计”任务书,熟悉问题,查阅文献,了解问题背景及相关知识。

(2)对设计问题进行需求分析,分析问题中所涉及的数据对象,划分功能,人机交互需求与数据文件读写等,并对问题进行形式化表示。

(3)基于上述需求分析,进行系统设计,明确程序的模块结构;设计数据结构(逻辑结构及其物理结构),参考并设计主要子问题的求解算法。

(4)程序实现,基于系统设计,制定相应的实施方案,编写各程序模块,完成程序编写与调试任务。**建议了解与练习编写程序安装与使用说明。**

(5)程序测试,设计测试用例对程序进行功能测试,性能测量及理论分析。

(6)程序优化,对设计方案中的结构,算法进行一定优化,测试与分析性能改善结果。

(7)设计总结,按规范化要求撰写“程序设计综合课程设计”报告。

(8)成果提交:将程序源代码/工程文件、可独立运行的可执行程序、简要程序安装和使用说明及“程序设计综合课程设计”报告电子版打包,文件夹名称格式为“专业班级-学号姓名”,如:网安 2202-U202114999 李某某。并将设计报告打印为纸质版(A4 双面打印),然后以班为单位在指定时间(一般在设计课结束后两周内)集体提交到指导老师。

2 设计问题：磁盘目录文件扫描和查询

2.1 问题概述

随着信息技术的飞速发展，数据存储在我们的生活中占据了越来越重要的位置。磁盘，作为现代计算机的主要存储媒介，存储了大量的个人、组织或公司的数据。这些数据中，目录文件系统是组织和管理数据的关键。然而，随着数据的增长和复杂性的提升，对磁盘目录文件的扫描和查询需求也日益增加。

然而，随着信息技术的不断发展，信息安全问题也日益突出。恶意攻击者可能会利用各种手段对磁盘目录文件进行非法访问、篡改或删除，导致数据泄露、系统瘫痪等严重后果。因此，需要了解和熟悉对磁盘目录文件进行扫描和查询的方法和效率，为今后提高文件系统的安全性和可靠性打下基础。

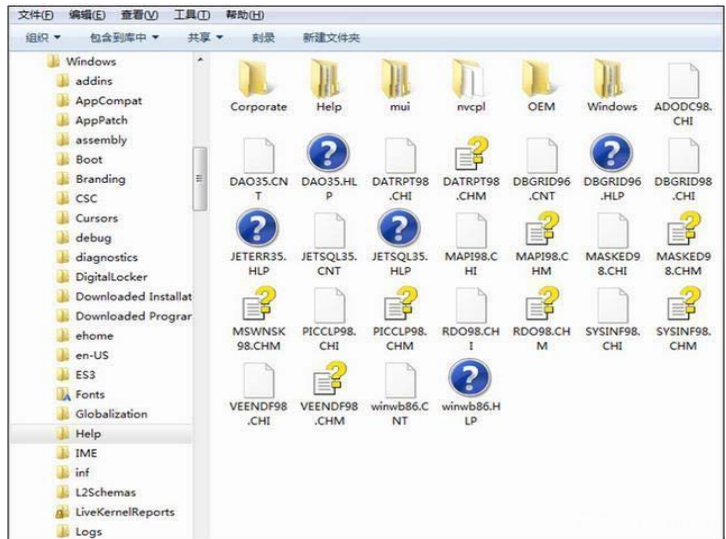


图 1 Windows 的文件系统

在 windows 系统中，默认系统目录安装在 c:\。对该目录中包含的各级子目录和文件进行扫描和统计，会遇上目录层次多、目录名和文件名长等情况，需要进行一些优化设计才能完成。通过遍历各级子目录和文件并进行统计与比较，可以检查出有些文件操作使得哪些子目录中的文件发生了变化，为后续采取相应的措施打下基础。

- 1) 采用合适的方式遍历 c:\windows 目录中的子目录和文件，分析文件目录有几层，完整的路径名和文件名需要多大的字符串空间等；
- 2) 设计关系数据库的表结构，有效地把扫描到的文件和子目录信息保存到数

数据库中，为后续更多的分析做准备；

3) 模拟一组对文件和目录进行地增加、修改和删除操作，有效维护在内存中的目录结构，或数据库中的文件信息，并检查哪些目录下的文件发生了变化；

4) 在实现上述功能的基础上，对于如何优化设计和优化实现方法上，提出建议方案并进行验证说明。

2.2 设计建议

系统大致可以由五个功能模块构成：读取系统文件目录下的各子目录和文件信息、将读取的信息写入数据库保存、在内存中构造目录结构、在内存/数据库中模拟文件操作、统计某个目录中文件属性的变化。

功能模块的基本描述如下：

1) 获取文件信息。

调用系统函数获取磁盘目录和文件的信息，需要使用遍历的方式逐一有序地获取。**Windows** 目录中的子目录层级和文件的数量与用户安装的具体应用有关，但是基本结构是类似的，估计每位同学使用的系统目录中，有 5 万以上的子目录（文件夹）和将近 20 万个以上的文件存在。因此功能模块在设计中需要考虑足够的变量空间，如果采用递归程序，需要估计到系统是否能够支持足够层次的递归调用。

文件的信息有许多，本次设计要求只需要考虑时间和字节数大小，其中时间是秒数，需要进行一些转换才能正确显示。

2) 文件信息保存到数据库。

可以使用手工方式建立数据库的表等内容，然后将需要保存的文件/目录信息以 **SQL** 语句的方式进行体现，使用有效执行 **SQL** 语句的方式将文件/目录信息存入数据库中。

如果 **SQL** 文件包含的语句数量过大，数据库系统可以无法执行，需要进行合理的分割才能执行。

3) 内存中重构目录结构。

后续进行文件的模拟操作和文件属性的统计，需要在内存或数据库中进行，则需要在内存中构建目录结构来保存必要的文件信息，这样对文件的属性进行修

改操作时不会影响磁盘中的实际文件。

需要采用合适的数据结构来保存目录/文件信息的多叉目录树结构，以及采用合适的遍历算法来实现文件信息的模拟操作和统计处理。

4) 模拟文件操作。

依据数据文件提供的文件/目录的模拟操作要求，需要找到对应的数据结点进行处理。如果因为每位同学的 windows 中文件的差异而找不到指定的操作文件，则按该项模拟操作为无效操作来处理。

需要注意，对于子目录的操作是指包含下级的子目录和文件。

5) 检查目录中文件属性的变化。

依据数据文件 mystat.txt 中指定的一组子目录进行统计，得到该目录中 A 文件（不包括子目录/文件夹）数量、B 文件字节数总和、C 最早时间的文件（含文件名、文件大小、文件时间）和 D 最晚时间的文件（含文件名、文件大小、文件时间）。

然后对于进行模拟操作后，需要检查这些指定目录中的文件统计数据有哪些发生了变化，即上面的 ABCD 有哪些发生了变化。

2.3 主要功能要求

（1 - 4.1 为基本内容，共 70 分，具体的分数分解见检查表）

1) 磁盘中目录/文件的扫描。（30 分）

利用 C 语言的库函数来获取目录属性和文件属性，去除一些无实际意义的文件和目录（文件夹），对于子目录还需要找出下一级的文件和存在的子目录。在访问每个子目录/文件的基础上，需要统计出：（1）统计出子目录数量；（8 分）。

（2）统计出文件总数量；（8 分）。（3）统计出目录层数；（6 分）。（4）找出最长的带全路径的文件名及长度。（8 分）。

需要注意，可以取文件的修改时间，显示时间时需要进行格式转换。

【考查遍历目录文件的算法，同时访问目录/文件属性的能力；】

2) 生成插入数据库的 SQL 语句。（20 分）

需要自行选择一个数据库平台，进行模式设计和操作功能的实现。然后可以使用在遍历目录文件生成的 SQL 插入语句来导入数据，以便在后续的统计和检查中使用数据库方式来实现。相应需要实现的工作有：（1）设计表结构，包含目

录表、文件表。(6分)。(2)生成可以导入的多个SQL文件。(8分)。(3)数据导入到数据库中。(6分)。

【考查数据库的模式设计、插入目录文件信息的SQL语句生成,以及具体实现数据导入数据库的能力;】

3)在内存中构造目录树,统计文件信息。(15分)

在内存中构造目录树,需要说明内存中的目录树的数据结构以及相关的处理算法等。(1)孩子兄弟树结构等有效结构。(5分)。(2)计算树的深度。(4分)。

(3)采用哪种遍历算法。(6分)。(建议采用非递归的层次遍历,构建队列辅助。)

【考查树结构的设计/数据库表的设计,以及非递归遍历的算法实现能力;】

4.1)统计指定目录中的文件信息。(5分)

读入数据文件,按照指定的目录来统计该目录中的文件属性信息,不包含其中子目录下

的文件信息。需要统计:最早时间的文件(含文件名、大小、时间)、最晚时间的文件、文件总数、总的文件大小。

如果选择在数据库中实现,需要编写存储过程/联机程序来读取数据文件和实现查询操作。

【考查树的遍历操作/数据库的查询操作的实现能力;】

2.4 扩展功能要求

(共20分)

4)文件/子目录的模拟操作,统计并检查变化。(20分)

读入数据文件,对内存中的目录树/数据库中的表进行目录/文件数据的修改,然后对指定的目录中的文件信息进行统计和比较。(该部分可以在内存的目录树中进行操作和比较,也可以在数据库中进行操作和比较。)

{基础功能部分,(1)统计文件信息。(5分)。}

(2)模拟文件操作(myfile.txt)。展示3条操作后的文件信息变化。(5分)。

即需要自己选择数据文件中的3条数据中的文件,在操作前查出文件的属性,在操作后再次查出文件的属性,体现对该文件属性的修改操作是成功的。

(3)再次统计文件信息,并检查有哪些变化。(5分)。即需要保存第一次的统计信息,然后在模拟文件操作后,再次得到统计信息,并进行对比,把发生变化的那些目录的统计信息的区别显示出来。

(4) 模拟目录操作(mydir.txt)。展示该目录和下层子目录的文件信息变化。
(5 分)。即需要自己选择数据文件中的 1 条数据中的文件，在操作前查出该目录和子目录下的文件属性，在操作后再次查出该目录和子目录下文件的属性，体现对该文件属性的修改操作是成功的。

(5) 第三次统计文件信息，并检查与原来初始的目录文件属性统计信息有哪些变化。(5 分)。

【考查对于子目录的属性修改会影响到下级所有子目录和文件的处理能力；】

2.5 升级功能要求

(完成 2.3 和 2.4 设计任务后可完成本关任务。共 10 分)

通关升级后的实现

5) 处理的优化建议和验证展示。(10 分)

根据前面各项自身实现的技术方案，提出其中的一些操作有哪些可以优化的方案，如：优化内存中的目录文件树，或者优化数据库表的设计，能够提高操作效率。(1) 给出可以进行实现的技术方案；(5 分)。(2) 对优化方案进行实现并演示。(5 分)。

2.6 约定说明

各位同学都在自己的 C 盘安装 windows 系统，并安装了常见的一些工具软件和应用软件，这样 windows 目录下的文件数量会比较大。

如果因为有其他特别的情况，使得提供的数据文件在自己的计算机上运行后没有对应的子目录和数据文件，可以将产生检查数据的工具程序在本机上在运行一次，以便形成相应的数据文件。

2.7 数据文件格式说明

1) mystat.txt 数据文件。指定一系列子目录进行文件属性的统计。

第一行“stat dirs”和最后一行的“end of dirs”表明当前文件是 mystat.txt，防止误读其他格式的数据文件；

目录的格式：

子目录的路径名\n （路径名结尾含符号\）

例子：

```
stat dirs
.....
c:\windows\UUS\x86\
.....
end of dirs
```

2) mydir.txt 数据文件。对指定的子目录进行模拟的操作。

第一行“selected dirs”和最后的“end of dirs”表明当前文件是 mydir.txt，防止误读其他格式的数据文件；

目录模拟操作的格式：

子目录路径名,D,time,size\n

（本次模拟的操作只有删除，以 D 表示，操作后对目录属性的时间和大小都没有意义，因此直接使用 time 和 size 替代。）

例子：

```
selected dirs
.....
c:\windows\SysWOW64\XPSViewer\zh-CN\,D,time,size
.....
end of dirs
```

3) myfile.txt 数据文件。对指定的文件进行模拟的操作。

第一行“selected dirs”和最后的“end of files”表明当前文件是 myfile. txt，防止误读其他格式的数据文件；

文件模拟操作的格式：

文件的路径名,D,time,size\n

（对文件的模拟操作有：D 删除，M 修改时间和大小属性，A 新增文件。其中，新增的文件带有最近的时间属性和大小属性，时间和文件的字节数都是用 long int 来表示。时间是从 1970 年 1 月 1 日开始的秒数。）

例子：

```
selected files
```

```
.....  
c:\windows\SystemApps\MicrosoftWindows.Client.CBS_cw5n1h2txyewy\SearchUx.MinUI.dll,M,1704466243,810496  
.....  
end of files
```

2.8 设计要求说明

在满足以上功能约定要求和数据格式基本要求的前提下，学生可以在界面、数据结构、数据文件、数据库表结构等方面自由发挥，可以设计相关辅助功能，也可以设计其他一些有意义的扩展功能：

课堂上将提供 2.7 部分的各类数据文件，在此基础上构成基本的程序运行数据环境，作为设计与开发过程中的功能设计、实现、测试以及验证检查的基础。

同时，使用基本相同的数据环境，还可以考察学生使用不同数据结构和处理算法时，所开发程序运行的性能之间是否有明显的差别。

2.9 评测说明

1) 评测步骤：

主要设计思路与实现方法的讲解，功能界面及提示，分项功能操作和执行时间统计，程序功能的测试情况，扩展功能介绍；

2) 运行评测：

检查磁盘目录文件扫描遍历的合理方案、高效的算法、保存文件信息的数据库表结构的设计，以及合适的变量空间设计等，能够依据模拟操作的数据要求进行正确的处理和统计分析。

3) 设计评测：

评价在内存中重构目录结构的数据结构、保存文件信息的关系表结构设计以及模拟文件操作、文件操作后影响指定目录文件的统计与检查功能等，还有上述各项算法的优化方案建议以及验证的展示。

3 总体要求

3.1 坚守学术诚信

鼓励创新，进行有一定特色的设计。严禁对程序与报告的抄袭行为（包括对网络资源及其他同学的设计），一经发现，课程设计成绩计 0 分，以考试抄袭舞弊行为处理。

3.2 程序规范

程序遵从一般性规范：

- (1) 源码依据模块组织到不同.h 与.c 文件中，不要将全部程序放到一个源文件中。
- (2) 变量尽量基于描述性命名，看其名知其意。
- (3) 函数头有统一注释，说明功能，输入输出与条件等。
- (4) 函数内部关键处理步骤处加上注释予以说明。

3.3 报告规范

按照网安学院课程设计报告的要求及本课程设计报告的格式规范与内容要求撰写设计报告，避免出现错别字及形式的不规范现象。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

二、需求与技术现状分析

三、程序总体设计(含模块结构图)

四、数据结构和算法详细设计

五、程序实现

(C 语言程序实现的简要说明，如开发环境、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，报告中只能作为附录内容之一)

六、程序测试及结果分析

七、复杂度分析

八、总结

(1) 通过课程设计了解到的社会信息在安全、发展等思政方面的情况，具体实现过程中对于科学家精神、工匠精神的体会等；

(2) 在整体设计和实现技术上的收获、体会，以及经验教训；

(3) 在思维模式上的收获和体会，例如：科学思维->工程思维，求解思维->设计思维，单一思维->综合思维，学习思维->创造思维等；

附录一：主要参考文献

附录二：主要源程序片段

（限 10 页内，有特色功能、关键算法、数据结构定义部分）

附录三：程序使用说明

3.4 课堂与考勤要求

要求按时到实验室完成程序设计综合课程设计，根据完成与验收情况由指导老师批准方可在其它场所查阅资料，撰写报告。课程设计课坚持记录考勤。

3.5 检查与验收

在设计课内/线上，全体同学需给指导老师或助教演示程序、解释程序、回答老师提问、验收或报告完成情况。

指导参考书目

- [1] 曹计昌, 卢萍, 李开. C 语言与程序设计. 电子工业出版社, 2013
- [2] 严蔚敏等. 数据结构 (C 语言版). 清华大学出版社,
- [3] Larry Nyhoff. ADTs, Data Structures, and Problem Solving with C++. Second Edition, Calvin College, 2005
- [4] 殷立峰. Qt C++ 跨平台图形界面程序设计基础. 清华大学出版社, 2014:192~197
- [5] 严蔚敏等. 数据结构题集 (C 语言版). 清华大学出版社
- [6] 王珊, 萨师煊. 数据库系统概论, 2014.9

附录 A 评价指标

评价指标	细化	满分	评价标准
程序设计实现 (70%)	功能实现 (50%)	100	成绩 = 2.3 主要功能要求(70 分)、2.4 扩展功能(20 分)，以上内容完成后，可以继续做通关升级后的 2.5 内容(10 分)
	设计特色 (15%)	15	特色不明显：10，有一定特色：12+，特色突出或有创意：14+
	代码规范(5%)	5	程序规范：基本 3，注释：4+，模块化且注释好：5，不规范：3-。
课程设计报告 (30%)	报告内容 (25%)	100	问题描述与分析：10，程序总体设计、数据结构、算法设计和理论分析：45，测试计划及测试分析：15，复杂度分析：10，总结（思政、思维、体会）：20。
	报告规范(5%)	100	基本规范：80，规范：80+，不规范：80-。
综合成绩 = 功能实现 × 50% + 设计特色 + 代码规范 + 课程设计报告 × 30%			

注：实验考勤原则上仅记录签到情况，不考虑任何请假情形。