NTFS 文件系统下文件恢复程序说明书

一、文件恢复问题概述及其重要性分析

要描述文件恢复问题，就不得不先提及数据恢复的定义。数据恢复是指通过正常途径不能恢复的数据通过一定的技术手段恢复的过程，其中不可恢复的数据是指主要是指一些意外丢失的数据，主要包括有人为误操作造成的数据丢失，即在使用计算机的过程中不小心删除了文件，或者不小心将分区进行格式化操作，导致的数据丢失；恶意程序的破坏造成的数据丢失，即有些病毒可能会造成的硬盘锁死、分区丢失或数据丢失；硬件故障造成的数据丢失等。而本组所设计的程序针对的问题属于第一类数据恢复问题，即我们旨在解决在 NTFS 文件系统下的计算机使用过程中由于人为误操作造成的文件删除的恢复问题，这个问题也是数字取证问题中的一个重要分支。

该取证问题的解决方案的重要性也是毋庸置疑的，当今世界离不开电脑，不管从事什么行业，都可能需要使用电脑，即使在家待业，也可能使用电脑在线聊天、看电影、写 Blog 等。我们的工作、我们的资料、我们的劳动成果等转换成 0 和 1 的符号，以文件的形式存储在电脑里，所以这些数据对于我们来说是最重要的。电脑坏了可以买，但是如果因为意外的操作导致文件被删除被覆盖而不能恢复，就没有地方买了。这些数据中极为珍贵。所以设计一个可在 NTFS 系统中恢复指定的被删除文件的程序是有重要意义的。

二、程序拟完成的功能

1. 可视化操作界面（GUI）

本程序拟设计一个简单的可视化操作界面，以便用户可清晰简洁地选择其需要扫描恢复的盘区，并将扫描后可恢复的文件列表通过可视化界面展示给用户供其选择。

2. 分区扫描

可选定任意磁盘分区以扫描该磁盘下的可恢复文件，并通过可视化界面呈现。

3. 数据恢复

指定扫描结果中需要恢复的文件，并将该文件恢复至指定文件夹。

三、程序的流程

1. 分区扫描模块

(1) 分区扫描通过扫描指定分区中 NTFS 文件系统中的 MFT 的文件记录来确定文件在分区中位置。

(2) 通过查看每一条文件记录偏移 0x16 处开始的两个字节（00H 表示文件被删除，01H 表示文件正在使用，02H 表示目录被删除，03H 表示目录正在使用），判断该文件是否为已删除文件。

(3) 若该文件为已删除文件，则通过分析 0x30 属性体获取文件名等信息。

2. 数据恢复模块

(1) 首先判断文件记录为 0x80 的属性体是否为常驻属性。

(2) 若是则分类获取数据内容。

(3) 然后在保存位置中新建一个文件，用已删除文件的文件名来命名，再将数据区的文件内容复制到新的数据区中。

3. 可视化模块

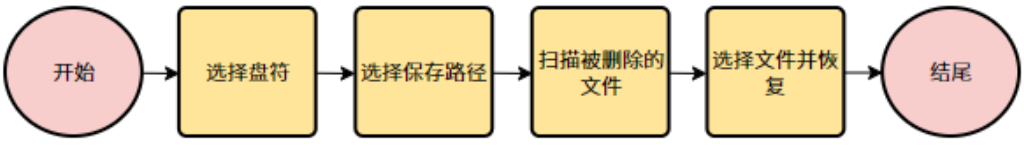
(1) 选择盘符。

(2) 选择保存路径。

(3) 扫描相应盘符下的已删除文件。

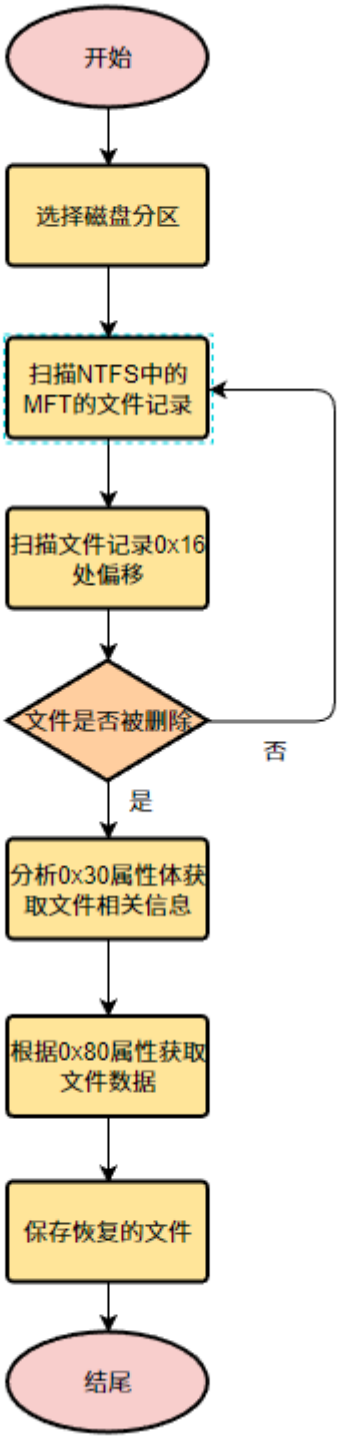
(4) 选择已删除文件，并恢复到指定位置。

4. 程序操作流程



图（1）程序操作流程

5.程序运行流程



图（2）程序运行流程

四、程序设计过程中遇到的难点

1. 解析数据结构

涉及的数据结构的较为复杂，需要准确地定位到所需要读取的位置，如 MFT 表的起始地址和其内部属性的位置，且属性长度不固定，存储形式不同（常驻属性和非常驻属性）。

2. 遍历文件记录项

在遍历 MFT 中的文件记录项时，起初我们是从 BPB 获取 MFT 的起始簇号算出起始位置，直接从头按照每 1024 字节（MFT 文件记录项一般占两个扇区，共 1024 字节）进行读取，很容易发生错误，进行中止，扫描获得的文件数量也不符合。查找资料得知，MFT 的第一个文件记录就是$MFT 自身，它记录着所有文件和目录的所有情况。我们对第一个文件记录项进行了分析。结合文件记录项的 0x80 属性，它记录文件的实际数据存放位置，且对于 MFT 该属性为非常驻属性。属性体中的一个或多个 RunList（记录实际数据存放的起始簇号或偏移和所占簇数的一种结构）记录了 MFT 的数据。经分析，一个磁盘中的 MFT 可能不连续的，我们应该根据 RunList 分别进行扫描获取所有已被删除文件对应的文件记录项，所以我们用数组记录了这个信息以方便遍历文件记录项。如下图（3）所示，元组中第一个为起始位置，第二个为簇数。图（4）（5）（6）分别为在不同 MFT 段所扫描到的以删除的文件。

```
[(3221225472, 12096), (93729030144, 22400), (48417996800, 14272)]
```

图（3）MFT 数据记录

```
第0个MFT段 5250 : 艺术气质四页01.docx
第0个MFT段 5251 : 艺术气质四页02.docx
```

图（4）第一个 MFT 段

```
第1个MFT段 2446 : Scenes.meta
第1个MFT段 2447 : ProjectVersion.txt
```

图（5）第二个 MFT 段

```
第2个MFT段 32486 : TransformRotate4.png
第2个MFT段 32487 : TransformSetParentOriginal.png
```

图（6）第三个 MFT 段

3. 恢复文件数据

文件记录项中 0x80 属性记录了文件数据，分为常驻和非常驻属性，在恢复数据是需要考虑到这个问题。对于常驻属性，根据 0x80 属性头中的属性体起始偏移值和文件真实长度来直接在属性体中进行数据；对于非常驻属性，需要根据 RunList 和数据真实大小进行数据读取并进行拼接。特别注意的是，需要真实大小的原因是最后一个 RunList 不一定全部用完，在读取时需要考虑到。

五、实验结果和截图

注意：要以管理员身份运行程序

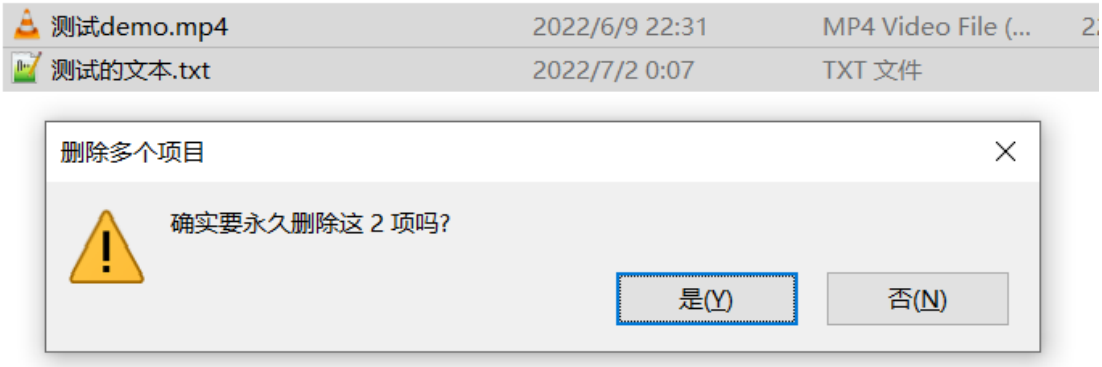
1. 新建文件

将测试视频文件复制到目标盘 G，并在盘 G 中新建文本文件“测试的文本.txt”，并写入文本：这是测试文本，测试文件恢复。

图（7）新建文件

2. 删除文件

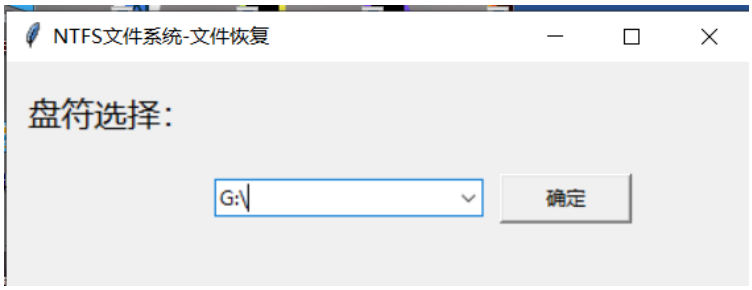
不能单纯删除放入回收站，而是要进行永久删除，即使用 shift+del 进行删除。



图（8）删除文件

3. 选择目标盘

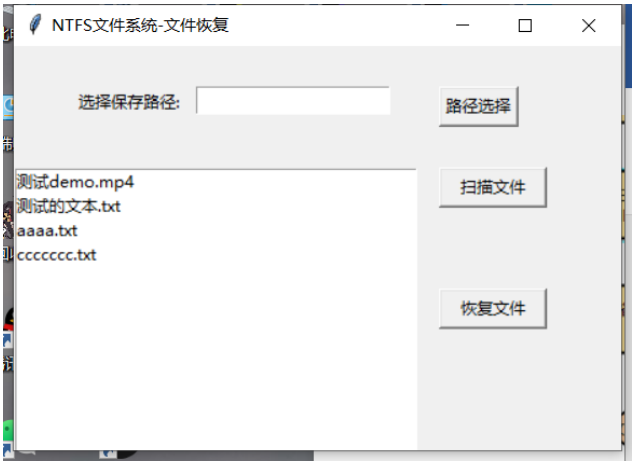
选择需要进行扫描的磁盘



图（9）选择目标盘

4. 进行扫描

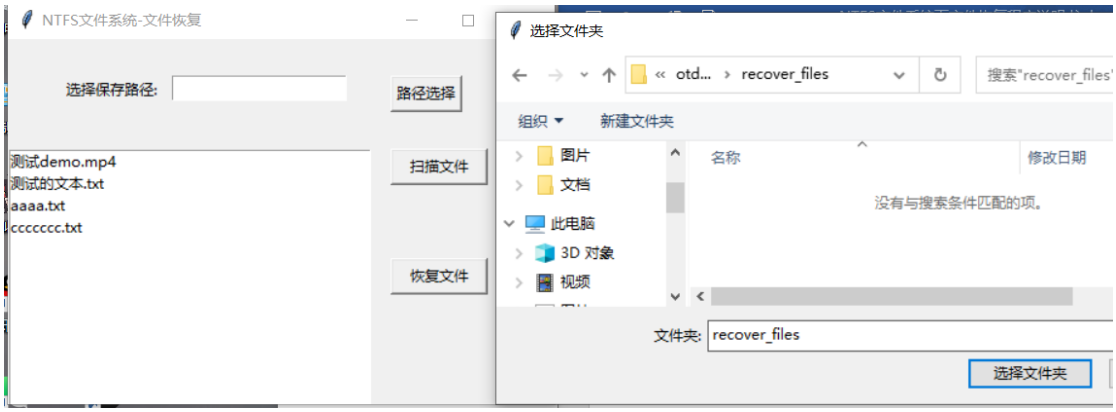
进行扫描后可以得到一个列表，可以看到，刚才删除的文件被扫描到，也包含以前被删除的文件。



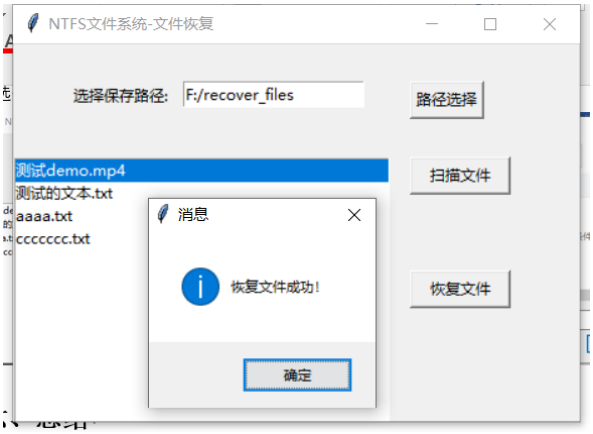
图（10）进行扫描

5. 选需要恢复的文件和保存路径

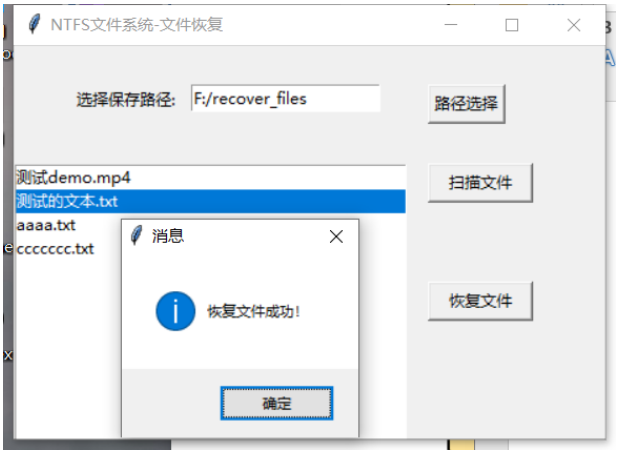
选择 F 盘下的 recover_files 文件夹作为恢复文件输出目录，分别选择恢复“测试 demo.mp4”、“测试的文本.txt”。



图（11）选择保存路径



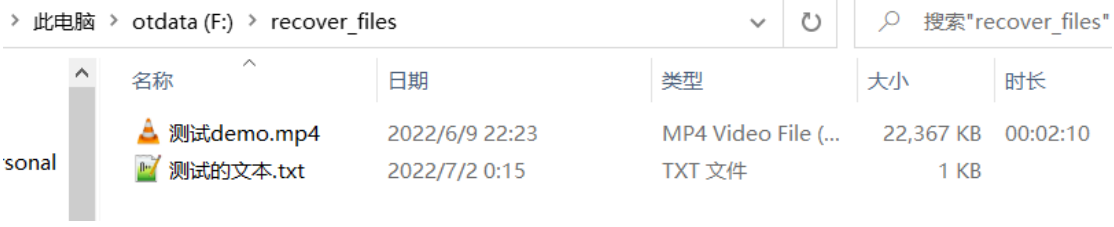
图（12）恢复“测试 demo.mp4”



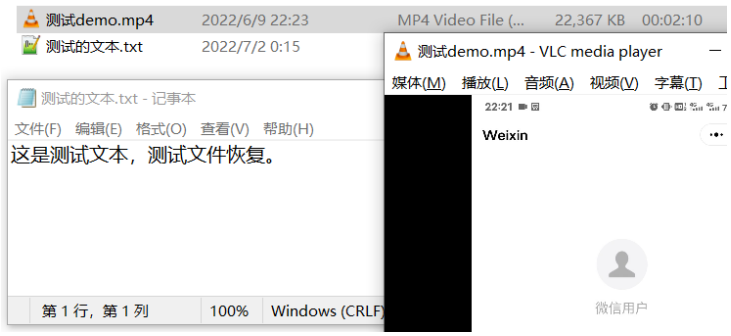
图（13）恢复“测试的文本.txt”

6. 查看恢复结果

可以看到已经成功恢复，分别打开文本编辑器和视频播放器进行查看。



图（13-1）恢复结果



图（13-2）恢复结果

六、总结

本程序最终完成了一个拥有可视化界面供用户操作、可选定任意磁盘进行扫描以得到该磁盘下的可恢复文件并以列表形式呈现给用户、可恢复用户指定的列表中的可恢复文件至指定文件夹的完整程序，完成度与最初对该程序的设计较为吻合。在完成该程序的过程中，本组成员也遇到了各种各样的困难，我们深刻体会到了小组合作的魅力、python 作为一门脚本语言的便捷性、NTFS 文件系统数据结构的严谨性，NTFS 文件系统删除覆盖文件的流程以及 NTFS 文件系统下文件恢复过程的魅力。而在这其中，我们把从数字取证课程中所学习到的知识、一次次实验积累的经验一步一步地整合转化为一个可运行，完成度高的程序成果的经历是最让我们受益匪浅的。这也是数字取证课程带给我们这个小组最宝贵的经验。

七、小组分工

成员	学号	备注	负责内容
李惠奕	2019052207	组长	核心代码编写，程序测试，程序说明书撰写
柯俊伟	2019052206		可视化界面代码编写，程序测试，程序说明书撰写
柯智耀	2019052205		代码编写，程序测试，程序说明书撰写
王宥竣	2019052208		代码编写，程序测试，程序说明书撰写