Computer Science and Application 计算机科学与应用, 2017, 7(12), 1245-1253
Published Online December 2017 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2017.712139

Hans汉斯

Design and Implementation of Data Recovery System Based on NTFS

Fujun Feng, Junping Yao, Xiaojun Li

Research Institute of High-Tech, Xi'an Shaanxi
Email: f_fjun@163.com, junpingy200225@163.com, xi_anlxj@126.com

Received: Dec. 3rd, 2017; accepted: Dec. 15th, 2017; published: Dec. 22nd, 2017

Abstract

Windows NTFS is a kind of file systems with good stability and security, and it is important for data security to research on data recovery based on NTFS. Analyzing the NTFS structure, a data recovery system based on NTFS is designed and implemented, and three modules are presented particularly, which include modules of disk analysing, partition scanning and data recovery. The data recovery module is the core of the recovery system, and the implementation process of dada recovery based on NTFS is provided. Through soft testing, the system proposed can recover the original data very well.

Keywords

NTFS, Data Recovery, MFT

基于NTFS的数据恢复系统设计与实现

封富君，姚俊萍，李晓军

西安高新技术研究所，陕西 西安
Email: f_fjun@163.com, junpingy200225@163.com, xi_anlxj@126.com

收稿日期：2017年12月3日；录用日期：2017年12月15日；发布日期：2017年12月22日

摘　要

Windows下的NTFS文件系统具有较好的稳定性和安全性，对NTFS下数据恢复软件的研究对数据的安全保护具有重要作用。分析了NTFS文件系统的结构，设计并实现了一个基于NTFS的数据恢复系统，详细阐述了系统的三个主要模块的实现过程，包括磁盘分析模块、分区扫描模块和数据恢复模块。数据恢复

文章引用: 封富君, 姚俊萍, 李晓军. 基于 NTFS 的数据恢复系统设计与实现[J]. 计算机科学与应用, 2017, 7(12): 1245-1253. DOI: 10.12677/csa.2017.712139

模块是恢复系统的核心模块，给出了NFTS数据恢复实现的具体流程。软件测试结果表明该软件能够很好的恢复原始数据。

关键词

NTFS，数据恢复，MFT

1. 引言

随着网络技术的不断发展，以及数据量的急剧增加，在大数据环境下，越来越多的用户习惯于把重要资料保存到计算机硬盘中，却不得不面对数据可能丢失的风险。数据恢复是信息安全的最后一道防线，是保护数据安全的重要方法。数据恢复技术就是因计算机系统遭到误操作、病毒侵袭、硬件故障、黑客攻击等事件后，将用户数据从各种“无法读取”的存储设备中拯救出来，从而将损失减到最小的技术[1]。

数据恢复根据故障可分为两大类：软恢复和硬恢复。由于磁盘故障导致的数据恢复称为硬恢复，例如磁道或盘片损坏。软恢复是指通过软件的方式进行的数据恢复，整个过程并不涉及硬件维修。产生的原因主要有误操作，如误格式化或误删除；恶意程序的破坏，如病毒感染；操作系统错误，如系统死机导致的数据丢失等。大多数的数据恢复属于软恢复。在数据恢复之前应该对数据存储介质进行全面的了解和检查，并根据实际情况制定科学合理的数据恢复方案。数据恢复涉及的领域较广，具有电子数字数据和数字文档存储的地方都会有数据存储、数据损坏和数据修复的问题，目前对计算机的数据恢复[2] [3] [4] [5] [6]和电子取证中的数据恢复[7] [8]研究较多，本文则设计并实现了一个数据恢复系统。

目前 Windows 都采用 NTFS 文件系统，NTFS 是微软公司开发的一种文件系统，具有较好的容错性和安全性，它最大的优点是稳定性好，适用于对数据安全性要求比较高的应用。因此对 NTFS 下的数据恢复软件的研究对数据的安全保护具有重要作用。

2. NTFS 文件结构

2.1. 数据存储结构

硬盘主要用来存储计算机上的各类数据，包括系统程序或用户文件，目前硬盘的存储量也越来越大，了解硬盘的结构对于研究数据恢复是非常有必要的。当磁盘旋转时，磁头若保持在一个位置上，则每个磁头都会在磁盘表面划出一个圆形轨迹，这些圆形轨迹就叫做磁道。磁盘上的每个磁道被等分为若干个弧段，这些弧段便是磁盘的扇区，每个扇区可以存放 512 个字节的信息，磁盘驱动器在向磁盘读取和写入数据时，要以扇区为单位。硬盘通常由重叠的一组盘片构成，每个盘面都被划分为数目相等的磁道，并从外缘的“0”开始编号，具有相同编号的磁道形成一个圆柱，称之为磁盘的柱面。磁盘的柱面数与一个盘面上的磁道数是相等的。由于每个盘面都有自己的磁头，因此，盘面数等于总的磁头数。

2.2. 主文件表 MFT

在 NTFS 文件系统中，将文件系统所需的全部数据像文件一样储存在硬盘上，如记录卷的分配状态

位图、文件、目录和系统引导程序等数据。这些系统文件称为元文件，这些数据称为元数据。一个文件在磁盘上的位置是通过主控文件表 MFT 来定位的。MFT 是 NTFS 文件系统的核心，它是一个关系数据库，由文件记录的数组组成，磁盘卷上的每一个文件都有一个文件记录，大的文件可能有多个记录。每个文件记录的长度是固定的(一般为 1 KB)，文件记录在 MFT 中从 0 开始编号，MFT 的前 16 个文件是元文件，以“$”开始，它们是隐藏文件。这 16 个文件是 NTFS 文件系统中唯一在 MFT 表中具有固定地址的文件，其他文件和目录的文件记录可以存放在 MFT 表中的任意地方。

MFT 的第一个文件记录就是$MFT 自身，它记录着所有文件和目录的所有情况，由于$MFT 文件本身的重要性，为确保文件系统结构的可靠，系统专门为它的起始部分记录准备了一个镜像文件($MftMirr)，也就是 MFT 中的第 2 个记录。

3. NTFS 下的数据恢复

在 NTFS 文件系统中删除一个文件时，数据所在区域并没有被覆盖，同时，其文件系统中的相关记录所在区域也没有被清除。只是在三个方面做了改变：该文件 MFT 头偏移 0x16H 处的一个字节改变；其父文件夹的索引根属性或者索引分配属性改变；在位图元数据文件中把该文件所占用的簇对应位置置 0。文件删除后的关键信息如下：

1) BPB 参数

BPB 参数是 NTFS 文件系统安装卷的依据，其含义为 BIOS 参数块。数据恢复时要用到相关的参数有 MFT 的起始簇号：MfrStartCluster、每簇扇区数：SeetorPerCluster，以及每扇区字节数：BytesPerSector。

2) MFT 变化

MFT 文件记录头是以 “46 49 4C 45”开始，即用 ASCII 码表示的“FILE”开始，当文件删除后不会改变。在文件记录头部中，偏移 0x16H 处的一个字节是标志字节，00H 表示删除的文件，01H 表示正常的文件，02H 表示删除的目录，03H 表示正常的目录，从偏移 0x2CH 开始的 4 个字节是文件号是 MFT 的记录编号，从 0 开始。

要恢复 NTFS 文件系统中所有被删除的文件，可以搜索所有被取消分配的 MFT 项，找到后根据项中的文件名属性和父目录的文件参考号确定文件的名字和路径。簇指针如果完好存在，且文件内容没有被覆盖，就可以完好的恢复出来。即使文件严重片段化，有多个文件碎片也不影响恢复。如果属性值是常驻属性，在 MFT 被重新分配前，数据是不会被覆盖的。如果文件不止一个 MFT 项存储它的属性，就需要对其他的 MFT 项进行恢复。

4. 数据恢复系统的设计与实现

4.1. 主要模块实现

本文设计的数据恢复系统主界面如图 1 所示。

在 NTFS 文件系统中，文件删除之后，系统所做的工作是：在文件记录头部中将标志字节置为 00H 或 02H，文件记录的其它属性均没有变化；对于有数据运行的文件，回收文件所占用的空间，不改变数据区的内容。由此，提出数据恢复策略，设计软件的主要功能模块为：磁盘分析模块、分区扫描模块和数据恢复模块。

4.1.1. 磁盘分析模块

磁盘分析的目的是对 NTFS 分区进行定位，主要分析磁盘的主引导记录 MBR、扩展引导记录 EBR，以及主文件记录 MFT 的相关信息。MBR 位于整个硬盘的 0 磁道 0 柱面 1 扇区。在 512 字节的主引导

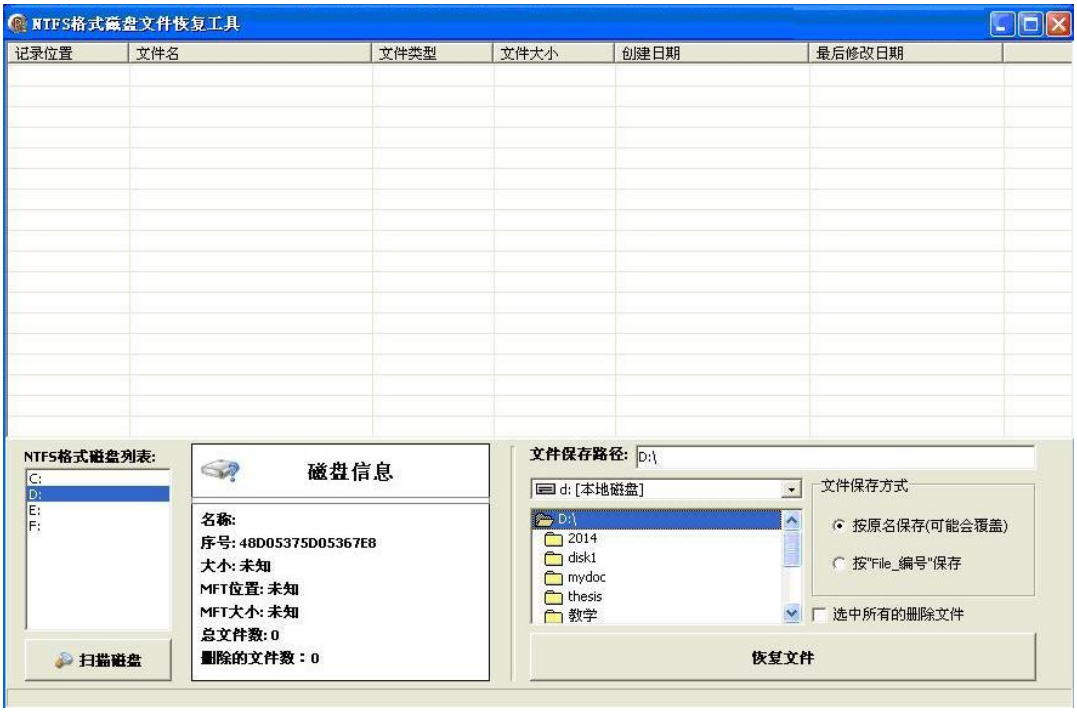


Figure 1. Main interface of data recovery system
图 1. 数据恢复系统主界面

扇区中，MBR 的引导程序占用其中的前 446 个字节，随后的 64 个字节就是 DPT，每 16 个字节定义一个分区，最多可以确定 4 个分区，其中第 5 个字节表示了该分区类型，其中 0x07H 表示 NTFS 分区，每个分区表项标示着一个分区信息或一个扩展分区表的位置，而扩展分区表中可能还有子扩展分区表，这就形成了一个链状结构，可以记录多个分区。

扩展分区沿用了基本分区所采用的 DPT 结构，把扩展分区的分区表所在的扇区称为 EBR，扩展分区的类型是 0x05H。所有扩展分区以链表的形式级联存放，后一个扩展分区的数据记录在前一个扩展分区的分区表中，但两个分区表的空间并不重叠。EBR 存储在扩展分区的第一个扇区中，且最多只能表示两个分区的数据项。

4.1.2. 分区扫描模块

分区扫描主要是扫描 NTFS 分区中 MFT 的文件记录，通过查看每一条文件记录偏移 0x16H 开始的两个字节，判断该文件是否为已删除文件，通过分析 30 和 10 属性获取文件名、文件创建时间和文件最后修改时间等信息。图 2 为 D 分区中所有被删除文件的扫描结果。

4.1.3. 数据恢复模块

当目录的属性值存放在 MFT 表的基本文件记录中，该属性就称为常驻属性。对于常驻属性，属性值存放在属性名的后面。如果一个目录的属性值太大，不能存放在一个文件记录中，那么，NTFS 将从数据区为该属性值分配存储空间。这些存储空间通常称为一个运行，用来存放属性值，存储在运行中的属性称为非常驻属性。

判断文件记录的 80 属性是否为常驻属性，分类获取数据内容，在其他分区中新建一个文件，用已删除文件的文件名来命名，然后将数据区的文件内容复制到新的数据区中。数据恢复流程图如图 3 所示。

数据恢复模块 Delphi 语言关键代码如下：

Figure 2. Partition scanning module
图 2. 分区扫描模块

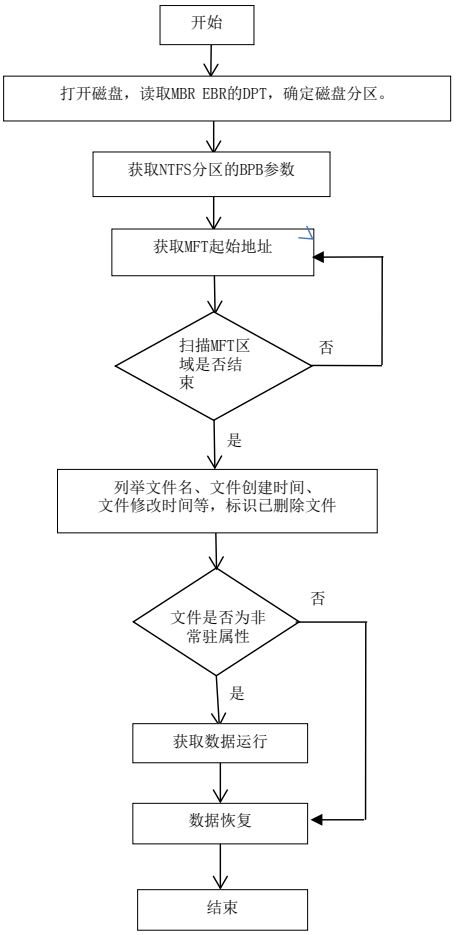


Figure 3. Flow diagram of data recovery
图 3. 数据恢复流程图

```
unit uRecoverThread;
interface
uses
  Classes,Windows,SysUtils,uNTFS;
type
  TRecoverThread = class(TThread)
  private
    { Private declarations }
    FdiskInfo    : TDISK_INFORMATION;
begin
  { Place thread code here }
  isRecovering := true;
  with MainForm do
  begin
    PB.Max := FFileCount;
    btnRecover.Caption := '停止恢复';
    btnScan.Enabled := false;
    for i := 0 to ListView1.Items.Count -1 do
    begin
      if StopRecover then
      begin
        break;
      end;
      if listview1.Items[i].Checked then
      begin
    btnRecover.Caption := '恢复文件';
    btnScan.Enabled := true;
  end;
```

4.2. 软件测试

该数据恢复软件测试过程为：

1) 在 D 分区下创建 123.doc 文件，并在备份文件夹中对原文件进行备份。

2) 将 123.doc 文件删除后清空回收站。

3) 打开数据恢复工具，在列表中选择待恢复文件所处的盘符 D，点击扫描磁盘按钮，对该盘符进行扫描，程序将扫描出已经删除的文件列表。

4) 在删除文件列表中查找需恢复的文件，并选定该文件，并将文件保存到 E 盘，图 4 为在磁盘扫描结果中选定待恢复文件。文件恢复时可以选择“按原名保存”和“按文件编号保存”，如果选择原名保存可能会覆盖原来的数据存储区域。图 5 分别为两种情况下的数据恢复文件，即 Dd161.doc 和 File_8049.doc。

5) 打开恢复后的文件 Dd161.doc，与保存的备份文件进行对比，内容完全一致，表明文件恢复成功，如图 6 所示。同样，文件 File_8049.doc 的内容与原文件也完全一致。

Figure 4. Result of disk scanning
图 4. 磁盘扫描结果

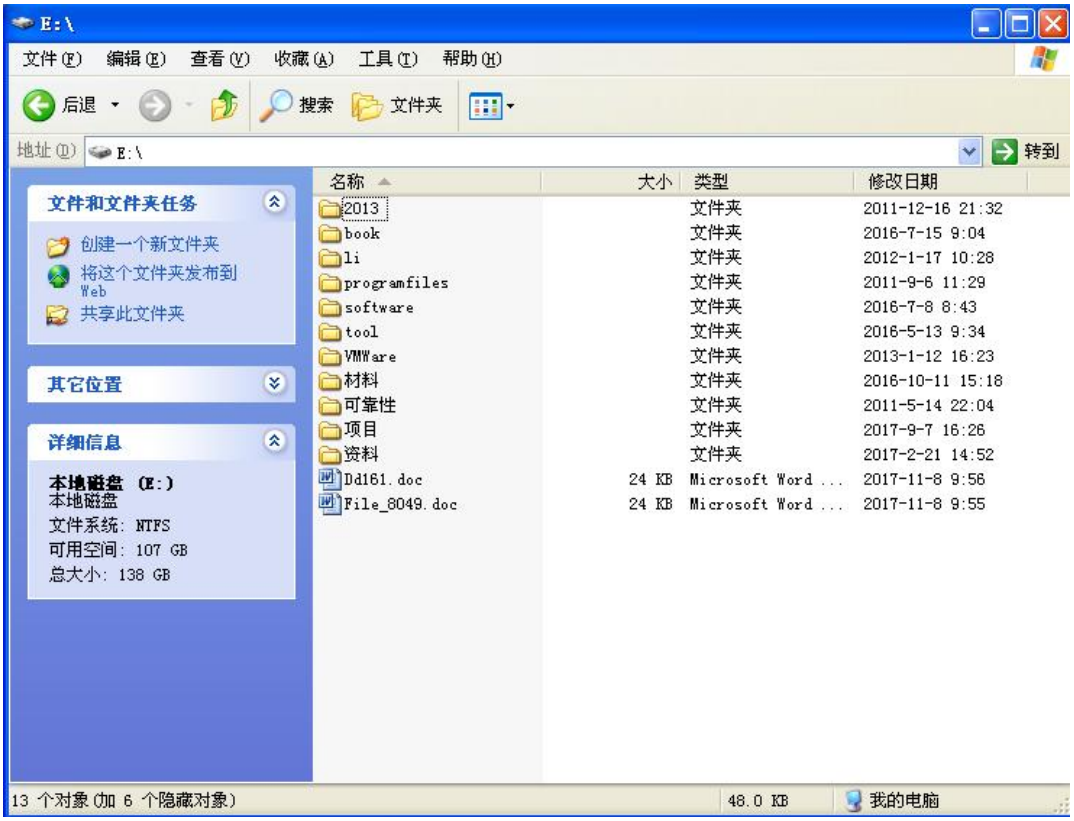
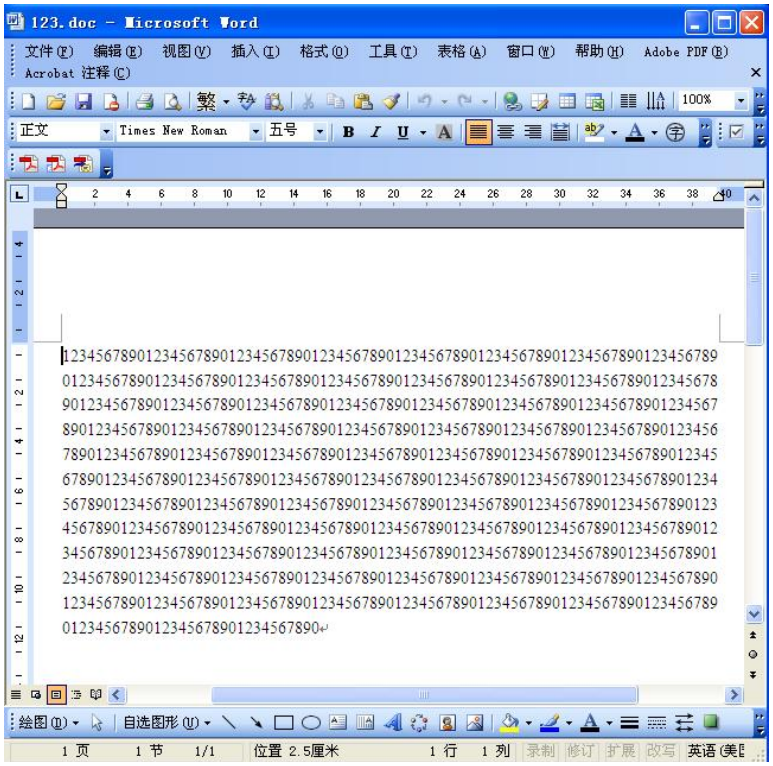
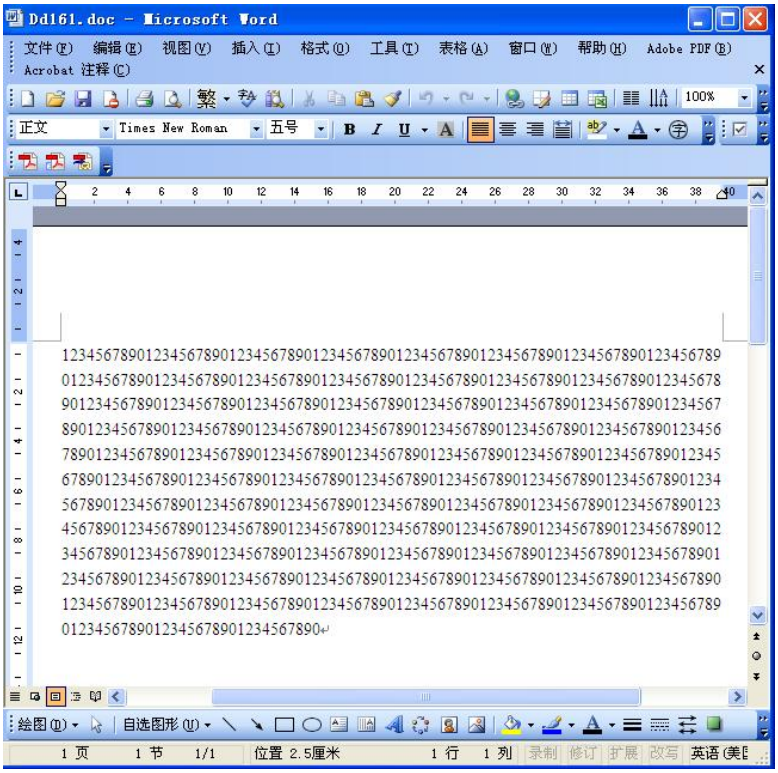


Figure 5. Result of data recovery
图 5. 数据恢复结果



(a)



(b)

Figure 6. Result comparison of data recovery. (a) Original contents; (b) Recovered contents
图 6. 数据恢复后的文件内容对比。(a) 原文件内容；(b) 恢复后的文件

以上测试为文件删除后的数据恢复，对于文件格式化后的恢复过程与此类似，不再赘述。对于其他类型的文件格式，如视频文件、音频文件、压缩文件等经测试后都能够很好的恢复原始数据。

5. 结束语

本文对 NTFS 文件系统进行研究，并基于 NTFS 设计了一个数据恢复系统，主要实现了三大功能模块：磁盘分析模块、分区扫描模块和数据恢复模块。经过测试文件删除后的恢复等一系列操作，能够实现基于 NTFS 文件系统文本文件、图像文件以及视频文件的恢复，实现了数据的安全保护。下一步将对系统的恢复速度进行优化，并完善恢复系统的更多功能。

参考文献 (References)

[1] 戴士剑, 涂彦晖. 数据恢复技术[M]. 北京: 电子工业出版社, 2014: 1-10.

[2] 张志, 陈欣欣, 李世强. 重要数据恢复平台建设研究[J]. 信息网络安全, 2016(s1): 146-148.

[3] 王薇. 计算机数据恢复技术及其应用分析[J]. 电子技术与软件工程, 2015(7): 207-209.

[4] 王彩霞. 数据恢复技术的分析与实践[J]. 信息与电脑, 2012(6): 88-92.

[5] 徐仙伟, 杨雁莹, 曹霁. Windows 系统中文件级数据恢复方法分析研究[J]. 皖西学院学报, 2014, 30(2): 24-27.

[6] 莫启, 代飞, 朱锐. 业务过程协同中数据恢复策略建模及分析[J]. 计算机应用研究, 2016, 33(2): 462-466.

[7] 戴芬, 李璐. 电子取证中的数据恢复技术研究[J]. 信息通信, 2016(1): 123-125.

[8] 马国富, 王祥珩. 数据恢复在电子数据取证与司法鉴定中的应用[J]. 河北大学学报, 2015, 36(5): 538-545.