CDF BOOT CAMP
电子数据取证技术训练营

实验

取证基础/文件过滤

目录

2 取证基础-文件过滤

2.1 实验内容

文件扩展名，可以理解为一种软件特有的格式定义。通过扩展名，Windows 可以帮助我们搜索到这些相同扩展名的文件。对于文件，我们一般还会进行分类，例如：文档类、图片类、视频类、邮件类、压缩类等等。取证软件可通过过滤来对所需要的文件进行快速查找。过滤可以通过文件名、文件类型、时间、大小、位置等方式，帮助我们更加快速地寻找到如“2016 年制作的大于 1GB 的视频”、“所有 2018 年复制到本地的 doc 文件”等等。过滤是取证调查中非常有效的数据分析方法。而 X-Ways Forensics 和法证通采用相同的过滤机制，是所有取证软件中过滤最直接、效果最好的工具。而 X-Ways Forensics 支持大量的属性过滤，需要全面掌握才能发挥出过滤的优势，缩短分析时间。

- 文件过滤：什么是过滤？怎么使用过滤
- 文件名和扩展名过滤：*.DOC, A*.DOC，?
- 文件类型库：对文件类型的定义，如图片类
- 文件类型描述：文件类型的创建软件,pst:OUTLOOK
- 排序：升序、降序
- 通过文件名过滤：常用办公文档、常用图片
- 通过文件类型过滤：所有邮件，所有图片
- 组合过滤：大于 1MB 的 PDF 文件，小于 4K 的文件等
- 保存过滤条件：所有大于 1KB,小于 100MB 的所有办公文档
- 修改文件类型库：
- 高级过滤条件：通过属性过滤，如$j 文件的过滤
- 排除和隐藏：路径中不包含 OFFICE 的文件；隐藏小于 4KB 的._文件。

2.2 实验目的

实际数据分析过程中，经常需要对某一个文件或某一类文件进行过滤，以便缩小范围，查找到我们所需要的准确数据，提高工作效率。例如，时间案件中，调查员经常需要快速过滤出当前案件中的 Office 文档、电子邮件，也可能会需要查找一个操作系统注册表文件或上网记录，或者要将案件中所有的图片查找出来。

本节实验，重点学习文件过滤，快速找到所需要的文件类型。掌握过滤的操作方法、理解组合过滤。结合相关知识点，达到可以利用系统信息、文件属性快速找到所需文件的目的。

2.3 实训资源

C01-CCFC-Windows XP.e01

2.4 背景知识

2.4.1 浏览递归

展开目录

X-Ways Forensics 中的目录需要层级展开，而且每一层都需要专门选择浏览，用户需根据需要选择指定分区来展开目录并浏览。

展开所有证据项下的数据

在案件目录窗口**右键点击**案件根目录，并选择需要展开的分区；展开所有证据后，可以列出分区下的所有的文件。

应用场景：我们需要统计整个磁盘所有分区下有多少个文件？显示所有磁盘中的被删除文件？预览所有磁盘中的图片的缩略图？都需要使用这个操作。

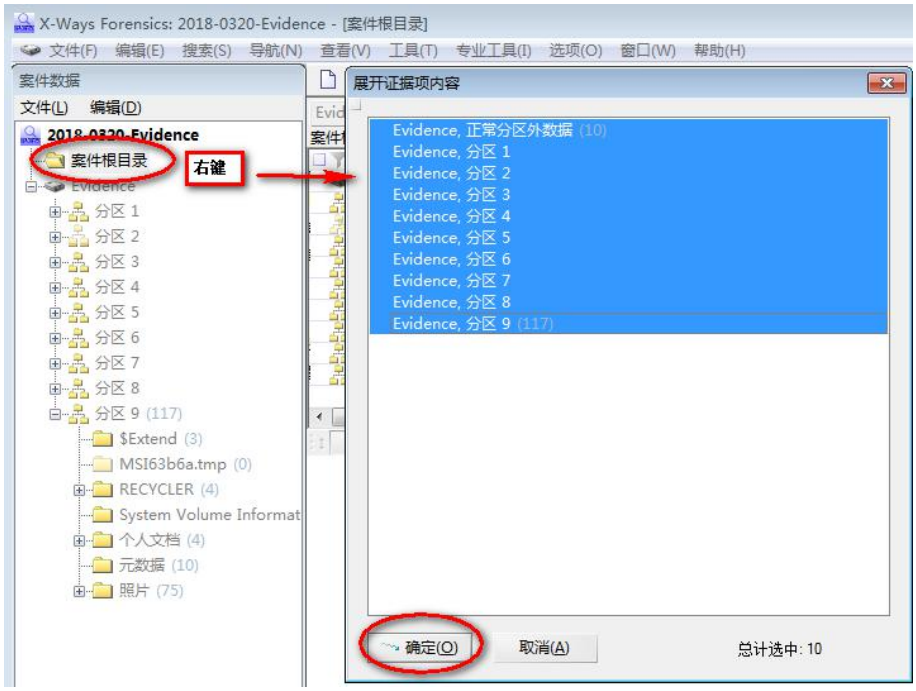


图 2　选择分区展开证据界面

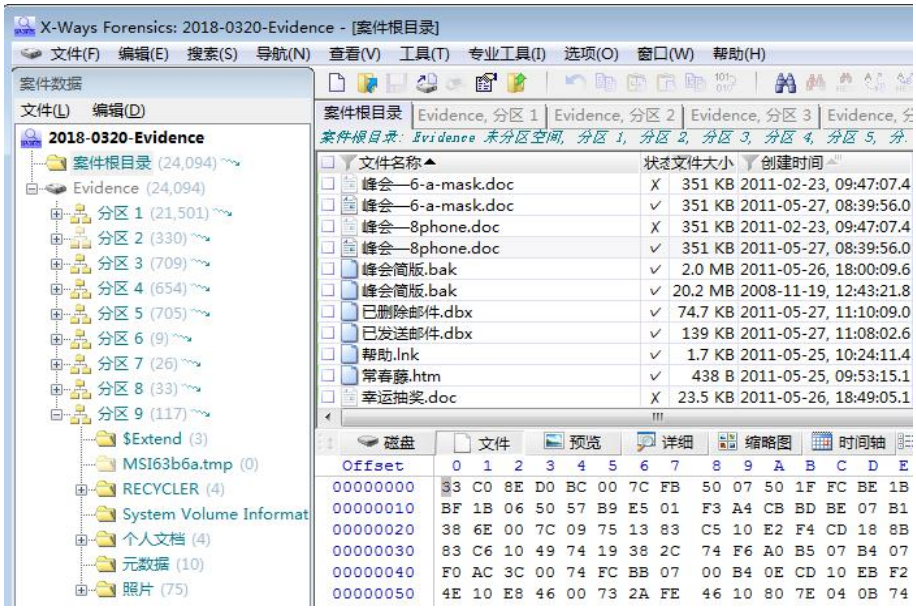


图 3 证据展开后的效果

展开某个分区

鼠标右键点击分区 1，选择“浏览递归”，则能查看该分区内的所有文件。

应用场景：查找 C 盘中的所有注册表文件；C 盘中的所有办公文件等。

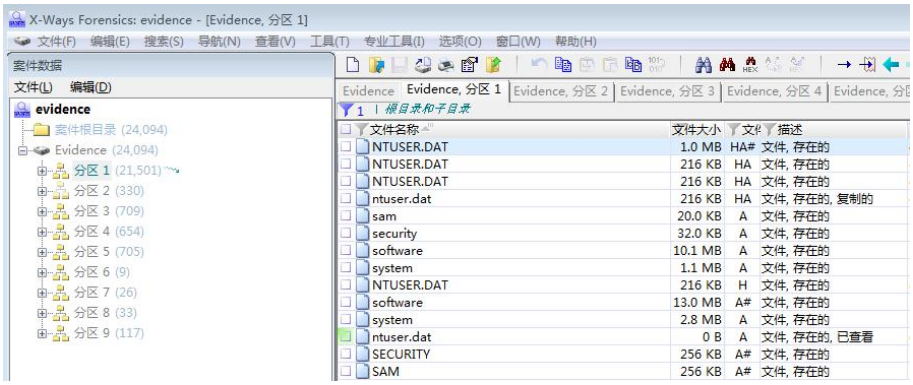


图 4 分区展开界面

右键点击某一个目录，可以显示该目录下的所有文件。

应用场景：查找 C 盘 Document and Settings 目录中的所有 doc 文件、Windows 目录下的注册表文件。

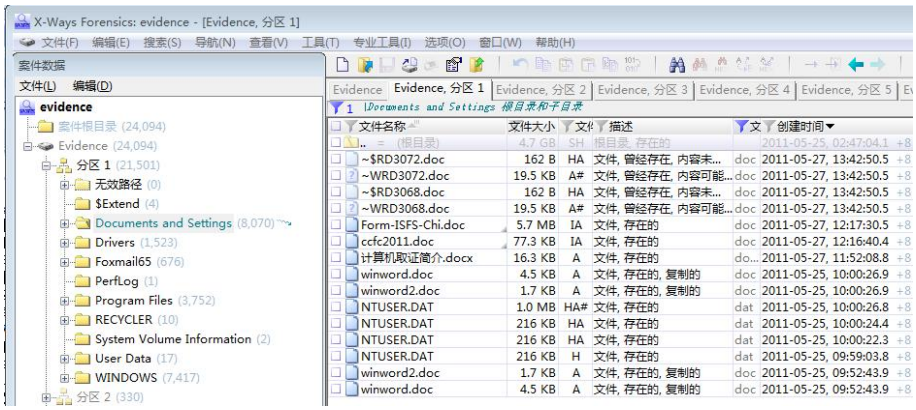


图 5　查看分区 1 某一个目录的所有文件

2.4.2 过滤

过滤，即按照设定的条件查找符合条件的数据。Myhex 具有强大的过滤功能，可以通过各种组合条件实现精确的数据查找

通俗来说，过滤的主要功能就像这个形象的漏斗一样，把想要的东西留下来，将不要的东西筛走。通过过滤可以将复杂的操作简单化，快速找到自己想找的文件。在“目录浏览及过滤设置”窗口中，所有带有漏斗的栏目都可以进行过滤操作。显示灰色漏斗的，表示未启用过滤选项；如果显示为蓝色漏斗的，表示当前已应用了过滤设置。

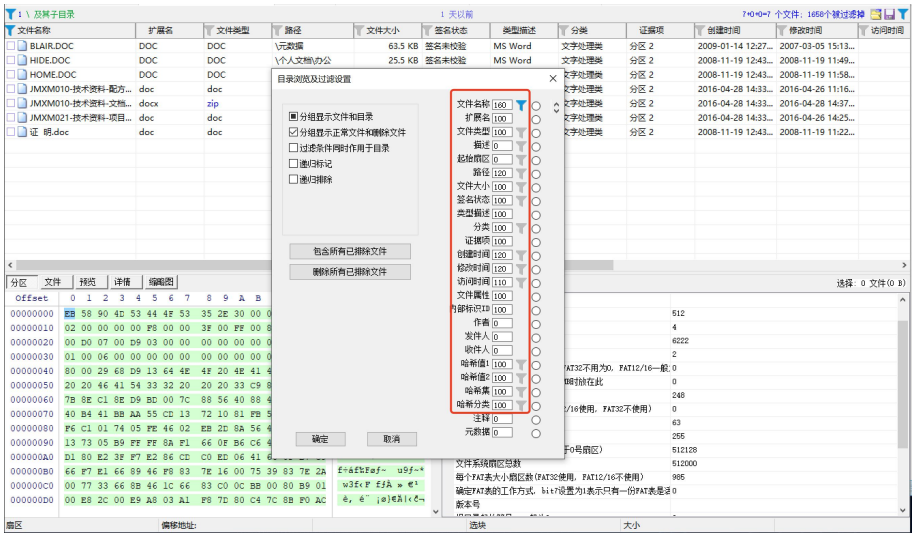


图 1 可用于过滤的列

例如：如果调查员想找到所有在 2008 年 11 月 19 日创建的文件，可以利用时间排序功能，将所有的文件全部按照升序或降序排列，然后找到创建时间是 2008 年 11 月 19 日的文件。但是，利用时间过滤则更加有效，可以直接将创建时间是 2008 年 11 月 19 日的所有文件显示出来，将不属于这个日期的文件隐藏掉。

2.4.3 文件名过滤

文件名称

每个文件都有一个名字，称为文件名，它由字母、数字或字符组成。文件名又可分割为主文件名和扩展文件名，就拿“数字取证.docx”为例，“数字取证”就是主文件名，主要说明文件的内容，docx 为扩展文件名，它主要说明文件的性质(在这里表示 word 文档)，中间的小数点为主文件名和扩展文件名的分隔符。在 DOS 下，文件名采用 8+3 结构，即：最长 8 位的文件名，由小数点分隔后再跟上最长 3 位的后缀名，如：READ.ME、SETUP.EXE。

通常，采用文件名进行过滤是最简单的方法。为了快速过滤某一类文件，或与某个字符相关的文件名或目录名称，Myhex 允许通过文件通配符 * 号或？配合过滤。当通配符位于文件名的最前面和最后面时，最多使用 2 个星号。

*.doc	查找所有扩展名是 doc 的文档
.doc	查找所有扩展名中包含 doc 的文档，例如*.docx
*.jpg	查找所有 JPG 图片
1.gif	查找文件名为 1.gif 的文件
峰会*.bak	文件名是中文峰会为起始字符，扩展名为 bak 的文件
技术资料	文件名中包含“技术资料”四个字符的所有文档

2.4.4 依据时间过滤

可按照设定的文件分类，对不同类型的文件进行过滤。通过此过滤方式，可以容易地将办公文档、图形图像、压缩文件，

Windows 时间属性很多。后续章节会详细介绍创建时间、修改时间、访问时间、记录更新时间、删除时间、内部创建时间等。

创建时间创建时间代表文件在一个位置生成的时间。

修改时间文件被最后编辑、写入数据的时间。

访问时间文件被访问的时间，是计算机系统本身对文件进行了某种操作的时间。最常见的行为是：文件打印、查看（打开并未保存）。此外病毒检测、文件备份、系统维护等操作都会改变文件访问时间。FAT 文件系统仅记录访问日期。NTFS 文件系统下的访问时间可以记录至秒。

记录更新时间 NTFS 文件系统 FILE 文件记录(FILE record)、Linux 文件系统索引节点(inode)中文件和目录的最后发生变化修改时间。这是文件系统数据结构中包含的数据时间信息。索引节点：在 Linux 文件系

统下，每个存储设备或存储设备的分区被格式化为文件系统后，包含两部份，一部份是索引节点，另一部份是块区(Block)。块区是用来存储数据用的，索引节点是用来存储数据的信息，包括文件大小、属性、归属的用户组、读写权限等。 索引节点为每个文件进行信息索引，所以就有了索引节点的数值。Linux 系统根据指令，能通过索引节点值快速地找到相对应的文件。

删除时间是一个最难判断的问题。我们很难从文件时间属性直接来判断某个文件到底是什么时候被删除的。因为文件系统并不直接记录文件的删除时间，某些 NTFS 元数据会记录删除时间，回收站可以记录某个文件是的删除时间。不同的取证分析工具对删除时间的认识和解析方法不同。X-Ways Forensics 中，可以显示 Linux 文件系统或 NTFS 文件系统（在对文件系统对$UsnJrnl:$J 文件解析后）某些目录和文件的删除时间。

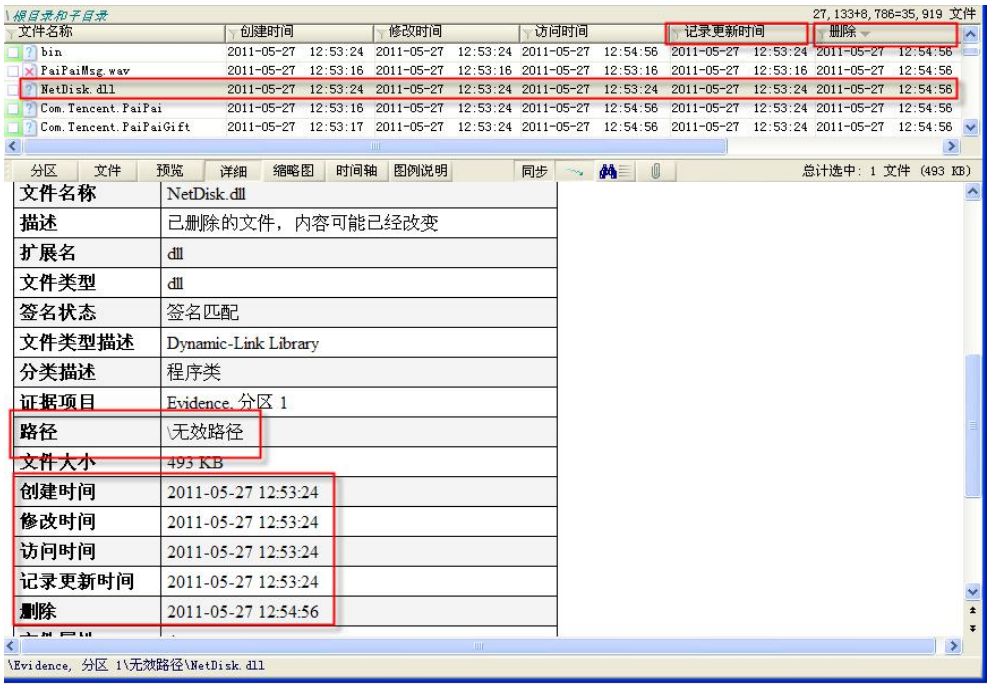


图 14 各种时间属性

内部创建时间文件元数据中记录的文件真正的创建时间。Microsoft Word 和 Excel，数码相机、手机拍摄的图片都包含有时间属性。内部创建时间通常不易被人为修改，也不会被文件系统自动修改。可以通过内部创建时间和其他时间属性一同分析判断用户行为。

图 15 提取内部创建时间

此外，某些应用程序还有可能保存一些其他的时间信息，例如文件打印时间。

图 16 元数据中包含的打印时间

2.4.5 磁盘快照

什么是磁盘快照？

为了实现数据的全面和自动化处理，提升工作效率，降低取证分析人员的工作量，X-Ways 提供了证据预处理功能，称作磁盘快照。

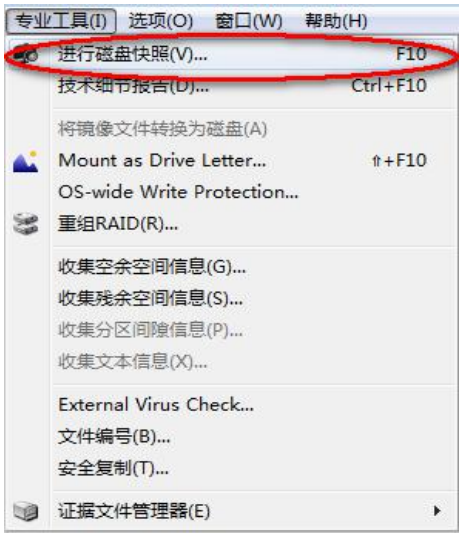


图 9 调用磁盘快照

磁盘快照功能包括：文件恢复、文件签名恢复、哈希校验、复合型文件提取、电子邮件内容提取等。

更新快照：是重新进行磁盘快照，将之前所有的解析结果全部清除。如果经过了几个小时解析了硬盘的所有数据之后，一定要慎重选择“更新快照”。否则您又要重新等待几个小时了。

在选定的证据中搜索：选择进行指定操作的证据项。例如在一个分区中，还是在九个分区中，或是在三个硬盘中进行指定的快照操作。

应用于所有文件和应用于所有标记的文件：如果我们只想针对所选的一类文件操作，可以将这些数据进行标记，然后在标记的这几个文件中进行操作。例如，可以提取所有“PPT”中的图片。则先标记所有的 PPT 再进行操作。

图 10 磁盘快照

2.5 实验步骤

2.5.1 实验 1:过滤指定名称的文件

查找所有 DOC 文档：点击“文件名称”右侧的灰色漏斗，输入过滤条件*.DOC，点击激活即可。此外，也可以同时输入多个过滤条件，如下图所示，可同时准确文件名为“index.dat”的文件。

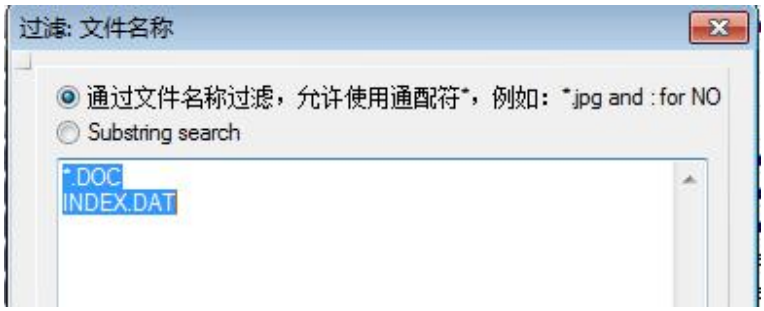


图 5 同时过滤多个条件

应用上述过滤后的结果。当前分区中符合上述 2 个条件的所有文件都被显示出来。

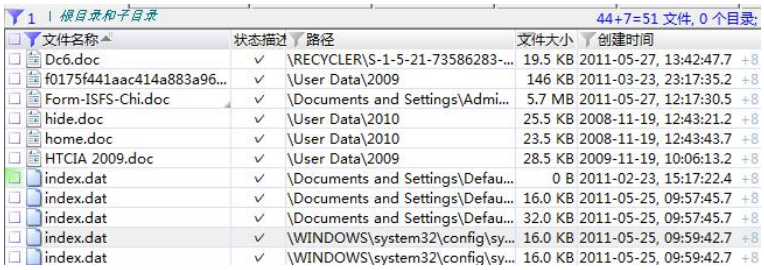


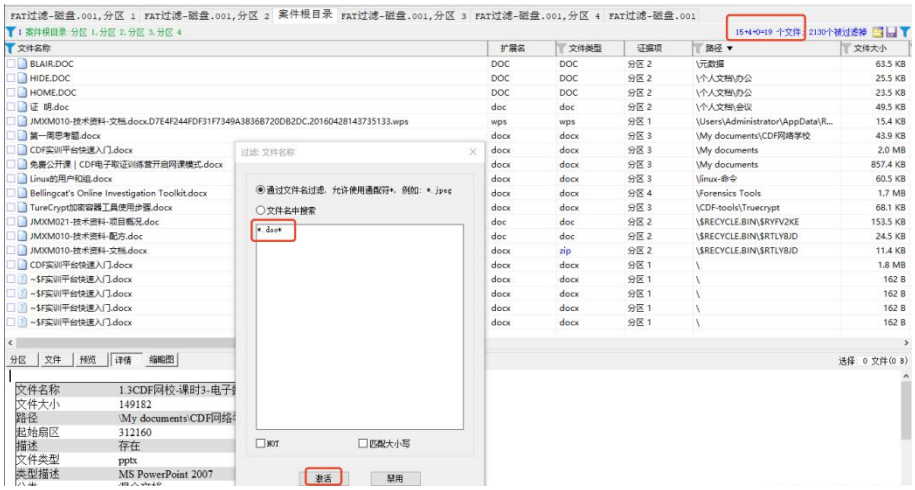
图 6 过滤结果

应用过滤之后，对任何目录操作都会自动应用此过滤。有时你可能会忘记了自己启用了过滤，会因为在当前目录下看不到文件而奇怪。其实，只要看到栏目上醒目的蓝色漏斗，就应该立刻想到，是因为启动了某个过滤条件而影响了文件的浏览。**需要取消某个过滤条件，**可调用目录浏览器过滤设置对话框，选择已经应用的过滤条件，点击“禁用”即可。也可单击两端的蓝色漏斗，直接取消所有过滤。

图 7 过滤条件被激活，没有当前过滤结果

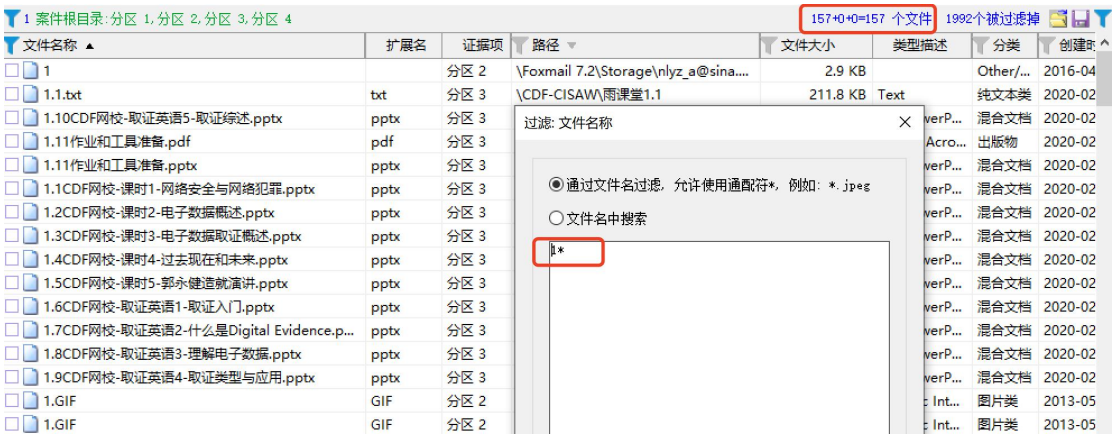
练习：

1. 查找所有分区中，扩展名为 doc 和 docx 的所有文档有多少个？

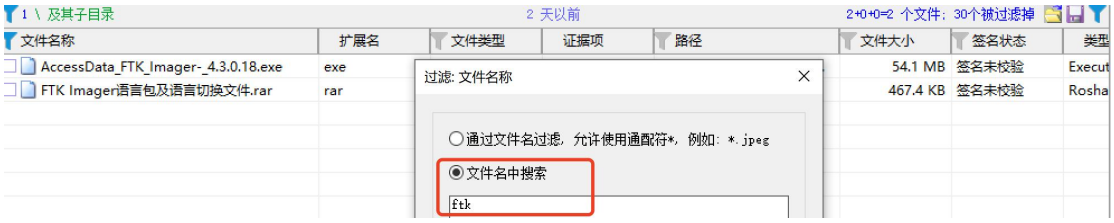


2. 查找分区 3 中，文件名为 1. png 的文件有几个？

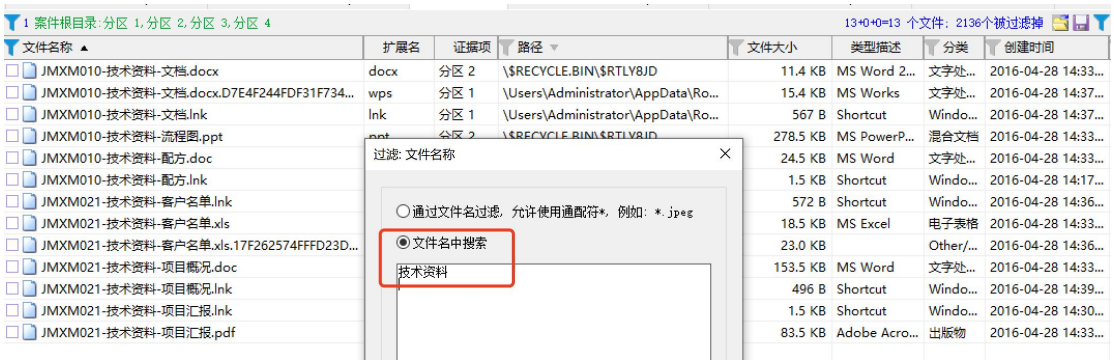
3. 查找所有分区中，文件名以 1 为起始字符的文件有多少个。



4 查找 FTK 软件保存在哪个分区？



5. 查找所有分区中，文件名包含“技术资料”四个字符的文件有多少个？



2.5.2 实验 2:查看“峰会简版.BAK”保存在哪个案件中

配合案例文件：C01-CCFC-Windows XP.e01

当一个案件中加载了多个证据文件时，很难快速掌握过滤到的文件在哪个分区下。通过“证据项目”过滤，可显示出文件或目录所属的证据磁盘。

图 11 查看“峰会简版.BAK”的保存位置

2.5.3 实验 3:去除 Windows 目录下的所有 TXT 文件

配合案例文件：C01-CCFC-Windows XP.e01

有时，过滤出的文件很多。我们会看到有些目录下的文件有用，是我们关注的；有些目录下的文件没用，且无用文件数量很多。此时，我们可以将无用的目录下的文件排除掉。此过滤可以配合**文件名称过滤、文件类型过滤和路径过滤。**

例如，我们希望过滤 foxmail、windows mail 等目录是否存在，可以在路径过滤中设施“输入 mail”即可查找所有名称中包含单词 mail 的路径

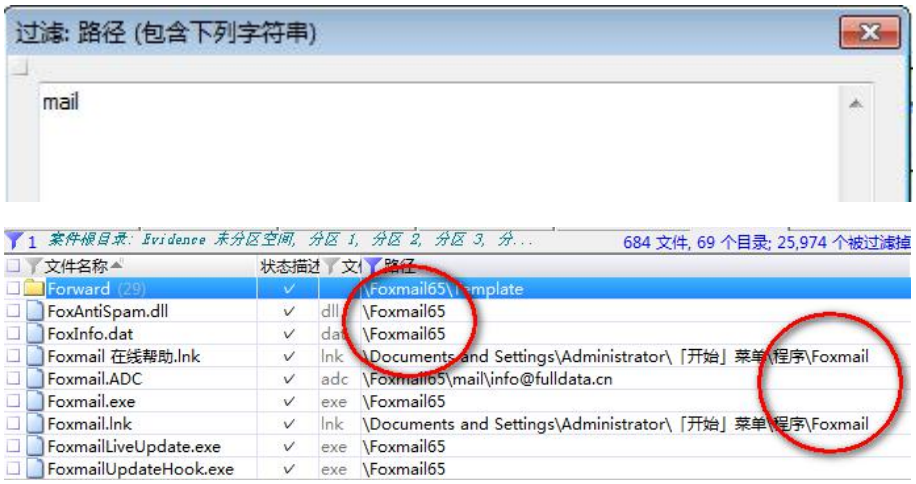


图 12 过滤路径中包含 mail 的文件

去除 Windows 目录下的所有 TXT 文件，可以先过滤 TXT 文件，再过滤出 windows 目录，并将过滤设置中的 Not 勾选。为了进一步去除无用的 TXT 文件，本例增加了 Driver 和 Cookie 两个目录，这样能够进一步减少无关的 TXT 文件数量。

图 13 路径中不包含 Windows、driver、cookie 的文件

2.5.4 实验 4:时间过滤查看所有删除数据中与 5 月 27 日相关的 Doc 文件

配合案例文件：C01-CCFC-Windows XP.e01

1.过滤分区 3 之中，2011 年 5 月 27 日 13 时之后创建的文件

通过过滤创建时间，选择“时间之后”，输入时间信息。激活。

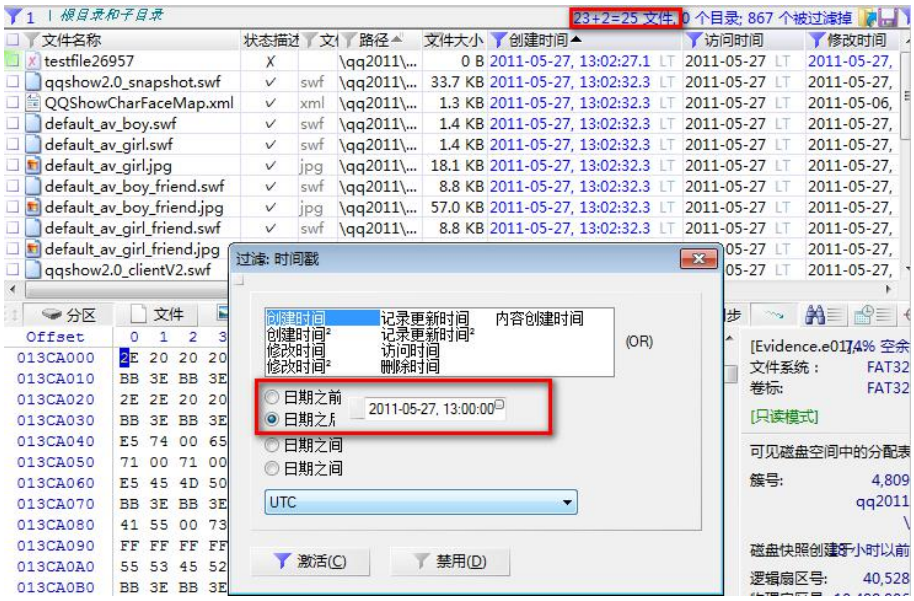


图 17　时间过滤

2.利用时间轴过滤

时间轴，将整个文件系统中的文件时间，以日历方式展示出来。我们列出所有分区中的文件，然后选择时间轴视图。

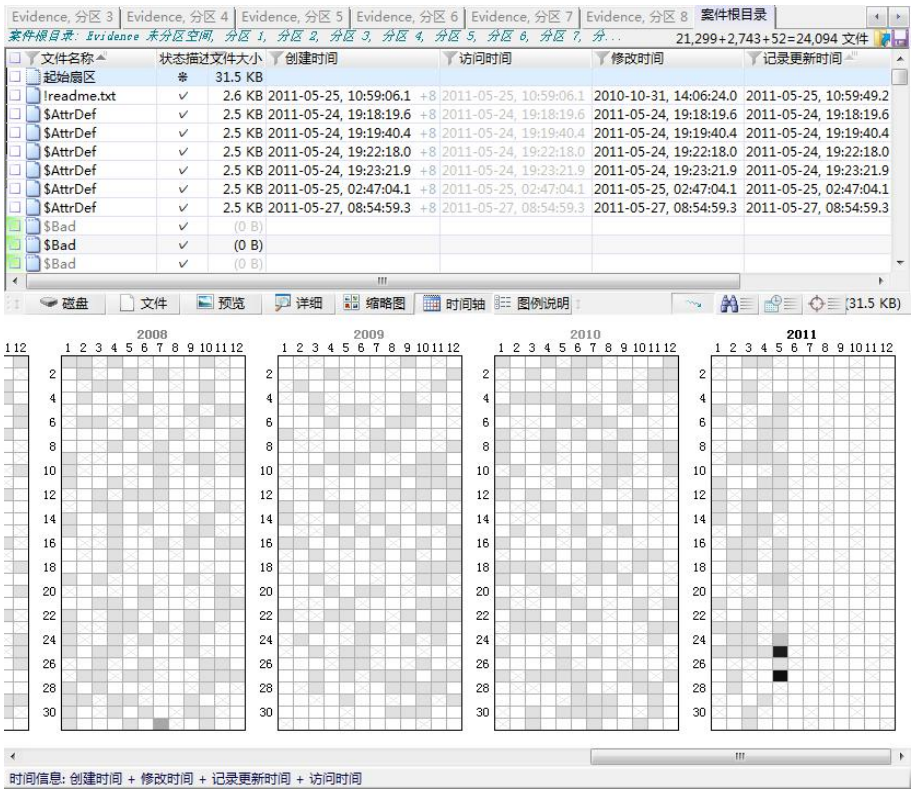


图 18　时间轴视图

从图中可以看到，黑色日期，是 2011 年 5 月 27 日和 5 月 25 日，说明这两天数据较多。我们过滤 5 月 27 日，可以看到在“创建、修改、访问、记录更新”时间中，属于 5 月 27 日的文件都显示出来。具有有 9404 个。

图 19　时间轴过滤结果

此时查看时间过滤，可以看到，时间轴过滤自动设置了如下的过滤条件。

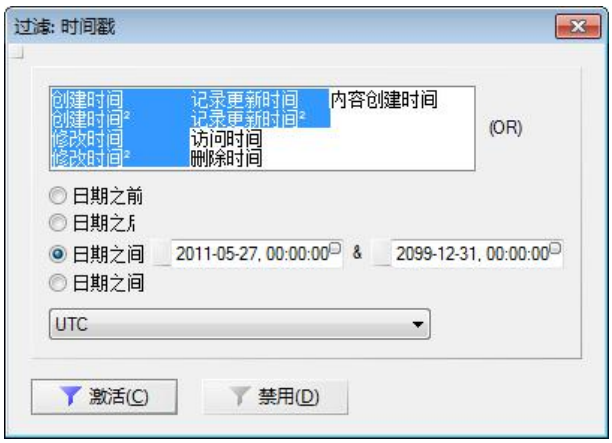


图 20　时间过滤设置

利用“描述过滤”，勾选“列出曾经存在的数据项目”，将只显示删除数据。通过文件名称过滤 DOC，可找到 4 个删除的 doc 文件。

图 21　时间过滤设置

2.5.5 实验 5:文件名过滤 20 题目

配合案例文件：C01-CCFC-Windows XP.e01。注意：本练习所得答案均在未进行磁盘快照状态下完成。如发现答案不符，请注意磁盘快照状态，并进行更新磁盘快照操作。

题目 1

查找“分区 1”中，所有扩展名为*.DOC 的文档，总数量有多少个？

A：29

B：4

C：33

D：53

题目 2

查看“分区 2”中，所有*.TXT 文件的数量？

A：17

B：18

C：9

D：1

题目 3

查找文件名以“富华”为起始字符的文件，这些文件出现在哪个分区？

问题描述：多选题，参考“证据项”列

A：分区 1

B：分区 2

C：分区 3

D：分区 4

E：分区 5

F：分区 9

题目 4

Internet Explore 的互联网历史记录文件(INDEX.DAT)保存在分区 1，共有 18 个。找到记录更新时间最晚的文件，所在文件夹的名称是？

问题描述：多选题

A：Content.IE5

B：History.IE5

C：Cookies

D：Recent

题目 5

查找当前案件中出现的所有的电子邮件和邮件客户端，下面哪个类型的文件确认包含电子邮件？

问题描述：多选题

A：Foxmail

B：Outlook Express

C：EML

D：MSG

题目 6

在“分区 4”和“分区 5”中查找*.ZIP 和*.RAR 文件，其中文件名相同的文件有几个？

A：7

B：8

C：12

D：20

题目 7

在所有分区中查找 Word（*.doc, *.docx），Excel(*.xls, *.xlsx)，PowerPoint(*.ppt, *.pptx)，*.RTF，*.PDF 文档，数量共计？

A：162

B：163

C：164

D：165

题目 8

查找出当前证据文件中，大于 10MB，且小于 100MB 的视频文件，其扩展名是？

A：AVI

B：WMV

C：RMVB

D：MPG

题目 9

所有分区中，大于 100KB，小于 2MB 的 PDF 文件，数量有几个？

A：10

B：35

C：11

D：16

题目 10

所有分区中，大于 100KB，小于 2MB 的 PDF 文件，出现在回收站中的文件有几个？

A：10

B：35

C：11

D：16

题目 11

分区 1 中，路径名称中包含英文 china mobile 的文件共有多少个？

A：204

B：607

C：34

D：23145

题目 12

在分区 1 中，路径名称中不包含英文 application、program、temporary、extensions 的 JPG 图片，其中最大的一个图片，文件名是？

A：back.jpg

B：碧昂丝.诺尔斯.jpg

C：02.jpg

D：凤姐副本.jpg

题目 13

传送”凤姐副本.jpg“的飞信用户 ID 是？

A：563532813

B：97156735

C：13810800118

D：971563725

题目 14

继续上题，该飞信用户收发了一些与计算机取证相关文件。这些文件涉及哪些扩展名？

问题描述：多选题

A：PDF

B：DOCX

C：PPT

D：PPTX

题目 15

请在所有分区中过滤文件名中包含 ISFS 的文件，其中内容为“会员申请表格”的文件有几份？

A：1

B：2

C：3

D：4

题目 16

文件内容为“会员申请表格”的几个文件中，最早出现在本地的是哪一个？

A：Form-ISFS-Chi.doc

B：Form-ISFS-中文版.doc

C：Form-ISFS-Chi.doc.lnk

D： Form-ISFS-中文版.doc.lnk

题目 17

根据文件时间属性分析，Form-ISFS-Chi.doc 和 Form-ISFS-中文版.doc，你认为哪一个文件可能被编辑过？

A：Form-ISFS-Chi.doc

B：Form-ISFS-中文版.doc

C：都未被编辑过

题目 18

关于 Form-ISFS-Chi.doc 和 Form-ISFS-中文版.doc 两个文件的关系，哪个描述正确？

A：后者是前者的副本，将前者拷贝到分区 8 根目录后改名为”Form-ISFS-中文版.doc“

B：用户编辑了前者，另存为 分区 8 根目录下”Form-ISFS-中文版.doc“

C：用户创建了一个新文件，将前者文件内容复制粘贴，生成后者

题目 19

所有回收站目录中，仍然保留有原始文件名的文件有多少个？

问题描述：排除 desktop.ini 和 INFO2 这两个文件

A：51

B：454

C：20

D：474

题目 20

文件 “内部传阅（机密）.doc”，曾保存于哪个目录下？

问题描述：提示：文件名过滤

A：C:\Documents and Settings\Administrator\Application Data\Foxmail\FoxTemp6.5

B：C:\Documents and Settings\Administrator\Recent

C：C:\User Data

D：C:\RECYCLER

2.5.6 实验 6:组合过滤 10 题目

配合案例文件：A01-FAT-DISK.001。注意：本练习所得答案均在未进行磁盘快照状态下完成。如发现答案不符，请注意磁盘快照状态，并进行更新磁盘快照操作。

题目 1

镜像中出现了大量的 PDF 文档。请问，哪个分区中不包含 PDF 文件？

问题描述：请结合”证据项“列进行判断

A：分区 1

B：分区 2

C：分区 3

D：分区 4

题目 2

继续上题。请问下面哪个 PDF 文件出现在回收站目录下。

A：JMXM021-技术资料-项目汇报.pdf

B：Application Compatablilty.pdf

C：USB Devices and Media Transfer Protocol.pdf

D：对网页浏览器隐身模式的电子数据检验-中英对照.pdf

题目 3

继续上题。通过分析发现硬盘中有两个文件是被彻底删除的，未在回收站中。这两个文件所在目录是：

问题描述：多选题。配合”描述“列分析或过滤。曾经存在。

A：\元数据

B：\CDF-作业\USB 取证相关

C：\?ersonal

D：\CDF-CISAW\期刊论文

题目 4

在所有文件名包含字母”CDF“的文件中，有一些文件是在 2019 年 11 月 24 日复制到分区 3 的。请问被复制的文件是以下哪种扩展名？

问题描述：提示：参考创建时间和修改时间

A：PPTX

B：DOCX

C：EXAM

D：PDF

题目 5

分区 2 中，有一些扩展名是 DOC 的 Word 文档是从其他位置复制过来的。请问共有几个文件？

问题描述：提示：利用描述过滤

A：6

B：7

C：9

D：10

题目 6

继续前题。下面哪些文字是 hide.doc 中的内容

问题描述：多选题

A：abcdefg

B：兹证明北京高科法证计算机服务有限公司为中国计算机法证技术研究会授权合作伙伴。

C：不经过仔细察看无法发现

D：hijklmn

题目 7

继续前题。下面关于 hide.doc 的描述哪些是正确的。

问题描述：多选题

A：通过文件系统分析，文件创建于 2008/11/19 11:45:00

B：通过文件元数据分析，该文件的创建时间是 2008/11/19 12:43:21

C：该文件总共编辑时间为 4 分钟

D：该文件的最后保存时间是 2008/11/19 11:49:20

题目 8

Investigating USB Devices Win8.pptx 这个文件中，包含有多少个 PNG 图片

问题描述：过滤出该文件，右键选择”浏览“

A：8

B：18

C：28

D：30

题目 9

雨课堂目录下包含有以下图片和 MP3 文件。其中大于 100KB，小于 200KB 的 MP3 文件总计容量有多少？

A：10MB

B：3.8MB

C：6.5MB

D：27MB

题目 10

所有的 MP3 文件中，最长的录制时间是多少秒？

A：39

B：48

C：50

D：60

2.5.7 实验 7:磁盘快照 10 题目

配合案例文件：C01-CCFC-Windows XP.e01。

题目 1

请过滤分区 4 中的 PDF 文件，计算这些文件的 SHA-1 哈希值。其中哈希值的第一字符是 6 的文件有：

问题描述：多选题

A：UFED Physical Analyzer Manual.pdf

B：UFED Brochure ENGLISH.pdf

C：UFED Ruggedized Data Sheet.pdf

D：3084841.pdf

题目 2

验证分区 4 所有文件的签名状态。其中签名匹配的文件有多少？

A：586

B：587

C：67

D：128

题目 3

分区 4 中，签名不匹配的 PWDUMP 文件的总容量是多少?

A：9.16MB

B：91.6MB

C：91.6GB

D：91.6KB

题目 4

分区 4 中，CCFC2011-ISFS-Sprite.ppt 文件中，总计有多少嵌入的图片？

A：6

B：7

C：8

D：9

题目 5

分区 4 中，包含有很多的 Thumbs.db 文件。其中哪一个路径下的缩略图库文件中包含的缩略图数量最多？

A：\ftp\china-forensic\image\logos

B：\ftp\china-forensic\images

C：\ftp\china-forensic\photo\speaker

D：\ftp\china-forensic\image

题目 6

分区 4 中，\ftp\china-forensic\photo\speaker 目录中的 Thumbs.db 文件中包含 38 个人物照片和 1 个文件夹图片。但是该目录下却只有 36 张人物照片。你认为哪些图片被删除了？

问题描述：多选题。提示：注意发现缩略图中包含，但 Speaker 目录下却不存在的人物照片。通过文件名对比不同。

A：Eddie Schwartz.jpg

B： image001.jpg

C：image002.jpg

D：image001.gif

题目 7

继续上题：你认为哪些文件是生成缩略图库之后被拷贝进来的？

问题描述：多选题。提示：注意发现 Speaker 目录下存在，缩略图中不包含人物照片名称。

A：xuzhiqiang.PNG

B：xzq.png

C：image002.jpg

D： image001.gif

题目 8

分区 4 中，解析文件元数据后，可以通过元数据进行过滤。请问下面哪些文件的作者是 a.belenko

问题描述：多选题。提示：磁盘快照，解析元数据，之后通过作者列过滤

A：complex.doc

B：simple.doc

C：complex2.doc

D： 011.doc

题目 9

请问下面哪些文件，最后保存文件的是 Sprite，且最后打开文件的也是 Sprite？

A：幸运抽奖.doc

B：Mac.doc

C：CCFC2011-ISFS-Sprite.ppt

D：011.doc

题目 10

分区 4 中，总计发现加密文件有多少个？

问题描述：提示：磁盘快照，加密文件检测

A：10

B：17

C：18

D：0