***Лекция 7.***

**Функции систем восстановления данных и файлов**

***План:***

***1. Особенности восстановления файлов***

***2. Восстановление файлов посредством анализа информации о файлах и папках***

***3.*** ***Восстановление файлов при помощи сканирования файлов известных типов (поиска файлов по сигнатурам)***

***Ключевые слова: раздел, логический диск, таблица раздела, схема раздела, метаданные, File Allocation Table, Master File Table, фрагментация файла, первая копия, виртуальная папка, поиск файлов по сигнатурам***

1. **Особенности восстановления файлов**

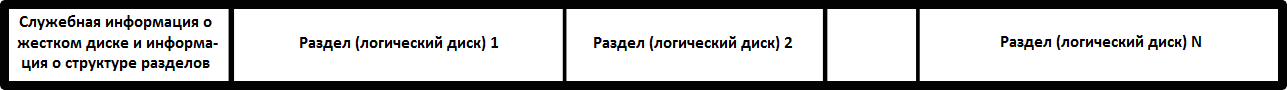
Знание особенностей расположения файлов на диске позволяет понять, как их можно восстановить в случае утраты.

В большинстве современных операционных систем жесткий диск делится на несколько независящих друг от друга частей ("разделов"). Эти разделы называются "логическими дисками". Логическим дискам назначаются буквы, также для них можно задать метку (для наглядности). Например, C: (Система) or D: (Данные).

Каждый раздел имеет свою файловую систему, которая не зависит от других разделов того же диска. Например, физический жесткий диск, на котором установлена ОС Windows, может состоять из двух логических дисков: одного с файловой системой NTFS, другого - с файловой системой FAT32.

Информация о разделах диска хранится в начале жесткого диска. Ее обычно называют "таблицей разделов" или "схемой разделов".

Структура жесткого диска



Служебная информация о жестком диске и информация о структуре разделов является "метаданными". Это информация о данных на диске (в отличие от самих данных).

Таким же образом каждый раздел или логический диск делится на две части: в первой находится информация о диске (структура папок, файловая система и т.д.), во второй - данные составляющие файлы.

Такое деление, начиная с метаданных, позволяет оптимизировать дисковое пространство, быстрее искать файлы, а также повысить надежность работы.

Структура логического диска



Служебная информация о диске - это информация о размере раздела, типе файловой системы и т.д. Для компьютера необходимо корректно найти нужные данные на разделе.

Информация о файлах и папках это файловые записи, содержащие имена файлов, размер, отметки даты/времени и другую техническую информацию. Также эта информация включает точное физические расположения (адреса) данных файлов на диске. На том же диске обычно имеется резервная копия этой информации.

На разных файловых системах данная информация хранится по-разному. Например, на файловой системе FAT она находится в Таблице Размещения Файлов (File Allocation Table), в то время как на файловой системе NTFS - в Главной Файловой Таблице (Master File Table (MFT)).

При необходимости прочесть файл, компьютер прежде всего обращается к информации о файлах и папках и ищет запись о данном файле. Далее ищется адрес файла и осуществляется переход к конкретному месту на диске, и затем уже читаются данные файла.

Для файлов, которые находятся на диске в одном месте (расположены рядом), все происходит достаточно просто. Однако файлы на диске могут быть фрагментированы, то есть занимать несколько несмежных областей. Это происходит достаточно часто, но большинство пользователей об этом не догадываются. Кроме того если посмотреть на файл в проводнике Windows или в Finder (Mac ОС), то всегда видно только один файл, так как все операции по сбору частей файла происходят внутри операционной системы.

Адреса всех фрагментов файла, хранимые в информации о файлах и папках, сразу же находятся при попытке его чтения. Данная информация и то, как она извлекается крайне важны при восстановлении файлов.

При удалении файла не происходит мгновенного разрушения его данных. Вместо этого вносятся некоторые изменения в информацию о файлах и папках, показывающих, что файл был удален.

В некоторых операционных системах файл просто помечается как удаленный, при этом сохраняются все метаданные о файле до тех пор, пока они не будут перезаписаны метаданными о новом файле.

Именно так происходит удаление файлов в файловых системах ОС Windows. В других операционных системах (например, Mac OS X) полностью разрушается файловая запись об удаленном файле.

Если сведения о файле в информации о файлах и папках в зависимости от ОС либо сохраняются, либо сразу удаляются при удалении файла, то сами данные файла во всех операционных системах остаются нетронутыми до тех пор пока данное место на диске не потребуется для записи другого файла. Если же на диск не записываются другие файлы, то информация о файле и его данные будут на нем сохранены.

Как уже отмечалось, в том месте диска, где хранится информация о данных файла, также содержится резервная копия информации о файлах и папках. При этом там может находиться и некоторая дополнительная информация о структуре файлов и папок расположенных в различных местах диска.

Если данные на диске перезаписаны, то их не удастся восстановить ни одной программой и ни одном из известных методов. Поэтому крайне важно чтобы до восстановления данных на диск не записывалась какая-либо информация. Есть два метода восстановления файлов, которые не были перезаписаны. Во всех утилитах восстановления используется либо один из них, либо оба.

***2.* Восстановление файлов посредством анализа информации о файлах и папках**

Это самый первый метод, используемый в программах восстановления данных, так как при его успешном применении восстанавливаются файлы с оригинальными именами, путями, отметками даты/времени, и сами данные.

Работа утилиты восстановления файлов начинается с попытки чтения и обработки первой копии информации о файлах и папках. В некоторых случаях (например, при [случайном удалении файла](https://www.r-undelete.com/ru/free-recovery-SDcards.html)) это единственное что требуется для восстановления файлов.

Если первая копия *информации о файлах и папках* сильно повреждена, то утилита сканирует диск и ищет вторую копию *информации о файлах и папках*. При этом также производится детальный поиск дополнительной информации о структуре папок и файлов, которая может находиться в области диска, где хранятся данные.

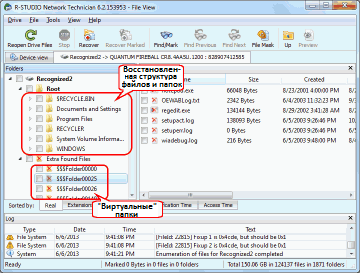
После этого вся найденная информация обрабатывается и воссоздается оригинальная структура папок и файлов.

Если файловая система диска серьезно не повреждена, то вполне вероятно удастся полностью восстановить структуру папок и файлов.

При сильном же повреждении файловой системы данный метод не позволит воссоздать полную структуру папок. В этом случае восстановленные файлы будут находиться в папках с присвоенными им виртуальными именами.

На рисунке показаны эти папки в программах [R-Studio](https://www.r-studio.com/ru/data-recovery-software/) и [R-Undelete](https://www.r-undelete.com/ru/).

Структура восстановленных файлов и папок и виртуальные папки



**3. Восстановление файлов при помощи сканирования файлов известных типов (поиска файлов по сигнатурам)**

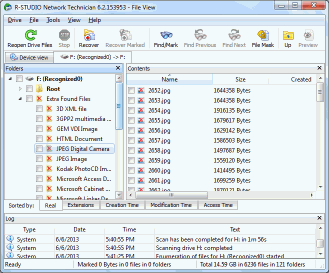
Если при помощи первого метода не удастся добиться желаемого результата, то следует произвести [поиск файлов по сигнатурам](https://www.r-studio.com/ru/creating-custom-file-type-r-studio.html). Этот метод позволяет восстановить больше данных, однако при этом не удается получить оригинальные имена файлов, отметки даты/времени или полную структуру папок и файлов на диске.

При поиске при сканировании файлов известных типов (поиск файлов по сигнатурам) анализируется содержимое диска и производится поиск по "файловым сигнатурам".

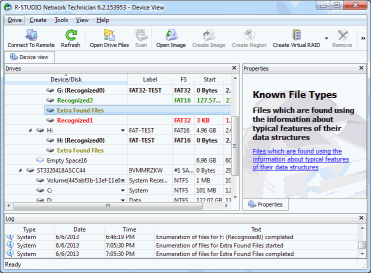
*Файловая сигнатура* это некий общий для определенного типа файлов шаблон данных, находящийся в начале или в конце файла. Почти каждый тип файлов имеет по крайней мере одну файловую сигнатуру. Например, все файлы типа png (portable network graphics, переносимая сетевая графика) начинаются с последовательности символов "‰PNG" и многие MP3 файлы начинаются с последовательности символов "ID3". Такие файловые сигнатуры позволяют отнести данные на диске к определенному типу файлов и далее восстановить.

После сканирования файлов известных типов программы поместят найденные файлы в категорию "Дополнительно Найденные Файлы (Extra Found Files)", где они будут структурированы по расширениям на основе идентифицированной файловой сигнатуры и где им будет присвоено некое шаблонное имя (порядковый номер).

Файлы, найденные на логическом диске путем сканирования файлов известных типов (Дополнительно Найденные Файлы (Extra Found Files)).

**

Файлы, найденные вне логического диска путем сканирования файлов известных типов (Дополнительно Найденные Файлы (Extra Found Files)).

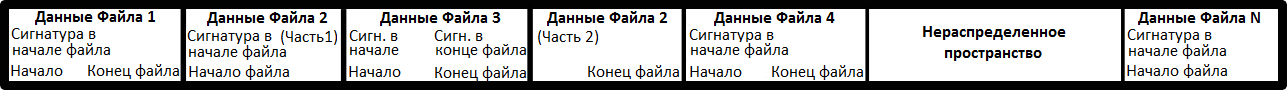


Несмотря на то, что при помощи данного метода можно получить наилучшие результаты при восстановлении данных с сильно поврежденной файловой системы, в нем имеются определенные ограничения.

Прежде всего, необходимо учитывать, что в некоторых типах файлов имеется файловая сигнатура в начале и в конце файла, в некоторых - только в начале (в конце файловой сигнатуры нет) и, наконец, в некоторых типах файлов вообще нет какой-либо различимой файловой сигнатуры.

Если утилита восстановления данных найдет файловую сигнатуру в начале и в конце файла, то он будет распознан и восстановлен. Если файл не имеет сигнатуры в конце, то утилита восстановления может восстановить его, допуская, что он заканчивается в начале следующего файла. И если файлы не имеют сигнатуры (например, зашифрованные диски, хранящиеся в файле-контейнере), то при помощи поиска файлов по сигнатурам не удастся получить какие-либо данные и на их месте будет показано нераспределенное дисковое пространство.

Также все это может быть осложнено еще и фрагментацией файлов. Более того, файлы, не имеющие сигнатуры в конце, после восстановления могут содержать в конце некую последовательность не относящихся к ним символов (т.е. "мусора").



В ситуации, изображенной на рисунке, файл 1 и файл 3 будут успешно восстановлены, а файл 2 и 4 - нет. Это поясняет следующая таблица.

Восстановление фрагментированных файлов

|  |  |  |
| --- | --- | --- |
| **Файл** | **Условие** | **Результат** |
| Файл 1 | Нет сигнатуры в конце файла, однако файл заканчивается в том месте, где начинается сигнатура в начале файла 2. | Файл успешно восстановлен |
| Файл 2 | Фрагментированный файл.  Файл 3 пересекается с файлом 2. | Файл не восстановлен. Утилита посчитает, что файл заканчивается в месте начала файла 3. Вторая часть файла 2 будет утрачена |
| Файл 3 | Смежный файл с сигнатурой в начале и в конце. | Файл успешно восстановлен. |
| Файл 4 | Нет сигнатуры в конце файла, за файлом следует нераспределенное пространство. | Файл не восстановлен. Утилита посчитает, что файл заканчивается в месте начала файла N, и нераспределенное пространство будет добавлено в конец файла 4. |

Помимо проблем с фрагментацией при поиске файлов по сигнатурам также можно получить "ложные" результаты.

Например, в любом файле могут находиться символы "ID3", которые при этом не будут являться файловой сигнатурой. Например, этот текст тоже содержит символы "ID3", но при этом это не MP3 файл. Поэтому при поиске файлов по сигнатурам часть данного текста может ошибочно распознаться как начало MP3 файла.

Профессиональные программы восстановления файлов (например, R-Studio) предоставляют пользователям возможность самим задавать файловые сигнатуры любой степени сложности, т.е. создавать пользовательский известный тип файла.

Как правило, на практике эти методы восстановления файлов применяются совместно: часть файлов восстанавливается при помощи первого метода, а оставшаяся - при помощи второго. Такие утилиты восстановления файлов, как R-Studio и R-Undelete, могут создавать образы дисков и одновременно сканировать данные.

*Случай 1: Восстановление файлов с жесткого диска с поврежденной служебной информацией*

Если диск не был соответствующим образом смонтирован или извлечен (например, из-за сбоя электропитания или ошибки пользователя), то часть или все метаданные на нем могут оказаться поврежденными или утраченными. При этом оказывается утраченной только исходная служебная информация о жестком диске и информация о структуре разделов, а оставшаяся часть данных на диске будет сохранена. В этом случае программы восстановления данных анализируют сохранившуюся на диске информацию о файлах и папках и восстанавливают все файлы и папки. Поиск известных типов файлов в таких случаях вряд ли понадобится. Это самый простой случай восстановления данных и, как правило, позволяет получать наилучшие результаты.

*Случай 2: Восстановление файлов с разбитого заново на разделы жесткого диска (физический диск)*

Случай заново разбитого на разделы диска во многом схож с предыдущим, его основное отличие состоит в том, что при создании нового раздела на диск записываются новые данные. При этом на диске перезапишется служебная информация о физическом диске. Однако остальная информация сохранится, в том числе и информация о файлах и папках. При помощи программ восстановления можно отсканировать диск, найти эту информацию и восстановить файлы и папки, которые не были затронуты данными нового раздела. Поиск известных типов файлов в таких случаях не понадобится.

*Случай 3: Восстановление файлов с переформатированного раздела (логический диск)*

Как правило, при переформатировании утрачивается больше данных, чем при разбиении диска на разделы. Все зависит от того, какое было выполнено форматирование. При быстром форматировании часть или все данные информации о файлах и папках перезаписываются, но при этом данные файлов сохраняются. При помощи программ восстановления можно отсканировать диск, найти то, что осталось от предыдущей файловой системы, и далее восстановить файлы и папки.

*Случай 4: Восстановление файлов с диска с поврежденной файловой системой*

Данный случай во многом зависит от того, насколько сильно повреждена файловая система. На диске имеются две копии информации о файлах и папках. Если повреждена только одна копия, то программа восстановления данных сможет прочесть данные из резервной копии и восстановить всю информацию и сами данные файлов. Если же повреждены обе копии, то шансы восстановления данных малы. Здесь, как и в случае 3, может помочь поиск известных типов файлов.

*Случай 5: Восстановление файлов, утраченных при их переносе на диске*

Если компьютер зависает или происходит какой-либо другой сбой во время дефрагментации диска или операции разбиения диска на разделы, то результаты восстановления данных могут не дать какого-либо положительного результата. Обычно это наихудший сценарий при восстановлении файлов. Информация о файлах и папках может выглядеть неповрежденной, однако метаданные при этом будут указывать на неверные физические адреса файлов, которые находились в процесса переноса во время сбоя в работе.

Например, данные могут находиться уже в другом месте, однако в информации о файлах и папках это отображено еще не будет. Или же в информацию о файлах и папках будет уже записана новая информация, но некоторые или все файлы еще не будут перенесены. В этом случае даже поиск известных типов файлов может не помочь, так как многие файлы, возможно, будут фрагментированы.

**Контрольные вопросы:**

1. Что такое логический диск?
2. Могут ли логические диски иметь разные файловые системы?
3. Что такое схема разделов?
4. Что включает в себя служебная информация о диске?
5. Как компьютер читает файл?
6. Что значит фрагментация файла?
7. Что происходит при удалении файла?
8. Назовите методы восстановления данных.
9. Как работает метод восстановления файлов посредством анализа информации в файлах и папках?
10. Понятие файловой сигнатуры. Как работает поиск файлов по сигнатурам?