

Глава 0

Принципы хранения информации

- ❑ Носители и накопители
- ❑ Случаи потери информации и принципы восстановления

В этой главе рассматриваются общие моменты. По компьютерной традиции нумеровать все с нуля, нумерация глав книги также начинается с нуля. Ко всему сказанному здесь повествование будет возвращаться вновь и вновь при рассмотрении самых разных случаев утраты и восстановления информации.

Носители и накопители

Информация, о восстановлении которой пойдет речь в этой книге, существует в двоичном виде на различных *устройствах хранения*, или *носителях*. С точки зрения обычного пользователя, носитель — это устройство, способное хранить информацию и выдавать ее через какой-либо *интерфейс*. Например, жесткие диски с интерфейсами SATA или IDE, flash-диски с интерфейсом USB и т. д. Пользователь, владеющий компьютером на более профессиональном уровне, называет их накопителями или *устройствами хранения информации*, а слово «носитель» употребляет лишь для определения одного из внутренних компонентов такого устройства.

Внутри устройства есть компонент, который и является собственно носителем, — диски (в просторечии — «блины») винчестера или кристалл flash-памяти. На нем физически записана информация как последовательность двоичных элементов. Это могут быть магнитные домены либо полупроводниковые элементы в одном из двух состояний. Поскольку активное использование устройств хранения связано с использованием магнитных дисков (хотя еще раньше появились барабаны и ленты), с носителями ассоциируется именно слово «диск».

Компонент, который обеспечивает передачу информации с носителя в интерфейс и обратно, назовем пока общим словом *контроллер*. При этом на носителе информация записана непрерывно, а по интерфейсу она передается определенными *словами*, или *блоками*. Структура двоичной записи на носителе может очень сильно отличаться от того, как она видится через интерфейс.

Мы обычно понимаем это как *аппаратный уровень* хранения данных. И носитель, и контроллер являются составляющими, специально созданными друг для друга, а их «взаимоотношения», как правило, — секрет производителя винчестера или flash-диска. Тем не менее пользователь, который хочет заниматься восстановлением информации, должен быть готов к тому, что ему придется работать со всеми элементами устройства хранения! Лишь в отношении компакт-дисков носитель и привод с контроллером выступают совершенно отдельными и самостоятельными компонентами, и это заметно облегчает задачу: неисправен контроллер — просто поменяйте привод!

Информация на носителе *структурирована*, что позволяет контроллеру по мере необходимости обращаться лишь к нужным участкам дорожки или матрицы. Эта

структура, как правило, тоже засекречена. Обычно считывание информации на глубинном физическом уровне возможно лишь с помощью «родного» или совершенно идентичного контроллера. Благодаря функциям контроллера, операционной системе компьютера носитель представляется уже набором *кластеров* или *логических блоков*.

На аппаратном уровне уже заложены меры, повышающие устойчивость информации. В первую очередь, это *избыточность* записи на пластинах винчестера или лазерном диске: одни и те же данные дублируются, снабжаются контрольными суммами. При повреждении отдельных элементарных ячеек утраченное содержимое восстанавливается за счет дубликатов. Этот процесс происходит непрерывно, и пока контроллер может восстановить данные на аппаратном уровне, мы таких потерь не замечаем вовсе.

Логическая структура информации, прежде всего, связана с файлами и файловой системой. *Файлом* называют законченную, непрерывную, именованную последовательность байтов, представляемую операционной системе компьютера. На носителе файл существует как некоторый набор кластеров с данными. Информация о том, какие кластеры составляют этот файл, является содержанием *файловой системы*. Физически информация о структуре данных записана там же, где и сами данные, и говорить о «физическом» или «логическом» повреждении информации на носителе можно лишь условно. Если потери затронули кластеры, где находится содержимое файла (область данных), мы обычно говорим о «физическом» дефекте; если же они коснулись области, где были записаны сведения о структуре, дефект представляется нам «логическим».

Следует учитывать, что и на носителе, и в файле всегда присутствует некая «полезная» уникальная информация — она-то и представляет главную ценность. Другая же часть информации является «служебной», то есть несет в себе структуру данных. При повреждении этой части обратиться к устройству хранения стандартными средствами операционной системы невозможно, но с помощью специальных программ удастся сделать его побайтную или поблочную копию — так называемые сырые данные (Raw data).

Случаи потери информации и принципы восстановления

Отдельные главы книги посвящены восстановлению информации с различных типов носителей. В каждой главе, кроме первой, будут рассмотрены сходные ситуации. От причин и места потери данных зависит тактика их восстановления.

Тем не менее при восстановлении есть один главный принцип. Для успешного восстановления данных нужно, прежде всего, максимально полно и бережно скопировать содержимое проблемного носителя «как есть» на какой-либо другой носитель. Поскольку мы почти никогда не знаем заранее, что послужило причиной потерь, безопаснее работать с такой копией, ведь любые лишние операции с проблемным носителем могут увеличить степень его повреждения.

- ❑ Потери данных на физическом уровне возникают, когда повреждаются сами элементы, хранящие информацию, а избыточность записи не может покрыть такую потерю. Примерами служат разрушение магнитного слоя пластин винчестера, царапины на отражающем слое лазерного диска или деградация ячеек flash-накопителя. Это самый серьезный случай: восстановление данных в принципе невозможно. Если затронуты элементы, на которых хранилась содержательная часть данных, то о них придется забыть. Если повреждены служебные данные, в некоторых случаях можно извлечь оставшуюся «полезную» часть, а затем попробовать воссоздать ее структуру.
- ❑ Примером аппаратных проблем, приводящих к недоступности данных, могут служить неисправности контроллеров, головок или механики жесткого диска, царапины на прозрачной стороне лазерного диска. При этом сама информация обычно не затронута, нужно лишь обеспечить доступ к ней. Восстановление начинается с устранения аппаратной проблемы.
- ❑ Нарушения логической структуры при полной аппаратной сохранности носителя могут возникнуть в результате некорректного изменения служебной информации. Как правило, это вина пользователя, программных ошибок или действий вредоносных программ. Причиной могут послужить и аппаратные сбои, например перепады напряжения в момент записи такой информации, «небезопасное» извлечение flash-диска во время обращения к нему. Восстановление сводится к извлечению сырых данных и воссозданию их структуры.

Частным случаем является корректное, но опрометчивое удаление файлов и папок либо форматирование дисков. Фактически при удалении файла сначала только модифицируются записи в файловой системе, а сами данные остаются на диске до тех пор, пока на их место не будут записаны новые. Успех восстановления зависит, прежде всего, от того, производилась ли запись новых данных, «затерших» удаленные, или нет.

- ❑ Повреждение информации внутри файла. Файл формально цел, но не может быть открыт предназначенной для этого программой: выдается сообщение об ошибке. Это может быть следствием любой из названных выше причин. С разбора именно такой ситуации начинается книга.

Средства, необходимые для восстановления, зависят от характера и причины потери данных. В простейшем случае это специализированные программы, работающие в среде MS-DOS или Windows, — такие средства доступны каждому, кто хочет попробовать свои силы в восстановлении информации. Аппаратно-программные комплексы включают в себя электронные схемы, подменяющие или дополняющие штатную электронику носителей информации, и программное обеспечение. Как правило, такие комплексы достаточно дороги, что и ограничивает их применение любителями. Ремонт гермоблоков жестких дисков связан с выполнением очень тонких манипуляций и требует серьезных практических навыков, специального инструмента и «чистой комнаты». Это, а также потребность в запасных частях для ремонта делают выполнение подобной работы возможным и рентабельным лишь в специализированных сервисных центрах. Тем не менее в книге будут рассмотрены некоторые моменты таких манипуляций с винчестерами.

Глава 1

Восстановление поврежденных файлов

- ☐ Структура файлов в программе Hexplorer
- ☐ Понятие формата файла
- ☐ Восстановление файлов собственными средствами прикладных программ
- ☐ Программы для восстановления файлов
- ☐ Полезные ссылки
- ☐ Резюме

В этой главе речь пойдет о восстановлении так называемых битых файлов. Файл может быть поврежден при его некорректном сохранении, особенно если вы пытались открыть, а затем сохранить файл, созданный кем-то в более новой версии прикладной программы. Такие файлы можно извлечь с поврежденного носителя. И хотя скопировать их удастся, при попытке открыть «спасенный» файл вы увидите сообщение об ошибке. После восстановления данных со случайно отформатированного диска среди найденных файлов наверняка окажется немало битых. Наконец, подобные файлы найдутся в результате анализа данных, сохраненных операционной системой при проверке и автоматическом исправлении ошибок жесткого диска, — такие файлы имеют расширение СНК.

Считается, что файл — именованная конечная и непрерывная последовательность байтов, несущая некую информацию. Для наших целей (восстановления данных) этого достаточно. В тех файлах, которые приходится восстанавливать на практике, такая последовательность обычно кодирует вполне осмысленную и завершенную информацию, например текст документа или растр (множество точек) изображения.

Напомним, что 1 байт = 8 бит, то есть каждый байт может принимать значения от 0 до $2^8 = 256$ в десятичной (dec) или $2^8 = FF$ в шестнадцатеричной (hex) форме записи. Существуют программы, которые позволяют просматривать и изменять содержимое любого файла как простой последовательности байт. Логично называть их байтовыми или двоичными (бинарными) редакторами. Поскольку такие программы обычно показывают значения байтов в шестнадцатеричном виде, за ними закрепилось название «HEX-редакторы».

Если открыть файл в любом HEX-редакторе, мы увидим такое представление его содержимого. О применении HEX-редакторов для восстановления содержимого файла будет сказано в конце главы, они помогут понять структуру файла и то, в чем заключается суть ее повреждений.

Структура файлов в программе Hexplorer

Программа ICY Hexplorer — простой, но достаточно функциональный бесплатный редактор двоичных данных. Дистрибутив ее доступен на сайте разработчиков <http://hexplorer.sourceforge.net>.

Рабочая область окна разделена по вертикали. После открытия файла в левой части рабочей области отображается содержимое файла по байтам в виде их шестнадцатеричных значений (рис. 1.1). В правой части окна те же самые байты интерпретируются как ASCII-значения текстовых символов.

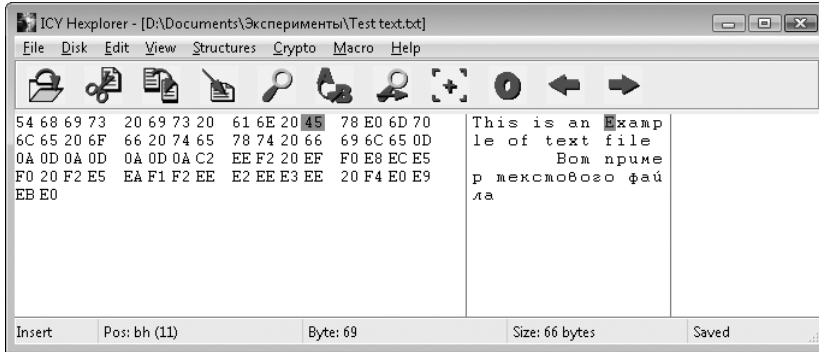


Рис. 1.1. Окно программы Hexplorer — открыт текстовый файл

При этом выделенный фрагмент одновременно подсвечивается в обеих половинах окна. Легко заметить, что означает каждый байт файла, если считать, что им закодирован какой-либо текстовый символ. При рассмотрении текстовых файлов такая интерпретация совершенно справедлива. Команды меню View (Вид) и Structures (Структуры) позволяют интерпретировать содержимое файла иначе, например как точечный рисунок, и просмотреть его в дочернем окне.

Понятие формата файла

Существует множество типов и форматов файлов. В общем виде формат — это сведения о том, как программа должна интерпретировать содержимое файла при его открытии. Иногда формат путают с расширением имени файла, но это совершенно разные понятия. Операционная система может использовать расширение имени файла, чтобы быстро определить, какой программой должен открываться данный файл. Как правило, расширение должно соответствовать формату, но это вовсе не обязательно. Сведения о формате находятся в самом файле — они помещаются в начало последовательности байтов.

Текстовый файл — самый простой формат. Каждый байт текстового файла является кодом определенного символа ASCII (буквы, цифры или знака препинания), а также символов пробела, начала и конца строки. Кроме значений текстовых символов, в текстовом файле действительно больше ничего нет! Структура любых других типов файлов гораздо сложнее.

В качестве примера откроем в HEX-редакторе файл рисунка в формате BMP (рис. 1.2). В самом начале файла дана последовательность 42 4D — в текстовой интерпретации это буквы BM. Затем следует довольно длинная последовательность

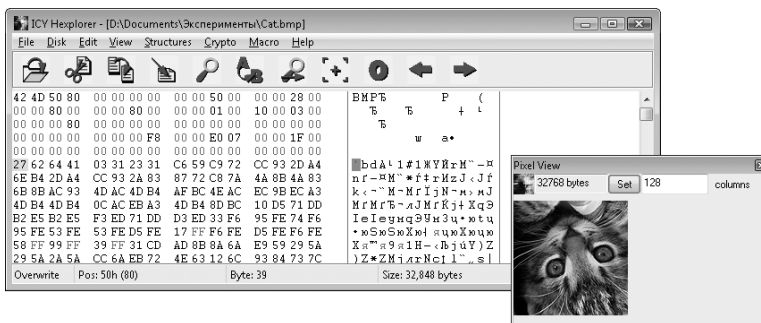


Рис. 1.2. Окно программы Hexplorer — открыт файл рисунка в формате BMP

байтов, в которой преобладают нулевые значения, а уже за ней начинается массив самых разных значений, который продолжается до самого конца файла.

Можно предположить, что начало файла — какая-то служебная информация, причем самые первые ее байты указывают на формат файла. Лишь затем следует то содержимое, которое программа, работающая с подобными файлами, должна интерпретировать как рисунок.

ПРИМЕЧАНИЕ



В Интернете существует ресурс, который неоднократно будет упоминаться — «Википедия», или «Свободная энциклопедия». Она содержит большое количество сведений и о форматах файлов, и о многих других понятиях, всю информацию о которых в одну книгу собрать просто невозможно.

В статье говорится, что BMP-файл действительно состоит из четырех частей:

- ❑ заголовок файла (BITMAPFILEHEADER). Первое поле заголовка (первые два байта) является сигнатурой — указанием на тип файла;
- ❑ заголовок изображения (BITMAPINFOHEADER) — необязательный компонент, он может отсутствовать;
- ❑ палитра, которая также может отсутствовать;
- ❑ само изображение — каждый байт описывает одну точку рисунка.

Здесь необходимо сделать важный вывод. Очевидно, что искажение или утрата данных, относящихся к первым трем частям, может привести к тому, что файл перестанет открываться предназначенной для этого программой, хотя вся значимая информация в нем осталась цела. В этом случае появится сообщение «Файл имеет неправильный формат или поврежден». Наоборот, утрата или искажение данных самого изображения никак не повлияет на открытие файла, но в открывшемся рисунке выпадут или будут искажены отдельные фрагменты.

В качестве эксперимента можно сделать несколько копий файла, намеренно испортить каждую из них в разных местах с помощью HEX-редактора, а затем сохранить. Открывая эти файлы стандартным средством просмотра Windows или программой Paint, пользователь убедится в справедливости сделанного ранее вывода.

Подобным образом устроен любой файл, хотя спецификация частей файла зависит от его типа. Описания почти всех форматов опубликованы, их можно найти в Интернете. Некоторые из детальных описаний занимают не один десяток страниц. Причем понятное описание форматов файлов Microsoft Office найти труднее всего, но все же существуют и общие положения.

- В начале файла всегда находится его заголовок (header). Сам заголовок тоже обладает внутренней структурой.
 - Заголовок начинается сигнатурой, или «магическим числом». Это фрагмент кода, который однозначно дает понять, что это за файл или какой программой он создан. Например, EXE-файл всегда начинается с символов MZ, архив RAR содержит в своем коде сигнатуру RAR.
 - Другие поля заголовка могут содержать информацию о размере файла, указывать, на каком байте заканчиваются заголовки и начинаются собственно данные, и так далее.
- В файле могут присутствовать специфические для данного формата заголовки, например для файлов изображений они указывают алгоритм сжатия, глубину цвета, размеры изображения в точках и т. д. Для звуковых файлов такими параметрами будут частота и разрядность оцифровки, число каналов, алгоритм сжатия и пр.
- В некоторых файлах могут присутствовать вспомогательные компоненты: палитра, кодовая таблица, внедренные данные шрифтов, макросов и т. д. В отдельных форматах, наподобие документов Microsoft Office или CorelDRAW, такие данные могут занимать больше места, чем все остальное. Достаточно посмотреть, каков размер «пустого» документа Microsoft Word 2003!
- Собственно уникальные данные документа. Иногда в сложных документах эти данные организованы в виде цепочек, или порций (chunks). В таком случае где-то в заголовках дается ссылка на начало и размер каждой порции.
- Указатель конца файла или контрольная сумма (check sum) подтверждают целостность файла. Они присутствуют далеко не во всех форматах.

Очевидно, любой поврежденный файл можно восстановить или хотя бы извлечь полезную информацию, открыв и отредактировав такой файл HEX-редактором. Если испорчена информация о формате, ее (теоретически) можно воссоздать,

а затем открыть файл «штатным» приложением. Если повреждена область самих данных, из файла можно извлечь только то, что осталось.

Восстановление или извлечение полезной информации из поврежденных файлов вручную — трудоемкая задача. По сути она схожа с программированием. Проблема даже не в том, чтобы разобраться в структуре формата, а в том, что для поиска и создания новой записи «нужных» байтов потребуется масса рутинных и утомительных пересчетов, проб и проверок. Поэтому на практике, если существует такая возможность, вначале целесообразно прибегнуть к двум другим способам.

Восстановление файлов собственными средствами прикладных программ

В некоторых прикладных программах изначально заложена возможность тестирования и восстановления поврежденных документов. Эти возможности довольно ограничены, но иногда и их оказывается достаточно. С другой стороны, кому как не разработчикам программ известны все тонкости используемых форматов, а также алгоритмы их создания и восстановления.

ВНИМАНИЕ



Перед любыми попытками восстановления битого файла необходимо сделать несколько его копий. Не стоит экспериментировать с единственным исходным экземпляром!

Восстановление документов средствами Microsoft Word

В программе Microsoft Word предусмотрена функция открытия документов с восстановлением. Кроме того, в полный состав пакета Microsoft Office входит *конвертер восстановления текста*. По умолчанию он не устанавливается.

Заметим, что есть еще и третья функция, связанная с использованием автосохранения документов Word в процессе работы приложения. Она скорее относится к области резервного копирования: копии документа периодически сохраняются в той же папке, что и открытый документ, либо в папке \Application Data\Microsoft\Word в профиле пользователя. Если приложение Microsoft Word аварийно завершило работу до корректного сохранения документа пользователем, при очередном его открытии вы увидите панель **Восстановление документов**, на которой показаны существующие резервные копии файла. Скорее всего, при восстановлении действительно битых файлов эта функция не работает.

Для установки конвертера восстановления текста необходимо повторно запустить программу установки Microsoft Office.

1. В начальном окне установки выберите пункт **Добавить или удалить компоненты**. Нажмите кнопку **Далее**.
2. В окне **Выборочная установка** установите флажок **Расширенная настройка приложений** и нажмите кнопку **Далее**. Появится окно **Расширенная настройка**.
3. Компонент **Конвертер восстановления текста** находится в группе **Общие средства Office** ► **Конвертеры и фильтры** ► **Конвертеры текстовых файлов**. Нажмите стрелку рядом с этим компонентом и выберите пункт **Запускать с моего компьютера**.
4. Следуя указаниям мастера, завершите установку выбранных компонентов.
5. После этого запустите программу Word. Зайдите в окно настроек (меню **Сервис** ► **Параметры**) и на вкладке **Общие** установите флажок **Подтверждать преобразование при открытии**. Сохраните настройки.

Теперь есть два пути. Для начала воспользуйтесь встроенной функцией восстановления документа.

1. В меню **Файл** выберите команду **Открыть**. Появится окно **Открытие документа**.
2. Выберите в нем файл для восстановления.
3. Нажмите на стрелку рядом с кнопкой **Открыть** и выберите команду **Открыть и восстановить** (рис. 1.3).

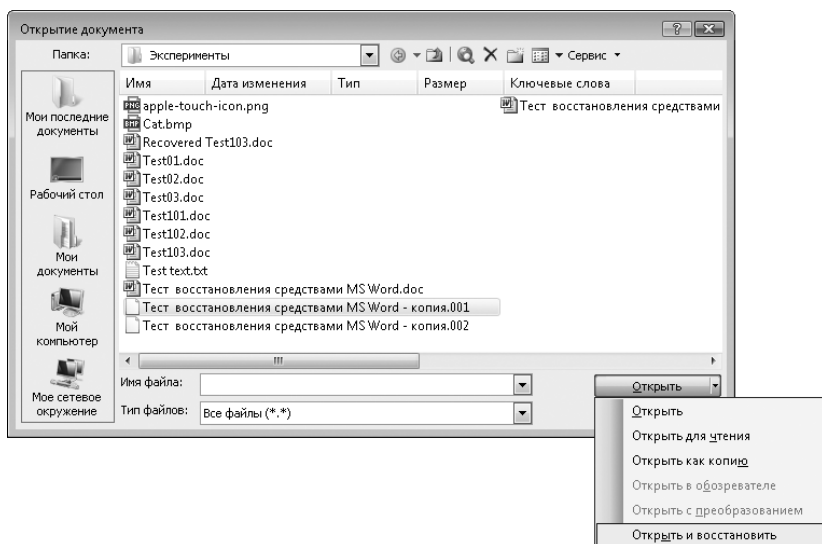


Рис. 1.3. Открытие документа с восстановлением

Возможно, документ успешно откроется. Открыв документ, сохраните его под другим именем. Еще лучше скопировать все содержимое и вставить его в новый документ. Однако эксперименты показывают, что таким образом можно исправить только самые незначительные дефекты. Чуть большими возможностями обладает конвертер восстановления текста.

1. В окне Открытие документа в раскрывающемся списке Тип файла выберите параметр Восстановление текста из любого файла.
2. Откройте нужный файл. В процессе открытия может появиться дополнительное окно Преобразование файла (рис. 1.4).
3. Выберите в этом окне вариант Восстановление текста из любого файла и нажмите кнопку ОК.

При таком восстановлении будет утеряно исходное форматирование документа — вместо него получится чистый текст (plain text). Тем не менее и эта функция срабатывает далеко не всегда. В таком случае следует восстановить документ специальными программами от сторонних разработчиков.

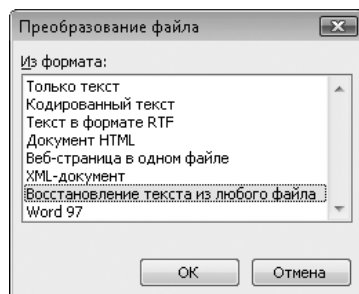


Рис. 1.4. Окно Преобразование файла

Восстановление архивов в программе WinRAR

В популярном архиваторе WinRAR (<http://www.rarlab.com>) заложена функция восстановления поврежденных архивов. Эта программа может восстанавливать архивы форматов RAR и ZIP. Тип архива WinRAR пытается определить автоматически по сигнатуре файла и по его расширению, но пользователь может указать тип архива и самостоятельно.

Формат RAR предусматривает внесение в архив избыточной информации — так называемой *информации для восстановления* сразу при его создании. Для этого при добавлении файлов в архив в окне Имя и параметры архива необходимо установить флажок Добавить информацию для восстановления (рис. 1.5). По умолчанию этот флажок снят. На вкладке Дополнительно задается относительный размер информации для восстановления в процентах от общего размера архива.

Добавление такой информации неизбежно увеличивает размер архива, но этой мерой предосторожности пользуются не всегда. Как правило, при сжатии файлов на первое место выходит минимальный размер получаемого архива. Создатели программы рекомендуют добавлять информацию для восстановления в архивы,

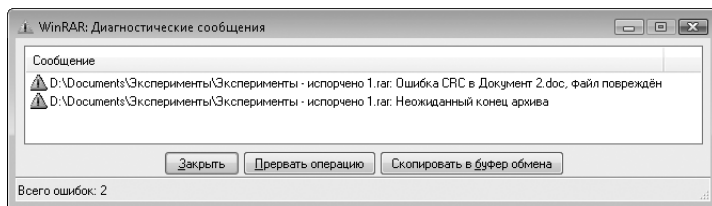


Рис. 1.6. Диагностические сообщения программы WinRAR

В частности, при повреждении заголовка потребуется указать, каков формат исходного архива: RAR или ZIP, а также является ли архив непрерывным. В ходе операции восстановления программа может спросить у пользователя, нужно ли считать файлом найденный в архиве элемент: Подозрительный элемент <возможное имя файла>. Добавить в архив: Да/Нет/Все. Ответьте **Да**, чтобы добавить этот элемент в `fixed.arcname.rar`.

Для устранения неисправностей архивов RAR программа WinRAR представляется оптимальным и основным средством. Для архивов ZIP она также дает неплохие результаты, но здесь конкуренцию составляют многочисленные специализированные программы.

Восстановление баз 1С:Предприятие

Базы данных, с которыми работают программы семейства 1С:Предприятие, состоят из множества отдельных файлов. Часть из них несет сведения о конфигурации, а другие (файлы с расширением DBF) являются изменяемыми файлами базы данных. Если ошибка затронула только конфигурационные файлы, проще всего не восстанавливать базу, а просто переписать эти файлы из дистрибутива конфигурации или любой резервной копии.

Повреждение базы чаще всего происходит в результате системных сбоев, внезапного отключения питания компьютеров или возникновения проблем с локальной сетью. Основной и прямой путь восстановления базы — тестирование и исправление ее средствами программы 1С, запущенной в режиме Конфигуратора. Главная особенность подобного восстановления в том, что Конфигуратор использует логику построения базы и некоторые данные могут быть восстановлены по связанным с ними записям в других, неповрежденных файлах.

Разумеется, перед любыми попытками восстановления следует создать резервную копию базы данных. Лучше всего сделать это простым копированием папки с базой в другой каталог.

1. Откройте программу 1С в режиме конфигуратора.
2. Запустите проверку открытой базы (меню **Администрирование** ► **Тестирование и исправление**).
3. При наличии свежей резервной копии можно сразу приступить к исправлению: выберите вариант **Тестирование и исправление**. В ходе тестирования показываются обнаруженные ошибки базы. Обратите внимание на имена файлов. При исправлении, в случае разночтений, программа может попросить указать, на основании каких данных следует восстанавливать конкретную запись или же ее следует удалить вообще.
4. После восстановления откройте базу в монопольном режиме и проверьте корректность записей и итогов. Вполне возможно, что какие-то записи придется исправить или добавить вручную.

Восстановление поврежденных баз возможно почти всегда. Дело лишь в числе утраченных записей. Восстановление отдельных файлов специальными программами практически не требуется — средства Конфигуратора 1С достаточно эффективно восстанавливают структуру файлов базы данных. В программу заложены средства архивирования, и другие способы резервного копирования всегда есть в распоряжении. Восстановление из последнего архива может оказаться быстрее тестирования и восстановления даже с учетом повторного ввода данных, появившихся с момента последнего архивирования.

Программы для восстановления файлов

Специализированные программы созданы для восстановления файлов почти любого известного формата. По числу программ на первом месте, безусловно, оказываются документы Microsoft Office. Простой поиск в Интернете выдает около трех десятков названий утилит для восстановления документов Word или таблиц Excel.

Работа подобных программ основана на том, что в файле ищутся присущие этому формату структуры (заголовки, таблицы и т. д.). Если структура найдена, она анализируется на предмет допустимости содержащихся в ней записей. Правила проверки известны, они прямо вытекают из описания формата. Если запись явно не соответствует допустимому или ожидаемому значению, то она исправляется. Точно так же могут генерироваться отсутствующие или неполные структуры. Конкретные алгоритмы работы у всех программ свои собственные, поэтому разными оказываются и результаты их работы в реальных ситуациях.

В отношении программ восстановления действует то же правило, что и в отношении антивирусов: чем больше приложений будет использовано в данном конкретном