

Архивирование данных

ПОНЯТИЯ АРХИВНОГО ФАЙЛА И АРХИВАТОРОВ

Архивные файлы – это документы, рисунки и др. файлы, которые специально сжаты (упакованы) с целью более рационального размещения на дискете или для передачи по электронной почте через интернет. При этом архивный файл занимает в несколько раз меньше места (иногда в 10 - 100 раз!) и может быть свободно размещен на дискете или более быстро отправлен по электронной почте (вместо 10мин. – 1-2 мин.)

Архивные файлы широко используются для передачи информации в Интернете и по электронной почте, причем благодаря сжатию информации повышается скорость её передачи. Это особенно важно при передаче информации по телефонной линии.

МЕТОДЫ АРХИВИРОВАНИЯ

Алгоритм Хаффмана.

Алгоритм основан на том факте, что некоторые символы из стандартного 256-символьного набора в произвольном тексте могут встречаться чаще среднего периода повтора, а другие, соответственно, – реже. Следовательно, если для записи распространенных символов использовать короткие последовательности бит, длиной меньше 1 байта, а для записи редких символов – более длинные, то суммарный объем файла уменьшится. Например буквы а,о,е,и – встречаются очень часто в русском тексте, объем каждой буквы равен 1 байт (8 бит), их можно заменить на цифры 0,1,2,3, которые можно разместить в 2-х битах. Т.е. сжатие будет равно 25%, т.е. сжатие в 4 раза.

Алгоритм Лемпеля-Зива.

Классический алгоритм Лемпеля-Зива – LZ77, названный так по году своего опубликования. Он формулируется следующим образом: «если в более раннем тексте уже встречалась подобная последовательность байт, то в архивный файл записывается только ссылка на эту последовательность (смещение, длина), а не сам текст».

Так фраза «КОЛОКОЛ_ОКОЛО_КОЛОКОЛЬНИ»[24] за кодируется в последовательность «КОЛО(-4,3)_О(-6,4)_(-7,7)ЬНИ»[13].

Коэффициент сжатие - 54%. Аналогично сжимается изображение. Большие области одного цвета заменяются на ссылку: (цвет, длина) Графические файлы сжимаются очень хорошо – в 100–200 раз!

ВИДЫ АРХИВОВ

Самораспаковывающиеся архивы (*sfx-файл*).

В тех случаях, когда архивация производится для передачи документа потребителю, следует предусмотреть наличие у него программного средства, необходимого для извлечения исходных данных из уплотненного архива. Если таких средств у потребителя нет, создают самораспаковывающиеся архивы. Самораспаковывающийся архив готовится на базе обычного архива путем присоединения к нему небольшого программного модуля. Сам архив получает расширение имени .EXE, характерное для исполнимых файлов. Потребитель сможет выполнить его запуск как программы, после чего распаковка архива произойдет на его компьютере автоматически.

Непрерывные архивы.

Очень высокий выигрыш в размере (10-50%), особенно при архивировании большого количества малых файлов, получается при использовании опции создания непрерывных архивов. В этом случае все данные будут архивироваться в виде одного большого файла, а не каждый по отдельности. Это значительно увеличивает эффективность сжатия информации, но данный метод имеет и свои недостатки. Во-первых, для распаковки нескольких файлов из такого архива понадобится больше времени, чем для извлечения из обычного. Тут будет иметь значение, какими по счету были заархивированы нужные файлы. Если они были в середине, то для их распаковки программе будет нужно распаковать все файлы упакованные до необходимых, пока она доберется до требуемых. Создание архивов таким способом также может повлечь за собой большие утраты, ведь если архив окажется, поврежден, будут потеряны все файлы, которые в нем находились. В случае же упаковки обычным способом, можно извлечь из поврежденного архива пусть не все, но большинство файлов.

Многотомный архив

Это архивный файл, который состоит из нескольких файлов (томов). Размер одного тома не превышает заданного размера.

Расширение многотомному архиву дается следующим образом:

1 том имеет расширение .гаг, остальные .г00, .r01, .г02 и т.д.

Для создания многотомного архива нужно:

- выбрать несколько исходных файлов;
- из контекстного меню выбрать пункт «Добавить в архив», указать имя, местоположение архива, а также размер каждого тома.

Для извлечения многотомного архива:

- скопировать все тома в одну папку;
- щелкнуть по первому тому и выбрать «Извлечь в...».

Архиваторы - это специальные программы, которые позволяют работать с архивными файлами, т.е. запаковывать и распаковывать архивные файлы.

Необходимость архивации связана также с необходимостью резервного копирования информации на диски и дискеты с целью сохранения ценной информации и программного обеспечения компьютера для защиты от повреждения и уничтожения (умышленного или случайного, под действием компьютерного вируса).

Архиваторы позволяют сжимать информацию в памяти компьютера при помощи специальных математических методов. При этом создается копия файла меньшего размера. В одном архиве может храниться сразу несколько файлов или даже несколько папок. Это даёт возможность разместить больше информации на диске или дискете.

ВИДЫ АРХИВАТОРОВ

WinZip

Самый распространенный формат архивов - Zip, и самый распространенный архиватор - WinZip. Невысокая степень сжатия компенсируется широкой известностью WinZip.

Не смотря на то, что степень сжатия у программы невелика, она обладает простым интерфейсом, а также мастерами для создания и распаковки архивов. Zip-архивы могут включать в себя комментарии и быть зашифрованными с использованием пароля. Поддержка многотомных архивов в самом формате Zip ограничена: их можно создавать только на дискетах.

Стандартная версия WinZip может создавать только простейшие самораспаковывающиеся (SFX) архивы. Чтобы запустить программу после извлечения файлов из архива или изменить текст в окне распаковщика, необходима утилита WinZip Self-Extractor.

WinRAR

Вечный конкурент программы WinZip. Данный архиватор особой любовью пользуется у россиян во многом благодаря тому, что создан соотечественником Леонидом Рошалем. На мировом рынке программных средств известен как архиватор с одним из лучших алгоритмов сжатия данных.

При упаковке данных можно задать степень сжатия архива или отключить сжатие. Помимо этого, программа позволяет защищать архивы паролем (можно указать и то, какие файлы из архива должны быть доступны, даже если пользователь не знает пароля), писать комментарии, создавать самораспаковывающиеся архивы. В WinRar присутствует возможность воспользоваться специальным алгоритмом мультимедиа-сжатия, оптимизирующим компрессию оцифрованного звука и графических файлов.

ВИДЫ АРХИВАТОРОВ

7Zip

Это ещё один архиватор от российских разработчиков. Благодаря более совершенному алгоритму сжатия качество архивирования данных у собственного формата *.7z опережает даже WinRAR. Правда, такой коэффициент сжатия достигается благодаря тому, что архиватор работает медленнее. В остальном 7-Zip действительно во многом превосходит более известных конкурентов, а потому предпочтение ему отдают все больше пользователей.

7-Zip дает возможность создавать архивы, защищенные паролем, а также шифровать имена файлов, которые содержатся в архивах. При этом используется 256-битный алгоритм AES. Именно этот алгоритм был в 2001-году выбран американским национальным институтом стандартов и технологий (NIST) в качестве криптографического стандарта на ближайшее время, поэтому можно сказать, что это - один из лучших алгоритмов шифрования на сегодняшний день.

WinAce

Несмотря на то, что WinAce выпускается довольно давно, добиться такой популярности среди пользователей, как WinRar, архиватору не удалось. Для того чтобы архиватор мог соперничать с конкурентом, он должен превосходить соперника по двум параметрам - скорости и степени сжатия. Размер архивов в формате ACE часто меньше чем после использования WinRar, однако не настолько, чтобы привлечь к себе внимание.

Поддерживается создание непрерывных архивов, проверка на целостность всех знакомых программе форматов, восстановление поврежденных ACE и ZIP-архивов и многое другое. Кроме того, можно конвертировать архивы различных форматов в ACE для уменьшения их размера. Программа обладает отличной интеграцией с Windows.

ВИДЫ АРХИВАТОРОВ

PowerArchiver

Очень мощный архиватор. Красивый, удобный и понятный интерфейс. PowerArchiver в отличии от WinRAR, 7Zip и WinAce не предоставляет своего собственного формата для сжатия данных однако он поддерживает большое количество различных архивов включая и RAR, ACE и 7-ZIP причем последний поддерживается полностью.

Встроенный просмотрщик позволяет "на лету" просматривать TXT, RTF и все популярные форматы графических файлов прямо в архиве. Кроме того, PowerArchiver предлагает пользователю 5 различных методов шифрования особо секретных данных.