**Тема: Создание архива данных. Извлечение данных из архива. Атрибуты файла и его объем**

**Как же происходит сжатие информации?**

Сжатие данных похоже на производство сухого молока или сухофруктов. То есть, это – процесс удаления воды, которую затем можно добавить, чтобы предать продукту первоначальный вид.

А какая в данных может быть вода? Это вода информационная. В данных встречается много повторов. Это можно использовать для сжатия данных.

Например, сжатие текстовых файлов происходит приблизительно так. Составляется таблица встречающихся в тексте слов и выражений. Затем всем словам и выражениям в этой таблице даются номера. И весь текст в файле заменяется номерами из таблицы слов и выражений. Такой способ позволяет уменьшить размер текстового файла в 2-3 раза. Иногда текст сжимается и в 10 раз, если в нём много повторов.

Программа, переводящая текстовый файл в "сжатый" вид, называется упаковщиком. А полученный в результате сжатия файл называется упакованным или сжжатым файлом.

Очень часто сжатые файлы называют архивами или архивными файлами, что, если подходить строго к терминологии, неверно. Изначально архивами назывались файлы, специально созданные во время процессов резервного копирования. В ходе такого процесса создавался один файл, который содержал в себе несколько исходных файлов и папок. Это и был архив. Никакого сжатия при этом не производилось. Подобная ситуация до сих пор существует в операционной системе Linux, где архивация данных и сжатие данных являются двумя независимыми процессами. В операционной системе MS-DOS, а затем и в MS Windows программы сжатия данных с самых первых своих версий стали поддерживать как сжатие, так и архивацию данных, то есть создавали сжатый файл, содержащий не один, а несколько исходных (архивируемых) файлов и папок. С той поры в этих операционных системах понятие "архивация" означает и архивацию (сбор в один файл-архив), и одновременное сжатие данных.

Поскольку архивный файл записан не в текстовом формате, с ним не могут работать редакторы текстов. Перед открытием архивного файла редактором текстов этот файл необходимо разархивировать. Разархивацию производит та же самая программа – архиватор. После разархивации текстовый файл приобретает точно такой же вид и размер, как и до архивации.

Архиваторы текстов могут архивировать также файлы программ. Только программы гораздо меньше сжимаются, чем текст.

Упаковщики, применяемые для сжатия текстов и программ, не могут эффективно сжимать звуковые, графические или видео файлы. Для их сжатия были разработаны другие, более сложные, алгоритмы. Правда, после распаковки полученные файлы немного отличаются от оригиналов (такое сжатие называется сжатием с потерями). Но этого не улавливает обычное человеческое ухо и не замечает обычный глаз на экране монитора.

**А как же другие, не текстовые данные?**

Рассмотрим графические файлы. Незаархивированная графика – это рисунок, состоящий из множества разноцветных точек. В таком формате для каждой точки рисунка или фотографии задаётся цвет. Графический файл такого рисунка имеет расширение «BMP». Но такие файлы обладают довольно солидным размером. Даже небольшая фотография в формате «BMP» будет иметь размер в несколько мегабайт. То есть, она не поместится на дискету, и по сети Интернет передать её будет нелегко.

Для уменьшения размера графического файла его сжимают специальными математическими методами. Таким образом графический файл можно уменьшить в размере в 20-30 раз. Сжатый графический файл нельзя полностью восстановить. При распечатывании на хорошем принтере будет заметно ухудшение качества рисунка. Но на экране монитора этого видно не будет.

Наиболее распространённые форматы сжатия изображения – это «GIF» и «JPEG» (или «JPG»). Создать такие форматы графических файлов может, например, графический редактор «ФотоШоп» или уже знакомый нам «ФайнРидер». Чем глубже сжатие рисунка, тем больше теряет он в качестве.

Гораздо позже появились эффективные методы сжатия звука. Если Вы зайдёте на обычный музыкальный компакт-диск, то обнаружите там файлы с расширением «CDA». В таких файлах содержится несжатый цифровой звук. Такие файлы имеют очень большой размер. На компакт-диск помещается 80 минут музыки в незаархивированном виде. Сжатие звуковых файлов потребовало специальных исследований человеческого слуха. Оказалось, что часть звуков можно убрать из звукового файла, и это будет незаметно для слуха человека.

Самый распространённый формат сжатого звука – это «MPEG3» (или «MP3»). Его создают специализированные редакторы звука и программмы-перекодировщики из других звуковых форматов. Проигрывание на компьютере сжатых звуковых файлов в формате «MP3» требует мощности процессора не менее 100 мегагерц. Можно сжимать звуковой файл в десятки раз. Но при большом сжатии начинает безвозвратно теряться качество звука.

Видеофильм объединяет в себе звук и графику. Помимо звука – это ещё 24 кадра на каждую секунду. Отсюда понятны огромные размеры файлов с видеофильмами и необходимость их сжатия.

При появлении компакт-дисков художественный фильм не помещался на один диск. Эту задачу решил сжатый формат «MPEG4». Показ сжатого видео ещё больше повышает требования к мощности процессора. Для проигрывания MPEG4 уже недостаточно 200 мегагерц.

**Архивация** (упаковка) — помещение (загрузка) исходных файлов в архивный файл в сжатом или несжатом виде.

Архивация предназначена для создания резервных копий используемых файлов, на случай потери или порчи по каким-либо причинам основной копии (невнимательность пользователя, повреждение магнитного диска, заражение вирусом и т.д.).

Для архивации используются специальные программы, архиваторы, осуществляющие упаковку и позволяющие уменьшать размер архива, по сравнению с оригиналом, примерно в два и более раз.

Архиваторы позволяют защищать созданные ими архивы паролем, сохранять и восстанавливать структуру подкаталогов, записывать большой архивный файл на несколько дисков (многотомный архив).

Сжиматься могут как один, так и несколько файлов, которые в сжатом виде помещаются в так называемый архивный файл или архив. Программы большого объема, распространяемые на дискетах, также находятся на них в виде архивов.

**Архивный файл** — это специальным образом организованный файл, содержащий в себе один или несколько файлов в сжатом или несжатом виде и служебную информацию об именах файлов, дате и времени их создания или модификации.

Выигрыш в размере архива достигается за счет замены часто встречающихся в файле последовательностей кодов на ссылки к первой обнаруженной последовательности и использования алгоритмов сжатия информации.

Степень сжатия зависит от используемой программы, метода сжатия и типа исходного файла. Наиболее хорошо сжимаются файлы графических образов, текстовые файлы и файлы данных, для которых степень сжатия может достигать 5 - 40%, меньше сжимаются файлы исполняемых программ и загрузочных модулей — 60 - 90%. Почти не сжимаются архивные файлы. Программы для архивации отличаются используемыми методами сжатия, что соответственно влияет на степень сжатия.

Для того чтобы воспользоваться информацией, запакованной в архив, необходимо архив раскрыть или распаковать. Это делается либо той же программой-архиватором, либо парной к ней программой-разархиватором.

**Разархивация** (распаковка) — процесс восстановления файлов из архива в первоначальном виде. При распаковке файлы извлекаются из архива и помещаются на диск или в оперативную память.

**Самораспаковывающийся архивный файл** — это загрузочный, исполняемый модуль, который способен к самостоятельной разархивации находящихся в нем файлов без использования программы-архиватора.

Самораспаковывающийся архив получил название SFX-архив (SelF-eXtracting). Архивы такого типа в обычно создаются в форме .ЕХЕ-файла.

Архиваторы, служащие для сжатия и хранения информации, обеспечивают представление в едином архивном файле одного или нескольких файлов, каждый из которых может быть при необходимости извлечен в первоначальном виде. В *оглавлении архивного файла* для каждого содержащегося в нем файла хранится следующая информация:

* + имя файла;
  + сведения о каталоге, в котором содержится файл;
  + дата и время последней модификации файла;
  + размер файла на диске и в архиве;
  + код циклического контроля для каждого файла, используемый для проверки целостности архива.

***Архиваторы имеют следующие функциональные возможности****:*

* 1. Уменьшение требуемого объема памяти для хранения файлов от 20% до 90% первоначального объема.
  2. Обновление в архиве только тех файлов, которые изменялись со времени их последнего занесения в архив, т.е. программа-упаковщик сама следит за изменениями, внесенными пользователем в архивируемые файлы, и помещает в архив только новые и измененные файлы.
  3. Объединение группы файлов с сохранением в архиве имен директорий с именами файлов, что позволяет при разархивации восстанавливать полную структуру директорий и файлов.
  4. Написания комментариев к архиву и файлам в архиве.
  5. Создание саморазархивируемых архивов, которые для извлечения файлов не требуют наличия самого архиватора.
  6. Создание многотомных архивов – последовательности архивных файлов. Многотомные архивы предназначены для архивации больших комплексов файлов на дискеты.