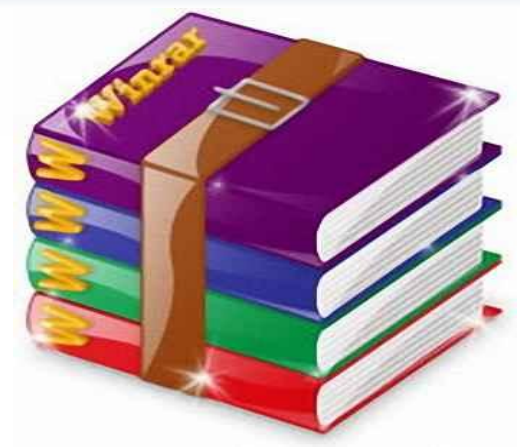


Архивация



Архивация

Архивация – это упаковка (сжатие) файла или группы файлов с целью, в большинстве случаев, уменьшить место, занимаемое ими на диске.

Архиватор – программа, позволяющая за счёт применения специальных методов создавать копии файлов меньшего размера, из которых можно затем извлечь файлы в их первоначальном виде (без искажений).



Объекты сжатия

✓ файлы



✓ папки



Архивация

✓ ДИСКИ(файлы на диске)



Локальный диск (C:)

Назначение архивации

Архивация файлов и папок используется для

- ✓ хранения,
- ✓ транспортировки,
- ✓ резервного копирования,
- ✓ «уплотнения» дисков,
- ✓ проверки целостности файла(ов)
- ✓ защиты от несанкционированного доступа
- ✓ обмена информацией по почте, сети Интернет.

“Уплотнение” дисков применяют для повышения эффективности использования их рабочего пространства.

Упаковка (архивация) – помещение (загрузка) исходных файлов в архивный файлов сжатом или несжатом виде. Распаковка (разархивация) – процесс восстановления. файлов из архива точно в таком виде, какой они имели да загрузки в архив. При распаковке файлы извлекаются из архива и помещаются на диск или в оперативную память.

Типы архиваторов



Файловые архиваторы

Упаковывают один или несколько файлов в единый архивный файл, размер которого, как правило, меньше, чем суммарный размер исходных файлов, но воспользоваться запакованными программами или данными, пока они находятся в архиве, нельзя, пока они не будут распакованы. Для распаковки архивного файла можно использовать как этот архиватор, так и любой другой, поддерживающий использованный алгоритм сжатия (архивации).

Программные архиваторы

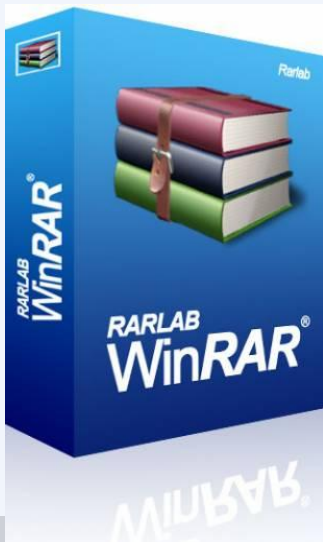
Позволяют упаковать за один прием один единственный файл - исполняемую программу EXE-типа, но зато так, что заархивированная программа будет сразу после ее запуска на исполнение самораспаковываться в оперативной памяти и тут же начинать работу.

Дисковые архиваторы

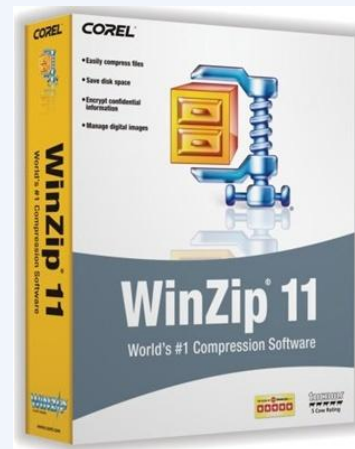
Позволяют программным способом увеличить доступное пространство на жестком диске. Типичный дисковый архиватор представляет собой резидентный драйвер, который незаметно для пользователя архивирует любую записываемую на диск информацию и распаковывает ее обратно при чтении. Однако операции чтения/записи файлов несколько замедляются, поскольку процессору требуется время для упаковки и распаковки.

Распространенные архиваторы

Существуют разные программы-архиваторы, отличающиеся форматом архивных файлов, скоростью работы, степенью сжатия файлов при помещении в архив, удобством использования. Наиболее распространенные программы-архиваторы имеют приблизительно одинаковые возможности и ни одна из них не превосходит другие по всем параметрам: одни программы работают быстрее, другие обеспечивают лучшую степень сжатия файлов. Даже если сравнивать программы только по степени сжатия, то среди них нет лидера: разные файлы лучше сжимаются разными программами.



WinRAR
WinZip
7-Zip



Архиваторы

1. Архиваторы используют разные алгоритмы сжатия. Алгоритм архивации определяется типом (расширением) архивного файла: **zip**, **arj**, **rar**, **7z**, **ice** и др. Например, архивы архиватора **Winrar** имеют расширение **rar**.
2. Для некоторых архиваторов можно выбирать степень «сжатия»: **store**(без сжатия – высокая скорость архивации), **fastest**, **fast**, **normal**, **good**, **maximum**(максимальное «сжатие», минимальная скорость архивации)
3. Самораспаковывающиеся архивы. Называются такие архивы – **SFX**-архивы (от **Self-eXtracting**). Имеют, как правило, расширение исполняемого файла – **exe**. Для распаковки надо просто запустить этот файл. Распаковщик не нужен
4. Как правило, архиваторы «умеют» работать не только со своим форматом, а «понимают» и другие форматы, т.е. могут разархивировать файлы, «сжатые» другими архиваторами
5. В некоторых случаях архиваторы могут восстановить поврежденный архив

Многотомный архив

Многотомный архив. Многотомный архив – это архив «раздробленный» на некоторое количество частей (томов), причем: первый том (файл!) многотомного архива получает расширение *.rar, а расширения последующих томов (файлов!) нумеруются как *.r00, *.r01, *.r02 и т.д. до *.r99.
Пример использования:

- ✓ Электронная почта, когда вложение (attachment) не может превышать заданный размер. Его приходится разбивать на несколько частей
- ✓ Архив не помещается на один носитель (диск, флэшка) и его записывают на несколько носителей

Популярные архиваторы

- 1. WinRAR** — самый популярный архиватор. Алгоритм сжатия формата RAR считается одним из лучших, а удобный интерфейс, поддержка всех востребованных форматов архивации и общее высокое качество не оставляет шансов конкурентам.
- 2. WinZIP** — программа, получившая всемирную популярность, как первый архиватор с графическим интерфейсом.
- 3. 7-Zip** — в первую очередь, этот архиватор привлекателен своим собственным форматом **7z**. Формат **7z** имеет наивысшую степень сжатия и отлично подходит для архивации больших объемов информации.
- 4. PowerArchiver** — мощный архиватор с продвинутым интерфейсом и широким ассортиментом дополнительных функций
- 5. WinACE** — Помимо поддержки ряда популярных форматов имеет свой собственный формат ACE

Непрерывный архив

- 1. Непрерывный архив** — это архив , упакованный специальным способом, при котором все сжимаемые файлы рассматриваются как один последовательный поток данных.
- 2. Непрерывная архивация** значительно увеличивает степень сжатия, особенно при добавлении в архив значительного количества небольших файлов с похожим содержимым. Однако следует иметь в виду, что у непрерывной архивации есть и некоторые **недостатки**:
 - ✓ чтобы извлечь **один** файл из непрерывного архива, приходится анализировать **все** предыдущие заархивированные файлы, поэтому извлечение отдельных файлов из непрерывного архива происходит медленнее, чем извлечение из обычного архива. Однако если из непрерывного архива извлекаются все файлы , то в этом случае скорость распаковки практически равна скорости распаковки обычного архива
 - ✓ **обновление** непрерывных архивов (т.е. добавление файлов в уже существующий архив или их удаление) происходит медленнее,

Алгоритмы «сжатия» - 1

основаны на поиске в файле повторяющихся «структур» (корреляции) и последующем более компактном представлении с целью получения минимального объема. Самым известным методом архивации файлов является **сжатие последовательностей одинаковых символов**. Например, внутри файла находятся последовательности байтов, которые часто повторяются. Вместо того, чтобы хранить каждый байт, фиксируется количество повторяемых символов и их позиция. Например, архивируемый файл состоит из 14-ти байт и состоит из символов: **V V V V L L L L L L A A A**. В шестнадцатеричной системе это будет выглядеть так:

0x42 0x42 0x42 0x42 0x4C 0x4C 0x4C 0x4C 0x4C 0x4C 0x4C 0x41 0x41 0x41.

Архиватор может представить этот файл в виде (16-тиричном):

0x04 0x42 0x07 0x4C 0x03 0x41.

Это значит, что с первой позиции 4 раза повторяется символ "V", затем семь раз повторяется "L" и наконец три раза повторяется "A". Для хранения файла в такой форме потребуется всего 6 байт, что на 8 байт меньше исходного.

Note! Количество информации в исходном файле и архивированном одинаково. Изменилось только лишь форма хранения [информации]. Описанный метод представляет собой самый простой способ уменьшения размера файлов.

Note! Сделайте вывод. От чего зависит степень уменьшения размера файла

Алгоритмы сжатия - 2

Более изощренный метод сжатия данных – это так называемый оптимальный префиксный код и, в частности, *кодирование символами переменной длины* (алгоритм Хаффмана <http://ppt4web.ru/informatika/szhatie-informacii-algoritm-khaffmana.html>). Код переменной длины позволяет записывать наиболее часто встречающиеся символы и группы символов всего лишь несколькими битами, в то время как редкие символы и фразы будут записаны более длинными битовыми строками. Например, в любом английском тексте буква E встречается чаще, чем Z, X и Q. Таким образом, используя специальную таблицу соответствия, можно закодировать каждую букву E меньшим числом бит и использовать более длинный код для более редких букв.

Некоторые архиваторы работают на основе алгоритма **Лемпела-Зива**. Эти архиваторы классифицируются как адаптивные *словарные кодировщики*, в которых текстовые строки заменяются указателями на идентичные им строки, встречающиеся ранее в тексте. Например, все слова какой-нибудь книги могут быть представлены в виде номеров страниц и номеров строк некоторого словаря.

Note! Упрощенное описание алгоритмов

Популярные алгоритмы «сжатия»

1

**Алгоритм
RLE**

2

**Алгоритм
KWE**

3

**Алгоритм
Хаффмана**

Алгоритм RLE

использует принцип выявления повторяющихся последовательностей и замены их простой структурой, в которой указывается код данных и коэффициент повтора.

При сжатии записывается последовательность из двух повторяющихся величин: повторяемого значения и количества его повторений.

Пример

Исходная последовательность: 3, 3, 12, 12, 12, 0, 0, 0, 0.

Сжатая информация: 3, 2, 12, 3, 0, 4.

Коэффициент сжатия: $6/9 * 100\% = 67\%$.

Алгоритм KWE

предполагает использование словаря, в котором каждому слову соответствует двухбайтовый код. Эффективность сжатия увеличивается с ростом объёма кодируемого текста.

Примечание. Каждый символ в слове (каждое знакоместо в текстовом файле) – один байт!

Алгоритм Хаффмана

предполагает кодирование не байтами, а битовыми группами. В нем можно выделить 3 основные этапа:

1. Выявляется частота повторения каждого из встречающихся символов.
2. Чем чаще встречается символ, тем меньшим количеством битов он кодируется.
3. К закодированной последовательности прикладывается таблица соответствия.

Степень сжатия зависит от



Степень сжатия определяется коэффициентом сжатия:

$$K_c = V_c / V_{и} * 100\%,$$

где $V_{и}$ – размер исходного файла,

V_c – размер «сжатого» файла

В первом приближении **K_c** составляет десятки процентов

Полезное

- ✓ **Архив** — это файл, содержащий в себе один или несколько файлов в сжатом (упакованном виде). Имена файлов архивов могут иметь расширения .zip, arj, rar, arc и др. в зависимости от типа архиватора, в котором они были созданы.
- ✓ **Архиватор** — специальная программа, с помощью которой можно “сжимать” файлы, увеличивая эффективность использования дискового пространства, просматривать содержание архивов, контролировать их целостность, распаковывать архивы и т.п.
- ✓ **Архивы** создаются с целью уменьшения объема исходного файла (экономится место на диске) и уменьшения времени передачи/приема (например, при пересылке по Интернету и электронной почте).
- ✓ Архиваторы позволяют закрыть доступ к файлу с помощью пароля
- ✓ Если будет поврежден архивный файл, архиватор при разархивации обнаружит этот факт

Обратимость сжатия

Для некоторых видов файлов допустима некоторая потеря информации. Например, для звуковых и видео файлов, файлов, содержащих изображение.

Этот метод называют методом сжатия с регулируемой потерей информации.

Форматы сжатия с потерей информации

- ✓ JPG- графика
- ✓ MPG- видео
- ✓ MP3- звук

Форматы сжатия без потери информации

- ✓ GIF, TIF, PCX
- ✓ AVI
- ✓ ZIP, ARJ, RAR

Программа WinRAR позволяет

- ✓ создавать rar-архивы и извлекать из них файлы. Степень сжатия может достигать 95 %;
- ✓ создавать самораспаковывающиеся архивы – архивный файл в этом случае будет иметь расширение .exe. Запускается как обычная программа. Файлы из такого архива могут быть извлечены на любом компьютере путём запуска программы;
- ✓ создавать многотомные архивы, что позволяет разместить большой архив на нескольких носителях ограниченной емкости;
- ✓ устанавливать пароль на архив – при этом посторонний сможет лишь просмотреть список файлов архива, но не сможет распаковать архив;
- ✓ восстанавливать поврежденные архивы (до определенного уровня повреждения);
- ✓ распаковывать архивы, созданные другими архиваторами



WinRAR
Ярлык
1 КБ

Вопросы

1. Найти структуру архивного файла

Вопросы

1. Что такое архивация?
2. Для чего необходима архивация?
3. Почему есть возможность уменьшать размер файлов?
4. Какие файлы не имеет смысла архивировать?
5. Почему перед пересылкой текстового файла по электронной почте имеет смысл предварительно его упаковать в архив?
6. Что такое архиватор?
7. Перечислить типы архиваторов.
8. Какие 3 принципа работы архиваторов Вам известны?
9. Что значит многотомный архив?
10. Что такое самораспаковывающийся архив?
11. Что такое степень сжатия файла?
12. Для чего нужно резервное копирование?
13. Что такое разархивация?



Напоследок об исполняемых файлах

Исполняемые файлы (примеры:)

- WINWORD.EXE** – Word
- POWERPNT.EXE** – Презентации
- OUTLOOK.EXE** – Mail (почта)
- EXCEL.EXE** – Excell (таблицы)
- MSACCESS.EXE** – Access (базы данных)