

Лабораторная работа № U8. Архивирование/сжатие информации

Есть два принципиальных типа алгоритмов:

- без потери информации при сжатии - используются для сжатия документов, программ, данных
- с потерей части информации – обычно для сжатия видеоданных, фотографий

Архивирование данных может применяться для:

- Упрощения передачи данных между компьютерами (например по почте или компьютерной сети) – один большой файл быстрее и проще передать чем много мелких файлов
- Хранения старых данных – например старых файлов с логами
- Удобства хранения – например архив «мой старый компьютер» или «дипломная работа»
- Создания резервных копий – можно создавать архивы важных данных ежемесячно без необходимости применения специализированного ПО и оборудования

Обычно считается что архивирование необходимо вызывать вручную, но у всех программных средств такого типа есть возможность запуска из командной строки, в автоматическом режиме.

В отличие от законченных и полнофункциональных программ в ОС типа Unix обычно используются отдельные программы:

1. Для сборки отдельных файлов в один «пакет» - обычно это программа TAR
2. Для сжатия файлов:
 - a. Gzip – «старая» и широко распространенная программа.
 - b. Bzip2 – относительно недавно появившаяся программа. Обеспечивает более высокую степень сжатия информации.
 - c. Xz - недавно появившийся архиватор. Самая высокая степень сжатия из трёх архиваторов.

Полезность архивирования на примере исходного кода программы GCC версии 4.6.1 – 71тысяча текстовых файлов, в несжатом виде занимают 410Мб. Сжатый GZIP: 89Мб, сжатый BZIP2: 68Мб.

У файлов, обработанных TAR-ом, расширение будет .tar, обработанных GZIP-ом, расширение будет .gz, если это был BZIP2, расширение будет .bz2

Если файлы сначала были обработаны TAR-ом, а потом сжаты BZIP2-ом, расширение будет .tar.bz2

Общие сведения об использовании программы TAR:

Назначение: собрать файлы в указанном каталоге в один большой файл или «разобрать» пакет на составляющие файлы.

Опция `-x` означает извлечение

Опция `-v` означает включение подробного режима

Опция `-z` сообщает tar'y о том, что файл .tar.gz нужно сначала пропустить через gunzip

опция `-c` сообщает tar'y, что нужно создать архив

опция `-p` права доступа сохраняет/восстанавливает

опция `-f` сообщает tar'y, что следующая строка в командной строке - это файл, с которым надо работать

Команда для создания архива, вся информация при этом выводится в стандартный поток вывода

`tar -cp logs` ...и может быть перенаправлена либо в программу-архиватор, либо в файл:

`tar -cp logs >file.tar`

«Распаковать» файл: **`tar -vxf file.tar`**

Кроме того TAR может задействовать один из архиваторов в системе при указании дополнительных опций : `-z` для gzip, `-j` для bzip2 и `-J` для xz

Общие сведения об использовании программы GZIP:

GZIP – архиватор, сжимает или раскрывает указанный файл.

Для сжатия необходимо использовать команду **`gzip file`**

В итоге будет создан файл file.gz, при этом исходный файл будет удален

Для разархивирования нужно использовать **`gunzip file.gz`**

Общие сведения об использовании программы BZIP2:

BZIP2 – архиватор, сжимает или раскрывает указанный файл, использование аналогично GZIP.

Задание на самостоятельную работу:

1. Заархивировать каталог /usr/include, архив создавать в домашнем каталоге. Архиватор gzip, bzip2 или xz – на свое усмотрение.
2. Заархивировать каталог /usr/bin, архив создавать в домашнем каталоге. Архиватор должен быть тем же что и для п1.
3. Определить исходный размер каталогов /usr/include и /usr/bin (для определения можно воспользоваться командой du), оценить степень сжатия двоичных и текстовых файлов.
4. Распаковать в домашнем каталоге архив, проверить объем распакованных директорий

PS. Не забыть удалить рабочие файлы, они очень большого размера!

Каталог/тип файлов	Исходный размер	Сжатый размер	Степень сжатия (исх/сжатый)
/usr/include/ - директория с либами (файлами разных форматов)	26 М	3,4 М	$26/3,4 = 7,65$
0/usr/bin/ - папка исполнительных файлов	125 М	48 М	$125/48 = 2,6$