

C2110 *UNIX and programming*

Lesson 13 / Module 1

PS / 2020 Distance form of teaching: Rev1

Petr Kulhanek

kulhanek@chemi.muni.cz

National Center for Biomolecular Research, Faculty of Science
Masaryk University, Kamenice 5, CZ-62500 Brno

Compression

- Lossless
- Lossy

Compression

Compression is a procedure that reduces the size of data (files). It is achieved by searching **redundant** or **irrelevant** information in the data, which is then stored more efficiently. According to the type of compression algorithm, data compression can be divided into two basic categories:

- **lossy compression** - there is an irreversible loss of some irrelevant information, which is usually tolerated when compressing video or audio data
- **lossless compression** - there is no loss of the original information, the compressed data can be restored to its original state, the compression rate is several times lower than the lossy compression

Recovering compressed data is called **decompression**.

A compression ratio indicates the quality of the compression. It is given as the ratio of the size of the original data (in bytes) to the size of the compressed data.

Lossy compression

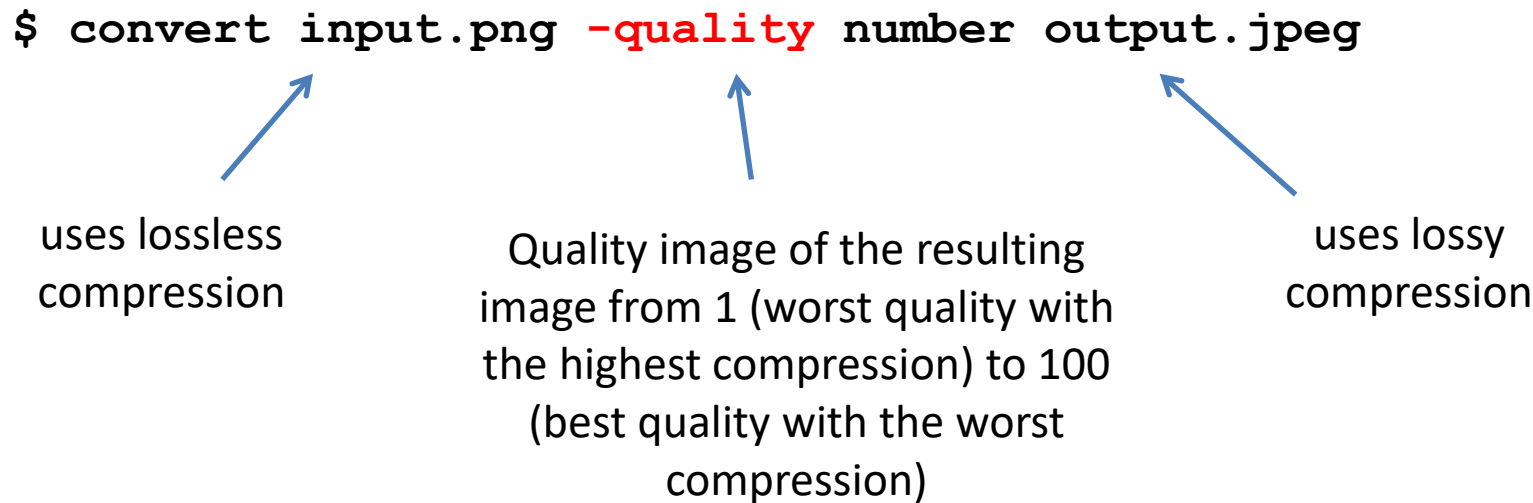
Lossy compression and decompression programs:

- **mplayer**
- **mencoder**
- **convert** (Image Magick)
- and more ...

Convert PNG (Portable Network Graphics) image to JPEG (Joint Photographic Experts Group) format:

```
$ convert input.png -quality number output.jpeg
```

uses lossless
compression



Quality image of the resulting
image from 1 (worst quality with
the highest compression) to 100
(best quality with the worst
compression)

uses lossy
compression

Exercise I

1. From the directory `/home/kulhanek/Documents/C2110/Lesson13` copy the image `test.png` to your home directory.
2. What is the size of the image file in bytes?
3. Do a lossy compression of the image to format `jpeg`. Use quality for compression 10, 50 and 90. Save the resulting images separately.
4. Compare visual quality compressed images (display command).
5. What is a compression ratio for quality 10 and 90?

Lossless compression

Programs for **lossless** compression and decompression:

- **gzip/gunzip**
- **bzip2/bunzip2**
- **zip/unzip**
- and more ...

Compression text file:

```
$ gzip file.txt
```

```
$ bzip2 file.txt
```

the resulting file will be named
file.txt.gz



the resulting file will be named
soubor.txt.bz2



Decompression of compressed data:

```
$ gunzip file.txt.gz
```

```
$ bunzip2 file.txt.bz2
```

Compression or decompression can be performed by sending the result **to standard output** (the original file then remains unchanged), e.g.:

```
$ bunzip2 -stdout soubor.txt.bz2 | wc
```

Exercise II

1. Copy the text file `bu6_f.log` to your home directory from `/home/kulhanek/Documents/C2110/Lesson13` directory.
2. What is the size of the file in bytes?
3. Perform lossless file compression using programs `gzip` and `bzip2`. Which program achieves a higher compression ratio?
4. Which program compresses the file faster?

Archives

- **Types**
- **Creating and unpacking archives**

Archives - tar

tar (abbreviation from tape archiver) is a collective name for a file format used to store many individual files, as well as for single-purpose programs that work with this format. The format itself originated in the early days of Unix and was later standardized under the POSIX standard. Originally it helped in archiving files on tape drives, but later its use expanded and today it is used simply where it is appropriate to merge **multiple files into one so that information about directory structure, access rights, and other attributes**, which the file system normally contains, **is preserved** for distribution or archiving purposes.

www.wikipedia.org

Unpack archive:

```
$ tar xvf archive.tar
```

if archive name contains .gz or .bz2 ending, the archive is automatically decompressed or compressed

Create archive:

```
$ tar cvf archive.tar directory/
```

```
$ cd directory
```

```
$ tar cvf /path/to/archive.tar *
```

Exercise III

1. What is the meaning of options cvf of tar command?
2. What is the meaning of options xvf of tar command?
3. Create an archive from files stored in the directory:
`/home/kulhanek/Documents/C2110/Lesson13/Archive`
4. What is the size of the file containing the archive?
5. Compress the archive. What is the compression ratio?
6. Unzip the archive to directory `/scratch/your_login/archive`