

C2110 Operační systém UNIX a základy programování

U13: Komprese a archivace dat

PS/2021 Prezenční forma výuky: Rev8

Petr Kulhánek

kulhanek@chemi.muni.cz

Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta
Masarykova univerzita, Kamenice 5, CZ-62500 Brno

Komprese

- **Bezeztrátová**
- **Ztrátová**

Kompresa

Kompresa (komprimace) je postup, kterým dochází k snížení velikosti dat (souborů). Docíluje se toho vyhledáváním **redundantních** nebo **nepodstatných** informací v datech, které jsou pak ukládány efektivněji. Podle typu kompresního algoritmu, lze kompresi dat rozdělit do dvou základních kategorií:

- **ztrátová komprese** – dochází k nevratné ztrátě některých nepodstatných informací, což je většinou tolerováno při kompresi obrazových či zvukových dat
- **bezeztrátová komprese** – nedochází k žádné ztrátě původní informace, komprimovaná data lze obnovit do původního stavu, míra komprese je několikanásobně nižší než u ztrátové komprese

Obnova komprimovaných dat se nazývá **dekomprese**.

Kompresní poměr udává kvalitu komprese. Udává se jako poměr velikosti původních dat (v bytech) vůči velikosti komprimovaných dat.

Ztrátová komprese

Programy pro ztrátovou kompresi a dekompresi:

- **mplayer**
- **mencoder**
- **convert** (Image Magick)
- a další ...

Převod obrázku ve formátu PNG (Portable Network Graphics) do JPEG (Joint Photographic Experts Group):

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

↑
využívá bezztrátové
komprese

↑
míra kvality výsledného obrázku
od 1 (nejhorší kvalita s největší
kompresí) po 100 (nejlepší kvalita
s nejhorší kompresí)

↑
využívá ztrátovou
kompresi

Cvičení 1

1. Z adresáře /home/kulhanek/Documents/C2110/01.Unix/U13 zkopírujte obrázek test.png do vašeho domovského adresáře.
2. Jakou velikost má soubor obrázku v bytech?
3. Proveďte ztrátovou kompresi obrázku do formátu jpeg. Ke kompresi použijte kvalitu 10, 50 a 90. Výsledné obrázky ukládejte každý zvlášť.
4. Porovnejte vizuální kvalitu komprimovaných obrázků (příkaz display).
5. Jaký je kompresní poměr pro kvalitu 10 a 90?

Bezeztrátová komprese

Programy pro **bezeztrátovou** kompresi a dekompresi:

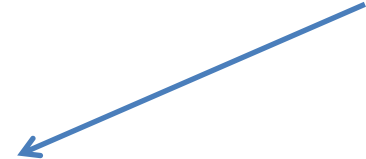
- **gzip/gunzip**
- **bzip2/bunzip2**
- **zip/unzip**
- a další ...

Komprese textového souboru:


```
$ gzip soubor.txt
```

```
$ bzip2 soubor.txt
```

výsledný soubor se bude jmenovat
soubor.txt.gz



výsledný soubor se bude jmenovat
soubor.txt.bz2



Dekomprese komprimovaných dat:

```
$ gunzip soubor.txt.gz
```

```
$ bunzip2 soubor.txt.bz2
```

Kompresi či dekompresi lze uskutečnit tak, že výsledek je poslán **do standardního výstupu** (originální soubor pak zůstává nezměněn), např.:

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

Cvičení 2

1. Z adresáře /home/kulhanek/Documents/C2110/01.UNIX/U13 zkopírujte textový soubor bu6_f.log do vašeho domovského adresáře.
2. Jakou velikost má soubor v bytech?
3. Proveďte bezeztrátovou kompresi souboru pomocí programů gzip a bzip2. Který z programů dosahuje vyššího kompresního poměru?
4. Který z programů komprimuje soubor rychleji?

Archívy

- Typy
- Vytváření a rozbalování archivů

Archívy - tar

tar (zkratka z **anglického tape archiver neboli páskový archivovač**) je souhrnný název jednak pro souborový formát sloužící k uložení mnoha jednotlivých souborů, jednak pro jednoúčelové programy, které s tímto formátem pracují. Formát samotný vznikl v počátcích Unixu a až později byl standardizován v rámci normy POSIX. Původně pomáhal při archivaci souborů na páskových jednotkách, ale později se jeho užívání rozšířilo a dnes je používán zkrátka tam, kde je vhodné pro účely distribuce či archivace sloučit **více souborů do jednoho tak, aby se zachovaly informace o adresářové struktuře, přístupových právech a dalších attributech**, které běžně souborový systém obsahuje.

www.wikipedia.org

Rozbalení archivu:

```
$ tar xvf archiv.tar
```

pokud jméno archivu obsahuje zakončení .gz nebo .bz2 tak je archiv automaticky dekomprimován nebo komprimován

Vytvoření archivu:

```
$ tar cvf archiv.tar adresar/
```

```
$ cd adresar
```

```
$ tar cvf /cesta/k/archiv.tar *
```

Cvičení 3

1. Jaký význam mají volby cvf příkazu tar?
2. Jaký význam mají volby xvf příkazu tar?
3. Vytvořte archiv ze souborů uložených v adresáři:
`/home/kulhanek/Documents/C2110/01.UNIX/U13/Archive`
4. Jakou velikost má soubor obsahující archiv?
5. Proveďte kompresi archivu. Jaký je kompresní poměr?
6. Archiv rozbalte do adresáře `/scratch/vas_login/archiv`