



FAKULTA PEDAGOGICKÁ
ZÁPADOČESKÉ
UNIVERZITY
V PLZNI

Bezeztrátová komprese

- **Komprese a dekomprese**
- **Dělení kompresních algoritmů**
- **RLE**
- **LZ**
- **Aritmetické kódování**

Proč komprimovat

- **Snaha o co nejlepší využití datových nosičů.**
- **Snaha o co nejlepší využití přenosových kanálů.**
 - Při přenosu zkomprimovaného souboru má rušení závažné důsledky.

- **Synonymem slova komprimace, zabalení.**
- **Speciální typ kódování.**
 - Nerovnoměrné kódy.
 - Prakticky se mohou kompresní algoritmy chovat heuristicky.
- **Proces, při kterém je vytvořen obraz původního souboru.**

- **Její úspěšnost může být dána**
 - **Kompresní poměr** – podíl délky zkomprimovaného souboru k délce původního souboru.
 - **Kompresní faktor** – převrácená hodnota předchozího.
 - **Kompresní zisk** – stonásobek přirozeného logaritmu z kompresního faktoru.

- **Bpc (Bits per character) či BPB (Bits Per Byte):**
 - Poměr počtu bitů ve zkomprimovaném souboru ku počtu symbolů (bytů) v původním souboru.
- **Kompresní algoritmy se testují na referenčních vzorcích daných např. Canterbury Corpusem.**

- **Synonyma jsou dekomprimace a rozbalení.**
- **Vstupem je zkomprimovaný soubor a výstupem soubor obnovený.**
- **Bezeztrátové kompresní algoritmy**
 - Při dekompresi je původní soubor zcela obnoven, žádná informace se neztrácí.

- **Podle poměru časové náročnosti komprese a dekomprese dělíme kompresní algoritmy na:**
 - symetrické x asymetrické.
- **Podle schopnosti kompresního algoritmu přizpůsobit se komprimovaným datům:**
 - adaptivní x neadaptivní.

➤ **Run Length Encoding – kódování proudů.**

– Proud – blok opakujících se symbolů ve vstupním souboru.

➤ **Proudy se hledají procházením po řádcích, sloupcích, dlaždicích či cik-cak.**

➤ **Použití např. u obrázků s malým počtem barev.**

- **Záznam o bloku či proudu ve zkomprimovaném souboru je vždy tvořen alespoň dvěma byty.**
- **První byte:**
 - První bit prvního bytu určuje, zda se jedná o záznam proudu opakujícího se bytu (1), či bloku neopakujících se bytů (0).
 - Zbývajících sedm bitů prvního bytu určuje délku proudu opakujícího se následujícího bytu, nebo délku bloku následujících neopakujících se bytů.

➤ **Kompresa**

- $65_{10} \ 65_{10} \ 65_{10} \ 65_{10} \ 65_{10} = 10000101_2 \ 01000001_2$
- $72_{10} \ 65_{10} = 00000010_2 \ 01001000_2 \ 01000001_2$

➤ **Dekomprese**

- $135_{10} \ 65_{10} \ 3_{10} \ 72_{10} \ 79_{10} \ 74_{10} \ 131_{10} \ 68_{10} \ 1_{10} \ 65_{10} =$
 $10000111_2 \ 01000001_2 \ 00000011_2 \ 01001000_2$
 $01001111_2 \ 01001010_2 \ 10000101_2 \ 01000100_2$
 $00000001_2 \ 01000001_2 = 65_{10} \ 65_{10} \ 65_{10} \ 65_{10} \ 65_{10}$
 $65_{10} \ 65_{10} \ 72_{10} \ 79_{10} \ 74_{10} \ 68_{10} \ 68_{10} \ 68_{10} \ 65_{10}$
- Kompresní poměr = 5:7

➤ **Modifikace předchozího.**

- Pro obrázky s paletou barev, kde jsou velké stejnobarevné plochy.

➤ **Typ bytu je určován podle hodnoty a závisí na tom, zda je nastaven na 1 určitý počet prvních bitů.**

- Pokud ano, následující byte se opakuje.
- Pokud ne, v bytu je již kód barvy.

➤ **Kompresse**

- Počítadlo nastaveno na $C0_{16} = 11000000_2$.
- $A0_{16} A0_{16} A0_{16} = 11000011_2 10100000_2 = C3_{16} A0_{16}$
- $88_{16} = 01000100_2 = 88_{16}$
- $DE_{16} = 11000001_2 11011110_2 = C1_{16} DE_{16}$

➤ **Dekomprese**

- Počítadlo nastaveno na $C0_{16} = 11000000_2$.
- $C1_{16} CC_{16} C5_{16} FE_{16} 18_{16} AF_{16} = 11000001_2$
 $11001100_2 11000101_2 11111110_2 00010100_2$
 $10101111_2 = CC_{16} FE_{16} FE_{16} FE_{16} FE_{16} FE_{16} 18_{16} AF_{16}$

Algoritmy rodiny LZ

- **Třída algoritmů pojmenovaná po tvůrcích původního z roku 1977.**
 - Lempel a Ziv.
- **Využívá se tzv. „posuvného okna“ rozděleného na prohledávané pole a komprimované pole.**
- **Jedna z prvních verzí se např. i dnes využívá u formátu PNG, další pak v ARJ, ZIP, RAR...**

- **Do výstupního souboru se zapisují trojice:**
 - Pozice kopírované sekvence z prohledávaného pole.
 - Délka sekvence.
 - Následující symbol.
- **Délka prohledávaného pole se pohybuje v kilobytech, délka komprimovaného pole spíše v B.**

- **Prohledávané pole nahrazeno dynamickým slovníkem opakujících se řetězců.**
- **Ve zkomprimovaném souboru dvojice:**
 - Ukazatel (většinou 12bitový) do slovníku již dříve nalezených řetězců.
 - Bezprostředně následující symbol.

- **Modifikace navržená Welchem.**
- **Nejprve se inicializuje slovník a není třeba posléze používat dvojice (stačí odkazy do slovníku).**
- **Využití např. u grafického formátu GIF.**
- **Slovník není součástí zkomprimovaného souboru.**

Aritmetické kódování

- **J. J. Rissanen v roce 1976.**
 - Vztahuje se na něj patent.
- **Vstupní soubor je převeden na číslo z intervalu $[0;1)$.**
- **Je třeba znát pravděpodobnost výskytu znaků ve vstupním souboru.**
 - Řadíme od nejmenší po největší.

Aritmetické kódování

- Z pravděpodobností výskytu jednotlivých prvků se vytvoří na intervalu $<0;1)$ podintervaly, které ho vyplní.
 - Pole **mez_d** obsahuje dolní meze podintervalů.
 - Pole **mez_h** obsahuje horní meze podintervalů.
 - Součást vstupního i výstupního souboru (+ počet znaků).

Aritmetické kódování

➤ Komprese

- Základní inicializace – $h = 1$; $d = 0$; $s = 1$.
- Cyklus postupně pro jednotlivé znaky:
 - $h := d + s \cdot \text{mez}_{h(\text{aktuálního symbolu})}$
 - $d := d + s \cdot \text{mez}_{d(\text{aktuálního symbolu})}$
 - $s := h - d$
- Po provedení cyklu pro poslední znak vybereme reprezentanta K z aktuálního intervalu $(d; h)$.

➤ Dekomprese

- Označíme K jako K_1 a podle intervalu, do něhož patří, dekomprimujeme první znak.
- Další znaky dekomprimujeme postupně na základě obecného vzorce pro i tého reprezentanta:

$$K_i = \frac{K_{i-1} - mez_{di-1}}{s_{i-1}}$$

- JIROUŠEK, R. a kol. Principy digitální komunikace. Voznice: Leda, 2006. 309 s. ISBN 80-7335-084-X
- SNÁŠEL, V., DVORSKÝ, J. Algoritmická matematika I. Olomouc : Univerzita Palackého, 1999. 256 s. ISBN 80-244-0013-8