

Komprese dat

Mgr. Kamila Cmolová

Komprese a její parametry

- **Komprese/komprimace** = proces **zakódování vstupní informace** pomocí určitého kódovacího schématu tak, aby **výstupní informace měla menší velikost** (zabírala méně místa).
- Kompresí dochází ke **zmenšení objemu dat**, ale zároveň **zachování informací** v nich obsažených.
- Komprese se provádí pomocí **komprimačních programů**.
- Opačný proces ke kompresi je **dekomprese**. Jedná se o **obnovení** již zkomprimovaných souborů do jejich **původní podoby**.

Kompresa a její parametry

- Mezi hlavní **parametry**, které ovlivňují kompresi, patří:
 - rychlost komprese,
 - rychlost dekomprese,
 - komprimační poměr.
- Kompresi můžeme dělit podle způsobu ukládání dat na:
 - **ztrátovou**
 - **bezztrátovou**.

Komprimační poměr

- Ukazatel **výkonnosti komprimace**
- U bezztrátové komprese je mnohem nižší než u ztrátové komprese (čím vyšší číslo, tím lepší výsledek komprimace).
- Vyjadřuje se jako **poměr velikosti nekomprimovaných dat ku velikosti dat komprimovaných**.
 - Soubor, který měl původně 200 KB a nyní 50 KB, byl tedy zkomprimován v poměru 4 : 1.
Procentuální hodnotu této komprese získáme, pokud **od čísla 1 odečteme obrácenou hodnotu poměru** ($1 - 1/4 = 0,75 > 75 \%$ úspora místa)

Ztrátová komprese

- Způsob ukládání dat, při němž dochází k **několikanásobnému zmenšení obsahu dat.**
- Nedůležité informace se **ztrácejí** (čím je komprese vyšší, tím více detailů je zahozeno).
- Ztracená data již **nelze** dekomprimací **vrátit zpět.**
- Využívá se u **obrázků, videí** či **zvukových** souborů.

Bezztrátová komprese

- Využívá se u textových a binárních souborů, kdy nesmí dojít k **žádné ztrátě dat**.
- Nedosahuje tak dobrého komprimačního poměru jako komprese ztrátová.
- Umožňuje přesnou **rekonstrukci** dat do původní podoby.
- Bezztrátové komprese využívají například formáty ZIP, RAR.
- Typy bezztrátové komprese:
 - RLE
 - LZW
 - Huffmanovo kódování

Bezztrátová komprese: RLE

- = Run-length Encording
- Jedna z **nejjednodušších** metod bezztrátové komprese (symetrická a velmi rychlá, s nižším kompresním poměrem).
- Funguje na principu, který se snaží **najít a odhalit posloupnosti** opakujících se znaků, které se v souboru musí vyskytovat hned za sebou, jinak bude komprimace neúčinná.
- Hodí se např. pro **jednoduché obrázky**, které mají málo barev nebo které obsahují větší plochy stejné barvy nebo zvukové soubory zachycující **řeč**.

Bezztrátová komprese: LZW

- Je to jedna z **nejrozšířenějších** kompresních metod, kterou používá spousta kompresních programů a grafických formátů obrázků.
- Na této metodě funguje formát GIF, PDF, TIFF, starší verze formátu ZIP či archivy ARC.

Bezztrátová komprese: Huffmanovo kódování

- Metoda založená na znalosti **pravděpodobnosti výskytu** jednotlivých **znaků**.
 - Nejčastější se vyskytujícím znakům je pak přiřazen krátký kód, méně často se vyskytujícím znakům zase delší kód.
- Tento princip použil již v roce 1800 Samuel Morse na zakódování písmen anglické abecedy (**Morseova abeceda**).
- Použití Huffmanova kódu je časté v kombinaci s jinými kompresními algoritmy, například při kompresi obrazu a videa ve standardech JPEG a MPEG.

Kompresní formáty

Ztrátové

- JPEG
- MP3

Bezztrátové

- GIF
- PNG

Kompresní formáty: JPEG

- Založen na principu **nedokonalosti lidského oka**, které není schopné rozeznat **drobné rozdíly** (nedůležité informace, barvy, zvuky apod. se ztratí, což znamená výraznou úsporu místa).
- Je možné dosáhnout komprimačního poměru **1:50 až 1:100** při zanedbatelné ztrátě informace.
- Není určen pro ukládání obrázků obsahujících malé množství barev, kontrastní barevné přechody, ostré hrany, písmo apod.
- Je naopak určen pro ukládání fotografií, naskenovaných dokumentů (které jsou již vlastním procesem skenování rozmazané), monochromatických rentgenových a ultrazvukových snímků atd.

Kompresní formáty: MP3

- Umožňuje zachovat poměrně **vysokou kvalitu** souborů, a přesto snížit velikost těchto souborů **až na setinu**.
- U mluveného slova tato metoda nedosahuje moc dobrých výsledků, protože může zkracovat pauzy mezi slovy či odstraňovat první nebo poslední slabiky.
- Tento formát se používá **poslechu hudby** (MP3 přehrávač)

Kompresní formáty: GIF

- Využívá **LZW** metodu.
- Velká nevýhoda tohoto formátu spočívá v **barevném omezení**, kdy na jeden obrázek připadá 256 současně použitých barev.
- Využívá se na jednoduchá **loga** a **ikony**.
- Na rozdíl od JPEG má **velmi ostré hranice**.
- Jeho hlavní předností je, že umožňuje vytvářet **animace**, a proto se často využívá u webové grafiky.

Kompresní formáty: PNG

- Alternativa formátu GIF (lepší výsledky).
- Nemá barevná omezení.
- Možnost práce i s **průhledností** obrázků.
- Chybí možnost vytvářet **animace**.

Komprimační a archivační programy

- **Archivace** = proces, kdy se pomocí **archivačních programů** dlouhodobě **uchovávají data** do archivů, aby se nijak nepoškodila (např. nechtěné smazání, porucha počítače).
- Archivační programy šetří **prostor** a **čas** pro uložení dat.
- Možnost jednotlivé archivy **heslovat nebo šifrovat**.
- Typy programů:
 - RAR, WinRAR
 - ZIP

Komprimační a archivační programy

RAR a WinRAR

- **Komerční** kompresní program.
- Umožňuje vytvářet **RAR** i **ZIP** archivy a rozbaluje archivy **ARJ**.
- Dokáže vytvářet **samorozbalovací** archivy a **dělené** archivy.
- Dále nabízí možnost **opravy poškozených souborů** a nechybí u něj ani funkce **šifrování**.
- **WinRAR** je přístupný pouze pro operační systém Microsoft Windows, zatímco **RAR** je využitelný i v jiných operačních systémech.
- RAR archivy je možné otevřít a **dekomprimovat**.

Komprimační a archivační programy

ZIP

- Podporuje komprimační formáty **ZIP**, **RAR**, **ARJ** apod.
- Formát **RAR** umí pouze **rozbalovat**.
- Dá se používat ve **všech operačních systémech**.
- Dosahuje jednoho z nejlepších kompresních poměrů mezi stávajícími archivačními programy.

Děkuji za pozornost ☺