

Komprimace a archivace dat

CO MÁME ZNÁT

- Orientace v souborovém systému (vytvoření adresáře, změna aktuálního adresáře, vytvoření textového souboru, prohlížení obsahu souboru apod.)
- Práce s kalkulačkou na počítači
- Programy pro zobrazení obsahu libovolného souboru (PSPad Hex Edit, od apod. – viz ukázka)
- Typy komprimací

MATERIÁLY

Programy: komprimační software (WinZIP, WinRAR, 7zip apod.)

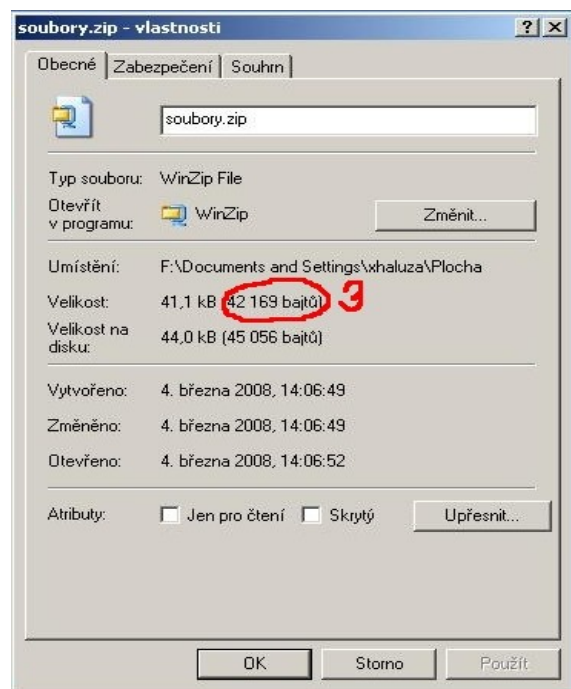
Soubory: archiv s pracovními soubory cviceni4.zip

VZOR

Mějme tři soubory, soubor1.doc, soubor2.doc a soubor3.doc. S využitím komprimačního softwaru byl vytvořen archiv soubory.zip obsahující tyto soubory. Výsledek komprimace zaneseme do tabulky.

Význam jednotlivých sloupců:

- velikost před komprimací – představuje součet velikostí původních souborů, tzv. hrubá data (*součet ve sloupci „1“*)
- velikost komprimovaných dat – představuje součet velikostí komprimovaných souborů, tzv. čistá data (*součet ve sloupci „2“*)
- komprimační poměr – vyjadřuje úspěšnost komprimace (různé způsoby: ve vzorovém příkladu 4 : 1 nebo na 23 % nebo o 77 %)
- režijní informace – informace o struktuře archivního souboru, která je přidána k čistým datům; vypočteme ji jako rozdíl velikosti archivu (*hodnota „3“*) a čistých dat („2“)



| Name | Type | Modified | Size | Ratio | Packed | Path |
|-------------|----------------------------------|----------------|--------|-------|--------|------|
| soubor1.doc | Dokument aplikace Microsoft Word | 3.3.2008 12:11 | 65 024 | 76% | 15 573 | |
| soubor2.doc | Dokument aplikace Microsoft Word | 3.3.2008 12:11 | 53 760 | 77% | 12 166 | |
| soubor3.doc | Dokument aplikace Microsoft Word | 3.3.2008 12:11 | 59 904 | 76% | 14 114 | |

| soubory v archivu (formát souborů) | velikost před komprimací | velikost kompri- movaných dat | komprimační poměr | režijní informace |
|---------------------------------------|-----------------------------|----------------------------------|----------------------|----------------------|
| * .DOC | 178 688 bajtů | 41 853 bajtů | 4 : 1 | 316 bajtů |

Poznámka: Jedná se o vzorový příklad, údaje v tabulce závisí na počtu a obsahu jednotlivých souborů, nelze je zobecnit na libovolné soubory stejného formátu!

ÚKOLY

- Vytvořte si na ploše pracovní adresář komprimace a nastavte si jej jako aktuální. Do tohoto adresáře zkopírujte pracovní archiv cviceni4.zip dle instrukcí cvičícího.
- Proveďte dekomprimaci souborů z archivu do pracovního adresáře.
- Vyzkoušejte si zkomprimovat libovolný jeden delší soubor. Zjistěte název souboru, který se vytvořil. Vytvořený archivní soubor si prohlédněte vhodným nástrojem pro zobrazení obsahu souboru.
 - Jaký vliv na úspěšnost komprimace má volba komprimačního softwaru?
 - Jaký vliv na úspěšnost komprimace má nastavení rychlosti a účinnosti algoritmu?
- Vytvořte textovým editorem alespoň tři velmi krátké textové soubory (délka do 15 znaků, obsah běžná česká slova). Tyto soubory zkomprimujte dohromady do jednoho archivního souboru. Porovnejte délky komprimovaných souborů a výsledného archivního souboru.
 - Jak se nazývá jev, ke kterému došlo? Jaká je příčina jeho vzniku?
- Stanovte čistý komprimační poměr pro různé typy datových souborů. Vytvořte tabulku podle vzoru a vyplňte ji pro různé skupiny souborů stejného formátu v pracovním adresáři.

| formát souborů | velikost před komprimací | velikost kompri- movaných dat | komprimační poměr | režijní informace |
|-------------------|-----------------------------|----------------------------------|----------------------|----------------------|
| * .DAT | | | | |
| * .BMP | | | | |
| * .JPG | | | | |
| * .GIF | | | | |
| * .TIF | | | | |
| * .PS | | | | |
| * .PDF | | | | |
| * .DOC | | | | |
| * .RTF | | | | |
| * .TXT | | | | |
| * .ZIP | | | | |

5. Jednotlivé soubory v pracovním adresáři rozmístěte do podadresářů podle jednotlivých formátů (BMP, GIF, JPG, ...) a vyzkoušejte si komprimaci celého podstromu (tj. včetně podadresářů). Všimněte si, jak se adresáře v archivu ukládají.
6. V existujícím archivu vyzkoušejte odstranit libovolný soubor a přidat nový soubor.
7. Zopakujte postup v úloze 4 s tím, že archivu nastavíte při komprimaci přístupové heslo. Porovnejte obsahy obou archivů.

KONTROLNÍ OTÁZKY

- Rozhodněte, zda komprimované soubory patří mezi textové nebo netextové soubory.
- Do kterého adresáře ukládají komprimační programy vytvářené soubory? Jak lze toto cílové místo změnit?
- Na čem závisí úspěšnost komprimačního algoritmu? Jakým způsobem může uživatel ovlivnit úspěšnost komprimace?
- Které další informace se ukládají do archivu při archivaci podstromu?
- Co vyjadřuje hrubý komprimační poměr? Co vyjadřuje čistý komprimační poměr? Proč je mezi oběma hodnotami rozdíl?
- Co je obsahem režijní informace v archivu? Jak lze režijní informaci zobrazit?
- Které typy souborů má smysl komprimovat a které ne?
- Jaký je smysl archivace? Čím se liší archivace od běžné komprimace?

DOPLŇUJÍCÍ ÚKOLY

1. Prostudujte si tři programy v jazyce Pascal (`komprimacecvicny[123].pas`). Vytvořte běhuschopnou podobu a tyto přeložené programy spusťte. Každý z těchto tří programů vytvořil jeden datový soubor (`komprimacecvicny[123].dat`). Zjistěte účinnost komprese pro tyto tři cvičné soubory a zhodnoťte, proč se výsledky tak výrazně liší.
2. Program `komprimacecvicny5.pas` přeložte a spusťte. Zjistěte, kolik souborů se vytvořilo, jakou mají délku, jaký mají obsah a odhadněte vhodnou metodu a účinnost komprese. Spočítejte, kolik dat je uloženo v těchto souborech, porovnejte s délkou původních souborů a přijměte závěr, o jaký druh komprimace se jedná.