

13. Állományok tömörítése

- Imertesse az adatok tömörítésének szükségességét! Milyen szerepet játszik a tömörítésben a redundancia?
- Ismertessen egy egyszerű példát tömörítésre!
- Mit nevezünk veszteségmentes illetve veszteséges tömörítésnek? Mondjon példát alkalmazásukra!
- Végezzen el egy tömörítést és egy kicsomagolást a számítógépen!

Adatok tömörítésének elvi lehetősége

Az adathalmazok általában redundánsak, terjengősek, nem a lehető legrövidebbek. Sokszor ugyanazt az információt rövidebben is le lehet írni, kódolni, ez adja az adatok tömörítésének elvi lehetőségét.

A tömörítő eljárások segítségével adatainkat olyan alakra hozhatjuk, amelynek kisebb az adatmennyisége, mint az eredetinek, gyakran csak töredéke (pl. tizede). Így kisebb helyet foglal el az adathordozón, rövidebb idő alatt továbbítható hálózaton. A felhasználáshoz általában vissza kell alakítani az eredeti formátumra.

A **veszteségmentes tömörítés** olyan kódolás, amelynek eredményeként létrejött kódolt (tömörített) jelhalmaz rövidebb, mint az eredeti, azaz kisebb az adatmennyisége, és a tömörített adathalmazból tökéletesen (bitről bitre) visszaállítható az eredeti, azaz információt nem veszünk.

- Így tömörítünk programokat, dokumentumokat, hiszen ott fontos a tökéletes visszaállítás lehetősége.
- A tömörítőprogramok hatékony eljárásokat alkalmaznak. Elemzik az állomány szerkezetét, és annak függvényében határozzák meg a tömörítési eljárást. A veszteségmentesen tömörítő programok általában kétfajta algoritmust használnak: az egyik generál egy *statisztikai modellt* a bemeneti adatokból, a másik pedig a modell felhasználásával bitsorozatot rendel a bemeneti adatokhoz oly módon, hogy a „valószínűbb” (tehát gyakrabban előforduló) adatoknak rövidebb bitsorozatot feleltessen meg, mint a „valószínűtlenebb” adatoknak. Sokszor csak az első algoritmust nevezik néven, a másodikat adottnak veszik vagy nem nevesítik.
- Példa egyszerű tömörítésre: Mivel mindenféle adatot digitálisan, számokkal kódolunk, elég foglalkozni egy tetszőleges számsorozat kódolásával. Pl. a 3555555227777 számsor 13 db egyjegyű számból áll. Kódoljuk ezt a következőképpen: Balról jobbra haladva adjuk meg a számot, és azt, hogy hányszor ismétlődik. Az eredmény: 31562274. Ez 8 darab szám, tehát 8/13 arányban tömörítettük az eredeti adatsort, ami egyértelműen visszaállítható. Képek esetén gyakori, hogy sok azonos színű pont van egymás mellett, így jelentős rövidülést érhetünk el.
- A tömörítés mértéke nemcsak a tömörítési eljárástól, hanem az adathalmaz tulajdonságaitól is függ. Például vannak-e ismétlődő adatok, vannak-e más szabályszerűségek, az egyes jelek, jelcsoportok milyen gyakorisággal fordulnak elő. Az is előfordulhat, hogy egy eljárás éppenséggel növelné az adathalmaz méretét. Pl. az előbb említett eljárást alkalmazva az 121212121 számsorozaton.

- Adataink archiválása során gyakran alkalmazunk tömörítést, illetve hálózaton is tömörítve küldünk (küldjük!) nagyobb méretű állományokat.
- Néhány ismert tömörítés: ZIP, ARJ, RAR

A **vesztesége tömörítés** olyan kódolás, aminek eredményeként létrejött kódolt (tömörített) jelhalmaz sokkal rövidebb, mint az eredeti, azaz kisebb az adatmennyisége, de a tömörített adathalmazból nem állítható tökéletesen vissza az eredeti, csak jó közelítéssel, azaz információt veszünk, hogy mennyit, azt gyakran magunk szabhatjuk meg.

- Ilyen tömörítést alkalmazunk például digitális fényképek, hangok, mozgóképek esetén, ha nem fontos a tökéletes visszaállítás lehetősége. A kis eltéréseket a szem, a fül, -a néző- általában észre sem veszi. Kétségtelen, hogy romlik a digitalizált, és tömörített anyag minősége. Néhány ismert veszteséges tömörítés: MP3 (hangoknál), JPEG (színes képeknél), MPEG (mozgóképeknél).
- Minél nagyobb a tömörítés mértéke, azaz minél kisebb a tömörített állomány, annál rosszabb lesz a kép- vagy a hangállomány minősége.

Tömörítés a gyakorlatban

- Tömörített mappák a Windows XP-ben
 - Létrehozása (Fájl→Új→Tömörített mappa) A tömörített mappába másolt fájlokat a rendszer tömöríti, ZIP kiterjesztéssel, a kimásolt fájlokat kicsomagolja
 - Jobb gomb a tömörítendő állományon →Küldés→Tömörített mappa, illetve jobb gomb a kitömörítendő állományon, →Az összes kibontása
- Egyes alkalmazásokból tömörítve menthetünk: Például a Paint rajzolóprogramban jpeg formátumba mentve tömöríthetjük a képet.
- Az önkicsomagoló állományok nemcsak a tömörített állományt tartalmazzák, hanem a kicsomagoló programot is.
- Könyvtárak tömörítése: az archívumba gyakran egész könyvtárszerkezeteket (a fájlokkal együtt) betömörítünk, így az összetartozó állományok együtt kezelhetők.