

12. Adatok tömörített tárolása (3.1)

Ismertesse az adatok tömörítésének lehetőségét és szükségességét!

Milyen előnyei, hátrányai lehetnek az adat tömörítésnek?

Mi a veszteségmentes tömörítés lényege? Mutasson rá egy egyszerű példát! Hol használják?

Mi a veszteséges tömörítés lényege? Hol használják?

Mutasson be egy általános fájl-tömörítő programot!

Az adathalmazok ismétlődő (redundáns) jeleket tartalmaznak, terjengősek, nem a lehető legrövidebbek, legtömörebbek. Sokszor ugyanazt az információt rövidebben is le lehet írni, kódolni, erre szolgálnak az adatok tömörítésének lehetőségei. A tömörítés a jelsorozat redundanciájának csökkentése, az adatok matematikai alapon történő átkódolása úgy, hogy ugyanazt az információt kevesebb jellel fejezzük ki. Az eredményként kapott adathalmaz mérete az átkódolás révén kisebb az eredeti adatokénál, gyakran csak töredéke. A felhasználásához általában vissza kell alakítani az eredeti formátumra, amihez külön programra van szükség. Az adatok visszaalakítását kicsomagolásnak nevezzük.

Előny

Az adatok tömörített tárolásának előnye, hogy kisebb helyet foglal az adat a számítógépen, az Interneten keresztül gyorsabban továbbítható, és lehetővé teszi az adat jelszavas védelmét. A tömörített állományok kiterjesztése jellemzően nem EXE vagy COM (eltekintve az önkicsomagoló állományoktól), ezért az ilyenekre "vadászó" vírusok nem fogják megfertőzni. A tömörített állományok megadott méretű részekre szeletelhetők, ez azt jelenti, hogy az archívum tartalma több különálló fájlra lesz elosztva.

Hátrány

A tömörített adatok érzékenyebbek a sérülésre, mint a tömörítés nélküliek. Sérülés esetén a kicsomagoló program jelzi, hogy hiányosak az adatok, visszaállítani nem tudja őket, esetleg a sérülés előtti részeket. Hátránya még, hogy lehetséges az adatok esetleges vesztese, és egy-egy fájl csak kicsomagolás után olvasható ill. módosítható.

A tömörítésnek alapvetően két típusa van, a veszteségmentes és a veszteséges tömörítés.

Veszteségmentes tömörítés

Olyan tömörítési eljárás, amely esetében a tömörített adatsorból az eredeti (tömörítés előtti) információ maradéktalanul reprodukálható.

Olyan esetekben használjuk ahol fontos az eredeti állapot visszaállítása. Így pl. szöveges dokumentumok, programok, adatbázisok stb. Erre jó példa egy Word dokumentum tömörítése. Milyen is lenne az a szöveg melynek tartalmát a kitömörítés után nem kapnánk vissza eredeti formájában. Veszteséges tömörítés nem alkalmazható egy dokumentumra. Helyigény szempontjából jó tömörítési eljárás lenne az, ha minden második karaktert törölnénk, de mi, vagy ki tudná ezt később visszaállítani ill. elolvasni.

Egy lehetséges megoldás a veszteségmentes tömörítésre az, hogy többször, egymás után ismétlődő elemeket le lehet rövidebben is írni. Ehhez választani kell egy *jelző elemet*, mely majd a kitömörítés során jelzi, hogy itt ismétlődő elemek következnek.

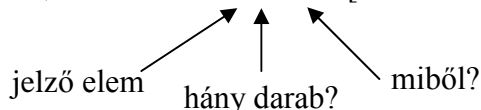
Egy tömörítetlen szövegben található 10 darab szóköz egymás után, ezt a következőképp lehet rövidebben leírni:

Tömörítés előtt:

Egyszer_volt,_hol_nem_volt_____ [Ez összesen 35 karakter]

Tömörítés után:

Egyszer_volt,_hol_nem_volt\$10_ [Ez összesen 29 karakter]



Természetesen nem csak szöveg esetén megengedhetetlen az, hogy karakterek elveszenek, hanem programok esetében is. Ezért programoknál szintén veszteségmentes tömörítést kell alkalmazni.

A tömörítés mértéke nem csak a tömörítési eljárástól, hanem az adathalmaz tulajdonságaitól is függ, például vannak-e ismétlődő adatok, az egyes jelek vagy jelcsoportok milyen gyakorisággal fordulnak elő. Képek esetében gyakori, hogy sok azonos színű pont van egymás mellett, így jelentős tömörítést érhetünk el.

Veszteségmentes tömörítés például a ZIP, ARJ, RAR,

Veszteséges tömörítés

Olyan tömörítési eljárás, amely esetében a tömörített adatsorból az eredeti (tömörítés előtti) információfolyam nem feltétlenül állítható elő.

A veszteséges tömörítés a veszteségmentes eljárásokhoz képest sokkal nagyobb tömörítési arányt tesz lehetővé, de csakis olyan adatok és adatfolyamok kódolásához használható, amelyek esetében nem követelmény az eredeti adatsor pontos visszaállítása, hanem kisebb eltérések megengedhetőek (pl. zeneállományok, képek, filmek)

A veszteséges tömörítés háttérét is valamilyen matematikai algoritmus képezi.

De miért „veszteséges” ez a tömörítés? Mi az, amit fel lehet áldozni, hagyni, hogy „elvessen”?

Képek veszteséges tömörítése

Paint - ben készítettünk egy színes rajzot. BMP - be mentettük, mely egy tömörítés nélküli formátum mérete: 55kbyte lett, majd elmentettük JPG - formátumba mely veszteséges tömörítési formátum mérete: 5kbyte lett.

Erős JPG tömörítést használva jól látszik a két kép közti különbség. A kép mérete kevesebb, mint tized részére csökkent, de ez „áldozatot” is jelent. A kontúrok elmosódtak, a színek veszítettek eredeti tartalmukból. Minél kisebb a tömörítés mértéke annál kevesebb veszteséggel kell számolnunk, viszont a kép mérete egyre jobban közelít majd az eredetihez.

Veszteséges tömörítést használunk digitális fényképek, hangok, mozgóképek esetén, mert ezeknél nem fontos a tökéletes visszaállítás lehetősége. A kis eltéréseket a szem nem érzékeli, bár valamit romlik a tömörített anyag minősége. Minél nagyobb mértékben tömörítünk egy adatot, annál rosszabb lesz annak minősége. A veszteségesen tömörítő programok esetében be lehet állítani a tömörítés mértékét, van, ahol állományméretben, van, ahol minőségben. Ezzel a tömörítési típussal sokkal nagyobb megtakarítást lehet elérni, míg a veszteségmentes eljárások az eredeti méretet csak harmadára-negyedére csökkentik, addig a veszteséges eljárások hatodára-tizedére. Veszteséges tömörítés például hangok esetében az MP3, színes képeknél a JPEG, mozgóképeknél pedig az MPEG.

Hanganyagok veszteséges tömörítése

A zenei CD lemezekon található hanganyag tömörítetlen formában van tárolva. Ha ezt számítógépünkre lemásoljuk szintén tömörítetlen az eredetivel megegyező méretű fájlt (fájlokat) kapunk eredményül. Ezek többnyire **WAV** kiterjesztésű fájlok.

Lehetőség van a hanganyag tömörítésére is, mely során az emberi fül számára kevésbé hallható részeket – melyek a zenei élvezetet kismértékben befolyásolják – elhagyják a zenéből. Legelterjedtebb ilyen tömörítési forma az **MP3**.

A tömörítés mértéke itt is meghatározható (mintavételi frekvencia kHz-ben, bitráta kbit/sec-ban). Amennyiben ezeket az értékeket nagyon lecsökkentjük, úgy egyre zajosabb lesz a hanganyag és értékes, hallható részek vesznek el. Tehát itt is kompromisszumot kell keresni a helyfoglalás és a hangminőség között. Tömörített hangformátumok: MP3, WMA, OGG stb.

Videóanyagok veszteséges tömörítése

Videók veszteséges tömörítésére, a képre és hangra leírtak egyaránt érvényesek, hiszen a videó többnyire képből és hangból áll össze.

Lehetőségünk van a videó képanyagát, képkockáinkét tömöríteni (M-JPEG tömörítés) vagy az elmozdulások különbsége alapján (AVI tömörítés). Ha egy 30 másodperces videóban egy tárgy ugyanott található, a 30 másodperc alatt nem mozdul el, akkor fölöslegesen mentenénk el a tárgy adatait minden egyes képkockában. A képkockák befoglaló méretével is csökkenthető a tárigény. Pl. a DVD felbontása 720x576 tömörítve, ez lehet 360x288. Ebben az esetben nem a kép széleiből „vágunk le”, hanem „összenyomjuk” a képet.

Tömörítés a gyakorlatban

A Windows XP-ben tömörített mappát a Fájl →Új →Tömörített mappa paranccsal hozhatunk létre. Ebbe a tömörített mappába másolt fájlokat a rendszer ZIP kiterjesztéssel tömöríti, a kimásolt fájlokat pedig kicsomagolja.

- Egyes alkalmazásokkal tömörítve is menthetünk. Például: Paint rajzolóprogramban JPEG formátumban mentve tömörítjük a képet.

- WINRAR (Általános fájl-tömörítő)

A WinRAR két különböző formátumú archívumot képes létrehozni: RAR-t és ZIP-et.

ZIP archívumok:

A ZIP formátum fő előnye a népszerűsége. Például az Interneten a legtöbb archívum ZIP archívum. Egy archívumban egy ZIP fájl maximális mérete 2 GB.

RAR archívumok

A RAR formátum lényegesen jobb tömörítést nyújt, mint a ZIP. A RAR másik fontos tulajdonsága a többkötetes archiválás támogatása.

A RAR formátumnak van néhány olyan fontos tulajdonsága, ami a ZIP-ben hiányzik, mint például

- a helyreállítási bejegyzés (ami lehetővé teszi a sérült adatok fizikai helyreállítását,)
- a fontos archívumok lezárása (az eseti módosulás megelőzése céljából.)

A RAR formátum gyakorlatilag korlátlan méretű fájlokat tud kezelni.

- **Önkitömörítő archívum** (nem csak a tömörített fájlt tartalmazza, hanem a kicsomagolóprogramot is)

Az SFX (Self-eXtracting) archívum olyan archívum, amely egyesítve van egy végrehajtható modullal, amit az archívumból történő fájlkicsomagolásra használunk a végrehajtás során. Így nem szükséges külső program egy SFX archívum tartalmának kicsomagolásához, elég lefuttatni. Mindemellett a WinRAR ugyanúgy tud dolgozni egy SFX archívummal, mint bármilyen más archívummal, így ha nem akar lefuttatni egy kapott SFX archívumot (például vírus lehet benne), a WinRAR-ral meglekintheti, vagy kicsomagolhatja a tartalmát. Az SFX archívumok általában .exe kiterjesztésűek, mint bármely más végrehajtható fájl. Az SFX archívumok használata sok kényelemmel jár, ha oda akarunk adni valakinek egy archívumot, de nem tudjuk, hogy van-e megfelelő archiválójuk a fájlok kicsomagolására. Az önkitömörítő archívumok mérete kicsit nagyobb, mint a zip vagy rar archívumok mérete.

- Fájlmenedzser programok

A fájlmenedzser programok, mint a Norton Commander és a Windows Commander képesek adattömörítésre, kibontásra sőt a tömörített állományok tartalmát közvetlenül meg tudják jeleníteni.