

# A tömörített adattárolás

## A tömörítés lényege és elve. Tömörítési módszerek (veszteséges és veszteségmentes):

Az adathalmazok általában redundánsak, nem a lehető legrövidebbek, legtömörebbek. **A tömörítő eljárások segítségével adatainkat olyan alakra hozhatjuk, amelynek kisebb az adatmennyisége, mint az eredetinek, így kisebb helyet foglal az adathordozón, és rövidebb idő továbbítani a hálózaton.**

A felhasználáshoz általában vissza kell alakítani az eredeti formátumra.

A tömörítés egy olyan eljárás, amelynek segítségével egy fájlból egy kisebb fájl állítható elő. Ha a tömörített fájlból teljes mértékben visszanyerhető az eredeti fájl ("bitről bitre"), akkor **veszteségmentes** (alkalmazásoknál, dokumentumoknál használjuk), ha nem akkor **veszteséges** (hangoknál, képeknél alkalmazzuk) tömörítésről beszélünk. Az első általánosan alkalmazható, a második csak akkor hogyha kis eltérések nem számítanak, pl. kép- és hangfájlok esetén.

A veszteséges tömörítési eljárások általában hatékonyabban, de nem minden esetben alkalmazhatóak (egy futtatható programot akár egy bitnyi eltérés is használhatatlanná tenne).

A tömörítés nagyon hasznos lehet, ha pl. helyet akarunk megspórolni a lemezen, vagy egy viszonylag lassú (pl. telefon) vonalon szeretnénk egy nagyobb fájlt elküldeni. Általános szabályként elmondható, hogy minél hatékonyabb egy tömörítési eljárás, annál több számolást igényel, vagyis lassabb. Nincs "optimális" tömörítés, mindig el kell dönteni, hogy érdemes-e több időt rááldozni arra, hogy a kapott fájl valamivel kisebb legyen.

A tömörítések többek között a fájlban levő ismétlődéseket és más redundanciákat használják ki, például azt, hogy egy szövegfájl minden egyes byte-ján 256 érték lenne ábrázolható, szemben az általában kihasznált kb. 70 betűvel és írásjellel. Azok a fájlok amelyek nem ennyire "rendezett", kevésbé tömöríthetőek. Például egy bináris fájl általában kevésbé tömöríthető mint egy szövegfájl, egy tömörített fájl pedig szinte egyáltalán nem tömöríthető tovább.

## Hang

A hangtömörítési eljárás digitális hangtechnikai, tágabb értelemben számítástechnikai fogalom. Ezen eljárások fő célja az, hogy a nagyon nagy méretű (nagy adatmennyiséget tartalmazó) audioanyagok méretét csökkentse, és így gazdaságosabbá tegye tárolásukat, illetve lehetővé tegye azok tárolását korlátozott kapacitású médiákon (mint amilyen a CD és a DVD).

- Veszteségmentes tömörítés

Az eljárások elve az, hogy a hanganyagok jellegzetességeire alapulva speciálisan ezen célra kialakított algoritmusokat használnak. A veszteségmentesen elérhető méretcsökkenés jelentősen kisebb, mint a veszteségesen elérhető, tipikus mértéke 15%–50%.

Eljárások pl:

- Free Lossless Audio Codec – FLAC
- WMA Lossless – WMA

- Veszteséges tömörítés

A hangtömörítési eljárások nagy része veszteséges tömörítés, vagyis a tömörítés folyamán információ vesz el: a cél az, hogy ez az információvesztés ne okozzon hallható minőségromlást, illetve a minőségromlás minél kisebb legyen.

A módszer lényege az úgynevezett pszichoakusztika, vagyis hogy az emberi fül nem minden létező hangot hall meg, illetve nem minden hangmagasságra egyformán érzékeny. Az eljárások megpróbálják a kevésbé hallott, vagy nem hallható részeket elhagyni, vagy beolvasztani a jobban hallható részekbe úgy, hogy ezzel adatcsökkenést tudjanak elérni.

Mivel a kódolások adatvesztéssel járnak, és az eltérő eljárások eltérő módon okoznak adatvesztést, minden ezen eljárásokkal kódolt hanganyag (veszteséges tömörítéssel való) újrakódolása vagy átalakítása jelentős minőségromlást eredményez.

Eljárások:

- Ogg Vorbis: egy szabad és nyílt hangtömörítési eljárás (kodek), amit a Xiph.org Alapítvány fejleszt.
- MP3: egy veszteséges tömörítésen alapuló zenei fájlformátum, jelenleg az egyik legelterjedtebb. Az MPEG-1-es szabvány nem ad pontos meghatározásokat egy MP3 tömörítő számára. Csak az algoritmus és a fájlformátum, mint körvonal adottak. A szabványt implementálóknak maguk kell meghatározni a megfelelő algoritmusokat a hangoknak az eredeti audióból történő elhagyására
- AAC (az Apple által favorizált formátum)
- Dolby AC-3
- Windows Media Audio – WMA

## Videó

Alapelv: **a különbségi kódolást alkalmazzák általában. Az tároljuk csak el, hogy az előzőhöz képest mi változott, csak a változásokat kell tárolni, így jelentős megtakarítást érhetünk el.**

**AVI** (Audio Video Interleaved): Video for Windows szolgáltatás szabványos fájlformátuma. Digitalizált videofilmeket tartalmaz, valamint a hozzá tartozó, egyfajta digitalizált hanggal. Ezek a videók lejátszhatók a Windows beépített médialejátszójával, egy OLE-kompatibilis felhasználói programmal vagy más olyan szoftver segítségével, amely közvetlenül képes a megfelelő lejátszási rutinokkal Windowsban kommunikálni.

**MOV** (Movie Files): Az Apple Quicktime for Windows programja ebben a formátumban tárolja a mozgófilmeket és digitalizált hangot is. A Windows alatti programokat speciálisan fel kell készíteni a Quicktime klippek lejátszására. Vannak konvertáló programok, amelyek az AVI és MOV formátumokat képesek egymásba átalakítani.

**MPEG** (Motion Picture Express Group, mozgóképes szakmai csoport): A videó-, és audiojelek be- és kitömörítésének szabványos technikáját definiálja. Abban hasonlít a JPG-re, hogy nagyon nagy mértékű tömörítést alkalmaz, ez azonban veszteséges. Kihhasználja, hogy a filmek rendszerint olyan képsorozatokat tartalmaznak, amelyekben számos képkocka hasonlít az előzőekre, illetve következőkre, ezért nem kell minden képkocka minden egyes képpontját tárolni. (Ennek eredményeként viszont a közönséges MPEG-film nem szerkeszthető!) A betömörítés nehezebb, ezért hosszadalmas: egy néhány perces film feldolgozása esetleg egy óráig is eltarthat, mert részletesen elemezni kell a képkockák tartalmát. A lejátszás természetesen (többé-kevésbé) valós idejű. A hosszadalmas műveletek gyorsítására különleges hardvereszközöket lehet használni. Mivel a lejátszás a gyakoribb, számos kereskedelmi grafikus kártya tartalmaz MPEG-lejátszó áramköröket, amelyeket a meghajtó program kezel.

### Rippelés:

A DVD-ken található anyagok általában CSS védelemmel (Content Scrambling System), ill. Macrovision védelemmel is el vannak látva. A videó-kódolásra alkalmazható programok nagy többsége (kivéve, pl. FlaskMPEG) ezek miatt nem tud közvetlenül a DVD-ről dolgozni (ill. nem adna megfelelő eredményt), ezért a DVD tartalmát vagy tartalmának egy részét a további föl- és átdolgozás előtt meg kell szabadítani a fentebb említett védelmekről. Erre szolgálnak az ún. ripper programok, melyek segítségével a DVD-lemezen levő "védett" anyag immár "védtelen" formában átkerül a merevlemezre. Ez az adatmennyiség jelentős, a legtöbb esetben több GB.