

Kompresia údajov

Komprimácia (kompresia, pakovanie, balenie) je proces, pri ktorom sa prekódujú údaje tak, aby **zaberali menší objem a dali sa opačným postupom odkódovať**.

Kompresia dát je vlastne efektívna reprezentácia zdroja. Presnejšie povedané, kompresia dát je reprezentácia zdroja v digitálnej podobe s čo najmenším možným počtom bitov. Toto sa dá dosiahnuť dvoma spôsobmi: stratová kompresia (lossy) znamená, že zdrojový signál nie je z dát 100% rekonštruovateľný a bezstratová kompresia (lossless) znamená, že sme schopní rekonštruovať zdrojový signál do jeho presnej podoby.

Bezstratová kompresia

Je to trieda kompresných algoritmov, umožňujúca spätnú rekonštrukciu komprimovaných dát do pôvodnej podoby bez straty informácie, narozdiel od stratovej kompresie, kde toto nie je možné.

Bezstratová kompresia je použitá v kompresných nástrojoch ako napríklad veľmi populárny formát **zip**, pôvodne unixové programy **bzip2**, **gzip** či **compress**. Populárne formáty sú aj **rar**, **StuffIt**, **7z**, **arj** a mnohé ďalšie.

Bezstratová kompresia sa používa všade tam, kde je potrebné, aby dekomprimované údaje boli identické s pôvodnými údajmi. Typické príklady sú spustiteľné programy alebo zdrojové kódy. Niektoré obrázkové formáty, napríklad **PNG** alebo **GIF**, používajú iba bezstratovú kompresiu.

Stratová kompresia

Je to kompresia, ktorá využíva algoritmy zabezpečujúce kompresiu (zníženie objemu dát) multimédia (zvuk, obraz, video) za cenu jeho kvality.

Stratová kompresia je metóda, ktorou po zkomprimovaní dát a ich rozkomprimovaní získame výsledok odlišný od originálu. Tento výsledok bude však tak blízko originálu, že sa dajú ešte na nejaký účel použiť (multimédia, streaming a v VOIP).

Stratová kompresia je založená na vynechávaní, niektorých málo viditeľných detailov obrazu. V praxi to znamená, že ak je niekde napríklad jedna svetlozltá bodka uprostred veľkého bieleho poľa, jednoducho sa vymaže. Ďalej ak je niekde tenká čiara medzi dvoma plochami, tak sa farba tejto čiary upraví tak, aby sa jej farba dala vypočítať zložením farieb plôch, ktoré obklopuje (zjednodušene povedané).

Stratová kompresia sa najčastejšie používa pre ukladanie obrazových a zvukových záznamov.

Napriek tomu, že sa časť informácie pri stratovej kompresii nenávratne stráca, je tento spôsob ukladania dát často veľmi výhodný. Strata niektorých informácií je totiž úplne vyvážená veľmi výrazným zmenšením komprimovaných dát. Zvyčajne je tak určitá (malá) strata kvality vyvážená výraznou úsporou miesta.

Vďaka stratovej kompresii možno na CD-ROM umiestniť aj 10 pôvodných nekomprimovaných zvukových CD, alebo tisíce obrázkov; na DVD sa vojde celovečerný film, vrátane mnohých zvukových sprievodov.

Takto komprimované dáta sa ľahšie ukladajú a ľahšie sa s nimi manipuluje.
Príklady použitia:

- digitálne televízne vysielanie
- filmové DVD
- prenášanie dát cez Internet a dátové siete
- ukladanie do multimedialných prehrávačov

Stratová kompresia je nepoužiteľná v prípade, kedy je potrebné uchovať presnú kópiu pôvodných údajov, napríklad text knihy, program alebo výsledky merania.

Princípy stratovej kompresie

Všeobecný princíp stratovej kompresie je jednoduchý. Po úvodnom predspracovaní preskupením alebo transformovaním dát tak, aby bolo možné ľahko oddeliť dôležité informácie od nedôležitých. Nepodstatné informácie sa potom potlačia oveľa výraznejšie ako dôležité a nakoniec sa výsledok skomprimuje niektorým z bezstratových kompresných algoritmov.

Algoritmus stratovej kompresie má teda dve podstatné časti - transformácia pôvodných dát a potlačenie menej dôležitých dát.

Transformácia pôvodných dát

K transformácii pôvodných alebo predspracovaných dát sa zvyčajne používa niektorá z ortonormálnych alebo takmer ortonormálnych transformácií. Príkladom takýchto transformácií sú DCT (diskrétna kosínusová transformácia), FFT (rýchla Fourierova transformácia) alebo DWT (diskrétna vlnková transformácia). Tieto transformácie prevedú pôvodné dáta do iných domén, napríklad z časovej do frekvenčnej. Väčšina dôležitých informácií je potom uchovaná v oveľa menšom objeme než pôvodne. Ak zvyšok dát nahradíme nejakými vopred známymi alebo vypočítateľnými dátami (niekedy sa na tento účel hodia samé nuly), dáta sa po spätnej transformácii budú veľmi dobre podobať dátam pôvodným.

Pri transformácii ešte nemusí dochádzať k degradácii pôvodných dát.

Potlačenie niektorých dát

V tejto časti kompresného algoritmu je rozhodujúci kvalitný psychovizuálny alebo psychoakustický model, ktorý určuje, aké dáta môžu byť potlačené alebo dokonca úplne odstránené. Pri kompresii obrazu sa posudzuje, ktoré frekvencie v obraze sú dôležité, aby človek na obrázku videl to, čo na ňom vidieť má. Podobne pri kompresii zvuku sa hľadajú frekvencie, ktoré človek nemôže vnímať. Problém pri kompresii zvuku je o to zložitejší, že ľudský sluch je veľmi citlivý aj na časové umiestnenie zvuku. Aj s tým musí dobrý psychoakustický model počítať.

Formáty využívajúce stratovú kompresiu JPEG, JPEG 2000, MPEG, MP3, Ogg Vorbis, WMA, AAC, VQF

Z princípu stratovej kompresie jednoznačne vyplýva, že nie je možné komprimovať touto metódou napríklad textový súbor. Po jeho dekompresii by sme nedostali čitateľné dáta.

Zdroje:

http://projects.hudecof.net/diplomovka/online/ucebnica/html/uvod_vyznam.html

http://sk.wikipedia.org/wiki/Stratov%C3%A1_kompresia

http://sk.wikipedia.org/wiki/Bezstratov%C3%A1_kompresia