



Kompresija podatkov

Informacijo lahko kategoriziramo v dve kategoriji:

- ✓ Informacija v obliki bloka (računalniške datoteke – tekstovne, slikovne, audio itd.),
- ✓ Informacija v obliki podatkovnega toka (informacija se nenehno generira v izvoru).

Večjo učinkovitost shranjevanja in pošiljanja podatkov dosežemo s *kompresijo podatkov*. S kompresijo kodiramo originalno informacijo, tako da je podana z manjšim številom bitov. *Kompresijsko razmerje* KR je definirano kot razmerje med številom bitov originalne informacije in številom bitov kompresirane informacije.



Uvod

Tip informacije	Kompresijska tehnika	Format	Originalna velikost	Kompresirana velikost	Uporaba
Tekstovna datoteka	zip	ASCII		KR = 2-6	Shranjevanje, prenos
Skeniran črno beli dokument	CCITT Group 3 fasc. standard	A4 (8.5x11 inč), 200x200 točk/inčo	256 kb	5-35 kb KR = 5-50	Faksimile, dokumentacija
Barvna slika	JPEG	8x10 inč 400x400 točk/inčo	38.4 Mb	1.2-8 Mb KR = 5-30	Shranjevanje, pošiljanje
Barvna fotografija	JPEG	3.2 milijona točk 2160x1440 točk	9.6 Mb	0.3-2 Mb KR = 5-30	Shranjevanje, pošiljanje
Glasovna		4 kHz	64 kbps		Digitalna telefonija
Glasovna	Residual-excited linear prediction	4 kHz	64 kbps	8-16 kbps	Digitalna celična telefonija
Audio	MPEG audio MP3	16-24 kHz	512-748 kbps	32-384 kbps	MPEG audio
Video	MPEG-2	720x480 točk/slika 30 slik/sek	249 Mbps	2-6 Mbps	Prenos videa
Video	MPEG-2	1920x1080 točk/slika 30 slik/sek	1.6 Gbps	19-38 Mbps	HDTV

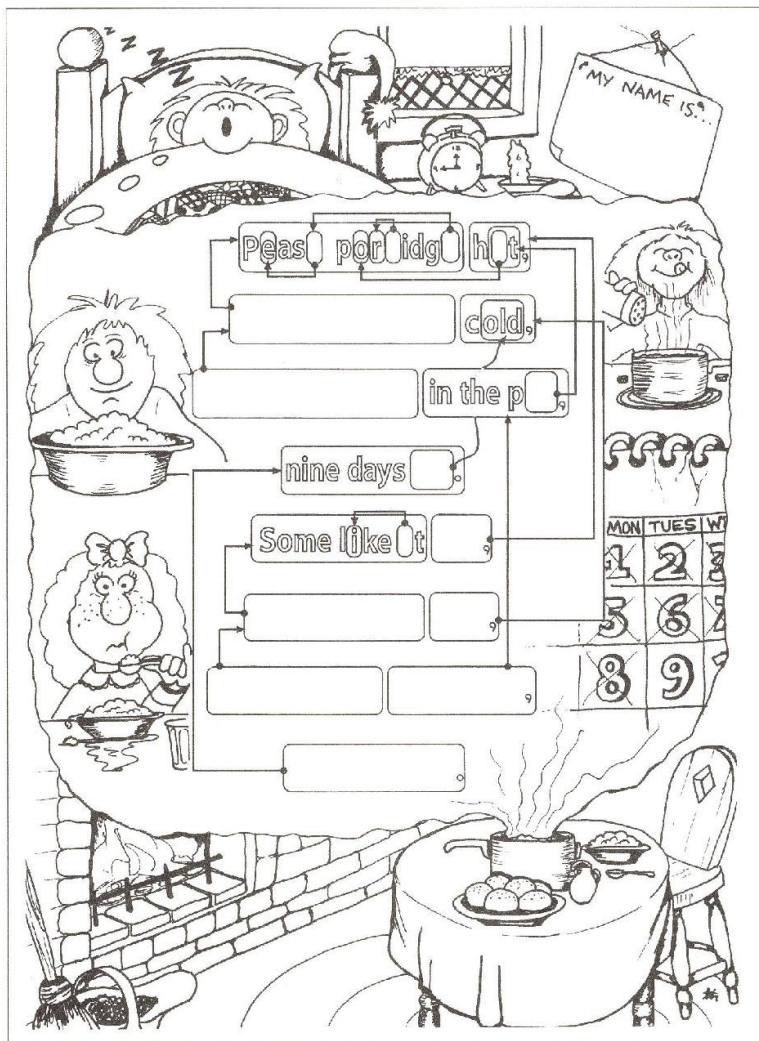


Kompresija brez izgube

Kompresija brez izgube omogoča ponovno generiranje originalne informacije. Uporablja se za učinkovitejšo izrabo spominskega prostora pri shranjevanju računalniških datotek, ki jih kreirajo različni programi in za učinkovitejši prenos podatkov.

Večina računalniških datotek je precej redundantnih, kar pomeni, da imajo precej podatkov, ki se neprestalno ponavljajo. Kompresijski program pa to redundanco odstrani. Koščke podatkov, ki se ponavljajo, kompresijski program uporabi samo prvič, v nadaljevanju, ko se ponovijo pa se nanje samo sklicuje s povezavo znotraj datoteke kot ilustrira slika.

Kompresija brez izgube





Kompresija brez izgube

Primer 1: Originalna informacija 79 znakov, kompresirana 74

*"Ask not what your country can do for you –
ask what you can do for your country."*

1. ask
2. what
3. your
4. country
5. can
6. do
7. for
8. you

"1 not 2 3 4 5 6 7 8 -- 1 2 8 5 6 7 3 4"



Kompresija brez izgube

Primer 1: Originalna informacija 79 znakov, kompresirana 59

*"Ask not what your country can do for you –
ask what you can do for your country."*

1. ask__
2. what__
3. you
4. r__country
5. __can__do__for__you

"1not__2345__--__12354"



Kompresija brez izgube

Zelo zanimiv je primer arhiviranja dokumentov. Ena stran skeniranega dokumenta formata A4 ima velikost 256 kb. Če uporabimo strategijo optičnega razpoznavanja teksta (OCR – Optical Character Recognition) in predstavimo dokument z standardnimi znaki kot je ASCII koda, pa je velikost ene strani dokumenta še samo okrog 4kb, kar je odvisno od števila znakov dokumenta. Vendar zaenkrat še ta tehnologija ni čisto zanesljiva in lahko pride do hudih napak. Zato je mnogo zanesljivejša uporaba kompresije z izgubo, ki omogoča rekonstrukcijo, ki ustreza originalnemu dokumentu.



Kompresija brez izgube

Pri kompresiji iščemo čim večje vzorce, ki se ponavljajo in jih pri ponovitvi zamenjamo z mnogo krajšim sklicevanjem na prvo pojavitev tega vzorca ali pa na slovar vzorcev – odvisno od načina kodiranja.

Kompresijsko razmerje je odvisno od mnogo faktorjev kot so vrsta datoteke, dolžina datoteke in kompresijske sheme. V tekstovnih datotekah se določene črke in besede pogosto pojavljajo v enakih vzorcih. Tako imajo veliko redundance in imajo veliko kompresijsko razmerje. Podobno je s programskimi jeziki, ker uporabljajo sorazmerno majhen nabor ukazov. Datoteke, ki vsebujejo pretežno različne podatke (kot so grafične ali glasbene datoteke) ne vsebujejo ponavljajočih vzorcev in niso primerne za kompresijo po opisanem postopku.



Kompresija z izgubo

Pri tehniki kompresije z izgubo pa se preprosto odstranijo manj pomembni biti informacije in datoteka se prekroji, tako da postane manjša. Ta tehnika se uporablja za zmanjševanje bitnih slik. Enostaven primer te vrste kompresije je, če spremenimo barvno globino bitne slike iz 24 bitov na 8 bitov (iz treh bytov na enega). Na ta način se slikovna informacija zmanjša na tretjino originalne velikosti. Količino informacij lahko zmanjšamo tudi z zmanjšanjem resolucije originalne slike. Če bomo npr. sliko reproducirali izključno na računalniškem monitorju ji lahko prilagodimo velikost na resolucijo računalniškega monitorja (npr. 1024x768 točk), ne da bi pri tem vplivali na kvaliteto prikaza.



Kompresija z izgubo - primeri

Velikost slike je 800x533 točk

- ✓ **Primer 1:** Format: bmp, globina barve: 8 bitov, kapaciteta: 418 kb
- ✓ **Primer 2:** Format: bmp, globina barve: 24 bitov, kapaciteta: 1250 kb
- ✓ **Primer 3:** Format: jpg, globina barve: 24 bitov, kapaciteta: 49 kb
- ✓ **Primer 4:** Format: jpg, globina barve: 24 bitov, kapaciteta: 31 kb
- ✓ **Primer 5:** Format: jpg, globina barve: 24 bitov, kapaciteta: 21 kb
- ✓ **Primer 6:** Format: jpg, globina barve: 24 bitov, kapaciteta: 9 kb

Primer 1



November, 2002 TR

12

Primer 2



November, 2002 TR

13

Primer 3



November, 2002 TR

14

Primer 4



November, 2002 TR

15

Primer 5



November, 2002 TR

16

Primer 6



November, 2002 TR

17



Kompresija z izgubo

V splošnem želimo, da se pri kompresiji ohrani barvna globina in originalna resolucija slike. Običajno večja površina slike (npr. nebo ali zelenica) deluje podobno. Vendar tudi ti deli slike vsebujejo točke, ki so različne, tako da so algoritmi za kompresijo brez izgube precej neučinkoviti. Da lahko takšno sliko zmanjšamo, je potrebno spremeniti barvo nekaterim točkam. Če ima slika veliko površino neba, program za kompresijo izbere modro barvo, ki bo uporabljena za vse točke določene površine. Glede na dobljeno informacijo program spremeni barvo omenjenih točk in naredi nov zapis datoteke. Na ta način se dosežejo velika kompresijska razmerja ne da bi se slika opazno spremenila. Seveda pa ne moremo več regenerirati originalne informacije potem, ko smo izvršili kompresijo.

Najbolj uporabljen standard za kompresijo slik je JPEG (Joint Photograph Expert Group). Uporablja se lahko tudi za kompresijo posamezne slike pri video datotekah. Ker se pri videu slike časovno izmenjujejo, vsebuje video informacija tudi časovno redundanco. Video sestavljajo scene. Vsaka scena je posneta v istem okolju in v istem časovnem obdobju. Zaporedne slike, ki to sceno reproducirajo, se torej med seboj le malo razlikujejo. Zato zadostuje, da kodiramo samo razlike med posameznimi slikami.



Kompresija z izgubo

Za kompresijo videa je najbolj uporabljan standard MPEG (Motion Picture Expert Group). MPEG-1 standard je bil razvit za produkcijo digitalnega videa VCR (videorekorder) kakovosti za hitrosti okrog 1.2 Mbps. Ta standard omogoča shranjevanje na CD ploščo in prenos preko digitalnih telefonskih linij. MPEG-2 standard je mnogo širši. Omogoča uporabo različnih velikosti slike (352x240 točk, 720x480, 1440x1152 in 1920x1080 točk). Glavni poudarek je na formatu 720x480, ki ustreza konvencionalni televiziji. Z uporabo MPEG-2 kompresije dobimo zelo kvaliteten video pri hitrostih 4 do 6 Mbps. Standard MPEG-4 pa je namenjen uporabi pri zelo nizkih bitnih hitrostih, kot je pri brezžičnih omrežjih.



Kompresija z izgubo

Za audio kompresijo je najbolj razširjeno MP3 (MPEG layer 3) kodiranje. Omogoča kompresijo, ki zreducira bitne hitrosti za faktor 12, pri zelo majhni izgubi kvalitete zvoka.