

ՅՈՒՆԵՍԿՈ ԵՎ ԵՐԵՎԱՆԻ ՄԱՐԶԻ ԿՐԹԱԿԱՆ ԵՎ ԿՐԹԱԿԱՆ ՎԵՐՈՒՄԸ

ՀԱՅԱՍՏԱՆԻ ԿՐԹԱԿԱՆ ԵՎ ԿՐԹԱԿԱՆ ՎԵՐՈՒՄԸ

ՈՒՆԵՍԿՈ ԵՎ ԵՐԵՎԱՆԻ ՄԱՐԶԻ

Otvoreno računarstvo

Kompresija podataka

- Uvod
- Kompresija bez gubitaka
- Kompresija s gubicima

Mario Žagar





Uvod

Gdje su problemi ?



- Sve veće i veće količine podataka pohranjene u raznim oblicima na raznim medijima na raznim mjestima,...
- Veliki troškovi arhiviranja, pohranjivanja
- Telekomunikacijski troškovi
- Rješenje?

Kompresija podataka

- Otvoreno računarstvo i kompresija podataka u čvrstoj su sprezi!



Kompresija - stenografija



ՌԻԵՃՅ ՄՐԻՋԱՐԻԵ

MARIO ŽAGAR

Mario Žagar

W f2

Što je kompresija?

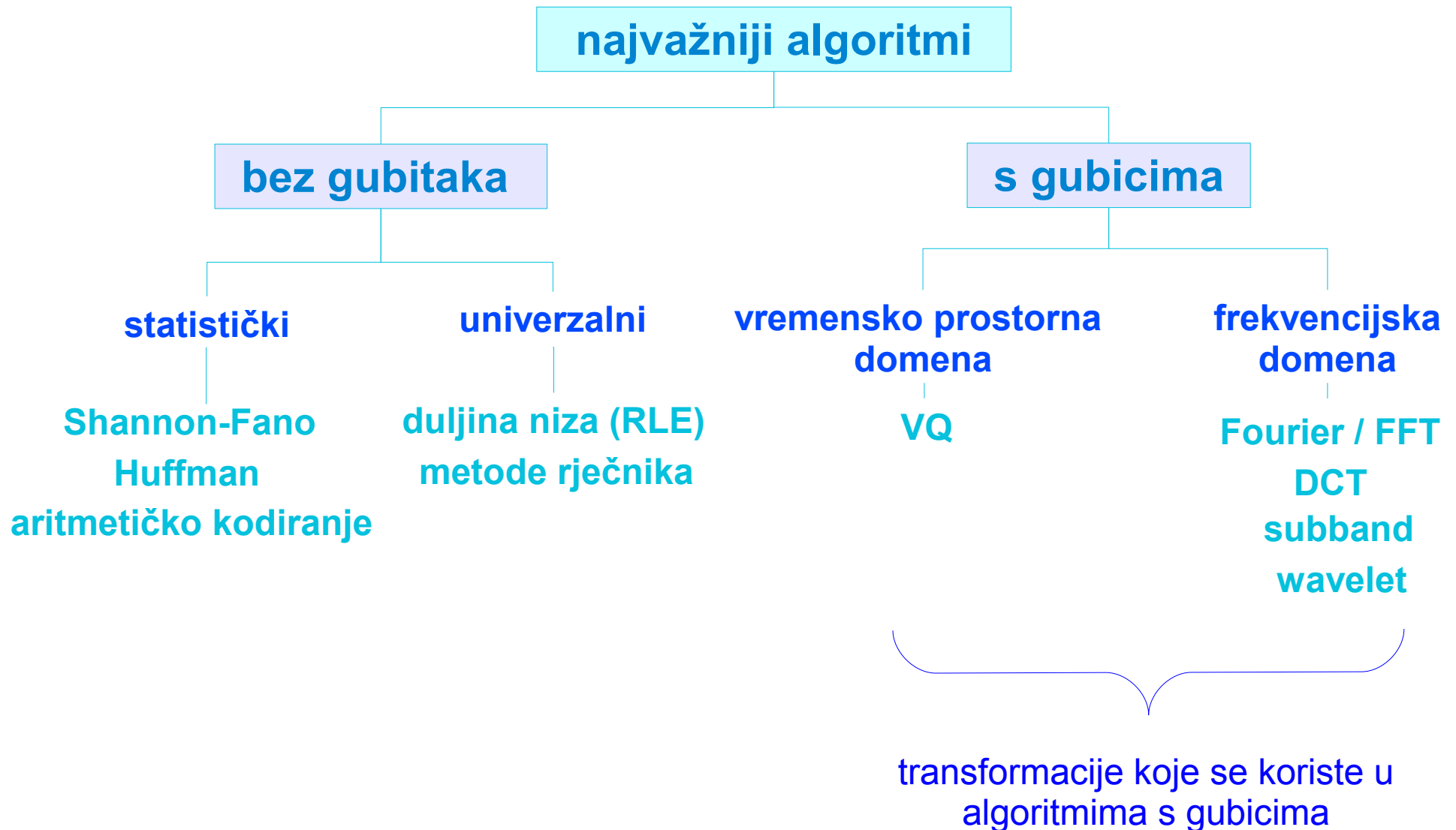


- Metoda za ...
- Zašto je kompresija podataka interesantna?
- Slova, grafikoni, slike, zvuk, video, trodimenzionalne slike, četverodimenzionalne slike,.....
- Gdje je kraj željama?

**KOMPRESIJA PODATAKA
ŠTEDI NOVAC**



Algoritmi kompresije





Kompresija bez gubitaka

Algoritmi bez gubitaka

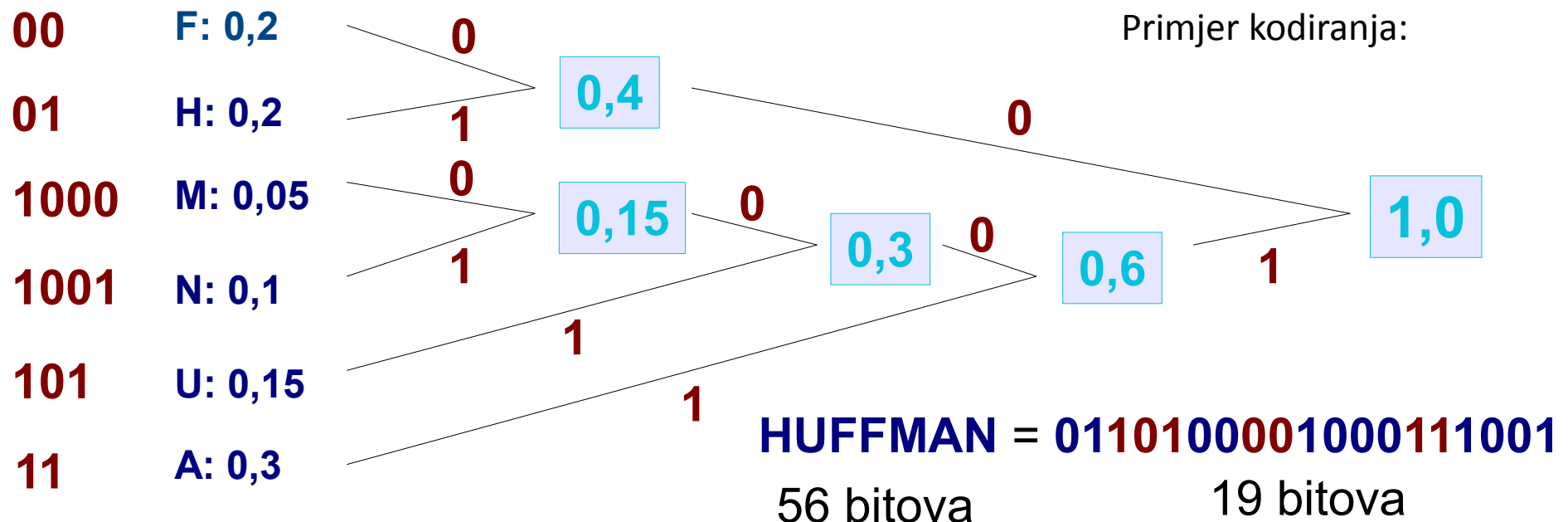


- Dekompresirani dokument istovjetan originalu
- Statistički i univerzalni
- Manji omjeri kompresije
- Primjene: medicina (slike, zvuk,), programi (kôd)
 - npr. bilo bi strašno “zanemariti” (izbaciti) neke mrlje na slici ili “izrezati” neke šumove iz audio zapisa rada srca ili pluća
 - da li su u programu bitne sve nule i jedinice?

Primjer: Huffmanov algoritam



- Huffman D., metoda objavljena 1952.g.
- Zasnovan na statističkim svojstvima
- Dokazano najkraći kôd promjenljive duljine (cjelobrojne duljine)
- Danas najčešće korišten kôd za entropijsko kodiranje
- Ako se koristi nad podacima s različitom razdiobom vjerojatnosti, može povećati količinu podataka





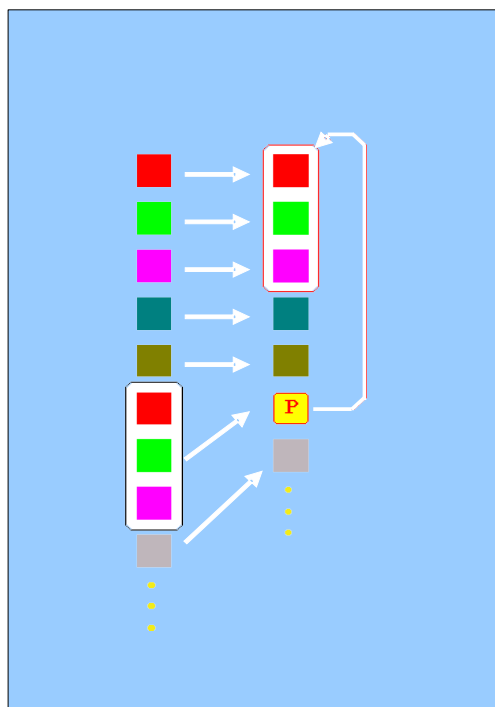
- 000000000000000000111111111111111111...000000.....0001111111.....111110000

11 Otvoreno računarstvo - Kompresija podataka

Metode rječnika



- Zasnovane na višestrukom pojavljivanju grupe simbola
- Stenografija !!
- Dva osnovna pristupa:



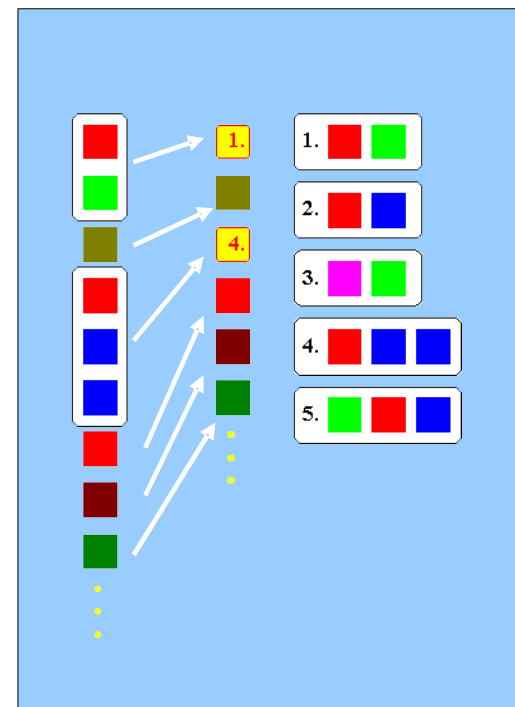
Grupa 1:

LZ77 (Lempel & Ziv, 1977)

LZSS (Storer and Szymanski, 1982)

Grupa 2:

LZ78 (Lempel & Ziv, 1978)
LZW (Lempel, Ziv, Welch 1984)



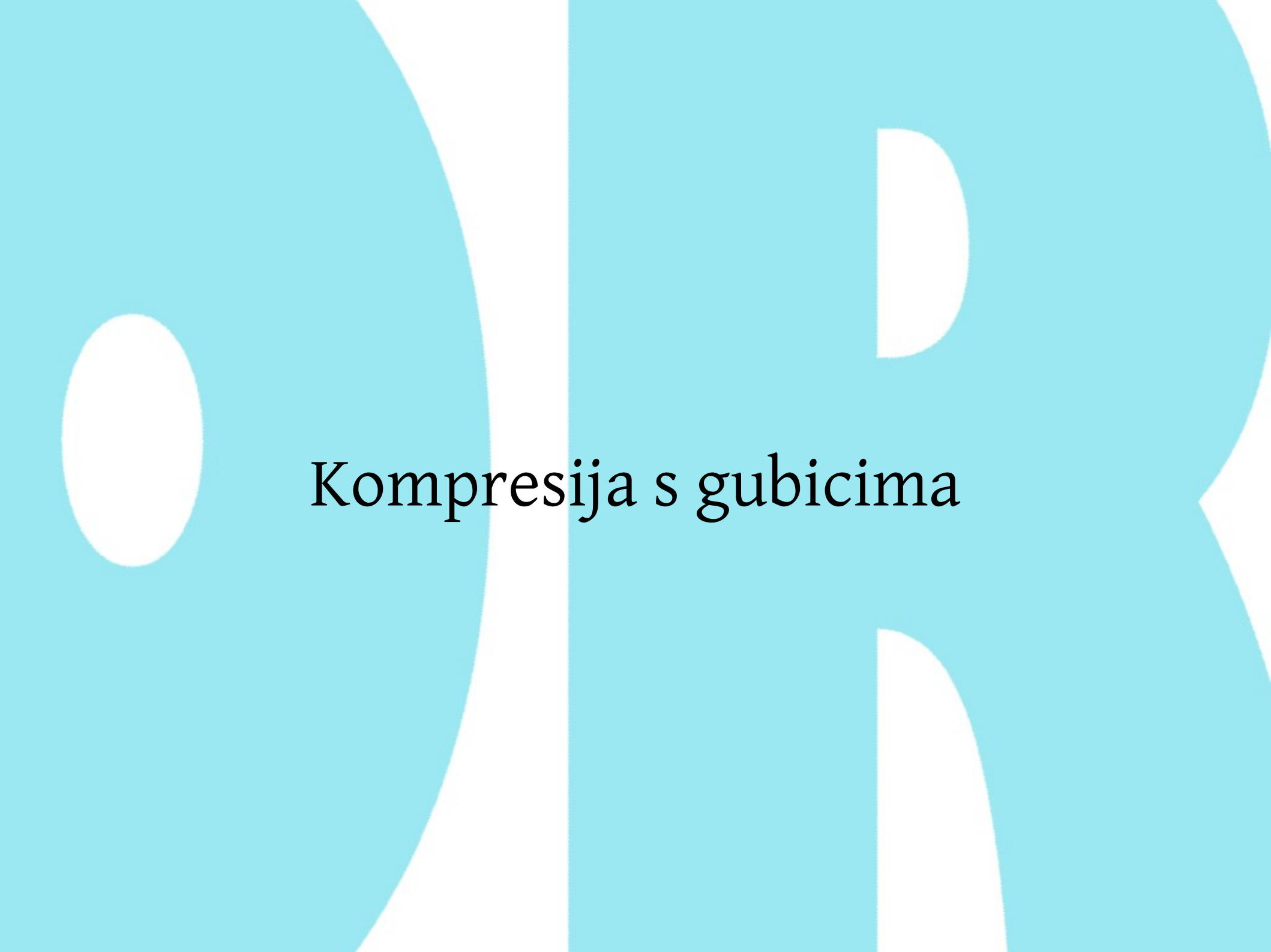
Primjer: JAR

- Java ARchive - kompresija i pakiranje programskih komponenata u Javi
 - Po uzoru na UNIX-ov TAR (Tape Archive)
 - iste opcije (-c, -v, -x, -f, -t, ...)
 - npr. jar -cvf ime...
 - Temelji se na ZIP (Winzip, gzip) i ZLIB formatima
 - osnova tih formata je algoritam za kompresiju datoteka *Deflate* - kombinacija LZ77 (metoda rječnika) i Huffmanovog algoritma
 - LZ77 za označavanje zajedničkih podnizova (niz okteta)
 - Huffman za učestalost pojavljivanja sekvence okteta u datoteci
 - *Deflate* nije patentiran i široko je korišten

Algoritmi bez gubitaka



- Osnova većine programa za kompresiju svih tipova podataka (pkzip, compress,)
- Najbolji algoritmi kombiniraju nekoliko osnovnih algoritama radi postizanja većeg omjera kompresije

The background features large, light blue stylized letters 'O' and 'R' that are partially cut off by the edges of the frame. The 'O' is on the left and the 'R' is on the right, with a vertical white line separating them.

Kompresija s gubicima

Algoritmi s gubicima



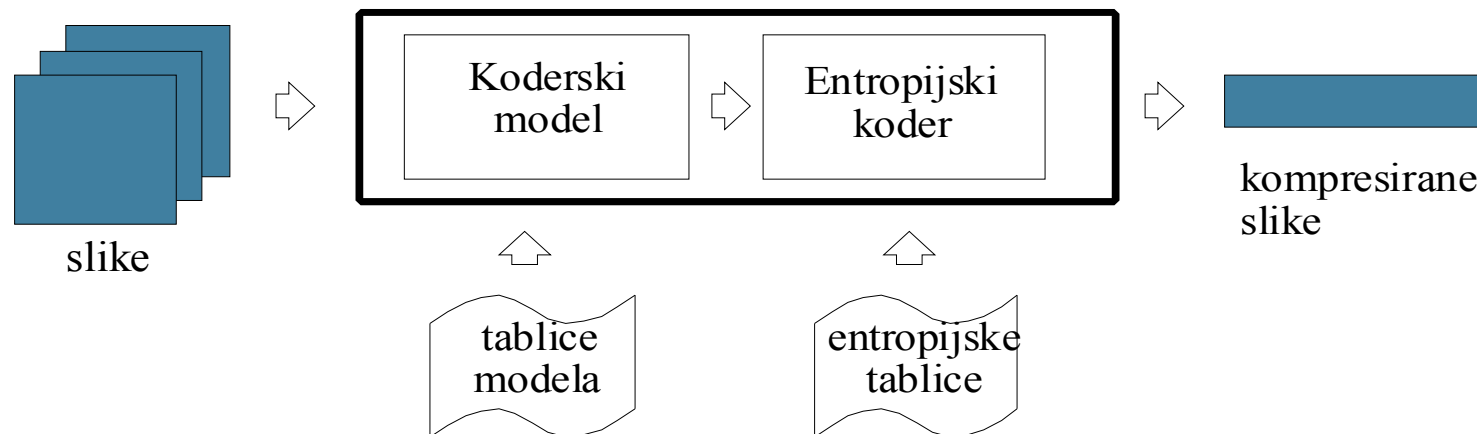
- Dekompresirani dokument sličan je originalu (razlike mogu ali i ne moraju biti prepoznatljive)
- Uglavnom služe za podatke na koje promjena neće značajno utjecati
- Vrlo kompleksni algoritmi
- Veliki omjeri kompresije !!!
- Primjene: slike, zvuk, mjerenja, ...

Data Compression
01110010AX\$5%66&
Data Compression
01110010AX\$5%66&
Data Compression
01110010AX\$5%66&
Data Compression

Primjer: JPEG



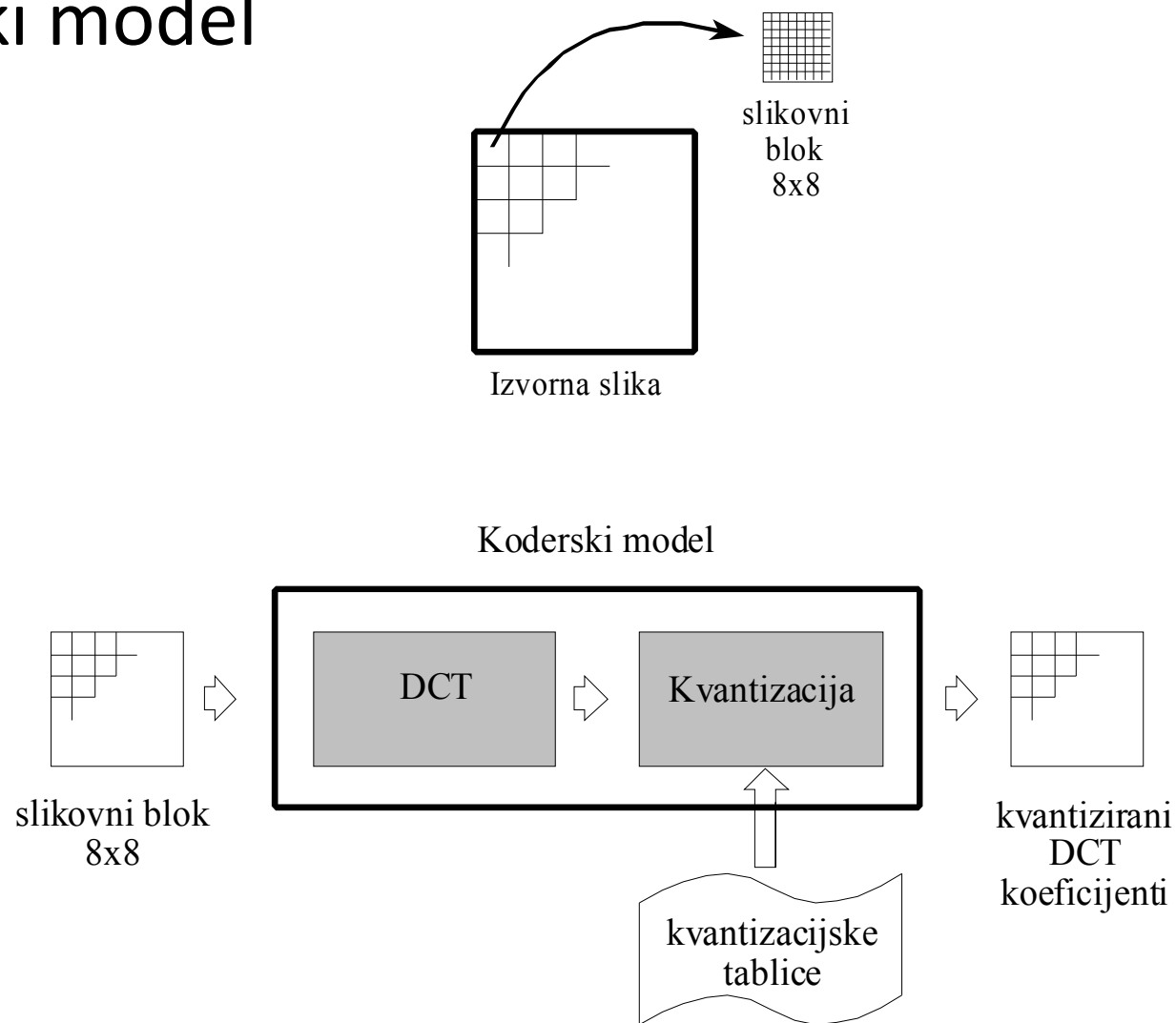
- Algoritam za kompresiju statičkih slika
- Zasniva se na nesavršenosti ljudskog oka
- Algoritam se sastoji od nekoliko koraka



JPEG



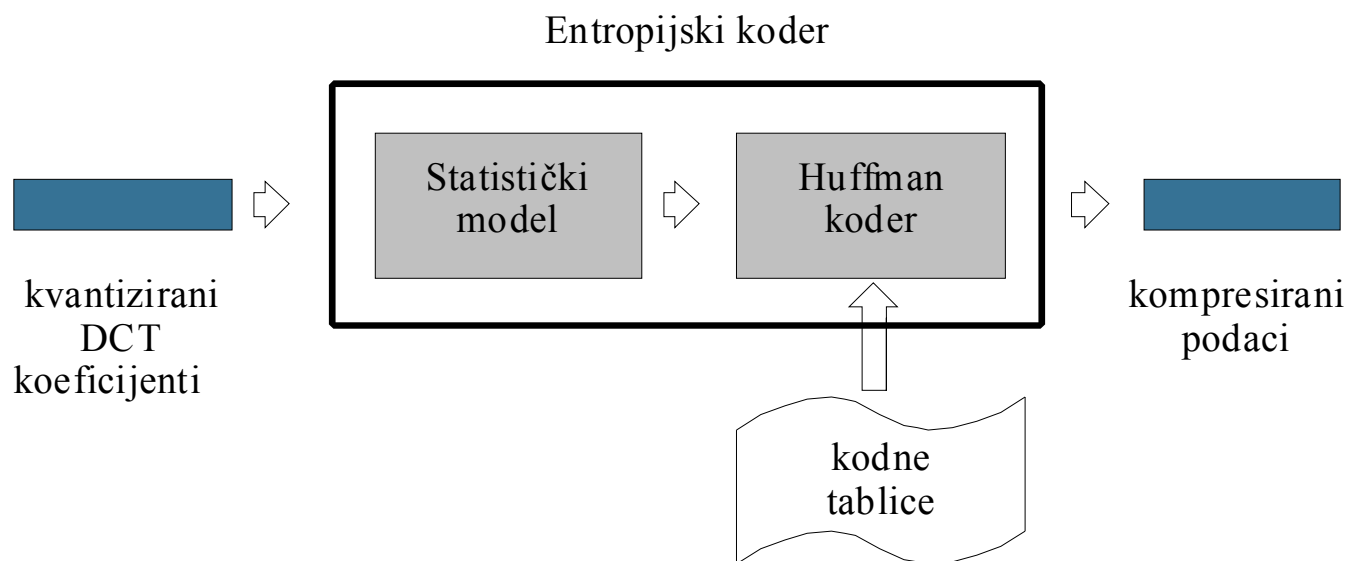
- Koderski model



JPEG



- Entropijski koder



JPEG



- Prepoznajete li original ?



Original: 184 kB



JPEG: 8kB

(23:1)

Pitanja?

- Grupa RASIP bavi se kompresijom podataka
 - Projekt:
 - Data Compression Reference Center (DCRC)
 - <http://compress.rasip.fer.hr>
 - Opisi različitih algoritama s puno više detalja