

Gospodarno kodiranje s postopkom LZW

Informacija in Kodi – Vaja 3

Asistent: as. dr. Klemen Grm

Gospodarno kodiranje

- Bolj učinkovit način zapisa sporočil
- Višanje informacijske vsebine -> več informacije v manj bajtih
- Zgled:

```
>>> print("aaaaaaaaaabbmbmbmbmbcd")  
aaaaaaaaaabbmbmbmbmbcd  
>>> print(10*"a"+10*"b"+"cd")  
aaaaaaaaaabbmbmbmbmbcd
```

Huffmanov kod

- Določitev binarnih kodnih preslikav na podlagi verjetnostni znakov
- Sortiranje po verjetnosti
- Grajenje drevesa predpon
- Zgled: postopek JPEG

Huffmanov kod - zgled

Vzemimo, da je podan vir brez spomina V , ki oddaja znake iz abecede $A = \{x_1, x_2, x_3, x_4\}$ in je moč abecede torej enaka $a = 4$. Vzemimo, da so verjetnosti oddaje znakov iz abecede A enake $p(x_1) = 0,81$, $p(x_2) = 0,09$, $p(x_3) = 0,09$ in $p(x_4) = 0,01$. Določite Huffmanov kod podanega vira, izračunajte povprečno dolžino kodnih zamenjav \bar{n} , entropijo vira $H(V)$ in uspešnost koda η !

Huffmanov kod - zgled

Vzemimo, da je podan vir brez spomina V , ki oddaja znake iz abecede $A = \{x_1, x_2, x_3, x_4\}$ in je moč abecede torej enaka $a = 4$. Vzemimo, da so verjetnosti oddaje znakov iz abecede A enake $p(x_1) = 0,81$, $p(x_2) = 0,09$, $p(x_3) = 0,09$ in $p(x_4) = 0,01$. Določite Huffmanov kod podanega vira, izračunajte povprečno dolžino kodnih zamenjav \bar{n} , entropijo vira $H(V)$ in uspešnost koda η !

- Domača naloga: Kod in η za V^2

Huffmanov kod - zgled

- Prejmemo s prejšnjim kodom izpeljan niz

011100100111010100

Dekodiraj znake izvirnega sporočila.

Postopek LZW

- Ne potrebujemo informacije o porazdelitvi znakov
- Sproti gradimo slovar podnizov sporočila:

Vzemimo niz dvojiških znakov

1011010100010...

Od začetka proti koncu niza iščemo najkrajše podnize, ki se v nizu prvič pojavijo. Dobimo

1, 0, 11, 01, 010, 00, 10...



Kod LZW - algoritem kodiranja

Algoritem gospodarnega kodiranja LZW temelji na ugotavljanju podnizov danega niza (na primer besedila, zakodiranega s kodom ASCII).

TABELE:

- ST* Tabela, ki hrani sporočilo (besedilo), ki ga želimo kodirati,
- KT* tabela s kodnimi zamenjavami sporočila iz tabele *ST*,
- PT* pomožna tabela – *slovar*, ki hrani na mestih med 0 in 255 znake abecede vira, od mesta 256 naprej pa različne nize znakov abecede vira.

VHOD: Tabela *ST*.

IZHOD: Tabela *KT*.

Kod LZW

- 1. korak** Preberi prvi znak s iz tabele ST .
- 2. korak** Dokler ne prideš do zadnjega znaka v tabeli ST , ponavljaj:

- t je naslednji znak v tabeli ST
- spni oba znaka (niza) v niz $u \leftarrow s||t$
- če** je u v slovarju PT ,

potem

- postavi $s \leftarrow u$,

sicer:

- preinesi naslov, ki ga ima s v slovarju PT , v tabelo KT ,
- vpiši u v slovar PT na prvo prosto mesto po naslovu 255,
- postavi $s \leftarrow t$.

konec-če

- 3. korak** Prenesi naslov, ki ga ima s v slovarju PT , v tabelo KT .

Kod LZW - primer

Tabela ponazarja postopek kodiranja in dograjevanja slovarja PT .

Sporočilo ST $TRALALALALA$						
s	t	u	Kod KT		Slovar PT	
			$(\cdot)_{10}$	$(\cdot)_2$	Vpis	Mesto
T	R	TR	84	01010100	TR	256
R	A	RA	82	01010100	RA	257
A	L	AL	65	01000001	AL	258
L	A	LA	76	01001011	LA	259
A	L	AL				
			PZ	11111111		
AL	A	ALA	258	100000010	ALA	260
A	L	AL				
AL	A	ALA				
ALA	L	$ALAL$	260		$ALAL$	261
L	A	LA		100000100		
LA		LA	259	100000011		

Kod LZW - algoritem dekodiranja

Algoritem za dekodiranje mora poleg dekodiranja dograjevati slovar PT , prav tako kot v primeru kodiranja.

VHOD: Tabela KT .

IZHOD: Tabela ST .

Začetek postopka

- 1. korak** Preberi kod c iz vhodne tabele KT (tabela s kodiranim sporočilom).
- 2. korak** V izhodno tabelo ST (tabela z dekodiranim sporočilom) vpiši niz znakov s , ki je v slovarju PT vpisan na mestu c .
- 3. korak** V slovar PT vpiši niz $s||t$, kjer je t prvi znak naslednjega niza s .

Konec postopka

Kod LZW - algoritem dekodiranja

Opisani algoritem ne more dekodirati niza znakov sporočila oblike $t||z||t$, kjer je t poljubni znak, z pa poljubni niz znakov.

Problem rešimo tako, da:

- a) vzamemo zadnji niz, ki smo ga dekodirali (vpisali v tabelo ST), to je niz oblike $t||z$,
- b) na njegov konec pripnemo prvi znak niza t (dobimo $t||z||t$) in ga
- c) vpišemo na naslednje prazno mesto v slovar PT .

Kod LZW - primer

Vzemimo, da želimo dekodirati niz kodnih zamenjav

84 82 65 76 PZ 258 260 259.

Mesta v slovarju med 0 in 255 so znaki, ki se lahko pojavijo v sporočilu. Znak PZ tolmačimo, kot je bilo opisano.

Kod <i>KT</i>	Sporočilo <i>ST</i>	Slovar <i>PT</i>	
		Vpis	Mesto
84	<i>T</i>		
82	<i>R</i>	<i>TR</i>	256
65	<i>A</i>	<i>RA</i>	257
76	<i>L</i>	<i>AL</i>	258
258	<i>AL</i>	<i>LA</i>	259
260		<i>ALA</i>	260
	<i>ALA</i>		
259	<i>LA</i>	<i>ALAL</i>	261

3. Laboratorijska vaja

- Naloga 1:
- Na podlagi LZW enkoderja implementiraj dekoder
- Preizkusi delovanje na besedilni datoteki
- Preveri z MD5 izvlečki

3. Laboratorijska vaja

- Naloga 2:
- Implementiraj izračun velikosti kompresiranega sporočila
- Preizkusi uspešnost kompresije na različnih vrstah datotek
- Ovrednoti razmerje med dolžino sporočila in uspešnostjo kompresije

