Sveučilište u Zagrebu
**Fakultet elektrotehnike i računarstva**

Međuispit iz predmeta **TEORIJA INFORMACIJE**, 27. studenog 2020.

**Pravilo bodovanja zadataka**

Svaki točno odgovoreni zadatak donosi 5 bodova, netočno odgovoreni 2 negativna boda, a neodgovoreni 0 bodova.

**Zadatak 1.** Promatrajte izvor na čijem se izlazu pojavljuju dva simbola, i to: točka (•) i crtica (–). Trajanje točke iznosi 0,2 s, trajanje crtice je tri puta dulje, a trajanje stanke između simbola iznosi 0,2 s. Vjerojatnost pojavljivanja točke je dva puta veća od vjerojatnosti pojavljivanja crtice. Izračunajte prosječnu brzinu generiranja informacije izvora u jedinici bit/s.

a) 0.4897 bit/s; b) 2,7552 bit/s; c) 0,9183 bit/s; d) 1,7219 bit/s; e) Ništa od navedenog.

*Postupa rješavanja:*

Dakle, trajanje simbola točka $t_1 = t(\bullet) = 0{,}2$ s, trajanje simbola crtica $t_2 = t(–) = 0{,}6$ s, a trajanje stanke $t_s = 0{,}2$ s. Neka je vjerojatnost pojavljivanja točke $P(\bullet) = p_1$, a vjerojatnost pojavljivanja crtice $P(–) = p_2$. S obzirom da je zadano $p_1 = 2p_2$, a mora vrijediti i jednakost $p_1 + p_2 = 1$, slijedi da je $p_2 = 1/3$, a $p_1 = 2/3$. Prosječna količina informacije po svakom simbolu određuje se proračunom entropije zadanog izvora:

$$H(X) = -\sum_{i=1}^{2} p_i \log_2 p_i = 0{,}9183 \frac{\text{bit}}{\text{simbol}}.$$

S obzirom da se iza svakog simbola generira i stanka, za prosječno trajanje generiranog simbola vrijedi:

$$T = p_1(t_1 + t_s) + p_2(t_2 + t_s) = p_1 t_1 + p_2 t_2 + t_s = 0{,}5333 \frac{\text{s}}{\text{simbol}}.$$

Konačno, prosječna brzina generiranja informacije u jedinici vremena iznosi:

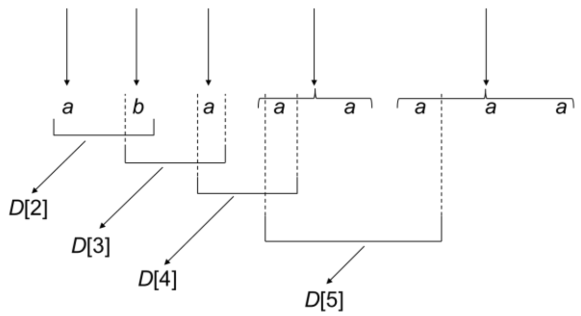
$$R = \frac{H(X)}{T} = \frac{0.9183}{0{,}5333} = 1{,}7219 \frac{\text{bit}}{\text{s}}.$$

**Zadatak 2.** Temeljem polaznog rječnika $D[0] = a$ i $D[1] = b$ dekodirajte primljenu poruku 0 1 0 4 5 koristeći algoritam LZW.

a) abaaaaa (7 znakova); b) abaaaaaaa (9 znakova); c) abaaaaaa (8 znakova);

d) abaaaa (6 znakova); e) Ništa od navedenog.

*Postupak rješavanja:*




Prošireni rječnik: $D[2] = ab$, $D[3] = ba$, $D[4] = aa$, $D[5] = aaa$. Dekodirana poruka: *abaaaaaa*.

**Zadatak 3.** Komunikacijskim kanalom prenose se četiri poruke generirane iz skupa $X$ koji sadrži četiri simbola, $X = \{x_1, \ldots, x_4\}$. Omjer vjerojatnosti njihovog pojavljivanja na izvoru, tj. na ulazu kanala iznosi: $P(x_1) : P(x_2) : P(x_3) : P(x_4) = 1 : 2 : 2 : 5$. Matrica kanala zadana je kao:

$$\begin{bmatrix} 0,1 & 0,2 & 0,2 & 0,5 \\ 0,1 & 0,2 & 0,5 & 0,2 \\ 0,1 & 0,5 & 0,2 & 0,2 \\ 0,5 & 0,1 & 0,2 & 0,2 \end{bmatrix}.$$

Odredite transinformaciju u zadanom kanalu.

a) 0.226 bit/simbol; b) 1,761 bit/simbol; c) 1,535 bit/simbol; d) 0,538 bit/simbol;

e) Ništa od navedenog.

*Postupa rješavanja:*

S obzirom na zadane omjere vjerojatnosti i na činjenicu da mora vrijediti da je zbroj vjerojatnosti svih simbola izvora jednak 1, slijedi

$$\begin{aligned} p(x_1) &= 0{,}1 \\ p(x_2) &= p(x_3) = 0{,}2 \\ p(x_4) &= 0{,}5 \end{aligned}$$

Nadalje, kako bio odredili vjerojatnosti simbola na izlazu kanala, $P(y_j)$, $j = 1, \ldots, 4$, trebamo odrediti matricu združenih vjerojatnosti

$$\left[P(x_i, y_j)\right] = \left[P(x_i) \cdot P(y_j \mid x_i)\right] = \begin{bmatrix} 0{,}01 & 0{,}02 & 0{,}02 & 0{,}05 \\ 0{,}02 & 0{,}04 & 0{,}1 & 0{,}04 \\ 0{,}02 & 0{,}1 & 0{,}04 & 0{,}04 \\ 0{,}25 & 0{,}05 & 0{,}1 & 0{,}1 \end{bmatrix}$$

Dakle, $[P(y_j)] = [0{,}3 \quad 0{,}21 \quad 0{,}26 \quad 0{,}23]$. Sada je moguće odrediti entropije $H(X)$, $H(Y)$ i $H(X, Y)$, te preko njih u konačnici transinformaciju $I(X; Y) = H(X) + H(Y) - H(X, Y)$:

$$H(X) = -\sum_{i=1}^{4} P(x_i) \log_2 P(x_i) = 1{,}761 \, \frac{\text{bit}}{\text{simbol}},$$

$$H(Y) = -\sum_{j=1}^{4} P(y_j) \log_2 P(y_j) = 1{,}9869 \, \frac{\text{bit}}{\text{simbol}},$$

$$H(X, Y) = -\sum_{i=1}^{4} \sum_{j=1}^{4} P(x_i, y_j) \log_2 P(x_i, y_j) = 3{,}5219 \, \frac{\text{bit}}{\text{simbol}}.$$

Konačno, $I(X; Y) = 0{,}226$ bit/simbol.

**Zadatak 4.** Poruka „aaaaabaaaaa*“ (navodnici nisu dio poruke, * je oznaka kraja poruke) kodira se algoritmom LZ77 tako da maksimalna duljina posmičnog prozora iznosi 1 simbol, a prozora za kodiranje 5 simbola. Koliko uređenih trojki (*pomak, duljina, sljedeći simbol*) generira navedeni algoritam kako bi kodirao poruku?

a) Pet; b) Šest; c) Tri; d) Četiri; e) Ništa od navedenog.

*Postupak rješavanja:*

U postupku kodiranja dobiju se četiri trojke: (0,0,a), (1,4,b), (0, 0, a) i (1,4,*).

**Zadatak 5.** Diskretni komunikacijski kanal opisan je matricom:

$$\left[P\left(y_j|x_i\right)\right]=\begin{bmatrix}0{,}8 & 0{,}2 & 0 & 0\\ 0{,}2 & 0{,}8 & 0 & 0\\ 0 & 0 & 0{,}8 & 0{,}2\\ 0 & 0 & 0{,}2 & 0{,}8\end{bmatrix}.$$

Odredite kapacitet takvog kanala.

a) 0.226 bit/s; b) 1,2781 bit/simbol; c) 1,535 bit/s; d) 0,538 bit/s; e) Ništa od navedenog.

*Postupa rješavanja:*

Očito se radi o slabo simetričnom kanalu. Svi su reci matrice kanala permutacije jedni drugih, a zbroj svih elemenata po svakom stupcu je jednak (iznosi 1). Dakle, kapacitet takvog kanala možemo odrediti koristeći izraz:

$$C=\log_2\left[\operatorname{card}\left(Y\right)\right]-H\left(Y|x\right)$$

Broj simbola na izlazu kanala iznosi card($Y$) = 4, a matricu $H(Y|x)$ računamo za jedan redak matrice kanala:

$$H\left(Y|x\right)=\sum_{j=1}^{4}P\left(y_j|x_i\right)\log\left(\frac{1}{P\left(y_j|x_i\right)}\right)=0{,}2\log_2 5+0{,}8\log_2 1{,}25=0{,}7219\,\frac{\text{bit}}{\text{simbol}}$$

Konačno, $C$ = 2 – 0,7219 = 1,2781 bit/simbol.

**Zadatak 6.** Zadan je informacijski izvor s memorijom modeliran markovljevim lancem s dva stanja. Taj je lanac opisan matricom prijelaznih vjerojatnosti:

$$\begin{bmatrix}5/8 & 3/8\\ 3/4 & 1/4\end{bmatrix}.$$

Odredite entropiju izvora.

a) 0,2704 bit/simbol; b) 0,9067 bit/simbol; c) 0,8115 bit/simbol; d) 0,6363 bit/simbol;

e) Ništa od navedenog.

*Postupak rješavanja:*

Stacionarne vjerojatnosti lanca $\boldsymbol{\pi} = [\pi_0, \pi_1]$ moguće je odrediti temeljem jednakosti $\boldsymbol{\Pi}^{\mathrm{T}}\boldsymbol{\pi} = \boldsymbol{\pi}$:

$$\begin{bmatrix}5/8 & 3/8\\ 3/4 & 1/4\end{bmatrix}^{\mathrm{T}}\cdot\begin{bmatrix}\pi_0\\ \pi_1\end{bmatrix}=\begin{bmatrix}\pi_0\\ \pi_1\end{bmatrix}.$$

Pomoćni izraz je $\pi_0 + \pi_1 = 1$. Dakle, $\pi_0 = 2/3$, $\pi_1 = 1/3$. Sada je moguće odrediti entropiju izvora:

$$H=-\sum_{i=0}^{1}\pi_i\sum_{j=0}^{1}p_{ij}\log\left(p_{ij}\right)=-\frac{2}{3}\left(\frac{5}{8}\log_2\frac{5}{8}+\frac{3}{8}\log_2\frac{3}{8}\right)-\frac{1}{3}\left(\frac{3}{4}\log_2\frac{3}{4}+\frac{1}{4}\log_2\frac{1}{4}\right)=0{,}9067\,\frac{\text{bit}}{\text{simbol}}$$

**Zadatak 7.** Zadan je skup $X$ koji sadrži dva simbola, $x_1$ i $x_2$. Simbol $x_1$ ima stvarnu (izmjerenu) vjerojatnost nastupa $P(x_1)$ = 0,6. Promatrač eksperimenta greškom zamijeni vjerojatnosti nastupa simbola $x_i$, i = 1, 2, jednu s drugom. Odredite relativnu entropiju stvarne prema pogrešnoj razdiobi vjerojatnosti simbola.

a) 0 bit/simbol; b) 0,351 bit/simbol; c) -0,234 bit/simbol; d) 0,117 bit/simbol;

e) Ništa od navedenog.

*Postupak rješavanja:*

Relativna entropija jedne prema drugoj razdiobi vjerojatnosti nastupa simbola, npr. $P(x_i)$ prema $Q(x_i)$, izračunava se izrazom:

$$D(P\|Q)=\sum_{i=1}^{n}P(x_i)\log_2\frac{P(x_i)}{Q(x_i)}$$

Dakle, u ovom konkretnom slučaju opisanom u tekstu zadatka vrijedi:

$$D(P\|Q)=\sum_{i=1}^{2}P(x_i)\log_2\frac{P(x_i)}{Q(x_i)}=0,6\cdot\log_2\frac{0,6}{0,4}+0,4\cdot\log_2\frac{0,4}{0,6}=0,117\,\frac{\text{bit}}{\text{simbol}}$$

**Zadatak 8.** Neki skup simbola $X$ sadrži pet simbola $x_i$, $i$ = 1, …, 5. Njihove su vjerojatnosti: $P(x_1)$ = 3/7, $P(x_2) = P(x_3) = P(x_4) = P(x_5)$ = 1/7. Zadani skup simbola potrebno je kodirati tehnikom Shannon-Fano te odrediti minimalnu srednju duljinu kodne riječi koju tako kreirani kôd može imati nad zadanim skupom simbola.

a) 2,143 bit/simbol; b) 2,286 bit/simbol; c) 1,571 bit/simbol; d) 1,857 bit/simbol;

e) Ništa od navedenog.

*Postupak rješavanja:*

Tehnika kodiranja Shannon-Fano ne daje uvijek optimalan kôd. U promatranom slučaju zadani skup simbola s pripadajućim vjerojatnostima moguće je kodirati na dva bitno različita načina:

```
     3/7   0                  3/7   0 0
1  ---------            2  -----------
     1/7   1 0 0              1/7   0 1
3  ------------         1  ---------
     1/7   1 0 1              1/7   1 0
2  -----------          2  -----------
     1/7   1 1 0              1/7   1 1 0
3   -----------         3  ------------
     1/7   1 1 1              1/7   1 1 1

        a)                        b)
```

Odgovarajuće srednje duljine kodnih riječi su $L_a$ i $L_b$:

$$L_a=\sum_{i=1}^{5}l_i\cdot P(x_i)=1\cdot\frac{3}{7}+4\cdot3\cdot\frac{1}{7}=\frac{15}{7}=2{,}143\,\frac{\text{bit}}{\text{simbol}}$$

$$L_b=\sum_{i=1}^{5}l_i\cdot P(x_i)=2\cdot\frac{3}{7}+2\cdot2\cdot\frac{1}{7}+2\cdot3\cdot\frac{1}{7}=\frac{16}{7}=2,286\,\frac{\text{bit}}{\text{simbol}}$$

Dakle, minimalna srednja duljina kodne riječi nad promatranim skupom simbola iznosi 2,143 bit/simbol.

**Zadatak 9.** Promatrajte skup $X$ koji sadrži 9 simbola, svaki s vjerojatnošću $p_i \in (0, 1)$, $i$ = 1, …, 9, zbroj svih vjerojatnosti $p_i$ iznosi 1. Sada fiksirajmo jednu vjerojatnost, neka je $p_1$ = 0,25. Odredite maksimum entropije $H(0{,}25, p_2, \ldots, p_9)$. *Napomena*: općenito, $H(p_1, \ldots, p_n)$ je način zapisa entropije diskretne slučajne varijable $X$ s $n$ simbola čije su vjerojatnosti $p_i$, $i = 1, \ldots, n$.

a) 3,1699 bit/simbol; b) 3,0613 bit/simbol; c) 3 bit/simbol; d) 0,8113 bit/simbol;

e) Ništa od navedenog.

*Postupak rješavanja:*

Općenito, ako vrijedi $1 \le r \le n$, $q_1 = p_1 + \ldots + p_r$ i $q_2 = p_{r+1} + \ldots + p_n$ tada je

$$H(p_1, \ldots, p_n) = H(q_1, q_2) + q_1 H\left(\frac{p_1}{q_1}, \ldots, \frac{p_r}{q_1}\right) + q_2 H\left(\frac{p_{r+1}}{q_2}, \ldots, \frac{p_n}{q_2}\right)$$

U našem slučaju $n = 9$, $r = 1$, $q_1 = 0{,}25$, a $q_2 = 0{,}75$. Sukladno tome vrijedi:

$$H(0{,}25, p_2, \ldots, p_9) = H(q_1, q_2) + q_1 \cdot H\left(\frac{p_1}{q_1}\right) + q_2 \cdot H\left(\frac{p_2}{q_2}, \ldots, \frac{p_9}{q_2}\right) =$$
$$= H\left(\frac{1}{4}, \frac{3}{4}\right) + \frac{1}{4} \cdot H(1) + \frac{3}{4} \cdot H\left(\frac{p_2}{q_2}, \ldots, \frac{p_9}{q_2}\right)$$

$H(1) = 0$, a očito je da će entropija $H(0{,}25, p_2, \ldots, p_9)$ imati maksimalni iznos kad je i entropija $H(p_2/q_2, \ldots, p_n/q_2)$ maksimalna. S obzirom da vrijedi:

$$\sum_{i=2}^{9} p_i = q_2 \to \sum_{i=2}^{9} \frac{p_i}{q_2} = 1,$$

možemo primijeniti opće poznato svojstvo da je entropija $H(\pi_1, \ldots, \pi_m)$, $\sum_{i=1}^{m} \pi_i = 1$, maksimalna kad vrijedi $\pi_i = 1/m$, za svaki $i = 1, \ldots, m$. Dakle, u našem specifičnom slučaju, kako bi ukupna entropija poprimila maksimalnu vrijednost, mora vrijediti:

$$\frac{p_2}{q_2} = \ldots = \frac{p_9}{q_2} = \frac{1}{8}$$

Sukladno tome vrijedi:

$$\max_{\{p_i | i=2,\ldots,9\}} H(0{,}25, p_2, \ldots, p_9) = H\left(\frac{1}{4}, \frac{3}{4}\right) + \frac{1}{4} \cdot H(1) + \frac{3}{4} \cdot H\left(\frac{1}{8}, \ldots, \frac{1}{8}\right).$$

Sada možemo odrediti maksimalnu vrijednost entropije kao:

$$\max_{\{p_i | i=2,\ldots,9\}} H(0{,}25, p_2, \ldots, p_9) = -\frac{1}{4}\log_2 \frac{1}{4} - \frac{3}{4}\log_2 \frac{3}{4} + \frac{3}{4} \cdot \log_2 8 =$$
$$= \frac{1}{2} + \frac{3}{2} - \frac{3}{4}\log_2 3 + \frac{9}{4} = \frac{17}{4} - \frac{3}{4}\log_2 3 = 3{,}0613 \frac{\text{bit}}{\text{simbol}}$$

**Zadatak 10.** Promatrani skup $X$ sadrži $N$ simbola $x_i$. Za vjerojatnosti tih simbola vrijedi:

$$1 > P(x_1) > P(x_2) > \ldots > P(x_{N-1}) \ge P(x_N) > 0, \sum_{i=1}^{N} P(x_i) = 1.$$

Simbole $x_i$ kodiramo Huffmanovim kodom. Svaki simbol pri tome kodiramo s $l_i$ bita. Odredite zbroj $\sum_{i=1}^{N} l_i$ , ako je $L = H(X)$.

a) $N\cdot(N + 1)/2$ [bit]; b) $N^2$ [bit]; c) $(N - 1)\cdot(N + 2)/2$ [bit]; d) $N\cdot(N - 1)/2$ [bit];

e) Ništa od navedenog.

*Postupak rješavanja:*

Ako vrijedi $L = H(X)$, tada mora vrijediti:

$$\sum_{i=1}^{N} l_i P\left(x_i\right) = -\sum_{i=1}^{N} P\left(x_i\right) \log_2 P\left(x_i\right)$$

iz čega slijedi da mora vrijediti: $l_i = -\log_2 P\left(x_i\right)$. S obzirom da su duljine kodnih riječi cjelobrojne, slijedi da vjerojatnosti $P(x_i)$ moraju biti oblika $2^{-k}$, $k \in \mathbf{N}$. Nadalje, s obzirom da općenito vrijedi

$$\sum_{i=1}^{M} 2^{-i} = 1 - 2^{-M}$$

te uzevši u obzir početni uvjet o međusobnim odnosima između vjerojatnosti $P(x_i)$, slijedi da vjerojatnosti simbola $x_i$ moraju zadovoljavati sljedeće:

$$P\left(x_i\right) = \begin{cases} 2^{-i}, & i = 1, \ldots, N-1 \\ 2^{-N+1}, & i = N \end{cases}.$$

Za takav skup vjerojatnosti $P(x_i)$ zadovoljen je i uvjet da njihov zbroj mora iznositi 1. Sukladno tome, vrijedi:

$$l_i = \begin{cases} i, & i = 1, \ldots, N-1 \\ N-1, & i = N \end{cases}$$

i u konačnici

$$\sum_{i=1}^{N} l_i = \sum_{i=1}^{N-1} i + \left(N-1\right) = \frac{\left(N-1\right)\cdot N}{2} + \left(N-1\right) = \frac{\left(N-1\right)\cdot\left(N+2\right)}{2} \left[\text{bit}\right].$$