

Codarea sursei

1. Cod unic decodabil, cod instantaneu.

Codul din figura 1 este unic decodabil ? Dar instantaneu?

Puteți găsi un cod instantaneu cu aceeași lungime a cuvintelor codului ?

s_1	\rightarrow	$c_1 = 01$
s_2	\rightarrow	$c_2 = 011$
s_3	\rightarrow	$c_3 = 10$
s_4	\rightarrow	$c_4 = 1000$
s_5	\rightarrow	$c_5 = 1100$
s_6	\rightarrow	$c_6 = 0111$

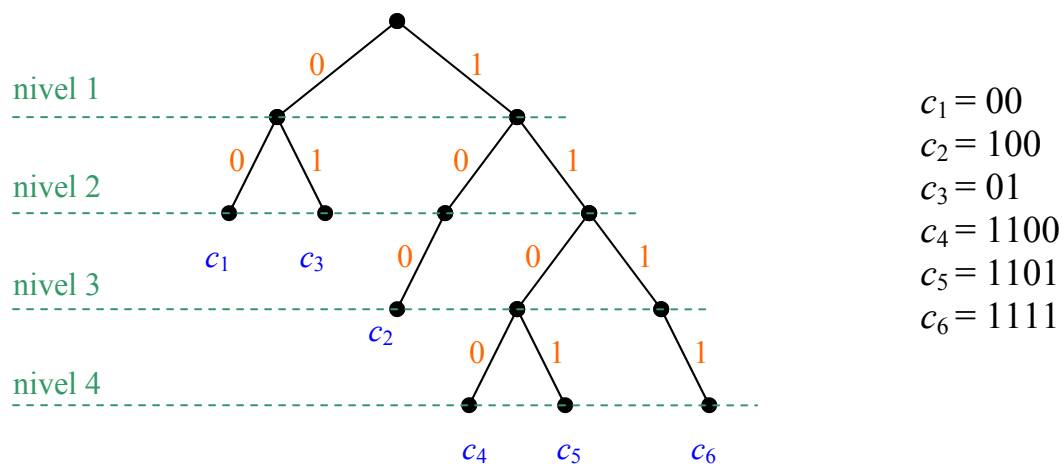
Figura 1.

Rezolvare

Codul din figura 1 este unic decodabil, dar nu este instantaneu. Se poate observa că cuvântul 01 (reprezentarea simbolului s_1) este prefix pentru reprezentarea binară a simbolului s_2 . De asemenea, și c_2 este prefix pentru c_6 . Dacă se recepționează secvența 011, nu putem ști dacă s-a transmis cuvântul c_2 sau dacă se transmite c_6 .

Un cod instantaneu cu aceleași lungimi ale cuvintelor de cod, se poate obține cu ajutorul arborelui, astfel:

Pentru un cuvânt de cod de lungime l_i , trebuie să parcurgem l_i nivele în arbore. Fiecărui simbol vom atribui câte o frunză din arbore. În cazul codurilor binare ($D=2$), dintr-un nod pornesc doi subarbori.



2. Cod unic decodabil.

Codul din figura 2 este unic decodabil?

Reprezentați două mesaje ale sursei primare, ACDF și ABFD, reprezentate cu aceleși cod.

A $\rightarrow c_1 = 010$
B $\rightarrow c_2 = 001$
C $\rightarrow c_3 = 101$
D $\rightarrow c_4 = 0001$
E $\rightarrow c_5 = 1101$
F $\rightarrow c_6 = 1011$

Figura 2.

Rezolvare

Acest cod nu este unic decodabil, deoarece se poate observa:

- cuvântul c_3 este sufix al cuvântului de cod c_5 ;
- cuvântul c_3 este prefix al cuvântului de cod c_6 .

Dacă se recepționează secvența de biți 1011101, aceasta provine fie din secvența de simboluri CE, fie din secvența FC.

C E
┌───┬───┬───┬───┬───┬───┐
1 0 1 1 1 1 0 1
└───┬───┬───┬───┬───┬───┘
F C

Cele două mesaje sunt codate astfel:

A C D F
┌───┬───┬───┬───┬───┬───┬───┬───┬───┬───┬───┬───┐
0 1 0 1 0 1 0 0 0 1 1 0 1 1
└───┬───┬───┬───┬───┬───┬───┬───┬───┬───┬───┬───┘

A B F D
┌───┬───┬───┬───┬───┬───┬───┬───┬───┬───┬───┬───┐
0 1 0 0 0 1 1 0 1 1 0 0 0 1
└───┬───┬───┬───┬───┬───┬───┬───┬───┬───┬───┬───┘

3. Cod binar instantaneu.

Construiți un cod binar instantaneu pentru următorul alfabet al sursei, cu lungimile cuvintelor de cod prestabilite:

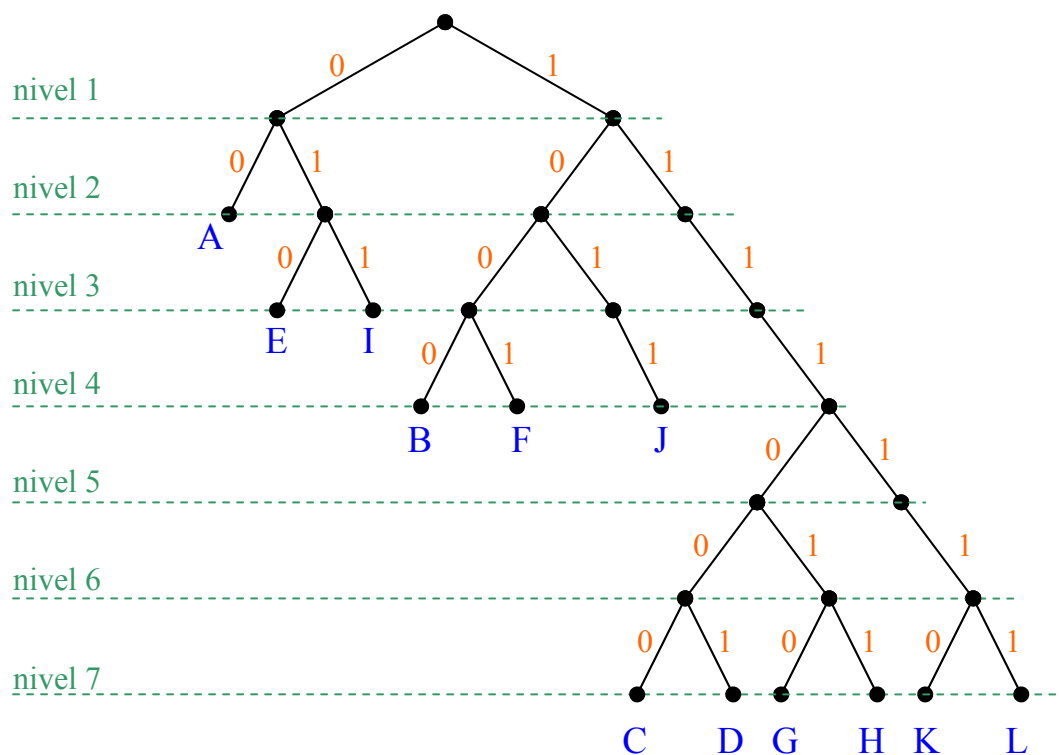
Simbol	A	B	C	D	E	F	G	H	I	J	K	L
Lungime	2	4	7	7	3	4	7	7	3	4	7	7

Rezolvare

Pentru construcția codului se folosește arborele asociat.

Simbolul de lungime l_i va fi reprezentat printr-o frunză pe nivelul l_i , iar cuvântul de cod corespunzător va fi format din mulțimea simbolurilor binare de pe arce, începând de la rădăcină și până la frunza respectivă.

În cazul *binar*, dintr-un nod vor pleca *doi* subarbori, iar alfabetul codului este $\{0,1\}$.



A = 00

B = 1000

C = 1111000

D = 1111001

E = 010

F = 1001

G = 1111010

H = 1111011

I = 011

J = 1011

K = 1111110

L = 1111111

5. Găsiți codurile Huffman binare pentru sursele din tabelul de mai jos.

Simbol	A	B	C	D	E	F	G	H
Prob. (I-a sursa)	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
Prob. (II-a sursa)	0.1	0.2	0.	0.3	0.05	0.1	0.05	0.1
Prob. (III-a sursa)	0.15	0.15	0.15	0.15	0.1	0.1	0.1	0.1

Calculați eficacitatea acestor coduri.

Rezolvare

Se aplică algoritmul de codare Huffman, care constă în aranjarea simbolurilor în ordine descrescătoare a probabilităților și gruparea ultimelor două simboluri într-o sursă restrânsă cu probabilitate egală cu suma probabilităților simbolurilor componente, până când se ajunge la ultima coloană cu numai două simboluri. Acestea sunt codate cu 0 și 1 de la sfârșit către început.

Sursa nr. I.

$p(A) = 1/8, p(B) = 1/8, p(C) = 1/8, p(D) = 1/8,$
 $p(E) = 1/8, p(F) = 1/8, p(G) = 1/8, p(H) = 1/8.$

111 A 1/8		00 r ₁ 2/8		01 r ₂ 2/8		10 r ₃ 2/8	
110 B 1/8		A 1/8		00 r ₁ 2/8		01 r ₂ 2/8	
101 C 1/8		B 1/8		A 1/8		00 r ₁ 2/8	
100 D 1/8		C 1/8		B 1/8		111 A 1/8	11 r ₄ 2/8
011 E 1/8		D 1/8		C 1/8	10 r ₃ 2/8	110 B 1/8	
010 F 1/8		011 E 1/8	01 r ₂ 2/8	D 1/8			
001 G 1/8	00 r ₁ 2/8	010 F 1/8					
000 H 1/8							

11 r ₄ 2/8		0 r ₅ 4/8		1 r ₆ 4/8
10 r ₃ 2/8		11 r ₄ 2/8	1 r ₆ 4/8	0 r ₅ 4/8
01 r ₂ 2/8	0 r ₅ 4/8	10 r ₃ 2/8		
00 r ₁ 2/8				

Codul Huffman este:

A – 111, B – 110, C – 101, D – 100, E – 011, F – 010, G – 001, H – 000.

Eficacitatea codului este: $\eta = \frac{\bar{l}_{\min}}{\bar{l}} = \frac{H(S)/\log D}{\bar{l}} = \frac{3/1}{3} = 1.$

Deoarece $\eta = 1$, codul este absolut optimal.

Sursa nr. II.

$p(A) = 0.1, p(B) = 0.2, p(C) = 0.1, p(D) = 0.3,$
 $p(E) = 0.05, p(F) = 0.1, p(G) = 0.05, p(H) = 0.1.$

01 D 0.3		D 0.3		D 0.3		D 0.3	
000 B 0.2		B 0.2		11 r ₂ 0.2		10 r ₃ 0.2	
100 A 0.1		001 r ₁ 0.1		B 0.2		11 r ₂ 0.2	
101 C 0.1		A 0.1		001 r ₁ 0.1		000 B 0.2	00 r ₄ 0.3
110 F 0.1		C 0.1		100 A 0.1	10 r ₃ 0.2	001 r ₁ 0.1	
111 H 0.1		110 F 0.1	11 r ₂ 0.2	101 C 0.1			
0010 E 0.05	001 r ₁ 0.1	111 H 0.1					
0011 G 0.05							

00 r ₄ 0.3		1 r ₅ 0.4		0 r ₆ 0.6
D 0.3		00 r ₄ 0.3	0 r ₆ 0.6	1 r ₅ 0.4
10 r ₃ 0.2	1 r ₅ 0.4	01 D 0.3		
11 r ₂ 0.2				

Codul Huffman este:

A – 100, B – 000, C – 101, D – 01, E – 0010, F – 110, G – 0011, H – 111

Entropia $H(S)$ a sursei S este:

$$H(S) = -0.3 \cdot \log 0.3 - 0.2 \cdot \log 0.2 - 4 \cdot 0.1 \cdot \log 0.1 - 2 \cdot 0.05 \cdot \log 0.05$$

$$= 0.5209 + 0.4644 + 1.3289 + 0.4322 = 2.7464 \text{ biti/simbol}$$

Lungimea medie a cuvintelor de cod este:

$$\bar{l} = \sum_{i=1}^8 l_i \cdot p(c_i) = 2 \cdot 0.3 + 3 \cdot 0.2 + 4 \cdot 3 \cdot 0.1 + 2 \cdot 4 \cdot 0.05 = 2.8 \text{ biti}$$

Eficacitatea codului este: $\eta = \frac{\bar{l}_{\min}}{\bar{l}} = \frac{H(S)/\log D}{\bar{l}} = \frac{2.7464/1}{2.8} = 0.98.$

Acest cod este optimal.

Sursa nr. III.

$p(A) = 0.15, p(B) = 0.15, p(C) = 0.15, p(D) = 0.15,$
 $p(E) = 0.1, p(F) = 0.1, p(G) = 0.1, p(H) = 0.1.$

000 A 0.15		11 r ₁ 0.2		10 r ₂ 0.2		01 r ₃ 0.3	
001 B 0.15		A 0.15		11 r ₁ 0.2		10 r ₂ 0.2	
010 C 0.15		B 0.15		A 0.15		11 r ₁ 0.2	
011 D 0.15		C 0.15		B 0.15		000 A 0.15	00 r ₄ 0.3
100 E 0.1		D 0.15		010 C 0.15	01 r ₃ 0.3	001 B 0.15	
101 F 0.1		100 E 0.1	10 r ₂ 0.2	011 D 0.15			
110 G 0.1	11 r ₁ 0.2	101 F 0.1					
111 H 0.1							

00 r ₄ 0.3		1 r ₅ 0.4		0 r ₆ 0.6
01 r ₃ 0.3		00 r ₄ 0.3	0 r ₆ 0.6	1 r ₅ 0.4
10 r ₂ 0.2	1 r ₅ 0.4	01 r ₃ 0.3		
11 r ₁ 0.2				

Codul Huffman este:

A – 000, B – 001, C – 010, D – 011, E – 100, F – 101, G – 110, H - 111

Entropia $H(S)$ a sursei S este:

$$H(S) = -4 \cdot 0.1 \cdot \log 0.1 - 4 \cdot 0.15 \cdot \log 0.15$$

$$= 1.3289 + 1.6423 = 2.9712 \text{ biti/simbol}$$

Lungimea medie a cuvintelor de cod este:

$$\bar{l} = \sum_{i=1}^8 l_i \cdot p(c_i) = 4 \cdot 3 \cdot 0.1 + 4 \cdot 3 \cdot 0.15 = 3 \text{ biti}$$

Eficacitatea codului este: $\eta = \frac{\bar{l}_{\min}}{\bar{l}} = \frac{H(S)/\log D}{\bar{l}} = \frac{2.9712/1}{3} = 0.99.$

Acest cod este optimal.

Probleme propuse

6. Găsiți codurile Huffman ternare și quaternare (coduri cu trei, respectiv cu patru simboluri) pentru sursele din tabel.

Simbol	A	B	C	D	E	F	G	H
Prob. (I-a sursa)	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
Prob. (II-a sursa)	0.1	0.2	0.	0.3	0.05	0.1	0.05	0.1
Prob. (III-a sursa)	0.15	0.15	0.15	0.15	0.1	0.1	0.1	0.1

7. Găsiți toate codurile Huffman binare pentru o sursă formată din simbolurile A, B, C, D, dacă A este de două ori mai frecvent decât B și B este de două ori mai frecvent decât oricare dintre C și D.