

Kódování

1. Existuje binární prefixové kódování číslic 0, 1, 2, ..., 9 takové, že číslice 9 je kódována nulou a ostatní číslice mají kód délky nejvýše pět?

Existenci kódu ověříme tak, že zjistíme, zda je splněna Kraftova nerovnost

$$n^{-d_1} + n^{-d_2} + \dots + n^{-d_{|K|}} \leq 1,$$

kde n je počet znaků kódové abecedy a d_i jsou délky kódových značek. Dosazením do nerovnice získáme

$$2^{-1} + 9 \cdot 2^{-5} = \frac{1}{2} + \frac{9}{32} = \frac{25}{32} < 1$$

Nerovnost platí. Pojďme vytvořit kód. Číslice 9 je kódována nulou. Kódy všech ostatních číslic tedy musí začínat jedničkou, pak máme ještě čtyři číslice k dispozici.

9 ... 0
0 ... 1_____
1 ... 1_____
...

Zkusme vystačit jen se třemi číslicemi za jedničkou.

9 ... 0
0 ... 1000
1 ... 1001
2 ... 1010
3 ... 1011
4 ... 1100
5 ... 1101
6 ... 1110
7 ... 1111
8 ...

Pro číslici 8 už by nezbyla žádná kódová značka. Musíme tedy využít celou délku kódové značky, tj. pět číslic. Kódová značka číslice 8 však nesmí být prefixem žádné jiné značky. Musíme tedy přidat pátou číslici i ke kódové značce číslice 7.

9 ... 0
0 ... 1000
1 ... 1001
2 ... 1010
3 ... 1011
4 ... 1100
5 ... 1101
6 ... 1110
7 ... 11110
8 ... 11111

2. Vytvořte tříznakový prefixový kód, který bude kódovat abecedu {A, B, C, D, E, F}. Dvě písmena budou kódována značkou délky 1, čtyři písmena značkou délky 2. Alternativně vytvořte kód, kde budou dvě písmena kódována značkou délky 2, dvě písmena značkou délky 2 a dvě písmena značkou délky 3.

Než se pustíme do vytváření kódu, ověříme Kraftovu nerovnost

$$2 \cdot 3^{-1} + 4 \cdot 3^{-2} = \frac{2}{3} + \frac{4}{9} = \frac{10}{9} > 1$$

Kraftova nerovnost neplatí, zadaný kód nejde vytvořit. Zkusme tedy alternativu

$$2 \cdot 3^{-1} + 2 \cdot 3^{-2} + 2 \cdot 3^{-3} = \frac{2}{3} + \frac{2}{9} + \frac{2}{27} = \frac{26}{27} < 1$$

Tedy je Kraftova nerovnost splněna. Můžeme vytvořit kód. Vezmeme (např.) první dvě písmena vstupní abecedy a přiřadíme jim kódy délky jedna – první a druhý znak z kódové abecedy.

A ... 0

B ... 1

Kódy dalších dvou písmen musí začínat třetím znakem kódové abecedy, aby kódy prvních dvou znaků nebyly jejich prefixem. Na druhou pozici (jsou to značky délky 2) dáme opět první, respektive druhý znak kódové abecedy.

A ... 0

B ... 1

C ... 20

D ... 21

Kódy zbylých dvou písmen (kódové značky délky 3) musí mít na první a druhé pozici třetí znak kódové abecedy. Na třetí pozici dáme opět první, respektive druhý znak kódové abecedy.

A ... 0

B ... 1

C ... 20

D ... 21

E ... 220

F ... 221

Kdybychom přidali ještě písmenu G s kódovou značkou 222, využili bychom všechny možné kódové značky a Kraftova nerovnost by vyšla přesně rovna jedné.

3. Je dána abeceda $\{A, B, C, D, E, F\}$ s pravděpodobnostmi výskytu znaků $\{2/8, 1/8, 1/8, 1/8, 2/8, 1/8\}$. Kódová abeceda je $\{0, 1, 2, 3\}$. Najděte kód s minimální střední délkou značky.

Obecný postup pro vytvoření kódu s minimální střední délkou značky je následující:

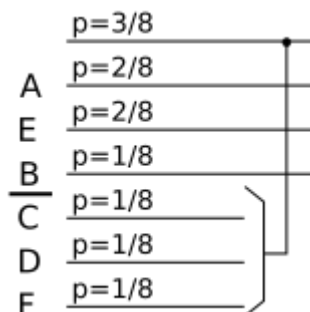
1. Seřadíme znaky vstupní abecedy podle pravděpodobnosti.
2. Rozdělíme znaky do skupin po $n-1$, kde n je počet znaků kódové abecedy.
3. Sloučíme poslední skupinu. Pravděpodobnost výskytu některého znaku ze skupiny je rovna součtu pravděpodobností znaků ve skupině. Sloučenou skupinu zařadíme podle pravděpodobnosti jako nový prvek mezi zbývající znaky.
4. Dále slučujeme (už bez ohledu na skupiny) po n znacích.

Slučování provádíme zdola nahoru. Mezi sloučené prvky pak shora dolů distribuujeme znaky kódové abecedy a vytváříme kódové značky.

Znaky tedy seřadíme podle pravděpodobnosti a rozdělíme je do skupin po třech

A	$p=2/8$
E	$p=2/8$
B	$p=1/8$
C	$p=1/8$
D	$p=1/8$
F	$p=1/8$

Rozdělení do skupin naznačuje tlustá čára. Vpravo od znaků si připravíme „dráty“ po kterých později budeme přidělovat kódové značky. Sloučíme poslední skupinu. Její pravděpodobnost je $3/8$, zařadíme ji tedy úplně nahoru



Dále bychom pokračovali ve slučování po čtyřech. Ale už zbývají jen poslední čtyři prvky. Jejich celková pravděpodobnost je 1. Pokud jste neudělali chybu, mělo by vám slučování po n prvcích vždy vyjít „akorát“, tedy že na konci vám zbyde přesně n prvků.

Nyní na pravé konce „drátů“ distribuujeme znaky kódové abecedy

	$p=3/8$	0
A	$p=2/8$	1
E	$p=2/8$	2
B	$p=1/8$	3
C	$p=1/8$	}
D	$p=1/8$	
F	$p=1/8$	

Už je jasné, že znak A má kódovou značku 1, znak E značku 2 a znak B značku 3. Číslice 0 je začátek kódových značek pro skupinu {C, D, F}. Za nulu na druhou pozici distribuujeme další znaky kódové abecedy

	$p=3/8$	0
A	$p=2/8$	1
E	$p=2/8$	2
B	$p=1/8$	3
C	$p=1/8$	00
D	$p=1/8$	01
F	$p=1/8$	02

Teď už má každé písmenu vstupní abecedy přiřazenou kódovou značku. Je vidět, že písmena s větší pravděpodobností A a E mají krátké kódové značky. I pro méně časté B zbyla krátká značka. Písmena C, D, F mají kódovou značku delší. Kód tedy často se vyskytující písmena kóduje úsporněji.

4. Je dána abeceda a pravděpodobnosti výskytu znaků v procentech:

{a, b, c, d, e, f, g, h, i, j}
 {20, 18, 15, 12, 10, 8, 7, 4, 4, 2}

Kódová abeceda je {0, 1, 2}. Najděte kód s minimální střední délkou značky.

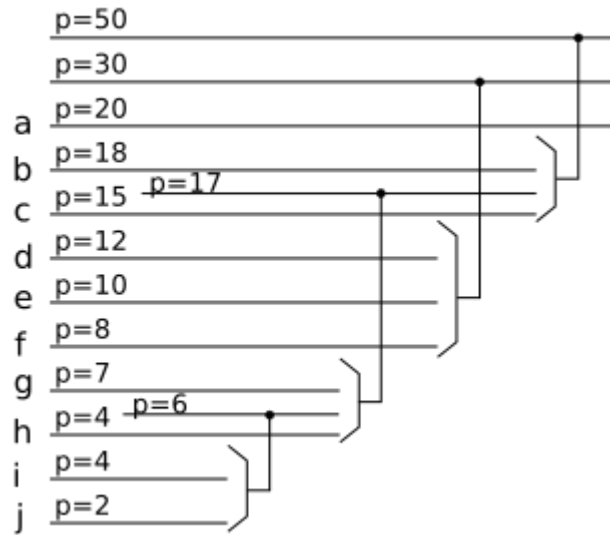
Postup řešení je obdobný, jako v předchozí úloze. Znaky už jsou seřazené. Rozdělíme je po dvou

a	<u>p=20</u>
b	<u>p=18</u>
c	<u>p=15</u>
d	<u>p=12</u>
e	<u>p=10</u>
f	<u>p=8</u>
g	<u>p=7</u>
h	<u>p=4</u>
i	<u>p=4</u>
j	<u>p=2</u>

Sloučíme poslední skupinu

a	<u>p=20</u>
b	<u>p=18</u>
c	<u>p=15</u>
d	<u>p=12</u>
e	<u>p=10</u>
f	<u>p=8</u>
g	<u>p=7</u>
h	<u>p=4</u> <u>p=6</u>
i	<u>p=4</u>
j	<u>p=2</u>

Dále pokračujeme ve slučování po třech



Distribuuje znaky kódové abecedy

