Aplikovaná informatika

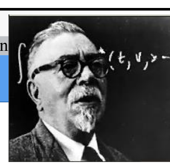
Podklady předmětu
Aplikovaná informatika
pro akademický rok 2013/2014
Radim Farana

2

Obsah

- Základní pojmy z Teorie informace,
  - jednotka informace,
  - informační obsah zprávy,
  - střední délka zprávy,
  - redundance.
- Přenosový řetězec.
- Základy pojmy z diskrétních kódů.
- Druhy kódů.
- Nejkratší kódy.
- Kódy konstantní změny (Grayovy kódy).

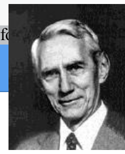
Kybernetika



Wiener, Norbert
* 26. 1. 1894 Columbia, Mo. USA
\+ 18. 3. 1964 Stockholm
http://en.wikipedia.org/wiki/Norbert_Wiener

- Wiener: Kybernetika je věda o řízení a sdělování v živých organismech a ve strojích.
- ale také: Kybernetika je věda o sběru, přenosu a zpracování informace.

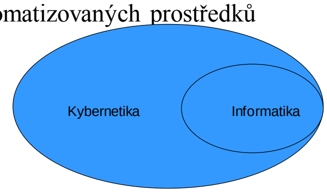
Informatika



Shannon, Claude Elwood

* 30. 4. 1916 Petoskey, Mich. USA
+ 24. 2. 2001 Medford, Mas. USA
http://www.ieee.org/web/aboutus/history_center/biography/shannon.html

- Informatika je věda o zpracování informace, zejména za pomoci automatizovaných prostředků



Informace

- Informací nazýváme abstraktní veličinu, která může být přechovávána v určitých objektech, předávána určitými objekty, zpracovávána v určitých objektech a použita k řízení určitých objektů. Jako objekt přitom chápeme živé organismy, technická zařízení nebo soustavy těchto prvků.
- také: Informace je sdělitelný poznatek, který má smysl a snižuje nejistotu.

Jednotka informace

- Jednotka informace je takové množství informace, které získáme potvrzením, že nastala jedna ze dvou stejně pravděpodobných možností.
- Označení: bit (binary digit) [b]

svítí : nesvítí
50 : 50

Násobky

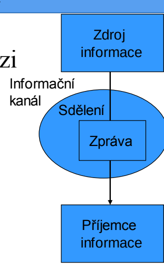
- Oproti desítkové soustavě jsou násobky odvozeny od binární soustavy

Násobek	Předpona	Symbol	Celý název	Odvozeno od
2^{10}	kibi	Ki	kilobinary: $(2^{10})^1$	kilo: $(10^3)^1$
2^{20}	mebi	Mi	megabinary: $(2^{10})^2$	mega: $(10^3)^2$
2^{30}	gibi	Gi	gigabinary: $(2^{10})^3$	giga: $(10^3)^3$
2^{40}	tebi	Ti	terabinary: $(2^{10})^4$	tera: $(10^3)^4$
2^{50}	pebi	Pi	petabinary: $(2^{10})^5$	peta: $(10^3)^5$
2^{60}	exbi	Ei	exabinary: $(2^{10})^6$	exa: $(10^3)^6$

Zavedla v roce 1998 celosvětová standardizační organizace IEC (International Electrotechnical Commission)

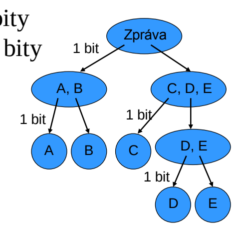
Informace, zpráva, sdělení

- Zprávu chápeme jako relaci mezi zdrojem a odběratelem, při které dochází k přenosu informace
- Sdělení je vhodným způsobem upravená zpráva, zejména pro potřeby přenosu.



Informační obsah zprávy

- Pravděpodobnost – informační obsah
- $P(x) = 0{,}5 \Rightarrow k(x) = 1$ bit
- $P(x) = 0{,}25 \Rightarrow k(x) = 2$ bity
- $P(x) = 0{,}125 \Rightarrow k(x) = 3$ bity
- $P(x) = 1/[2^{k(x)}]$
- $k(x) = -\log_2 P(x)$ [bit]



Entropie zdroje informace

- Shannon, Wiener: … informace představuje míru organizace, entropie míru neorganizovanosti …
- $H(z) = \sum P(i).k(i)$ [bit]

Zpráva	$P(i)$	$k(i)$	$P(i).k(i)$
A	0,250	2	0,500
B	0,250	2	0,500
C	0,250	2	0,500
D	0,125	3	0,375
E	0,125	3	0,375
		$H(z)$ =	2,250

Pro zdroj se shodnou pravděpodobností všech zpráv:

$H(z) = n.[(1/n).\log_2(n)]$

$H(z) = \log_2(n)$

pro $n = 5$

$H(z) = 2{,}322$

Jednotky

- Volba základu logaritmu je tedy pouze otázkou volby konstanty | měrné jednotky (viz norma IEC/ISO 80000, Díl 13).
  - U binárních logaritmů je jednotkou shannon [Sh].
  - U přirozených logaritmů jednotka nat [nat].
  - U dekadických logaritmů hartley, [Hart]

  - 1 Sh ≈ 0,693 nat ≈ 0,301 Hart
  - 1 nat ≈ 1,433 Sh ≈ 0,434 Hart
  - 1 Hart ≈ 3,322 Sh ≈ 2,303 nat

Informace, data, znalost

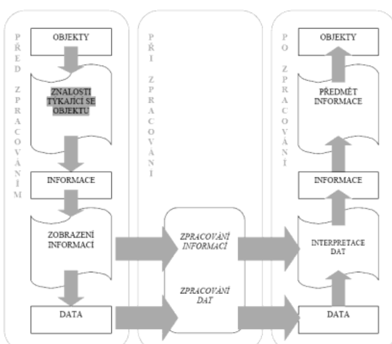
- Data jsou vyjádření skutečností formálním způsobem tak, aby je bylo možné přenášet nebo zpracovat .
- Znalost je to, co jednotlivec vlastní (ví) po osvojení dat a po jejich začlenění do souvislostí. Je výsledkem poznávacího procesu za předpokladu uvědomělé činnosti.

ČSN ISO/IEC 2382-1:1998

Informační technologie - Slovník - Část 1: Základní termíny

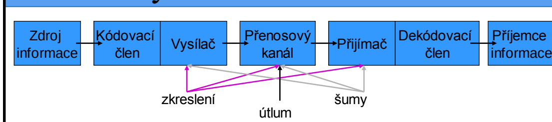
- Informace: Poznatek (znalost) týkající se jakýchkoliv objektů, např. faktů, událostí, věcí, procesů, myšlenek nebo pojmů, které mají v daném kontextu specifický význam.
- Data: Opakovaně interpretovatelná formalizovaná podoba informace vhodná pro komunikaci, vyhodnocování nebo zpracování.
- Zpracování informací: Systematické provádění operací s informacemi, zahrnující zpracování dat a případně i datovou komunikaci a automatizaci kancelářských prací.
- Datová komunikace: Přesun dat mezi funkčními jednotkami podle souboru pravidel řídících přenos dat.
- Funkční jednotka: Entita technického nebo programového vybavení, nebo obou, schopná vyhovět danému účelu.
- Automatizace kancelářských prací: Integrace kancelářských prací pomocí systému zpracování dat.
- Systém zpracování dat: Jeden nebo více počítačů, periferních zařízení a programů použitých pro zpracování dat.
- Zpracování dat: Systematické provádění operací s daty např. aritmetické nebo logické operace s daty nebo třídění dat, sestavování nebo kompilace programů a dále operace s textem např. úprava, třídění, slučování, ukládání, vyhledávání, zobrazování nebo tisk

ČSN



http://www.fi.muni.cz/~xkubasek/eis/01.html

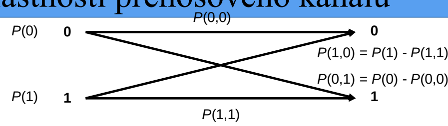
Přenosový řetězec



- Přenosový kanál
  - spojitý (analogový),
  - diskrétní (v úrovni) neboli kvantovaný,
  - číslicový (diskrétní v čase).
- Rychlost přenosu informace

  $v_p = k(x) / t \quad [\text{bit.s}^{-1}]$

Vlastnosti přenosového kanálu



- Přenosový kanál
  - bezšumový, P(0,1) = P(1,0) = 0
  - šumový, podle výskytu chyb:
    - bezpamětový (chyby jsou náhodné),
    - paměťový (chyby jsou shlukové).
  - šumový, podle vlivu šumu:
    - symetrický, P(0,1) = P(1, 0),
    - nesymetrický.

Kód

- Popis přiřazení kódových slov jednotlivým zprávám (kódová kniha).
- Kódové slovo je posloupnost znaků použité abecedy.
- Abeceda je množina znaků (binární abeceda $Z_2 = \{0, 1\}$)
- Minimální délka kódového slova:
  $N^*(x) = -\log_2(P(x))$ [bit]

Charakteristiky kódu

- Střední délka kódového slova:
  $L = \Sigma\, P(i).N(i)$ [bit]
- Redundance
  $R = L - H$ (protože $L > H$, neboť $N(i) \geq N^*(i)$)

Zpráva	$P(i)$	$k(i)$	$P(i).k(i)$	$N(i)$	$P(i).N(i)$		
A	0,250	2	0,500	4	1,000		
B	0,250	2	0,500	4	1,000		
C	0,250	2	0,500	5	1,250		
D	0,125	3	0,375	6	0,750		
E	0,125	3	0,375	7	0,875		
		$H(z)$ =	2,250	L =	4,875	R =	2,625

Vlastnosti kódu

- prosté kódování: různým zprávám odpovídají různá kódová slova,
- jednoznačná dekódovatelnost: ze znalosti zakódované zprávy lze jednoznačně určit zprávu zdrojovou,
- Kód K : A → B musí být prostým zobrazením.

Problém dekódování

Zpráva 

Kód A	0	01	001	111

Posloupnost zpráv (kódových slov): 00100101111 nelze jednoznačně dekódovat

Kód B	0	01	011	111

Posloupnost zpráv (kódových slov): 00101101111 lze jednoznačně dekódovat?

Ano, ale jen „odzadu“, po přijetí celé posloupnosti zpráv.

Kód C	0	10	110	111

Posloupnost zpráv (kódových slov): 01011010111 můžeme dekódovat on-line.

Důvod? Žádné kódové slovo není začátkem jiného kódového slova (prefixem).

Druhy kódů

- Prefixový kód je prosté kódování u kterého žádné kódové slovo není začátkem jiného kódového slova.
- Blokový kód je prosté kódování u kterého mají všechna kódová slova stejnou délku (počet znaků). Protože musí být prostým zobrazením, je nutně také prefixovým kódem.

Použití kódů

- Nejkratší (optimální) kódy $R \rightarrow 0$, $L \rightarrow \min$,
- Bezpečnostní kódy
  - detekční kódy (odhalují chyby),
  - samoopravné kódy (opravují chyby),
- Speciální kódy
  - kódy konstantní změny (Grayův kód),
  - čárové kódy,
  - alfanumerické kódy,
  - číselné kódy (datové formáty), …

Kraftova nerovnost

- Prefixový kód sestrojený nad *n*-prvkovou kódovou abecedou s délkami kódových slov $d_1, d_2, \ldots, d_r$ existuje právě tehdy, když platí ***Kraftova nerovnost***, tj.

$$n^{-d_1} + n^{-d_2} + \cdots\cdots + n^{-d_r} \leq 1$$

Mc Millanova věta

- Každé jednoznačně dekódovatelné kódování splňuje Kraftovu nerovnost.
- Z těchto dvou vět vyplývá, že každé jednoznačně dekódovatelné kódování je prefixové, ale pokud není, existuje jiné kódování nad stejnou kódovou abecedou s danými délkami kódových slov, které již prefixové bude.

Příklad

- Máme za úkol sestavit binární prefixový kód číslic 0, 1, ..., 9 s délkami kódových slov: $d_1 = d_2 = 2, d_3 = \cdots = d_7 = 3, d_8 = d_9 = 4$
  Kraftova nerovnost je rovna

$$2 \cdot \frac{1}{4} + 6 \cdot \frac{1}{8} + 2 \cdot \frac{1}{16} = \frac{11}{8} > 1$$

prefixový kód s těmito délkami slov neexistuje.

Příklad

- Zvolíme-li $d_1 = d_2 = 2, d_3 = \cdots = d_7 = d_8 = d_9 = 4$ potom Kraftova nerovnost je

$$2 \cdot \frac{1}{4} + 8 \cdot \frac{1}{16} = 1$$

a tedy takový prefixový kód existuje. Může vypadat například takto

Číslice	Kódové slovo
0	00
1	01
2	1000
3	1001
4	1010
5	1011
6	1100
7	1101
8	1110
9	1111

Nejkratší kódy

- Pokud má $R \to 0$, neboli $L \to \min$, pak
- $N(i) \to N^*(i)$ pro $i = 1, 2, \dots n$.
- Hledáme vhodný algoritmus konstrukce takového kódu:
  - Huffmanův kód (1952),
  - Shannonův kód.

Algoritmy se liší, stejně tak i dosažené výsledky, Huffmanův kód se snáze algoritmizuje a tedy také realizuje

Huffman, David A.

* 1925, USA

\+ 7. 10. 1999 California, USA

http://www.ucsc.edu/currents/99-00/10-11/huffman.html

Huffmanův kód

Triviální případ

Zpráva	*P*(*i*)	kód
1	>	0
2	<	1

Redukovaná abeceda

Zpráva	*P*(*i*)	redukce	kód	expanze
1	»	1	0	0
2	>	2,3	1	10
3	<			11

Příklad

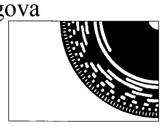
Zpráva	A	B	C	D	E
P(*i*)	0,4	0,3	0,1	0,1	0,1
				0	1
1.redukce	0,4	0,3	0,2	0,1	
			0	1	
2.redukce	0,4	0,3	0,3		
		0	1		
3.redukce	0,6	0,4			
znaky	0	1			
kód	1	00	011	0100	0101

Postup:
- seřazení podle pravděpodobnosti,
- postupná redukce a oprava pořadí,
- přiřazení znaků 0, 1 a zpětná expanze.

Problémy:
- definice pořadí zpráv pro stejnou *P*(*i*),
- zařazení skupin se stejnou *P*(*i*),
- pořadí přiřazení znaků 0, 1.

Kódy konstantní změny

- Každá dvě sousední kódová slova se liší o konstantní počet bitů (Hammingova vzdálenost je konstantní).
- Grayovy kódy pro $d = 1$ známé z absolutních snímačů polohy.
- Představuje Hamiltonovu kružnici na *n*-dimenzionální kostce.



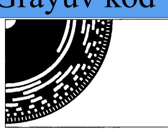
Gray, Frank

U.S. patent 2,632,058 z r. 1953 (Bell Laboratories)

Baudot, Jean- Maurice Émile

* 11. 9. 1845, + 28. 3. 1903 použil v telegrafu v roce 1878

Grayův kód



dekadické číslo	binární číslo	binární podoba					Gray Code				binární vyjádření
0	0000										0000
1	0001										0001
2	0010										0011
3	0011										0010
4	0100										0110
5	0101										0111
6	0110										0101
7	0111										0100
8	1000										1100
9	1001										1101
10	1010										1111
11	1011										1110
12	1100										1010
13	1101										1011
14	1110										1001
15	1111										1000
	$a_3a_2a_1a_0$	a_3	a_2	a_1	a_0		g_3	g_2	g_1	g_0	$g_3g_2g_1g_0$

Karnaughova mapa pro bit g_0

a_3a_2 \\ a_1a_0	00	01	11	10
00	0	1	0	1
01	0	1	0	1
11	0	1	0	1
10	0	1	0	1

- Realizace
- Převod binárního čísla do Grayova kódu a zpět

$g_0 = b_1b_0$ OR $b_1b_0 = b_1$ XOR b_0

$g_1 = b_2$ XOR b_1

$g_2 = b_3$ XOR b_2

$g_3 = b_3$

Inverzní převod

$b_3 = g_3$

$b_2 = b_3$ XOR g_2

$b_1 = b_2$ XOR g_1

$b_0 = b_1$ XOR g_0

Příklad

- Máme za úkol sestavit kódovou knihu pro převod pětibitového binárního kódu na Grayův kód.

Řešení

- Protože nemáme k dispozici předpis pro převod mezi binárním a Grayovým kódem znázorníme graficky rozložení jedniček binárního kódu a přiřadíme jim obdobně jedničky Grayova kódu.

Binární kód

dekadické číslo	binární číslo	binární podoba						Grayův kód					binární podoba
		a_4	a_3	a_2	a_1	a_0		g_4	g_3	g_2	g_1	g_0	
0	00000												
1	00001												
2	00010												
3	00011												
4	00100												
5	00101												
6	00110												
7	00111												
8	01000												
9	01001												
10	01010												
11	01011												
12	01100												
13	01101												
14	01110												
15	01111												
16	10000												
17	10001												
18	10010												
19	10011												
20	10100												
21	10101												
22	10110												
23	10111												
24	11000												
25	11001												
26	11010												
27	11011												
28	11100												
29	11101												
30	11110												
31	11111												

Nejvyšší bit

- Víme, že významově nejvyšší bit je u poloviny slov nulový a u poloviny jedničkový a to shodně s binárním kódem. Přiřadíme tedy hodnoty.

Nejvyšší bit

dekadické číslo	binární číslo	binární podoba						Grayův kód					binární podoba
		a_4	a_3	a_2	a_1	a_0		g_4	g_3	g_2	g_1	g_0	
0	00000												
1	00001					■							
2	00010				■								
3	00011				■	■							
4	00100			■									
5	00101			■		■							
6	00110			■	■								
7	00111			■	■	■							
8	01000		■										
9	01001		■			■							
10	01010		■		■								
11	01011		■		■	■							
12	01100		■	■									
13	01101		■	■		■							
14	01110		■	■	■								
15	01111		■	■	■	■							
16	10000	■						■					
17	10001	■				■		■					
18	10010	■			■			■					
19	10011	■			■	■		■					
20	10100	■		■				■					
21	10101	■		■		■		■					
22	10110	■		■	■			■					
23	10111	■		■	■	■		■					
24	11000	■	■					■					
25	11001	■	■			■		■					
26	11010	■	■		■			■					
27	11011	■	■		■	■		■					
28	11100	■	■	■				■					
29	11101	■	■	■		■		■					
30	11110	■	■	■	■			■					
31	11111	■	■	■	■	■		■					

Druhý bit

- Významově druhý bit bude o souvislé skupiny poloviny slov jedničkový a to tak, aby se jedničky s vyšším bitem shodovaly u poloviny kódových slov. Tak bude zajištěn minimální počet změn u tohoto bitu. Přiřadíme tedy jedničky do tohoto bitu.

Druhý bit

dekadické číslo	binární číslo	binární podoba						Grayův kód					binární podoba
		a_4	a_3	a_2	a_1	a_0		g_4	g_3	g_2	g_1	g_0	
0	00000												
1	00001												
2	00010												
3	00011												
4	00100												
5	00101												
6	00110												
7	00111												
8	01000												
9	01001												
10	01010												
11	01011												
12	01100												
13	01101												
14	01110												
15	01111												
16	10000												
17	10001												
18	10010												
19	10011												
20	10100												
21	10101												
22	10110												
23	10111												
24	11000												
25	11001												
26	11010												
27	11011												
28	11100												
29	11101												
30	11110												
31	11111												

Třetí bit

- Podobně budou jedničky u třetího bitu tvořit dvě skupiny slov opět posunuté o polovinu délky skupiny.

Třetí bit

dekadické číslo	binární číslo	binární podoba						Grayův kód					binární podoba
		a_4	a_3	a_2	a_1	a_0		g_4	g_3	g_2	g_1	g_0	
0	00000												
1	00001												
2	00010												
3	00011												
4	00100												
5	00101												
6	00110												
7	00111												
8	01000												
9	01001												
10	01010												
11	01011												
12	01100												
13	01101												
14	01110												
15	01111												
16	10000												
17	10001												
18	10010												
19	10011												
20	10100												
21	10101												
22	10110												
23	10111												
24	11000												
25	11001												
26	11010												
27	11011												
28	11100												
29	11101												
30	11110												
31	11111												

Čtvrtý bit

- Čtvrtý bit bude obsazen čtyřmi sekvencemi jedniček délky čtyři znaky, opět posunutých o polovinu délky, tedy dva znaky.

Čtvrtý bit

dekadické číslo	binární číslo	binární podoba						Grayův kód					binární podoba
		a_4	a_3	a_2	a_1	a_0		g_4	g_3	g_2	g_1	g_0	
0	00000												
1	00001					■							
2	00010				■						■		
3	00011				■	■					■		
4	00100			■						■	■		
5	00101			■		■				■	■		
6	00110			■	■					■			
7	00111			■	■	■				■			
8	01000		■						■	■			
9	01001		■			■			■	■			
10	01010		■		■				■	■	■		
11	01011		■		■	■			■	■	■		
12	01100		■	■					■		■		
13	01101		■	■		■			■		■		
14	01110		■	■	■				■				
15	01111		■	■	■	■			■				
16	10000	■						■	■				
17	10001	■				■		■	■				
18	10010	■			■			■	■		■		
19	10011	■			■	■		■	■		■		
20	10100	■		■				■	■	■	■		
21	10101	■		■		■		■	■	■	■		
22	10110	■		■	■			■	■	■			
23	10111	■		■	■	■		■	■	■			
24	11000	■	■					■		■			
25	11001	■	■			■		■		■			
26	11010	■	■		■			■		■	■		
27	11011	■	■		■	■		■		■	■		
28	11100	■	■	■				■			■		
29	11101	■	■	■		■		■			■		
30	11110	■	■	■	■			■					
31	11111	■	■	■	■	■		■					

Pátý bit

- Poslední bit bude obsazen osmi skupinami jedniček délky dva znaky, opět posunutými oproti vyššímu bitu o polovinu délky, tedy jeden bit.

Aplikovaná informatika 43

Pátý bit

dekadické číslo	binární číslo	binární podoba						Grayův kód					binární podoba
		a_4	a_3	a_2	a_1	a_0		g_4	g_3	g_2	g_1	g_0	
0	00000												
1	00001												
2	00010												
3	00011												
4	00100												
5	00101												
6	00110												
7	00111												
8	01000												
9	01001												
10	01010												
11	01011												
12	01100												
13	01101												
14	01110												
15	01111												
16	10000												
17	10001												
18	10010												
19	10011												
20	10100												
21	10101												
22	10110												
23	10111												
24	11000												
25	11001												
26	11010												
27	11011												
28	11100												
29	11101												
30	11110												
31	11111												

Aplikovaná informatika 44

Kódová slova

- Nyní již zbývá pouze zápis kódových slov v obvyklé podobě binárních znaků.

Aplikovaná informatika 45

Kódová slova

dekadické číslo	binární číslo	binární podoba						Grayův kód					binární podoba
		a_4	a_3	a_2	a_1	a_0		g_4	g_3	g_2	g_1	g_0	
0	00000												00000
1	00001												00001
2	00010												00011
3	00011												00010
4	00100												00110
5	00101												00111
6	00110												00101
7	00111												00100
8	01000												01100
9	01001												01101
10	01010												01111
11	01011												01110
12	01100												01010
13	01101												01011
14	01110												01001
15	01111												01000
16	10000												11000
17	10001												11001
18	10010												11011
19	10011												11010
20	10100												11110
21	10101												11111
22	10110												11101
23	10111												11100
24	11000												10100
25	11001												10101
26	11010												10111
27	11011												10110
28	11100												10010
29	11101												10011
30	11110												10001
31	11111												10000

Závěr

- Tímto postupem můžeme sestavit Grayův kód libovolné délky.