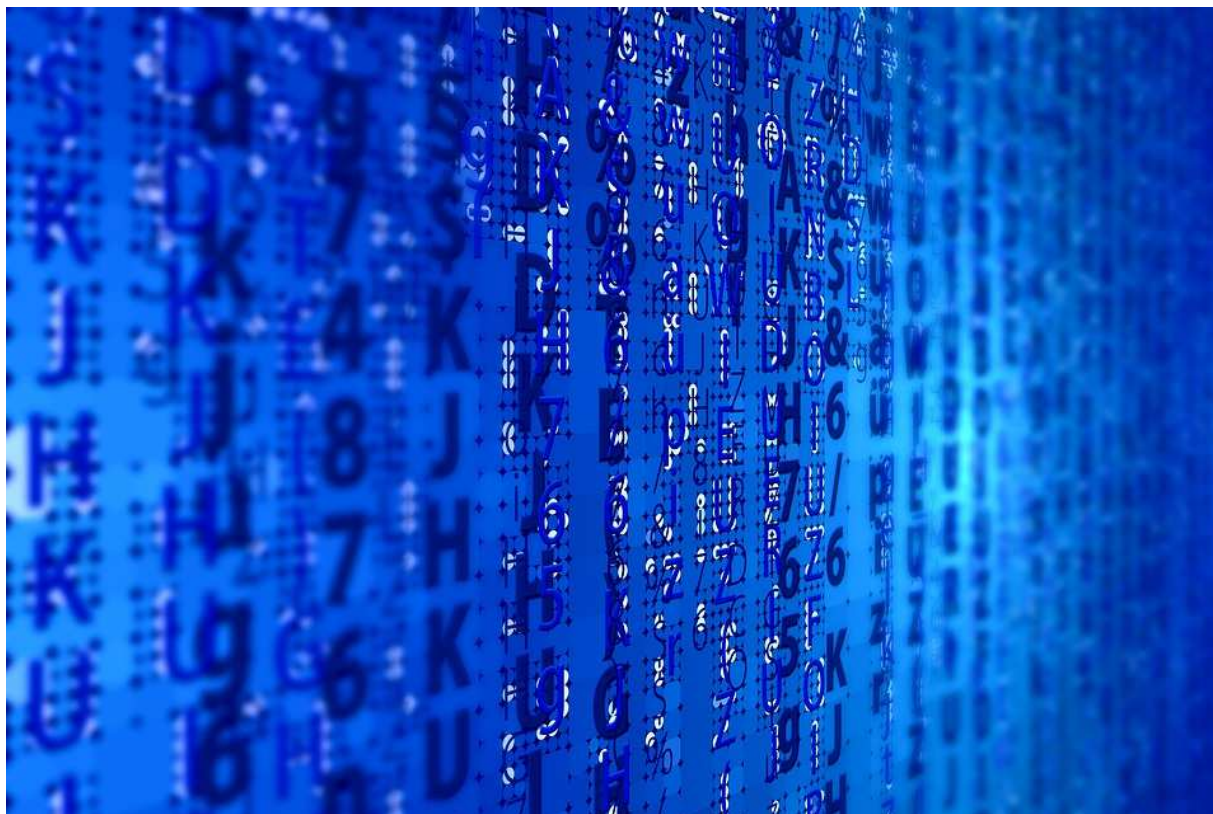


Kódování dat

doc. Ing. Josef Chaloupka, Ph.D.



Kódování dat

- 1) **Uložení dat** – ASCII, uložení čísel, dvojková – šestnáctková soustava

- 2) **Kompresce dat**

-ztrátová

MP3 - psychoakustický model ucha, JPG, MPEG4...

kompresní poměr (1:10) původní data (10MB)
komprimovaná data (1MB)
úspora datového toku 90% (např. MP3)

kompresce binárního obrazu – značkování – dvě značky

uint8, obraz 80 x 100 obrazových bodů

původní data 8000B (bit)
komprimovaná data 6B
úspora datového toku 99,93% (99,4%)

-bezztrátová

Huffman, aritmetické kódování, RLE, LZW...

reálná data – úspora datového toku reálná data max. 40-50%

- 3) **Šifrování dat** - kryptografie
- 4) **Samoopravné kódy** – najít chybu, opravit

Kódování čísel v PC

- Celá čísla (8, 16, 32, 64... bit.) – zápis desítková, dvojková, šestnáctková soustava

10^3 kilo (k), 10^6 mega (M), 10^9 giga (G), 10^{12} tera (T), 10^{15} peta (P), 10^{18} exa (E), 10^{21} zetta (Z), 10^{24} yotta (Y)

disk 100GB – $100 \cdot 2^{30} \text{B} = 107374182400 \text{B}$, $100000000000 \text{B} - 93\%$

- Záporná čísla - dvojkový doplněk

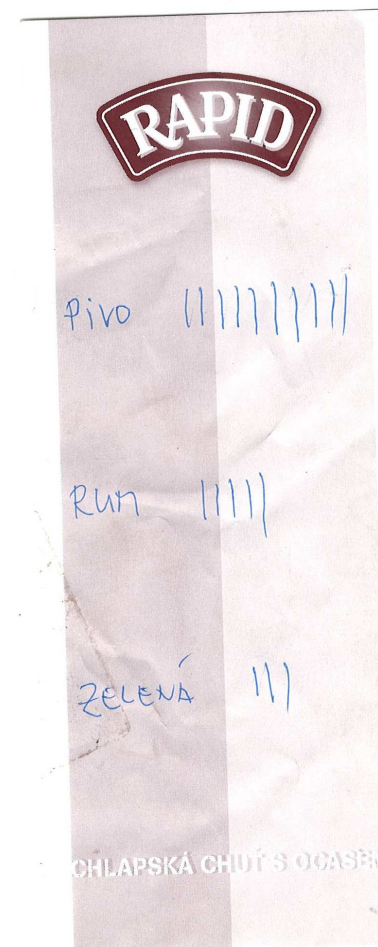
- Desetinná čísla

pevná x plovoucí řádová čárka

$$3700 = 3,7 \cdot 10^3$$

$$0,037 = 3,7 \cdot 10^{-2}$$

$x = (-1)^S M_x \cdot B^{E_x}$ M_x : mantisa, E_x : exponent, B : základ 2 nebo 10
standard IEEE 754 (1985) – $B = 2$, IEEE 854 – $B = 2$ nebo 10
přesnost (platnost): single: 7 desetinných míst
 double: 16 desetinných míst



Římské číslice

● I = 1, V = 5, X = 10, L = 50, C = 100, D = 500, M = 1000

● VI = 6, CX = 110

● IV = 4 (Římané IIII), XC = 90

● Římané 999: DCCCCLXXXXVIII
CMXCIX, IM nelze



MDCCCLXVII

1: I	26: XXVI	51: LI	76: LXXVI
2: II	27: XXVII	52: LII	77: LXXVII
3: III	28: XXVIII	53: LIII	78: LXXVIII
4: IV	29: XXIX	54: LIV	79: LXXIX
5: V	30: XXX	55: LV	80: LXXX
6: VI	31: XXXI	56: LVI	81: LXXXI
7: VII	32: XXXII	57: LVII	82: LXXXII
8: VIII	33: XXXIII	58: LVIII	83: LXXXIII
9: IX	34: XXXIV	59: LIX	84: LXXXIV
10: X	35: XXXV	60: LX	85: LXXXV
11: XI	36: XXXVI	61: LXI	86: LXXXVI
12: XII	37: XXXVII	62: LXII	87: LXXXVII
13: XIII	38: XXXVIII	63: LXIII	88: LXXXVIII
14: XIV	39: XXXIX	64: LXIV	89: LXXXIX
15: XV	40: XL	65: LXV	90: XC
16: XVI	41: XLI	66: LXVI	91: XCI
17: XVII	42: XLII	67: LXVII	92: XCII
18: XVIII	43: XLIII	68: LXVIII	93: XCIII
19: XIX	44: XLIV	69: LXIX	94: XCIV
20: XX	45: XLV	70: LXX	95: XCV
21: XXI	46: XLVI	71: LXXI	96: XCVI
22: XXII	47: XLVII	72: LXXII	97: XCVII
23: XXIII	48: XLVIII	73: LXXIII	98: XCVIII
24: XXIV	49: XLIX	74: LXXIV	99: XCIX
25: XXV	50: L	75: LXXV	100: C

Arabské číslice

- Indo – (0 až 9) arabské, cca 300 let .př.n.l., poziční systém
- 0 – Peršané cca 6 století
- Desetinná tečka cca 10. století Arábie
- Od 10. století v Evropě (Španělsku), Fibonacci (kniha Liber Abaci, 1202)

BRAHMI		—	=	≡	+	μ	ε	γ	ς	7
HINDU	०	१	२	३	४	५	६	७	८	९
ARABIC	•	١	٢	٣	٤	٥	٦	٧	٨	٩
MEDIEVAL	0	1	2	3	4	5	6	7	8	9
MODERN	0	1	2	3	4	5	6	7	8	9

Kódování znaků v PC - ASCII

- **Kódování češtiny:** ISO 8859-2 (latin2, 1987), windows (CP)1250 (odlišné Š, Ť, Ž)
ve win 3.1, bratři Kamenický – MS DOS, 128 až 171 české a slovenské znaky

000	00		043	2B	+	086	56	U	129	81	ü	172	AC	¼	215	D7	
001	01	☐	044	2C	,	087	57	W	130	82	é	173	AD	½	216	D8	¼
002	02	☐	045	2D	-	088	58	X	131	83	â	174	AE	¾	217	D9	½
003	03	♥	046	2E	.	089	59	Y	132	84	ä	175	AF	»	218	DA	Γ
004	04	♦	047	2F	/	090	5A	Z	133	85	à	176	B0	▨	219	DB	■
005	05	♣	048	30	0	091	5B	[134	86	á	177	B1	▨	220	DC	■
006	06	♠	049	31	1	092	5C	\	135	87	ç	178	B2	▨	221	DD	■
007	07	•	050	32	2	093	5D]	136	88	ê	179	B3		222	DE	■
008	08	☐	051	33	3	094	5E	^	137	89	ë	180	B4	└	223	DF	■
009	09		052	34	4	095	5F	`	138	8A	è	181	B5	└	224	E0	α
010	0A		053	35	5	096	60		139	8B	í	182	B6		225	E1	β
011	0B	ø	054	36	6	097	61	a	140	8C	î	183	B7	π	226	E2	Γ
012	0C	♀	055	37	7	098	62	b	141	8D	ï	184	B8	π	227	E3	Π
013	0D		056	38	8	099	63	c	142	8E	ñ	185	B9		228	E4	Σ
014	0E	⌘	057	39	9	100	64	d	143	8F	ä	186	BA		229	E5	σ
015	0F	*	058	3A	:	101	65	e	144	90	é	187	BB		230	E6	μ
016	10	►	059	3B	;	102	66	f	145	91	æ	188	BC		231	E7	τ
017	11	◄	060	3C	<	103	67	g	146	92	æ	189	BD		232	E8	⊗
018	12	‡	061	3D	=	104	68	h	147	93	ô	190	BE		233	E9	θ
019	13	!!	062	3E	>	105	69	i	148	94	ö	191	BF		234	EA	Ω
020	14	¶	063	3F	?	106	6A	j	149	95	ò	192	C0		235	EB	δ
021	15	§	064	40	@	107	6B	k	150	96	û	193	C1		236	EC	ω
022	16	—	065	41	A	108	6C	l	151	97	ù	194	C2		237	ED	œ
023	17	±	066	42	B	109	6D	m	152	98	ÿ	195	C3		238	EE	€
024	18	↑	067	43	C	110	6E	n	153	99	õ	196	C4		239	EF	∅
025	19	↓	068	44	D	111	6F	o	154	9A	ü	197	C5		240	F0	≡
026	1A		069	45	E	112	70	p	155	9B	ç	198	C6		241	F1	±
027	1B	←	070	46	F	113	71	q	156	9C	£	199	C7		242	F2	≥
028	1C	↳	071	47	G	114	72	r	157	9D	¥	200	C8		243	F3	≤
029	1D	↔	072	48	H	115	73	s	158	9E	℞	201	C9		244	F4	∫
030	1E	▲	073	49	I	116	74	t	159	9F	ƒ	202	CA		245	F5	∫
031	1F	▼	074	4A	J	117	75	u	160	A0	á	203	CB		246	F6	÷
032	20		075	4B	K	118	76	v	161	A1	í	204	CC		247	F7	≈
033	21	†	076	4C	L	119	77	w	162	A2	ó	205	CD		248	F8	°
034	22	"	077	4D	M	120	78	x	163	A3	ú	206	CE		249	F9	·
035	23	#	078	4E	N	121	79	y	164	A4	ñ	207	CF		250	FA	·
036	24	\$	079	4F	O	122	7A	z	165	A5	ň	208	D0		251	FB	√
037	25	%	080	50	P	123	7B	<	166	A6	ä	209	D1		252	FC	∞
038	26	&	081	51	Q	124	7C	!	167	A7	é	210	D2		253	FD	∞
039	27	'	082	52	R	125	7D	>	168	A8	è	211	D3		254	FE	■
040	28	<	083	53	S	126	7E	~	169	A9	í	212	D4		255	FF	
041	29	>	084	54	T	127	7F	Δ	170	AA	ı	213	D5				
042	2A	*	085	55	U	128	80	ç	171	AB	½	214	D6				

Kódování znaků v PC - Unicode

- Unicode (od 1993), UTF8: 1B pro ASCII, 2 až 4B pro jiné znaky, UTF16: 16bit., UTF32, výhoda: jednoznačnost, nevýhoda: hodně dat, český text 4x více než UTF8

Α α Β β Γ γ Δ δ
Ε ε Ζ ζ Η η Θ θ
Ι ι Κ κ Λ λ Μ μ
Ν ν Ξ ξ Ο ο Π π
Ρ ρ Σ σ ς Τ τ Υ υ
Φ φ Χ χ Ψ ψ Ω ω

řecká abeceda

द बी सटग जौन .
ही टोलड हिज मम .
शी टोलड हिम नौट
टु कराइ .

indické (hindi) písmo

ا ب ت ث ج ح خ
ب ا
د ذ ر ز س ش ص
ض ط ظ ع غ ف ق
ك ل م ن ه و ي
b 'a
z r d d
q f ġ ' z t d
y w h n m l k

arabské písmo

א ב ג ד ה ו
Vav Hey Dalet Gimmel Bet Aleph
ז ח ט י כ
Kaf Yod Tet Chet Zayin
ל מ נ ס ע פ
Pey Ayin Samech Nun Mem Lamed
צ ק ר ש ת
Tav Shin Resh Qof Tsade

hebrejské písmo

Kódování znaků v PC - Unicode

<u>n</u>	<u>wa</u>	<u>ra</u>	<u>ya</u>	<u>ma</u>	<u>ha</u>	<u>na</u>	<u>ta</u>	<u>sa</u>	<u>ka</u>	<u>a</u>	
ん	わ	ら	や	ま	は	な	た	さ	か	あ	a
<u>n</u>	<u>wa</u>	<u>ra</u>	<u>ya</u>	<u>ma</u>	<u>ha</u>	<u>na</u>	<u>ta</u>	<u>sa</u>	<u>ka</u>	<u>a</u>	
		り		み	ひ	に	ち	し	き	い	i
		<u>ri</u>		<u>mi</u>	<u>hi</u>	<u>ni</u>	<u>chi</u>	<u>shi</u>	<u>ki</u>	<u>i</u>	
		る	ゆ	む	ふ	ぬ	つ	す	く	う	u
		<u>ru</u>	<u>yu</u>	<u>mu</u>	<u>fu</u>	<u>nu</u>	<u>tsu</u>	<u>su</u>	<u>ku</u>	<u>u</u>	
		れ		め	へ	ね	て	せ	け	え	e
		<u>re</u>		<u>me</u>	<u>he</u>	<u>ne</u>	<u>te</u>	<u>se</u>	<u>ke</u>	<u>e</u>	
	を	ろ	よ	も	ほ	の	と	そ	こ	お	o
	<u>o*</u>	<u>ro</u>	<u>yo</u>	<u>mo</u>	<u>ho</u>	<u>no</u>	<u>to</u>	<u>so</u>	<u>ko</u>	<u>o</u>	

Japonština (80 000 znaků, UTF32)

[illegible]

korejština

橄鍵渥王湖陸笑習豹
 光獅漉少蜥寬男楓
 鏡靄謙猿母山婚鼻新
 裸近甥北今梟痛蹊民
 勘者繪豚姬純毒兔力
 詩苓雨霓胤寔理怡憶
 憇還謎冒川砂岩悲犧

Čínština (UTF32)

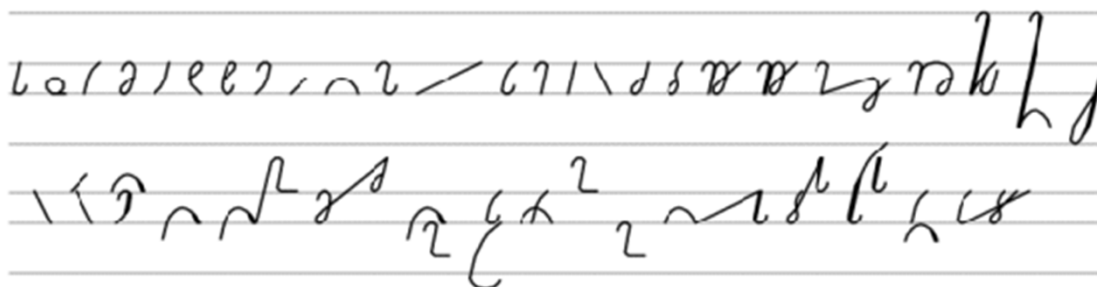
Vietnamese alphabet and pronunciation

A a	Ă ă	Â â	B b	C c	D d	Đ đ	E e	Ê ê
a	á	â	bé/bò	xé/cò	dé/dò	đé/dò	e	ê
[a]	[a/â:]	[â]	[b/ʔb]	[k]	[z]	[d/ʔd]	[e]	[e]
G g	H h	I i	K k	L l	M m	N n	O o	Ô ô
gè/gò	hát	ngán	ca	(e)-jò	(em)-mò	(en)-nò	o	ô
[ʒ/ʔ]	[h]	[i]	[k]	[l]	[m]	[n]	[o]	[o]
Œ œ	P p	Q q	R r	S s	T t	U u	Ú ú	V v
σ	pé/pò	cu/quy	(e)-nò	és-sí/sò	té/tò	u	ur	vé/vò
[œ]	[p]	[k]	[z]	[s]	[t]	[u]	[v]	[v]
X x	Y y							
ich-xò/xò	i dâi/cò-rét							
[x-ç]	[i:]							

vietnamština

Multimédia – komprese textu

- Těsnopis – stenografie
grafický, strojový (akordy - stisk více kláves najednou)
běžné písmo – 20-40 slov za minutu, těsnopis 80-100 slov za minutu (rekord 200)
průměrná řeč cca 70 slov za minutu



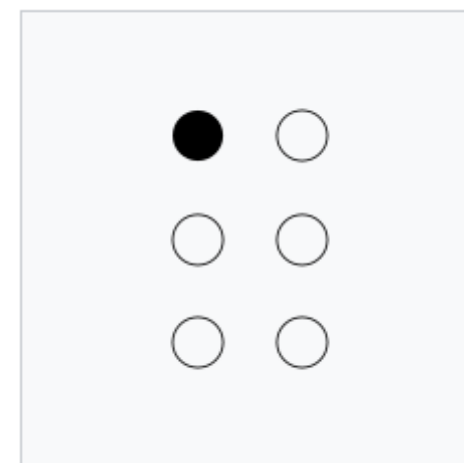
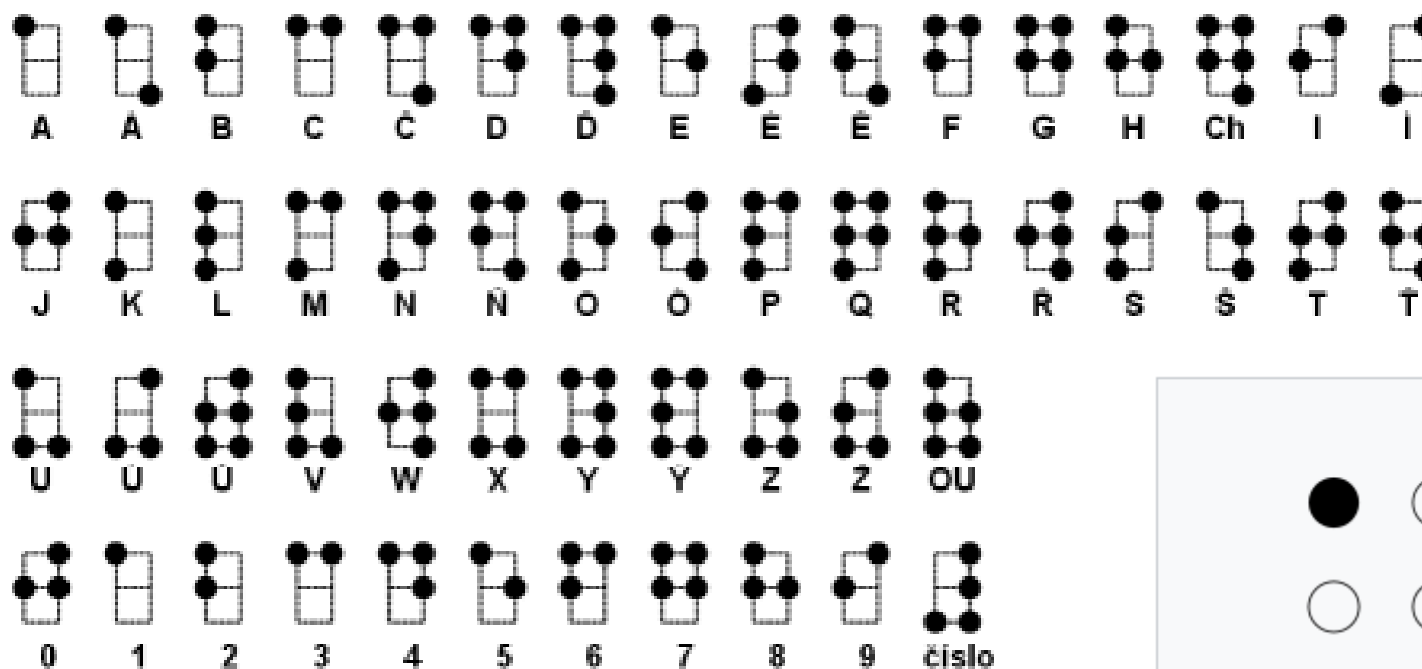
- Zkratky: atd. (a tak dále), ssp (se soudružským pozdravem)
- Náhrada textu opět, zpět – o5, z5
nevhodné: stojí, stopa – 100jí, 100pa

Náhrada slov znakem, který se v textu nevyskytuje
Kuli kuli kuli (Kuliové kovali kouli) # # #

- Ztrátové komprese nejsou vhodné
- Statistický přístup, slovníkový: LZW

Braillovo písmo

- Optimalizované pro čtení hmatem, Louis Braille (1825), úprava vojenského písma pro čtení ve tmě



Morseova abeceda

- Samuela Morse (asistent Alfred Vail) 1844, 1918 mezinárodní abeceda, 60 až 250 znaků za minutu x SMS, 3 znaky

A	•-	J	•---	T	-	0	-----
B	-•••	K	-•-	U	••-	1	•-----
C	-•-•	L	•-••	V	•••-	2	••-----
D	-••	M	--	W	•--	3	•••---
E	•	N	-•	X	-••-	4	••••-
F	••-•	O	---	Y	-•--	5	•••••
G	--•	P	•--•	Z	--••	6	-••••
H	••••	Q	--•-			7	--•••
CH	----	R	•-•			8	---••
I	••	S	•••			9	-----•

- @ 2003 .--.-.
- Není prefix – Huffmanův kód: A (0), R (10), B (111), K (1101), D (1100)

Čárový kód

- Šířka čar a mezer, 1949 (1952 patent), 1976 EAN

- EAN (European Article Number), **EAN-13**

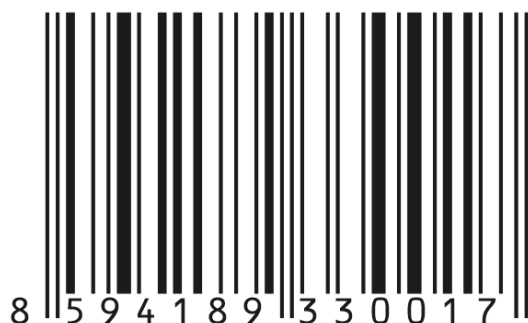
- 2 okrajové znaky – START, STOP, dělicí znak

- 13 čísel, první čísla 859 – ČR, 400–440 Německo

- Každé číslo 7 jednotek

- První část: 7 (6 + 1): kód země, kód výrobce

- Druhá část: 6 čísel (CCCCCC)



Kódovací tabulka EAN			
znak	sada A	sada B	sada C
0			
1			
2			
3			
4			
5			
6			
7			
8			
9			
okrajové znaky		dělicí znak	

číslo	kombinace
0	AAAAAA
1	AABABB
2	AABBAB
3	AABBBA
4	ABAABB
5	ABBAAB
6	ABBBAA
7	ABABAB
8	ABABBA
9	ABBABA

Čárový kód

- **EAN 13:** druhá část – 6 čísel (5: kód výrobku, 1: kontrolní součet)
- Kontrolní součet pro 12 číslic – dobře naskenovaný kód (8 5 9 4 1 8 9 3 3 0 0 1)
- 1) Součet hodnot na sudých pozicích, krát 3: $(5 + 4 + 8 + 3 + 0 + 1) * 3 = 63$
- 2) Součet hodnot na lichých pozicích $(8 + 9 + 1 + 9 + 3 + 0) = 30$
- 3) Součet 1) a 2), zaokrouhlení na celé desítky: $30 + 63 = 93 = 90$
- 4) Kontrolní číslo: rozdíl zaokrouhlené a vypočtené hodnoty: $90 - 93 = -3$

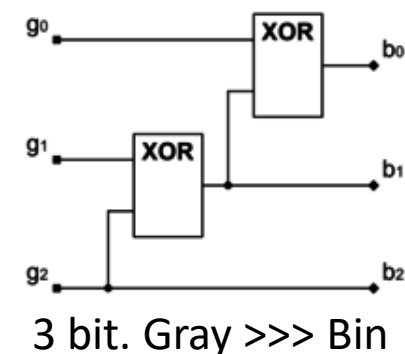
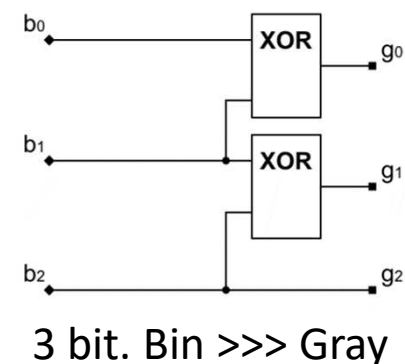


Grayův kód

- Změna sousedních slov (výrazů) pouze v jednom prvku (pouze v jednom bitu u binární reprezentace), Emile Baudot 1875, patent 1953 Frank Gray, AD převodníky, mechanické kodéry, rotační snímač absolutní polohy, oprava chyb v digitální komunikaci - digitální televize
- Číslicová technika – změna více logických hodnot současně obtížná
- Převod binární – Grayův kód: a) binární číslo s největší váhou se ponechá beze změn, b) další číslo se invertuje, pokud je před ním 1

DEC	BIN	Gray
0	000	000
1	001	001
2	010	011
3	011	010
4	100	110
5	101	111
6	110	101
7	111	100

X	Y	XOR
0	0	0
0	1	1
1	0	1
1	1	0



Filtr Move-To-Front (MTF)

- 1980 (B. Ryabko), 1986 (J.K. Bentley) Přesuň na začátek, po provedení Burrowsovy-Wheelerovy transformace

- Dlouhé sekvence symbolů nahrazeny nulami, málo se vyskytuje: velké číslo

- Nahrazení symbolů ze vstupního řetězce jejich indexy ze zásobníku

- zásobník s čísli (pořadí, ASCII...), *i*-tá pozice – kód *i* z pořadí, ASCII

- Aktuálně kódovaný znak přesunut na počátek (dopředu)

- A až Z (26 čísel v zásobníku, + české znaky 42 písmen)

- ANANAS bude zakódován jako 1, 14, 2, 2, 2, 19

- A – 1 (poté: ABCDEFGHIJKLMNO...)

- N – 14 (poté: NABCDEFGFGHIJKLMO...)

- A – 2 (poté: ANBCDEFGFGHIJKLMO...)

- N – 2 (poté: NABCDEFGFGHIJKLMO...)

- A – 2 (poté: ANBCDEFGFGHIJKLMO...)

- S – 19 (poté: SNABCDEFGFGHIJKLMO...)

1	A	14	N
2	B	15	O
3	C	16	P
4	D	17	Q
5	E	18	R
6	F	19	S
7	G	20	T
8	H	21	U
9	I	22	V
10	J	23	W
11	K	24	X
12	L	25	Y
13	M	26	Z

Filtr Move-To-Front (MTF)

ANANAS – ASCII

- A – 65, (poté A = 0)
- N – 78, (poté N = 0, A = 1)
- A – 1 (poté A = 0, N = 1)
- N – 1 (poté N = 0, A = 1)
- A – 1 (poté A = 0, N = 1)
- S – 83 (poté S = 0, A = 1...)

000	00		043	2B	+	086	56	U	129	81	ü	172	AC	¼	215	D7	
001	01	☉	044	2C	,	087	57	W	130	82	é	173	AD	½	216	D8	⌈
002	02	☼	045	2D	-	088	58	X	131	83	â	174	AE	¾	217	D9	⌋
003	03	♥	046	2E	.	089	59	Y	132	84	ä	175	AF	»	218	DA	┌
004	04	♦	047	2F	/	090	5A	Z	133	85	à	176	B0	▤	219	DB	█
005	05	♠	048	30	0	091	5B	[134	86	á	177	B1	▥	220	DC	▩
006	06	♣	049	31	1	092	5C	\	135	87	ç	178	B2	▧	221	DD	▯
007	07	•	050	32	2	093	5D]	136	88	ê	179	B3	▨	222	DE	▰
008	08	■	051	33	3	094	5E	^	137	89	ë	180	B4	▩	223	DF	▱
009	09		052	34	4	095	5F	_	138	8A	è	181	B5	▭	224	E0	α
010	0A		053	35	5	096	60	`	139	8B	í	182	B6	▮	225	E1	β
011	0B	ø	054	36	6	097	61	a	140	8C	î	183	B7	▯	226	E2	Γ
012	0C	♀	055	37	7	098	62	b	141	8D	ï	184	B8	▰	227	E3	Π
013	0D		056	38	8	099	63	c	142	8E	ä	185	B9	▱	228	E4	Σ
014	0E	¶	057	39	9	100	64	d	143	8F	å	186	BA	▲	229	E5	σ
015	0F	*	058	3A	:	101	65	e	144	90	é	187	BB	△	230	E6	μ
016	10	►	059	3B	;	102	66	f	145	91	æ	188	BC	▴	231	E7	τ
017	11	◄	060	3C	<	103	67	g	146	92	æ	189	BD	▵	232	E8	Ϙ
018	12	‡	061	3D	=	104	68	h	147	93	ô	190	BE	▶	233	E9	θ
019	13	!!	062	3E	>	105	69	i	148	94	ö	191	BF	▷	234	EA	Ω
020	14	¶	063	3F	?	106	6A	j	149	95	ð	192	C0	▸	235	EB	δ
021	15	§	064	40	@	107	6B	k	150	96	û	193	C1	▹	236	EC	ω
022	16	—	065	41	A	108	6C	l	151	97	ù	194	C2	▹	237	ED	ø
023	17	±	066	42	B	109	6D	m	152	98	ÿ	195	C3	▹	238	EE	€
024	18	↑	067	43	C	110	6E	n	153	99	õ	196	C4	▹	239	EF	η
025	19	↓	068	44	D	111	6F	o	154	9A	ü	197	C5	▹	240	F0	≡
026	1A		069	45	E	112	70	p	155	9B	ç	198	C6	▹	241	F1	±
027	1B	←	070	46	F	113	71	q	156	9C	£	199	C7	▹	242	F2	≥
028	1C	└	071	47	G	114	72	r	157	9D	¥	200	C8	▹	243	F3	≤
029	1D	↕	072	48	H	115	73	s	158	9E	₹	201	C9	▹	244	F4	ƒ
030	1E	▲	073	49	I	116	74	t	159	9F	ƒ	202	CA	▹	245	F5	÷
031	1F	▼	074	4A	J	117	75	u	160	A0	á	203	CB	▹	246	F6	+
032	20		075	4B	K	118	76	v	161	A1	í	204	CC	▹	247	F7	∞
033	21	!	076	4C	L	119	77	w	162	A2	ó	205	CD	▹	248	F8	°
034	22	"	077	4D	M	120	78	x	163	A3	ú	206	CE	▹	249	F9	·
035	23	#	078	4E	N	121	79	y	164	A4	ñ	207	CF	▹	250	FA	·
036	24	\$	079	4F	O	122	7A	z	165	A5	ñ	208	D0	▹	251	FB	√
037	25	%	080	50	P	123	7B	<	166	A6	æ	209	D1	▹	252	FC	∞
038	26	&	081	51	Q	124	7C	!	167	A7	æ	210	D2	▹	253	FD	z
039	27	'	082	52	R	125	7D	>	168	A8	¿	211	D3	▹	254	FE	■
040	28	<	083	53	S	126	7E	~	169	A9	₹	212	D4	▹	255	FF	
041	29	>	084	54	T	127	7F	△	170	AA	₹	213	D5	▹			
042	2A	*	085	55	U	128	80	Ç	171	AB	½	214	D6	▹			

Filtr Move-To-Front (MTF)

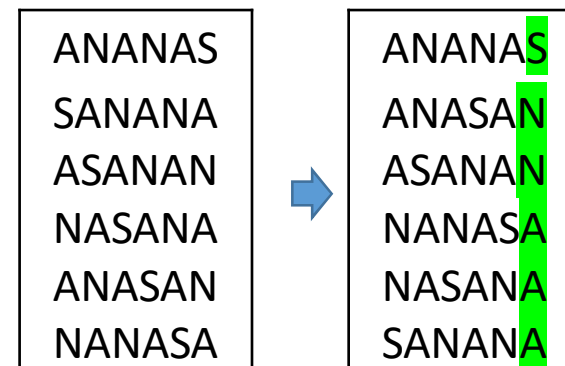
● Dekódování

- 1 - A (poté: ABCDEFGHIJKLMNOP...)
- 14 - N (poté: NABCDEFGHIJKLMO...)
- 2 - A (poté: ANABCDEFGHIJKLMO...)
- 2 - N (poté: NABCDEFGHIJKLMO...)
- 2 - A (poté: ANABCDEFGHIJKLMO...)
- 19 - S (poté: SNABCDEFGHIJKLMO...)

1	A	14	N
2	B	15	O
3	C	16	P
4	D	17	Q
5	E	18	R
6	F	19	S
7	G	20	T
8	H	21	U
9	I	22	V
10	J	23	W
11	K	24	X
12	L	25	Y
13	M	26	Z

Burrows-Wheelerova Transformace (BWT)

- Michael Wheeler (1984) + David Burrows (1994)
- Přeskupí symboly tak, že shodné budou s vysokou pravděpodobností vedle sebe, pak můžeme použít RLE...
- Bzip2 – BWT >>> MTF >>> Huffmanovo kódování
- Vstupní řetězec se zpracovává po blocích o konstantní velikosti, větší bloky – větší efektivita, blok cca desítky tisíc symbolů (Block Sorting Algorithm)
- 1) blok o délce N (např. ANANAS), N cyklických posuvů o jeden symbol doprava >>> matice N * N, řádky => cyklické posuny
- 2) lexikografické (podle abecedy) setřídění matice
- 3) poslední sloupec je výsledek transformace
- 4) přidáme číslo řádku v setříděné matici, kde se nachází vstupní řetězec: **SNNAAA, 1. ŘÁDEK**



Burrows-Wheelerova Transformace (BWT)

- Dekódování: SNNAAA, 1. ŘÁDEK

1	sort	2	sort	3	sort	4	sort	5	sort	6	sort
S	A	SA	AN	SAN	ANA	SANA	ANAN	SANAN	ANANA	SANANA	ANANAS
N	A	NA	AN	NAN	ANA	NANA	ANAS	NANAS	ANASA	NANASA	ANASAN
N	A	NA	AS	NAS	ASA	NASA	ASAN	NASAN	ASANA	NASANA	ASANAN
A	N	AN	NA	ANA	NAN	ANAN	NANA	ANANA	NANAS	ANANAS	NANASA
A	N	AN	NA	ANA	NAS	ANAS	NASA	ANASA	NASAN	ANASAN	NASANA
A	S	AS	SA	ASA	SAN	ASAN	SANA	ASANA	SANAN	ASANAN	SANANA

- VÝSTUP: ANANAS

BWT + MTF

● BWT Dekódování: SNNAAA, 1. ŘÁDEK

● MTF:

S – 83 (poté: S = 0)

N – 78 (poté: N = 0, S = 1)

N – 0

A – 65 (poté: A = 0, N = 1, S = 2)

A – 0

A – 0

pro Huffmanovo kódování:

0 – 3x

65 – 1x

78 – 1x

83 – 1x

U delších řetězců – vyšší počet 0

000	00		043	2B	+	086	56	U	129	81	ü	172	AC	¼	215	D7	
001	01	☉	044	2C	ˆ	087	57	W	130	82	é	173	AD	ı	216	D8	⌈
002	02	☼	045	2D	–	088	58	X	131	83	â	174	AE	«	217	D9	⌋
003	03	♥	046	2E	.	089	59	Y	132	84	ä	175	AF	»	218	DA	┌
004	04	♦	047	2F	/	090	5A	Z	133	85	à	176	B0	▤	219	DB	█
005	05	♠	048	30	0	091	5B	[134	86	á	177	B1	▥	220	DC	■
006	06	♣	049	31	1	092	5C	\	135	87	ç	178	B2	▧	221	DD	▩
007	07	•	050	32	2	093	5D]	136	88	ê	179	B3	▨	222	DE	▪
008	08	◼	051	33	3	094	5E	^	137	89	ë	180	B4	▩	223	DF	▬
009	09		052	34	4	095	5F	˘	138	8A	è	181	B5	▭	224	E0	α
010	0A		053	35	5	096	60	˙	139	8B	ı	182	B6	▮	225	E1	β
011	0B	ø	054	36	6	097	61	a	140	8C	î	183	B7	▯	226	E2	Γ
012	0C	♀	055	37	7	098	62	b	141	8D	ì	184	B8	▰	227	E3	Π
013	0D		056	38	8	099	63	c	142	8E	ä	185	B9	▱	228	E4	Σ
014	0E	℥	057	39	9	100	64	d	143	8F	ß	186	BA	▲	229	E5	σ
015	0F	✱	058	3A	:	101	65	e	144	90	é	187	BB	△	230	E6	μ
016	10	▶	059	3B	;	102	66	f	145	91	æ	188	BC	▴	231	E7	τ
017	11	◀	060	3C	<	103	67	g	146	92	æ	189	BD	▵	232	E8	ϖ
018	12	‡	061	3D	=	104	68	h	147	93	ô	190	BE	▶	233	E9	θ
019	13	!!	062	3E	>	105	69	i	148	94	ö	191	BF	▷	234	EA	Ω
020	14	¶	063	3F	?	106	6A	j	149	95	ò	192	C0	▸	235	EB	δ
021	15	§	064	40	@	107	6B	k	150	96	û	193	C1	▹	236	EC	ω
022	16	–	065	41	A	108	6C	l	151	97	ù	194	C2	►	237	ED	ø
023	17	±	066	42	B	109	6D	m	152	98	ÿ	195	C3	▻	238	EE	€
024	18	↑	067	43	C	110	6E	n	153	99	õ	196	C4	▼	239	EF	π
025	19	↓	068	44	D	111	6F	o	154	9A	ü	197	C5	▽	240	F0	≡
026	1A		069	45	E	112	70	p	155	9B	ç	198	C6	▿	241	F1	±
027	1B	←	070	46	F	113	71	q	156	9C	£	199	C7	▹	242	F2	¿
028	1C	↵	071	47	G	114	72	r	157	9D	¥	200	C8	▹	243	F3	≤
029	1D	↗	072	48	H	115	73	s	158	9E	℞	201	C9	▹	244	F4	ƒ
030	1E	▲	073	49	I	116	74	t	159	9F	ƒ	202	CA	▹	245	F5	÷
031	1F	▼	074	4A	J	117	75	u	160	A0	á	203	CB	▹	246	F6	÷
032	20		075	4B	K	118	76	v	161	A1	í	204	CC	▹	247	F7	≈
033	21	!	076	4C	L	119	77	w	162	A2	ó	205	CD	▹	248	F8	°
034	22	"	077	4D	M	120	78	x	163	A3	ú	206	CE	▹	249	F9	·
035	23	#	078	4E	N	121	79	y	164	A4	ñ	207	CF	▹	250	FA	˙
036	24	\$	079	4F	O	122	7A	z	165	A5	ñ	208	D0	▹	251	FB	√
037	25	%	080	50	P	123	7B	<	166	A6	æ	209	D1	▹	252	FC	∞
038	26	&	081	51	Q	124	7C	ı	167	A7	æ	210	D2	▹	253	FD	∞
039	27	'	082	52	R	125	7D	>	168	A8	¿	211	D3	▹	254	FE	■
040	28	<	083	53	S	126	7E	~	169	A9	ı	212	D4	▹	255	FF	
041	29	>	084	54	T	127	7F	Δ	170	AA	ı	213	D5	▹			
042	2A	*	085	55	U	128	80	Ç	171	AB	½	214	D6	▹			