

# Informatika – Datové formáty

Radim Farana  
Podklady předmětu Informatika  
pro akademický rok 2007/2008

---

---

---

---

---

---

---

---

## Obsah

- Datové formáty (datové typy).
- Textové formáty, vlastnosti zdroje zpráv.
- Číselné formáty, číselné soustavy.
- Přesnost uložení čísel.
- Numerické chyby.

---

---

---

---

---

---

---

---

## Datové formáty (datové typy)

- Jednoduché
  - Textové (alfanumerické)
  - Číselné
    - celočíselné
    - neceločíselné
      - s pevnou řádovou čárkou
      - s pohyblivou řádovou čárkou
  - Datum a čas
- Složené
  - Pole
  - Záznam

---

---

---

---

---

---

---

---

## Textové kódy

- **blokové kódy** s různou délkou kódového slova a mohutností abecedy,
- 5 bitů – nejstarší kódy, Baudotův kód, kód CCITT, dvě znakové sady po 32 znaků,
- 7 bitů – American Standard Code for Information Interchange – ASCII – UNIX,
- 8 bitů – ASCII-2, znakové sady (1250) pro různé jazyky,
- 16 bitů – Unicode.

---

---

---

---

---

---

---

---

## Baudotův kód

binary	hexa	LTRS	FIGS	FIGS (US Bell)
00000	00		– unused –	
00001	01	E	B	
00010	02	LINE FEED	LINE FEED	
00011	03	A		
00100	04	SPACE	SPACE	
00101	05	B		
00110	06	F	B	BELL
00111	07	G	F	
01000	08	CAR. RET.	CAR. RET.	
01001	09	D	E	
01010	0A	R		
01011	0B	V	BELL	
01100	0C	N		
01101	0D	C		
01110	0E	–		
01111	0F	K		
10000	10	L	B	
10001	11	Z		
10010	12	E		
10011	13	W	B	
10100	14	H		STOP
10101	15	V	B	
10110	16	P	B	
10111	17	O		
11000	18	D	B	
11001	19	B	F	
11010	1A	S	A	
11011	1B	FIGS	FIGS	
11100	1C	M		
11101	1D	X		
11110	1E	V		
11111	1F	LTRS	LTRS	



**Baudot, Jean-Maurice-Émile**  
 \* 11. 9. 1845 Magnéux, France  
 + 28. 3. 1903 Soissons, France  
<http://www.converter.cz/typo/baudot.htm>



Diagram showing the 5-bit binary representation of the Baudot code for letters and symbols. The letters are arranged in a grid with their corresponding 5-bit binary strings.

---

---

---

---

---

---

---

---

## Kód CCITT

CCITT	znak	kód	CCITT	znak	kód
1	A	11000	17	Q 1	11101
2	B ?	10011	18	R 4	01010
3	C :	01110	19	S !	10100
4	D vZ	10010	20	T 5	00001
5	E 3	10000	21	U 7	11100
6	F	10110	22	V =	01111
7	G	01011	23	W 2	11001
8	H	00101	24	X /	10111
9	I 8	01100	25	Y 6	10101
10	J zv	11010	26	Z +	10001
11	K (	11110	27	návrat válce	00010
12	L )	01001	28	posun řádku	01000
13	M .	00111	29	číslicová změna	11011
14	N ,	00110	30	písmenová změna	11111
15	O 9	00011	31	mezera	00100
16	P 0	01101	32		00000

---

---

---

---

---

---

---

---

# ASCII, ISO 8859-1 (Latin-1)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
20		!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
80				f	„	„	„	„	„	„	„	„	„	„	„	„
90			‘	’	“	”	*	—	—	—	™	§	>	œ		ÿ
A0	ı	€	£	¤	¥	¦	§	¨	©	ª	«	»	¼	½	¾	¿
B0	◊	±	²	³	´	µ	¶	·	,	»	¼	½	¾	¿		
C0	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D0	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E0	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F0	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ		

---

---

---

---

---

---

---

---

# Abeceda zdroje

- Z<sub>26</sub> – Mezinárodní telegrafní abeceda:  
ABCDEFGHIJKLMNOQRSTUVWXYZ,
- Z<sub>27</sub> – Telegrafní abeceda s mezerou,
- Z<sub>41</sub> – Česká abeceda velkých písmen:  
AĀBCČĎĎĚĚĚFGHIĪJKLMNŇŌÓPQRŘŠŠ  
ŤUÚŮVWXYZŽ,
- Z<sub>42</sub> – Česká abeceda s mezerou,
- Z<sub>128</sub> – Abeceda kódu ASCII znaků (7 bitů),
- Z<sub>256</sub> – Abeceda kódu ASCII-2 (8 bitů).

---

---

---

---

---

---

---

---

# Vlastnosti zdroje

Pravděpodobnosti výskytu jednotlivých znaků české abecedy Z<sub>42</sub>

A	0,054	F	0,002	Ň	0,015	ř	0,007	Ž	0,009
Ā	0,021	G	0,002	O	0,068	U	0,030	-	0,163
B	0,014	H	0,020	Ó	0,000	Ú	0,003		
C	0,019	I	0,034	P	0,027	Ů	0,002		
Č	0,008	ı	0,025	Q	0,000	V	0,039		




---

---

---

---

---

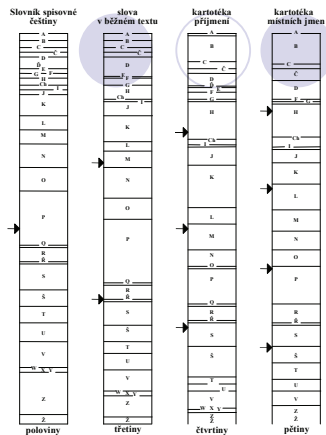
---

---

---

## Frekvence znaků v češtině

- kartotéky,
- soubory znaků,
- analýza jazyka zdroje zpráv.




---

---

---

---

---

---

---

---

---

---

---

---

## Určení jazyka zdroje

- Index koincidence  
 $I(C) = \sum P(i)^2$
- Využití v kryptoanalýze

Jazyk	$I(C)$
angličtina	0,066895
dánština	0,070731
finština	0,073796
francouzština	0,074604
holandština	0,079805
němčina	0,076667
italština	0,073294
ruština	0,056074
španělština	0,076613
slovenština	0,060270
čeština	0,063600

---

---

---

---

---

---

---

---

---

---

---

---

## Textové formáty

- Znak, Char, Character (1 B)
- String – posloupnost Byte (do 255 B)
  - s pevnou délkou
  - s proměnlivou délkou
- Memo (až 64 kB – adresa 16 bitů).




---

---

---

---

---

---

---

---

---

---

---

---

## Číselné formáty

- Celočíselné
  - Bit (logický datový typ)
  - Byte (1 B, rozsah 0 – 255)
  - Integer (2 B)
  - Long Integer (4 B)
- Racionální, podíl dvou celých čísel
- Neceločíselné
  - s pevnou řádovou čárkou
  - s pohyblivou řádovou čárkou

---

---

---

---

---

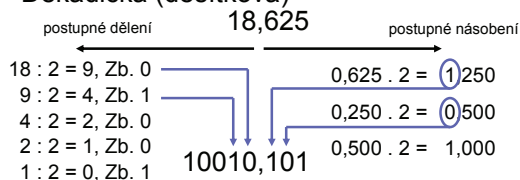
---

---

---

## Číselné soustavy

- Binární (dvojková)
- Oktalová (osmičková, 1 znak = 3 b)
- Šestnáctková (hexadecimální, 4 b)
- Dekadická (desítková)




---

---

---

---

---

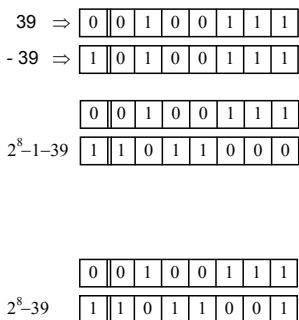
---

---

---

## Číselné kódy

- Přímý kód
  - znaménko
  - absolutní hodnota
- Inverzní kód
  - záporné +  $2^n - 1$
- Doplnkový kód
  - záporné +  $2^n$




---

---

---

---

---

---

---

---

## Realizace sčítání

- Včetně znaménkových bitů.

$$\begin{array}{r}
 \text{znaménko} \\
 39 = \quad 0 \mid 0 \ 1 \ 0 \ 0 \ 1 \ 1 \ 1 \\
 -39 = \quad 1 \mid 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \\
 \hline
 1 \mid 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \\
 \text{přenos}
 \end{array}$$

- Problém přetečení rozsahu zobrazitelných čísel (přeplnění) – zdvojení znamének

znaménko	znaménko
39 = 0 0   0 1 0 0 1 1 1	127 = 0 0   1 1 1 1 1 1 1
-39 = 1 1   1 0 1 1 0 0 1	1 = 0 0   0 0 0 0 0 0 1
1   0 0 0 0 0 0 0 0	0   0 1 0 0 0 0 0 0
přenos   stejná hodnota znaménka ukazuje na správný výsledek	přenos   nesejná hodnota znaménka ukazuje na přetečení rozsahu

---

---

---

---

---

---

---

---

---

---

## Čísla s pevnou řádovou čárkou

- Omezená velikost celé části.
- Omezená přesnost (počet desetinných míst).

$$\begin{array}{l}
 \text{zn. } 2^2 \ 2^1 \ 2^0 \ 2^{-1} \ 2^{-2} \ 2^{-3} \ 2^{-4} \\
 2,625 \Rightarrow \boxed{0 \mid 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0} = 2,625 \\
 \uparrow \\
 \text{poloha řádové čárky} \\
 2,6 \Rightarrow \boxed{0 \mid 0 \ 1 \ 0 \ 1 \ 0 \ 0 \ 1} = 2,5625
 \end{array}$$

---

---

---

---

---

---

---

---

---

---

## Čísla s pohyblivou řádovou čárkou

- Tvar  $m \cdot 2^e$ , kde je  $m$  – mantisa  $|m| < 1$
- $e$  – celočíselný exponent
- Normování mantisy  $0,5 \leq m < 1$

$$2,625 = 0,65625 \cdot 2^{+2}$$

zn. $2^{-1}$	zn. $2^{-2}$	zn. $2^{-3}$	zn. $2^{-4}$	zn. $2^1$	zn. $2^0$
0	1	0	1	0	0
mantisa				exponent	

$$\boxed{0 \ 1 \ 0 \ 1 \ 0 \ 0 \ 1 \ 0} = 2,5$$

- Skrytý bit (výsledek normování)
- Uložení hodnoty 0, exponent +  $2^{n-1}$

---

---

---

---

---

---

---

---

---

---

## Číselné kódy

- Přímý kód

$$x_{pf} = \begin{cases} x & x \geq 0 \\ 1-x & x < 0 \end{cases} \quad |x| < 1$$

- Inverzní kód

$$x_{inv} = \begin{cases} x & x \geq 0 \\ 2-2^{-n}+x & x < 0 \end{cases} \quad |x| < 1$$

- Doplnkový kód

$$x_{dop} = \begin{cases} x & x \geq 0 \\ 2+x & x < 0 \end{cases} \quad -1 \leq x < 1$$

---

---

---

---

---

---

---

---

## Přesnost uložení čísel

- Počítačové epsilon - nejmenší číslo, zobrazitelné v daném číselném kódu, pro které platí  $1,0 + \varepsilon > 1,0$
- Norma uložení IEEE 754 (formát single)



- Typické datové formáty
  - Single 4 B – 7 desetinných míst
  - Double 8 B – 15 desetinných míst
  - Extended 10 B – 19 desetinných míst

---

---

---

---

---

---

---

---

## Datum a čas

- Datum

- celá část čísla,
- počet dnů od referenčního datumu (1. 1. 1900).

- Čas

- desetinná část čísla,
- 1 den = 24 hodin => 1 hodina = 1/24

- Výpočty s datумы

- Problém testování rovnosti datumu a času

---

---

---

---

---

---

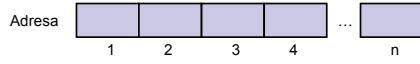
---

---

## Složené datové typy

- Pole (posloupnost prvků stejného datového typu).

- identifikace pořadím prvku,
- u vícerozměrného pomocí mapovacích funkcí



- Záznam (množina prvků různých datových typů).

- identifikace názvem prvku

---

---

---

---

---

---

---

---