

Úvod do programů pro zpracování textu část první ASCII a kódování znaků

Mgr. Pavel Hrubý

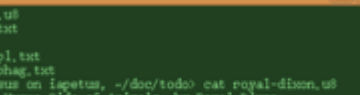
A series of horizontal lines of varying lengths and colors (light blue and white) extending from the right side of the slide.

Rozdělení programů pro zpracování textu

- **Textové editory** byly první aplikační programy, které byly na počítačích spouštěny. Původně sloužili pouze k psaní textu a zdrojových kódů dalších aplikačních programů.
 - Od lepšího textového editoru očekáváme více, než je pouhé psaní textu. Pro obchodní dokumenty, publikace, letáky a další tiskoviny očekáváme, že lze různě "tvarovat" písmo, zvětšovat a zmenšovat, měnit barvy písma i podkladu, zvýrazňovat a vkládat různé obrázky a jiné grafické objekty. Program by měl automaticky nacházet běžné překlepy a také je opravovat.
- Toto je vlastnost dokonalejších textových editorů, které se také nazývají **textové procesory**. Jedním z nich je textový procesor MS WORD, který výše uvedené možnosti obsahuje.
- Typy programů, které umožňují připravit hotové dokumenty ke konečné tiskové úpravě se nazývají programy **DTP** (Desk Top Publishing).

Textový editor a plain text

- **Prostý text** (též čistý nebo holý text, anglicky **plain text**) je v informatice označení uložení textu do souboru ve formě, kdy jsou uloženy pouze jednotlivé znaky (resp. čísla odpovídající kódům použité znakové sady, např. **ASCII** nebo **Unicode**) bez formátovacích informací.
- **Hlavní výhodou** prostého textu je jeho přímá čitelnost bez nutnosti interpretovat obsah speciálním programem, možnost upravovat jakýmkoliv jednoduchým textovým editorem (na rozdíl od formátovaného textu nebo binárních souborů).



```

xterm
royal-dixon.us
MP.txt
swl
swpl.txt
szzbhag.txt
cursor on lapetus, -/doc/todo: cat royal-dixon.us
The Human Side of Animals, by Royal Dixon;

HONEYEATS ARE THE MOST MUSICAL OF ALL ANIMALS. WHEN THEY
CONSPIRATE FOR "CONCERTS," AS SOME OF THE TRIBES DO, THE
AIR IS FILLED WITH WEIRD STRAINS OF HONEY-EY-MUSIC.

CATS, UNLIKE DOGS, ARE VERY FOND OF MUSIC, AND IT HAS
BEEN PROVED THAT THEIR MUSIC-SENSE CAN BE DEVELOPED TO A
REMARKABLE DEGREE.

cursor on lapetus, -/doc/todo:

```

The screenshot shows a web browser window with the address bar displaying 'http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.' and the page title 'http://www.w3.org/1999/xhtml'. The browser's menu bar includes 'Soubor', 'Úpravy', 'Formát', 'Zobrazení', and 'Nápověda'. The main content area displays the source code of an XHTML document. The code includes a DOCTYPE declaration, a public ID, a language attribute, a head section with a title 'Moxio', a meta tag for content type, and several links for editing, RSS, and shortcuts. It also includes multiple style sheet imports for various modules like 'node', 'poll', 'modules', 'system', and 'themes'.

```

<?xml version="1.0" encoding="UTF-8" lang="cs" xml:lang="cs"
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional."
<html xmlns="http://www.w3.org/1999/xhtml" lang="cs" xml:lang="cs">
<head>
<title>Moxio1.c2 | software, poradenství a webové stránky
<meta http-equiv="Content-type" content="text/html; chars
<link rel="EditURI" type="application/rsd+xml" title="RSD" hr
<link rel="alternate" type="application/rss+xml" title="Moxio
<link rel="shortcut icon" href="/misc/favicon.ico" type="imag
<style type="text/css" media="all">@import "/modules/
<style type="text/css" media="all">@import "/modules/node/nod
<style type="text/css" media="all">@import "/modules/poll/po
<style type="text/css" media="all">@import "/modules/system/d
<style type="text/css" media="all">@import "/modules/system/s
<style type="text/css" media="all">@import "/modules/user/use
<style type="text/css" media="all">@import "/themes/moxio1/

```

Kódování textu - ASCII

- ASCII je anglická zkratka pro **American Standard Code for Information Interchange** („americký standardní kód pro výměnu informací“).
- V podstatě jde o kódovou tabulku, která definuje znaky anglické abecedy, a jiné znaky používané v informatice.
- Jde o historicky nejúspěšnější znakovou sadu, z které vychází většina současných standardů pro kódování textu přinejmenším v euro-americké zóně.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0																
1												chinese 5 prefix				
2	SP	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8	ů	č	d'	ě	ň	ř	š	ť	ž	Č	Ď	Ě	Ň	Ř	Š	Ť
9	Ž	ř	í	í												
A		i	ç	£	¤	¥		§	"	©	ª	«	¬	-	®	¯
B	°	±	²	³	'	μ	¶	·	¸	¹	º	»	¼	½	⅓	¿
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Kódování textu - ASCII

- Tabulka obsahuje tisknutelné znaky: písmena, číslice, jiné znaky (závorky, matematické znaky (+ - * / % atd.), interpunkční znaménka (, . : ; atd.), speciální znaky (@ \$ ~ atd.)), a řídicí (netisknutelné) kódy, které byly původně určeny pro řízení periferních zařízení (např. tiskárny nebo dálnopisu).
- ALT kódy – pomocí klávesy levý ALT+kód znaku v desítkové soustavě se při nastavení ASCII kódování znak zobrazí

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0																
1												chinese 5 prefix				
2	SP	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8	ů	č	d'	ě	ň	ř	š	ť	ž	Č	Ď	Ě	Ň	Ř	Š	Ť
9	Ž	ř	í	í												
A		i	ç	£	¤	¥	¦	§	¨	©	ª	«	¬	-	®	¯
B	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	⅓	¿
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Kódování textu - ASCII

- **Základní ASCII tabulka má dohromady 128 znaků** (na pozici 0–127).
- Tato sada se nazývá standardní sada znaků ASCII (ASCII standard character set).
- Hodnoty 0-31 a 127 jsou použity pro netisknutelné řídicí kódy, **hodnoty 32–126 jsou použity pro prezentaci písmen, číslic a oddělovacích znaků**.
- Sadu znaků (128 - 255) **nazýváme rozšířenou sadou znaků ASCII** (ASCII extended character set).

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0																
1												chinese 5 prefix				
2	SP	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8	ů	č	d'	ě	ň	ř	š	ť	ž	Č	Ď	Ě	Ň	Ř	Š	Ť
9	Ž	ř	í	í												
A		i	ç	£	¤	¥	¦	§	¨	©	ª	«	¬	-	®	¯
B	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	⅓	¿
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Kódování textu - ASCII

- Pro psaní textu v základním ASCII můžeme využít alt kódy od 0 do 127
- Pro znaky 0 – 31
- ALT+2 ☺
- ALT+3 ♥
- ALT+4 ♦
- ALT+5 ♣
- ALT+6 ♠
- ALT+7 •
- ALT+8 ▣
- ALT+9 ○
- ALT+10 ◼
- ALT+11 ♂
- ALT+12 ♀
- ALT+13 ♪
- atd.

Matematický doplněk I. – číselné soustavy (1)

Mezi nejčastěji používané poziční číselné soustavy patří:

- **dvojková** (binární) – přímá implementace v digitálních elektronických obvodech (použitím logických členů), čili interně ji používají všechny moderní počítače
- **osmičková** (oktální, oktalová)
- **desítková** (decimální, dekadická) – nejpoužívanější v běžném životě
- **šestnáctková** (hexadecimální, $r=16$) – používá se v oblasti informatiky, pro číslice 10 až 15 se používají písmena A až F

Matematický doplněk I. – číselné soustavy (2)

Základní převody:

Dekadická	Binární	Hexadecimální
0	0	0
1	1	1
2	10	2
3	11	3
4	100	4
5	101	5
6	110	6
7	111	7

8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F

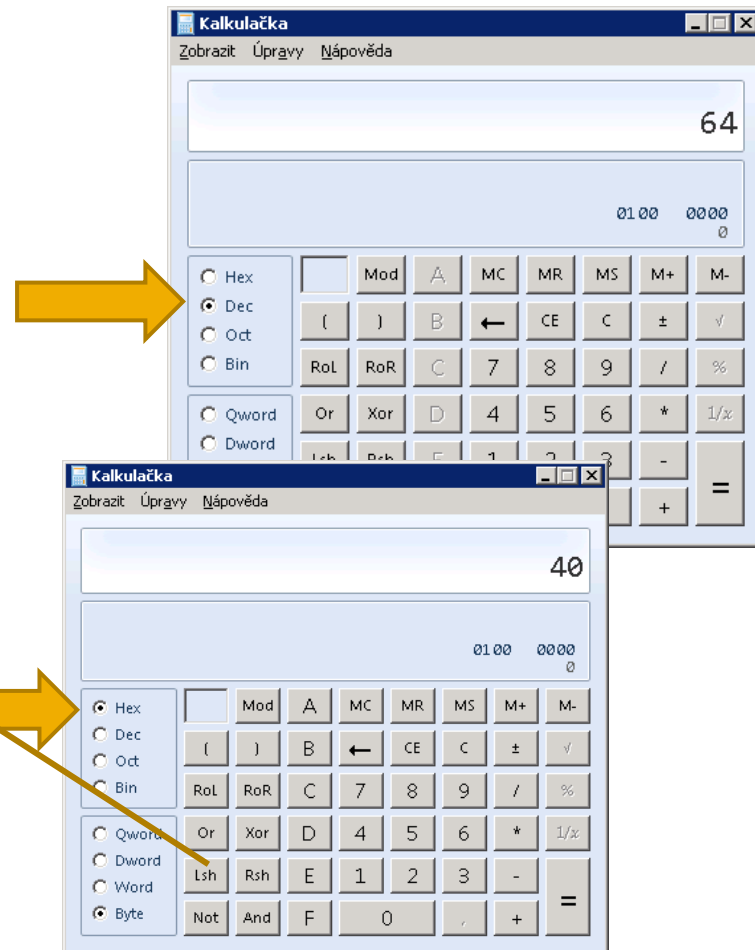
IT1A3-1 Matematický doplněk I. – číselné soustavy (3)

- Vytvořte nový textový dokument v Poznámkovém bloku. Soubor uložte pod jménem Soustavy.txt. Do souboru si překopírujte následující zadání a proveďte převody.
- Každý převod bude na jednom řádku např. $128D = 80H$
- Převeďte z desítkové do šestnáctkové soustavy
 - 240, 128, 64, 100, 1024,...
- Převeďte ze šestnáctkové do desítkové soustavy
 - 5A, AA, ABC, ABCD, ...
- Převeďte do dvojkové soustavy
 - 17, 32, 63, 100, 101, ...

Matematický doplněk I. – číselné soustavy (4)

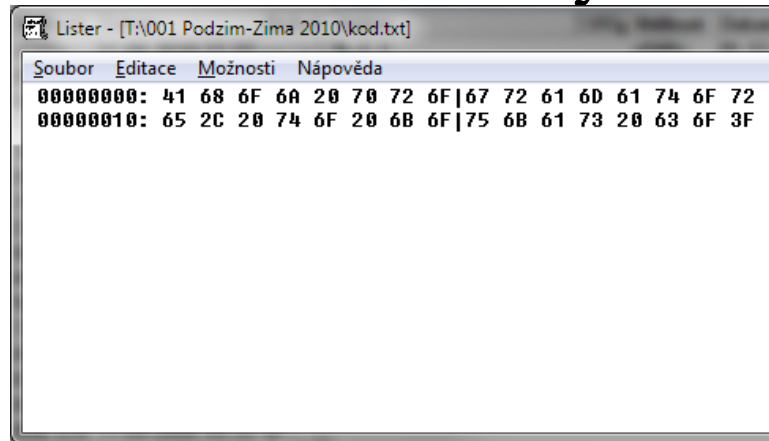
- Programátorská kalkulačka
- $64D = 40H$

POZOR
nutno se pro větší čísla
přepnout nejlépe na
QWORD



IT1A3-2 úkol – dekodování ASCII

- Přepište do normální abecedy kód v ASCII:

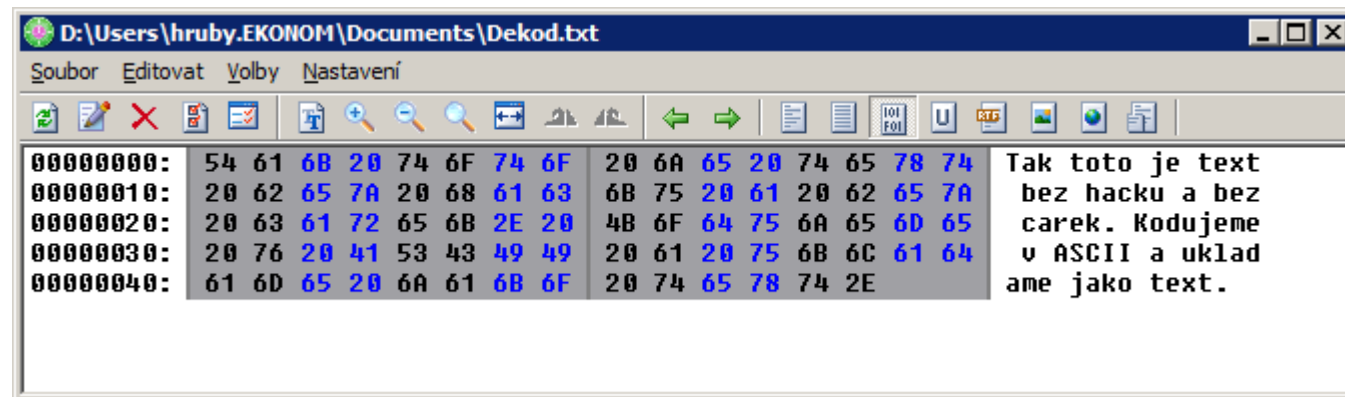
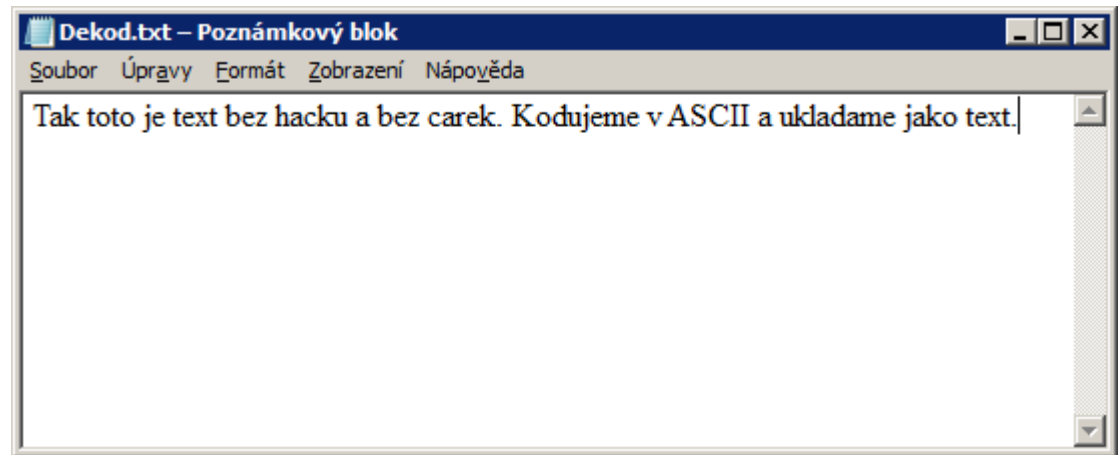


```
Lister - [T:\001 Podzim-Zima 2010\kod.txt]
Soubor Editace Možnosti nápověda
00000000: 41 68 6F 6A 20 70 72 6F|67 72 61 6D 61 74 6F 72 |
00000010: 65 2C 20 74 6F 20 6B 6F|75 6B 61 73 20 63 6F 3F |
```

- Vytvořte v poznámkovém bloku soubor Dekod.txt a do něj napište vyluštěný kód.
- K dekodování použijte tabulky ze snímku 6

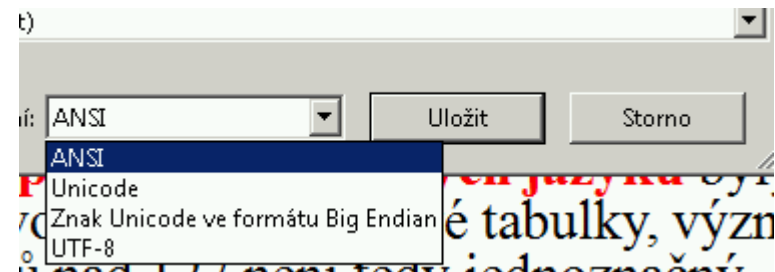
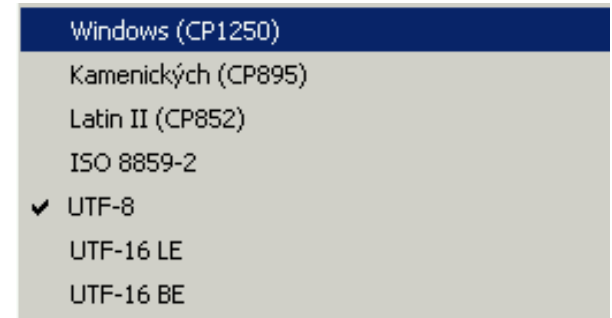
Zobrazení kódu dokumentu

- Kód dokumentu nelze přímo v systému MS WINDOWS zobrazit.
- Použijeme prohlížeč, který je součástí např. programu FreeCommander (v nabídce Zobrazit –F3)



Národní specifika ASCII

- Kód ASCII je podle původní definice sedmibitový, obsahuje tedy **128 platných znaků**.
- Pro potřeby dalších jazyků a pro rozšíření znakové sady se používají osmibitová rozšíření ASCII kódu, která obsahují dalších 128 kódů.
- Takto rozšířený kód je přesto příliš malý na to, aby pojmul třeba jen evropské národní abecedy.
- Pro **potřeby jednotlivých jazyků** byly vytvořeny různé kódové tabulky, význam kódů nad 127 není tedy jednoznačný.
- Systém kódových tabulek pro národní abecedy vytvořila například organizace ISO.



Osmibitová kódování češtiny

1 znak = 1 Byte

kódování	komentář
Windows-1250	kód používaný firmou Microsoft v operačních systémech Windows pro kódování středoevropských jazyků
ISO 8859-2	standard ISO, používaný např. v operačním systému Linux
CP852 (Latin2)	kód stanovený firmou IBM používaný např. v operačním systému DOS. Česká Windows jej dodnes využívají při zadávání speciálních znaků pomocí alt-kódů. Například když stisknete klávesu alt, vyťukáte číslo 248 a klávesu alt pustíte, objeví se na obrazovce znak pro stupeň - °. V tabulce znaků je uveden pod hexadecimálním kódem F8, což je dekadicky 248.
Kód Kamenických	jiné kódování vynalezené v Čechách používané v DOSu pro potřeby češtiny a slovenštiny (již se nepoužívá)
KOI8-ČS	kódování definované v rámci RVHP (již se nepoužívá)

Kódování češtiny

ISO-8859-2 nebo ISO Latin 2

- ISO Latin 2 je specifikováno **normou ISO-8859-2** (z roku 1987), která jej doporučuje používat pro **češtinu** a další jazyky.
- V MS WINDOWS je naleznete pod označením **Středoevropské jazyky (ISO)**

Text pro testování kódování češtiny:
Příliš žluťoučký kůň úpěl ďábelské ódy.

Kódování češtiny

Windows-1250 (CP1250, WinCS)

- **MS Windows ve verzi CS** používají kódování CP1250, které obsahuje všechny tisknutelné znaky sady ISO Latin 2.
- Čtrnáct znaků je však na **jiných pozicích** (z toho osm se používá v češtině a slovenštině).
 - **Není tedy pravda, že CP1250 je (jednosměrně) zaměnitelná s ISO-8859-2. (Rozdíl je patrný např. u "š")**
- V operačním systému MS Windows se označuje jako **ANSI** také některé národní kódování podporované systémem, např. **CP-1250 pro český jazyk.**

Kódování češtiny

CP852 (IBM852, PC Latin 2, PC L2)

- Kódování PC Latin 2 (zkráceně PC L2) se používá na PC.
- Toto kódování má všechny tisknutelné znaky sady ISO-8859-2 (ISO Latin 2), diakritická písmenka jsou však na jiných pozicích.
- Většina českých a slovenských uživatelů zná toto kódování pouze pod názvem Latin 2 a ani netuší, jak moc se **PC Latin 2** a **ISO Latin 2** liší.

Kódování češtiny

ISO-8859-1 (též Latin-1, Západoevropské ISO)

- Kódování ISO-8859-1 **neobsahuje všechny potřebné české znaky**, pouze některá mezinárodní písmenka mají shodný (podobný) vzhled s některými písmenky používanými i v češtině - především se jedná o dlouhé á, é ap.
- Této částečné podobnosti lze občas využít a napsat alespoň "cástečne" český text.

Kódování češtiny ANSI

- WINDOWS
CP 1250

00000000:	50 F8 ED 6C 69 9A 20 9E	6C 75 9D 6F 75 E8 6B FD	Příliš žluťoučký kůň úpěl ďábels ké ódy...
00000010:	20 6B F9 F2 20 FA 70 EC	6C 20 EF E1 62 65 6C 73	
00000020:	6B E9 20 F3 64 79 2E 0D	0A	

- ISO8259-2

00000000:	50 F8 ED 6C 69 B9 20 BE	6C 75 BB 6F 75 E8 6B FD	Příliš žluťoučký kůň úpěl ďábels ké ódy.
00000010:	20 6B F9 F2 20 FA 70 EC	6C 20 EF E1 62 65 6C 73	
00000020:	6B E9 20 F3 64 79 2E		

- Kamenických

00000000:	50 A9 A1 6C 69 A8 20 91	6C 75 9F 6F 75 87 6B 98	Přiliš žluťoučký kůň úpěl ďábels ké ódy.
00000010:	20 6B 96 A4 20 A3 70 88	6C 20 83 A0 62 65 6C 73	
00000020:	6B 82 20 A2 64 79 2E		

- Latin2

00000000:	50 FD A1 6C 69 E7 20 A7	6C 75 9C 6F 75 9F 6B EC	Přiliš žluťoučký kůň úpěl ďábels ké ódy.
00000010:	20 6B 85 E5 20 A3 70 D8	6C 20 D4 A0 62 65 6C 73	
00000020:	6B 82 20 A2 64 79 2E		

V červeném kroužku je ukázka zakódování znaku š v různých kódováních.

UNICODE

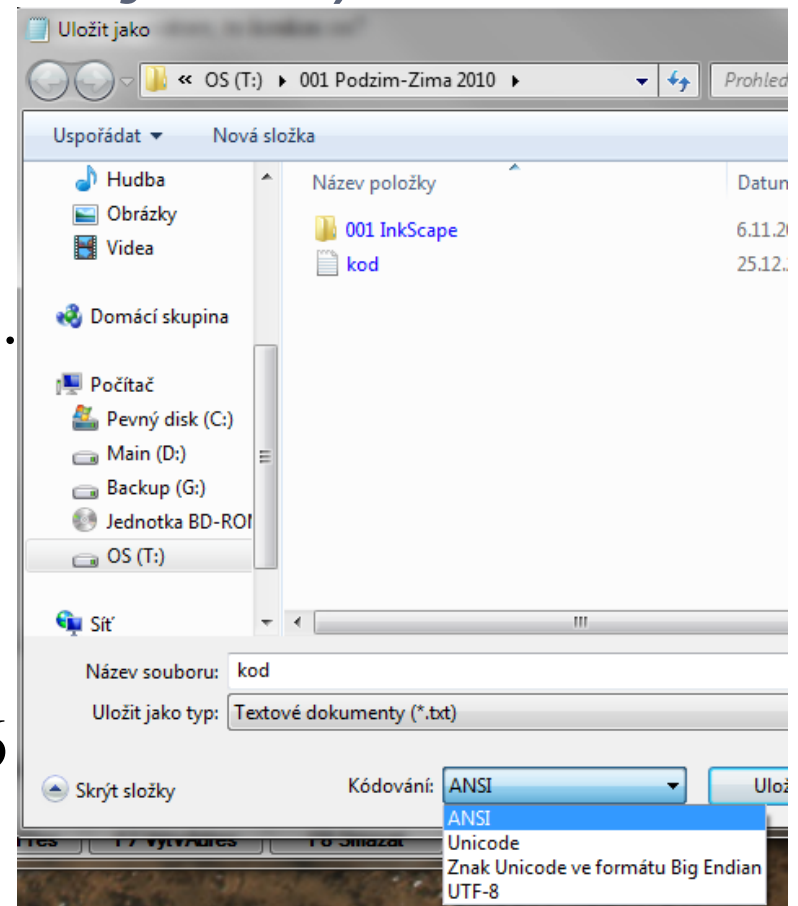
Unicode je tabulka znaků všech existujících abeced, která v současnosti obsahuje více než 245 000 znaků.

Vývoj Unicode začal v roce 1987 a byl dovršen v roce 1991 založením Unicode Consortium.

Kódování češtiny (vícebajtové)

UTF-8

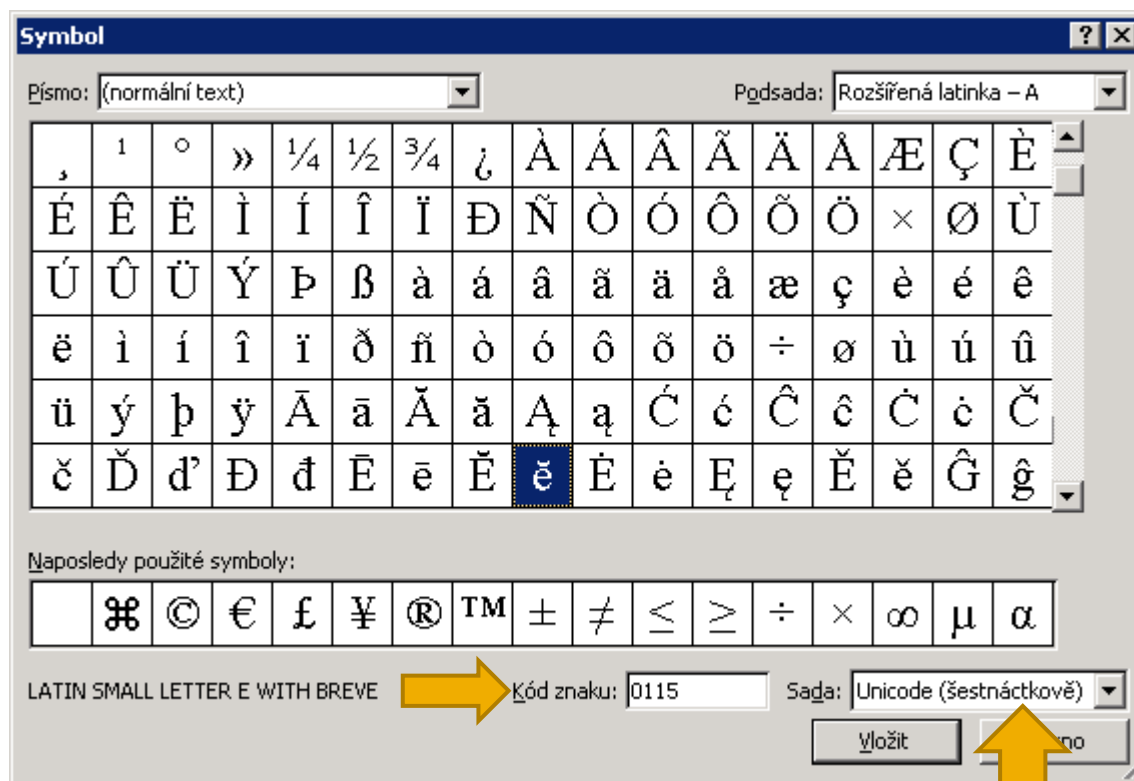
- Zápis UTF-8 je nejčastějším zápisem znakové sady Unicode.
- **Unicode** je narozdíl od výše zmíněných znakových sad určeno pro všechny světové jazyky najednou, protože znakům přiřazuje čísla až do 16 miliónů (zapisuje se většinou **dvěma bajty**).
- Jde o nejmodernější kódování.



Znaky Unicode

- Každý znak má **jednoznačný číselný kód** a svůj název.
- Navíc Unicode definuje u každého znaku některé základní vlastnosti jako např. zda se jedná o písmeno, symbol atd., zda je písmeno velké či malé atp.
- Tabulka Unicode poskytuje prostor pro 1 114 112 znaků s kódy 016 až 10FFFF16.
- Tento prostor se dělí na 17 částí, každý o velikosti 216.
- První část se nazývá **Basic Multilingual Plane (BMP)** a obsahuje znaky běžně používaných abeced.

Kódování v tabulce Symbol



Podrobněji o Unicode

- Existuje několik různých způsobů, jak znaky Unicode kódovat.
- Základní kodování, definovaná přímo ve standardu Unicode, jsou:
- UTF-32,
- UTF-16,
- UTF-8.
 - Kodování UTF-32 a UTF-16 mají každá své varianty podle používaného pořadí bajtů.
 - Bud' je **napevno stanoveno pořadí little-endian**, resp. **big-endian**, nebo se toto pořadí určuje podle tzv. **byte order mark (BOM)**, speciální značky umístěné na začátku textu.

UTF-32

- V kódování UTF-32 (též označováno jako UCS-4) je každý znak reprezentován přímo 32bitovým číslem. Jedná se tedy o principiálně velmi jednoduché kódování (jeho hlavní výhodou je stejná délka všech znaků), které však má poměrně vysoké nároky na paměť.
- Při serializaci do posloupnosti bajtů se podle endianness rozlišují varianty UTF-32BE (big-endian), UTF-32LE (little-endian) a UTF-32 (nestanovené, může být určeno pomocí BOM).

UTF-16

- V kódování UTF-16 se znaky BMP reprezentují jedním 16 bitovým číslem, znaky mimo BMP jsou reprezentovány párem 16bitových čísel (tzv. **surrogate pair**).
- Pro surrogate pairs se používají čísla v rozsahu D80016–DFFF16, přičemž odpovídající znaky v BMP ($U+D800 - U+DFFF$) jsou rezervovány pro tento účel, nemohou se proto v původním textu vyskytnout.
- I zde se podle endianness rozlišují varianty **UTF-16BE** (big-endian), **UTF-16LE** (little-endian) a UTF-16 (nestanovené, může být určeno pomocí **BOM**).
- UTF-16 je **výrazně úspornější** než UTF-32, ale ztrácí jeho výhodu pevné šířky znaku – některé znaky jsou široké 2 bajty, některé 4. Přesto se jedná o kódování **používané jako základní** ve velkém množství operačních systémů a dalšího software (např. **Microsoft Windows**, .NET Framework atd.).

UTF-8

- V UTF-8 se znaky kódují různě dlouhou (1–4 bajty) posloupností bajtů podle jejich pozice v Unicode.
- Znaky ASCII (U+0000 – U+007F) jsou kódovány jedním bajtem, identicky jako v ASCII, znaky v rozsahu U+0080 – U+07FF (kde jsou také **všechny znaky s diakritikou používané v české abecedě**) jsou kódovány dvěma bajty, znaky U+0800 – U+FFFF jsou kódovány třemi bajty, znaky mimo BMP jsou kódovány čtyřmi bajty.
- UTF-8 se **často se používá pro přenos dat**, neboť je prostorově úsporné (znaky běžných písem jsou kódovány krátkými posloupnostmi), odolné proti chybám a zpětně kompatibilní s ASCII. Jeho nevýhodou jsou však odlišné délky jednotlivých znaků.

IT1A3-3 Cvičení 1

- Rozeznajte znakovou sadu ve které je kódován následující text (využijte program, který není součástí MS WINDOWS např. PsPad Editor)
 - text1 – soubor text1.txt
 - text2 – soubor text2.txt
 - text3 – soubor text3.txt
 - text4 – soubor text4.txt
 - text5 – soubor text5.txt
- Soubory si stáhněte a rozzipujte do vhodné složky.
- Najděte soubor Odpovedi.txt a vyplňte jednotlivé řádky. Soubor uložte a odevzdejte.

IT1A3-4 Cvičení 2

- Převeďte (proved'te konverzi) textu (souboru Text1250.txt ve formátu CP1250) do následujících znakových sad
 - UTF8 – uložte jako TextUTF8.txt
 - UNICODE – uložte jako TextUNICODE.txt
 - ISO-8859-2 – uložte jako Text8859.txt
 - ANSI – uložte jako TextANSI.txt
 - UNICODE Big Endian – uložte jako TextUNIENDI.txt
- Pro splnění úkolu si vyberte si vhodný program.
- Všechny soubory odevzdejte.

Samostatně si zdůvodněte odpovědi na otázky

- ANSI je jiné označení pro kódování
 - CP1250
 - UNICODE
 - UTF-8
- V UNICODE se znak kóduje počtem bytů
 - 1B
 - 2B
 - 1B, 2B, 3B nebo 4B

Kódovací tabulky

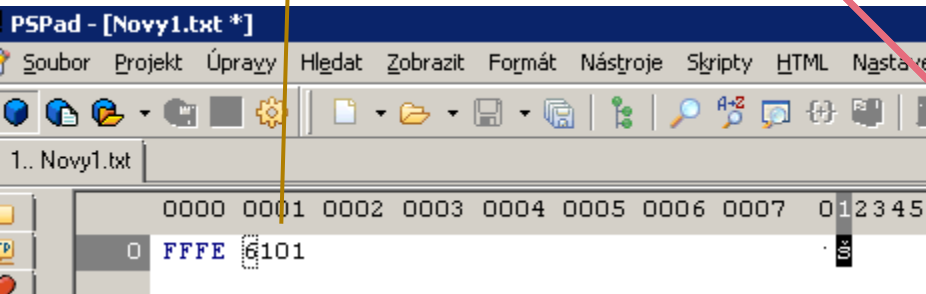
Windows-1250

	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x	Řídící znaky															
1x																
2x	SP	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	ZK
8x	€	NZ	,	NZ	„	...	†	‡	NZ	‰	Š	<	Š	Ť	Ž	Ž
9x	NZ	'	'	"	"	•	—	—	NZ	™	š	>	š	ť	ž	ž
Ax	NBSP	ˆ	ˆ	Ł	×	À	!	Š	ˆ	©	§	«	¬	SHY	®	Ž
Bx	°	±	.	†	'	µ	¶	·	¸	9	§	»	Ł	ˆ	!'	ž
Cx	Ř	Á	Â	Ä	Ä	Ł	Ć	Ç	Č	É	Ę	Ě	Ě	Í	Î	Ď
Dx	Đ	Ň	Ň	Ó	Ô	Ö	Ö	×	Ř	Ů	Ú	Ů	Ů	Ý	Ť	ß
Ex	í	á	â	ä	ä	í	ć	ç	č	é	ę	ě	ě	í	î	d'
Fx	ď	ň	ň	ó	ô	ö	ö	÷	ř	ů	ú	ů	ů	ý	ť	·

Převod Win1250 do UNICODE

- Znak š je kódován (hex.)
- ve 1250 ...**9A** (1byte)
- v UNICODE ...**0161** (2 byte)

POZOR v PSPad se byte zobrazují v opačném pořadí



Windows-1250

Hex	Znak	Unicode	Hex	Znak	Unicode	Hex	Znak	Unicode	Hex	Znak	Unicode
0x80	€	U+20AC	0xA0		U+00A0	0xC0	Ř	U+0154	0xE0	í	U+0155
0x81	NZ		0xA1	˘	U+02C7	0xC1	Á	U+00C1	0xE1	á	U+00E1
0x82	,	U+201A	0xA2	˙	U+02D8	0xC2	Â	U+00C2	0xE2	â	U+00E2
0x83	NZ		0xA3	Ł	U+0141	0xC3	Ã	U+0102	0xE3	ã	U+0103
0x84	„	U+201E	0xA4	»	U+00A4	0xC4	Ä	U+00C4	0xE4	ä	U+00E4
0x85	...	U+2026	0xA5	Ą	U+0104	0xC5	Í	U+0139	0xE5	í	U+013A
0x86	†	U+2020	0xA6	ı	U+00A6	0xC6	Ć	U+0106	0xE6	ć	U+0107
0x87	‡	U+2021	0xA7	§	U+00A7	0xC7	Ç	U+00C7	0xE7	ç	U+00E7
0x88	NZ		0xA8	ˆ	U+00A8	0xC8	Č	U+010C	0xE8	č	U+010D
0x89	%	U+2030	0xA9	©	U+00A9	0xC9	É	U+00C9	0xE9	é	U+00E9
0x8A	Š	U+0160	0xAA	Ş	U+015E	0xCA	Ę	U+0118	0xEA	ę	U+0119
0x8B	<	U+2039	0xAB	«	U+00AB	0xCB	Ê	U+00CB	0xEB	ê	U+00EB
0x8C	Š	U+015A	0xAC	¬	U+00AC	0xCC	Ë	U+011A	0xEC	ë	U+011B
0x8D	Ť	U+0164	0xAD	SHY	U+00AD	0xCD	Í	U+00CD	0xED	í	U+00ED
0x8E	Ž	U+017D	0xAE	®	U+00AE	0xCE	Î	U+00CE	0xEE	î	U+00EE
0x8F	Ž	U+0179	0xAF	Ž	U+017B	0xCF	Ď	U+010E	0xEF	ď	U+010F
0x90	NZ		0xB0	ˆ	U+00B0	0xD0	Ð	U+0110	0xF0	ð	U+0111
0x91	'	U+2018	0xB1	±	U+00B1	0xD1	Ñ	U+0143	0xF1	ñ	U+0144
0x92	'	U+2019	0xB2	˙	U+02DB	0xD2	Ň	U+0147	0xF2	ň	U+0148
0x93	"	U+201C	0xB3	†	U+0142	0xD3	Ó	U+00D3	0xF3	ó	U+00F3
0x94	"	U+201D	0xB4	˙	U+00B4	0xD4	Ô	U+00D4	0xF4	ô	U+00F4
0x95	•	U+2022	0xB5	µ	U+00B5	0xD5	Õ	U+0150	0xF5	õ	U+0151
0x96	—	U+2013	0xB6	¶	U+00B6	0xD6	Ö	U+00D6	0xF6	ö	U+00F6
0x97	—	U+2014	0xB7	˙	U+00B7	0xD7	×	U+00D7	0xF7	×	U+00F7
0x98	NZ		0xB8	ˆ	U+00B8	0xD8	Ř	U+0158	0xF8	ř	U+0159
0x99	™	U+2122	0xB9	ą	U+0105	0xD9	Ů	U+016E	0xF9	ů	U+016F
0x9A	š	U+0161	0xBA	ş	U+015F	0xDA	Ú	U+00DA	0xFA	ú	U+00FA
0x9B	˙	U+203A	0xBB	»	U+00BB	0xDB	Û	U+0170	0xFB	û	U+0171
0x9C	ś	U+015B	0xBC	Ł	U+013D	0xDC	Ü	U+00DC	0xFC	ü	U+00FC
0x9D	ť	U+0165	0xBD	˘	U+02DD	0xDD	Ý	U+00DD	0xFD	ý	U+00FD
0x9E	ž	U+017E	0xBE	ı	U+013E	0xDE	Ť	U+0162	0xFE	ť	U+0163
0x9F	ž	U+017A	0xBF	ž	U+017C	0xDF	ß	U+00DF	0xFF	˙	U+02D9

Kódování 1250 a 8859-2

Srovnání kódů

Hex	CP-1250	ISO 8859-2	Hex	CP-1250	ISO 8859-2	Hex	CP-1250	ISO 8859-2	Hex	CP-1250	ISO 8859-2
0x80	€	ŘZ	0xA0			0xC0		Ř	0xE0		ř
0x81	NZ	ŘZ	0xA1	˘	Ą	0xC1		Á	0xE1		á
0x82	,	ŘZ	0xA2		ˆ	0xC2		Â	0xE2		â
0x83	NZ	ŘZ	0xA3		Ł	0xC3		Ã	0xE3		ã
0x84	„	ŘZ	0xA4		»	0xC4		Ä	0xE4		ä
0x85	...	ŘZ	0xA5	Ą	Ł	0xC5		Í	0xE5		í
0x86	†	ŘZ	0xA6	!	Ś	0xC6		Ó	0xE6		ó
0x87	‡	ŘZ	0xA7		§	0xC7		Ç	0xE7		ç
0x88	NZ	ŘZ	0xA8		˙	0xC8		Č	0xE8		č
0x89	‰	ŘZ	0xA9	©	Š	0xC9		É	0xE9		é
0x8A	Š	ŘZ	0xAA		Ş	0xCA		Ê	0xEA		ê
0x8B	<	ŘZ	0xAB	«	Ť	0xCB		Ë	0xEB		ë
0x8C	Ś	ŘZ	0xAC	¬	Ž	0xCC		Ě	0xEC		ě
0x8D	Ť	ŘZ	0xAD			0xCD		Í	0xED		í
0x8E	Ž	ŘZ	0xAE	®	Ž	0xCE		Î	0xEE		î
0x8F	Ž	ŘZ	0xAF		Ż	0xCF		Ď	0xEF		ď
0x90	NZ	ŘZ	0xB0		˚	0xD0		Đ	0xF0		đ
0x91	'	ŘZ	0xB1	±	ą	0xD1		Ň	0xF1		ň
0x92	'	ŘZ	0xB2		˛	0xD2		Ñ	0xF2		ñ
0x93	"	ŘZ	0xB3		ı	0xD3		Ó	0xF3		ó
0x94	"	ŘZ	0xB4		ˆ	0xD4		Ô	0xF4		ô
0x95	•	ŘZ	0xB5	µ	İ	0xD5		Õ	0xF5		õ
0x96	—	ŘZ	0xB6	¶	ś	0xD6		Ö	0xF6		ö
0x97	—	ŘZ	0xB7	·	˘	0xD7		×	0xF7		×
0x98	NZ	ŘZ	0xB8		˙	0xD8		Ř	0xF8		ř
0x99	™	ŘZ	0xB9	ą	ş	0xD9		Ů	0xF9		ů
0x9A	š	ŘZ	0xBA		ş	0xDA		Ú	0xFA		ú
0x9B	>	ŘZ	0xBB	»	ť	0xDB		Ů	0xFB		ů
0x9C	ś	ŘZ	0xBC	Ł	ż	0xDC		Ů	0xFC		ů
0x9D	ť	ŘZ	0xBD		˙	0xDD		Ý	0xFD		ý
0x9E	ż	ŘZ	0xBE	İ	ž	0xDE		Ť	0xFE		ť
0x9F	ż	ŘZ	0xBF		ż	0xDF		ß	0xFF		·

