

## Kapitola: Zobrazení dat v počítači

### Cvičení 6 – Kódování národních znaků

Témata: Plain text a jednobytová kódování, vícebytová kódování. Kódové tabulky. Binární zobrazení souboru. Rozpoznání kódu souboru podle tabulek nebo experimentálně. Změna kódování – `cstocs`, editor, webový prohlížeč.

#### Co máme znát

Uložení znaků v počítači, znakový a řetězcový datový typ jazyka Pascal. Pojmy **znakový kód**, **ASCII**, **textový formát dat**, **obyčejný text**, **rozšířený text**, **kódování národních znaků**.

Umět pracovat s webovým prohlížečem, s účtem na serveru akela, s příkazovým řádkem ve Windows, s editorem PSPad, s editorem Notepad.

#### Ověřte, že rozumíte vstupním pojmům a vztahům odpověďmi na tyto otázky:

- Co je to *znakový kód*?
- Co je to *jednobytové kódování znaků*?
- Kolik znaků zobrazuje základní kód *ASCII*?
- Co jsou to *řídící znaky*? Na kterých pozicích se nacházejí?
- Co je to *zobrazitelný znak*?
- Na kterých pozicích je mezera, číslice, velká písmena, malá písmena v kódu *ASCII*?
- Jak lze charakterizovat *textový formát dat*?
- Co je *obyčejný* a co *rozšířený text*?
- Co je to *vícebytové kódování národních znaků*? Jaké vícebytové kódy znáte?
- V jakém kódování se zobrazují národní znaky v příkazovém řádku na serveru akela?
- V jakém kódování jsou zobrazeny národní znaky v okenních aplikacích Windows?
- V jakém kódování jsou zobrazeny národní znaky v příkazovém řádku Windows?

#### Materiál

**Pomůcky:** kódové tabulky *ASCII*, jednobytových kódování národních znaků, Unicode.

**Programy:** Editor PSPad, příkaz `od`, webový prohlížeč, příkaz `cstocs`, příkazový řádek Windows.

**Soubory:** Sada souborů s různým kódováním národních znaků.

#### Úkoly

1. Zjistěte, jak pomocí kódových tabulek jednoduše přepočítáte šestnáctkové číslo v rozsahu 0–255 na desítkovou hodnotu a opačně.
2. Předpokládejte kód znaku A7 (hex.). Vypište, jakému znaku tento kód odpovídá v různých kódováních.
3. **Určení kódování souboru I.**  
Zobrazte soubor `text.c1` hexadecimálně v editoru PSPad. Zjistěte, jaký znak má být na 9. bytu, má-li se jednat o slovo „vyšetření“. Zjistěte kód tohoto znaku. Najděte tabulku, v níž se pod nalezeným kódem nachází správný znak.
4. U některých znaků se může stát, že naleznete více tabulek, v nichž je pod stejným kódem stejný správný znak. V tom případě nelze na základě tohoto znaku jednoznačně určit kódování. Najděte alespoň dvě dvojice kódů, které mají některé stejně kódované národní znaky.

**5. Určení kódování souboru II.**

Experimentální určení kódu podle prostředí – v příkazovém řádku na serveru akela zobrazte obsah souborů `text.c1` až `text.c5` a podle výsledku určete kódování alespoň jednoho z nich.

6. Předchozí úlohu proveďte v příkazovém řádku Windows.

7. Předchozí úlohu proveďte zobrazením souborů `text.c1` až `text.c5` pomocí editoru Notepad.

8. Seznamte se s programem `cstocs` na serveru akela – spusťte jej bez parametrů a zjistěte, jak se zadávají různá kódování, vstupní soubor a kam je směřován výstupní překódovaný text.

**9. Určení kódování souboru III.**

Pomocí programu `cstocs` určete experimentálně kódování některého ze souborů `text.c1` až `text.c6`.

**10. Určení kódování souboru IV.**

Seznamte se s možností nastavení kódování souboru v editoru PSPad. Vhodným nastavením kódování dokažte správnost zjištěných výsledků z předchozích úloh na souborech `text.c1` až `text.c5`.

**11. Vícebytová kódování**

Zobrazte binární (hexadecimální) podobu souborů `text.c6` až `text.c8`. Rozhodněte, v jakých kódováních pravděpodobně jsou.

12. Zjistěte v binárním zobrazení, jaký je rozdíl mezi soubory `text.c7` a `text.c8`.

13. Zkuste možnosti zobrazení souborů `text.c7` a `text.c8` v příkazovém řádku a v editoru Notepad.

**14. Určení kódování souboru V.**

Zobrazte vybraný soubor webovým prohlížečem a určete jeho kódování vhodným nastavením zobrazení.

**15. Překódování**

Zjistěte možnosti změny kódování národních znaků v editoru Notepad, ve webovém prohlížeči a v editoru PSPad. Zaznamenejte do přehledné tabulky, jaká kódování jsou těmito prostředky podporována a porovnejte možnosti s programem `cstocs`. Do poznámky v tabulce запиšte název souboru, který obsahuje dané kódování.

Kód	Notepad	Web	PSPad	cstocs	Poznámka
kam					
pc2					
koi8					
il2					
1250					
utf8					
utf16					

**Shrnutí a kontrolní otázky**

1. Porovnejte možnosti, výhody a nevýhody pěti uvedených způsobů zjištění kódu národních znaků v zadaném souboru.
2. Jaké kódování národních znaků je správně zobrazitelné v příkazovém řádku Windows?
3. Jaké kódování národních znaků je správně zobrazitelné editorem Notepad?
4. Jaká jsou podle vašeho názoru nejčastější kódování národních znaků z uvedených 8 příkladů?

5. Jaká kódování jsou považována za perspektivní a proč?
6. Jaká kódování národních znaků jsou dnes již prakticky opuštěna a z jakého důvodu?
7. Jaká kódování jsou ještě používána, ale jejich význam již klesá a proč?
8. Jaké kódování národních znaků není v příkladech zastoupeno?