

Kapitola: Zobrazení dat v počítači

Cvičení 8 – Konverze dokumentních formátů

Témata: dokument a jeho formy, konverze, formáty RTF, XML, HTML, MIME, DOC, text. Kódování národních znaků v různých prostředích.

Co máme znát

Textový a binární souborový formát, kódování národních znaků, pojem dokument, pojem neformátovaný text, rozpoznávání souborového formátu, pojem konverze, standardní dialog pro otevření a uložení souboru, práce se styly v programu Word.

Umět pracovat s účtem na serveru akela, s příkazovým řádkem ve Windows, s programem Word.

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

- Jak můžeme pracovně definovat pojem *dokument*?
- V jakém tvaru mohou být v dokumentu vloženy formátovací značky?
- Jaké dokumenty v textovém formátu znáte?
- Jaké dokumenty v binárním formátu znáte?
- Co znamená *konverze souborového formátu*?
- Jaké služby programového vybavení se používají pro konverze souborových formátů?
- Jaké specializované konverzní programy znáte?
- Jakým konverzním programem lze změnit kódování národních znaků?
- V jaké položce standardního dialogu pro otevírání nebo ukládání souborů určujeme formát vstupního nebo výstupního souboru?
- Co je to *asociace* souborového formátu a aplikace v prostředí operačního systému typu Windows?
- Zjistěte v asociační tabulce, jakou aplikací bude zpracován soubor s rozšířením XML.

Materiál

Programy: Program MS Word (verze 2003), editor PSPad, příkaz `od`, příkaz `more`, příkazový řádek Windows.

Soubory: Vstupní cvičné soubory.

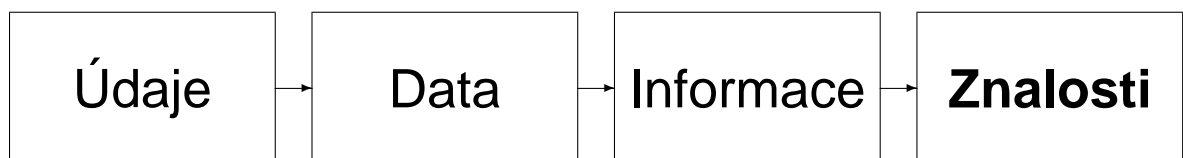
Úkoly

1. Cvičné soubory rozbalte do připraveného adresáře, jehož obsah lze sledovat jak z unixového, tak z windowsového prostředí (sdílený disk `h:`).
2. V průběhu práce budeme vyplňovat tabulku se záznamy o vlastnostech zdrojového a cílového formátu pro každou konverzi.
3. Spusťte program Word a zjistěte, jaké vstupní formáty umožňuje zpracovat.
4. Zjistěte, jaké výstupní formáty je schopen program Word generovat.
5. Otevřete v programu Word soubor `vstup1.doc`. Prohlédněte si jeho obsah a formátování.
6. Konvertujte soubor `vstup1.doc` postupně do formátu XML, kompletní webová stránka, zjednodušená webová stránka, webový archiv, RTF. Výsledné soubory pojmenujte podle vzoru v tabulce.
7. Prozkoumejte důkladně výsledky uvedených pěti konverzí. Při které konverzi dochází ke ztrátě některých informací? Upozorní na tuto skutečnost program Word? Jakým způsobem?

Tabulka 1: Tabulka pro zápis výsledků

Soubor	Text (P, E)/Bin	Délka	Konvertován na	Text (P, E)/Bin	Délka
vstup1.doc			vystup11.xml vystup12.html vystup13.html vystup14.mht vystup15.rtf vystup16.txt vystup17.txt		
vstup2.txt			vystup21.rtf vystup22.doc		
vstup3.rtf			vystup31.rtf vystup32.doc		

8. Vyšetřete formát EMZ. V asociační tabulce zjistíte, jakou aplikací jej lze otevřít, proveďte jeho zpracování a zobrazte v odpovídajícím programu jeho obsah.
9. Do původního dokumentu vstup1.doc dokreslete pomocí nástrojové lišty Kreslení následující schéma:



Zjistěte, v jakém obrazovém formátu můžete toto schéma získat po konverzi.

10. Konvertujte dokument vstup1.doc do neformátované textové podoby, a to pro použití v příkazovém řádku Windows a pro použití v systému Unix na serveru akela. Vyzkoušejte efekt voleb „Přidat konce řádků“ a „Povolit nahrazování znaků“. Zkontrolujte tvar a správnost obou výsledků zobrazením obsahu každého souboru v příslušném prostředí, v binárním zobrazení zkontrolujte správné tvary konců řádků.
11. Otevřete v programu Word soubor vstup2.txt. Je jeho obsah interpretován správným způsobem?
12. Proveďte potřebné úkony k tomu, aby se soubor vstup2.txt podařilo správně otevřít.
13. Proveďte potřebné náhrady, aby místo konců odstavců byly jen mezislovní mezery a místo prázdných řádků konce odstavců.
14. Definujte sazební styl **Základ** s parametry: písmo Palatino Linotronic 12 pt, sazba do bloku, odstavcová zarážka 24 pt, řádkování přesně 15 pt. Přiřaďte tento styl celému textu.
15. Konvertujte upravený soubor vstup2.txt do souboru formátu RTF a DOC. Porovnejte délky souborů formátu TXT, RTF a DOC pro soubor vstup1 a pro soubor vstup2.

16. Prohlédněte si v příkazovém řádku Windows nebo Unix soubor `vstup3.rtf`. V jakém kódování národních znaků je soubor zapsán?
17. Otevřete programem Word soubor `vstup3.rtf`. Ověřte, že všechny národní znaky jsou zobrazeny správně. Proveďte uložení tohoto souboru ve formátu RTF (pozor na změnu jména souboru) a ve formátu DOC. Porovnejte délky souborů `vstup3.rtf`, `vystup31.rtf` a `vystup32.doc`. Vysvětlete zjištěné hodnoty.

Shrnutí a kontrolní otázky

1. Jaké souborové formáty je možné číst nebo generovat programem Word?
2. Které formáty dokumentů ukládají obrazová data úsporným způsobem?
3. Které formáty dokumentů obsahují kompletní informace o dokumentu?
4. Jakým způsobem lze z dokumentu programu Word získat obrázek v některém obecném přenositelném obrazovém formátu?
5. Jaké možnosti má program Word pro čtení neformátovaných textů a jaké pro jejich generování?
6. Jakým postupem lze ze vstupního neformátovaného textu vytvořit základ pro správně vytvořený dokument?
7. Je formát RTF vždy ve formě obyčejného textu (plain text)?