

Kapitola: Zobrazení dat v počítači

Cvičení 7 – Rozpoznávání formátu souborů

Témata: binární souborový formát, textový souborový formát, rozpoznávání souborového formátu, obyčejný text, rozšířený text, konec řádku a jeho reprezentace v různých systémech, frekventované konkrétní souborové formáty.

Co máme znát

Textový a binární souborový formát, obyčejný a rozšířený text, kódování národních znaků, binární zobrazení souboru, některé konkrétní souborové formáty, pojem dokument, vztah souborového formátu a operačního systému.

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

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

- Co je to souborový formát dat?
- Jak lze pracovně definovat *textový souborový formát*?
- Jak lze poznat podle řídicích znaků pro konec řádku operační systém, v němž textový souborový formát vznikl?
- Jakým způsobem lze zobrazit *binární obsah* souboru v prostředí systému Unix?
- Jakým programem lze zobrazit *binární obsah* souboru v prostředí systému Windows?
- Jakým příkazem lze zobrazit *textový obsah* souboru v prostředí systému Unix a jakým v prostředí Windows?
- Co je to *plain text* (obyčejný text)?
- Co je to *extended text* (rozšířený text)?
- Co víte o souborových formátech CSV, PS/EPS, HTML, RTF, TXT, PAS, XML, SVG?
- Co víte o souborových formátech CDR, DOC, DOCX, JPG, XLS, XLSX, PNG, PDF, ZIP?
- Co je to dokument? Které z uvedených formátů chápeme jako dokumenty?
- Co je to *asociace formátu s aplikací* v systému Windows?

Materiál

Programy: Editor PSPad, příkaz `od`, příkaz `file`, příkaz `more`, příkazový řádek Windows.

Soubory: Sada souborů různých formátů.

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 cvičení budeme vyplňovat následující tabulku. U každého cvičného souboru bude určeno, zda je textový (obyčejný, nebo rozšířený), nebo binární, z jakého operačního systému pochází, obsah souboru, konkrétní formát a která aplikace je tomuto formátu přiřazena.

Soubor	Text (P, E)/Bin	Původ v OS	Obsah	Formát – aplikace
soubor.01				
soubor.02				
soubor.03				
soubor.04				
soubor.05				
soubor.06				
soubor.07				
soubor.08				
soubor.09				
soubor.10				
soubor.11				
soubor.12				
soubor.13				
soubor.14				

- U všech souborů rozhodněte, zda jsou binární, nebo textové. Vlastnost zjistíte příkazem `od`, zobrazením v hexadecimálním režimu editoru PSPad, případně zobrazením v editoru Notepad. Porovnejte uvedené metody zjišťování.
- U všech textových souborů analyzujte konce řádků a podle toho vyplňte sloupec „Původ v OS“. U binárních souborů tuto položku zatím nevyplňujte.
- Pokuste se zjistit formát binárních souborů zobrazením jejich začátku v binární podobě.
- Pokuste se zjistit formát binárních souborů pomocí příkazu `file`. Porovnejte výsledky těchto dvou zjišťování.
- Do sloupce „Obsah“ v tabulce doplňte informaci, zda se jedná o dokument.
- Soubor formátu MIME otevřete v programu Word (vstupní typ = mht). Podle zobrazené podoby prozkoumejte, jak jsou v tomto souboru uloženy jednotlivé části dokumentu.
- Soubor formátu CSV otevřete v programu Excel. Jaké oddělovače polí a jaké kódování národních znaků soubor CSV má? Proč jsou v prvním řádku souboru použity uvozovky?
- Dekomprimujte soubory, u nichž příkaz `file` zjistil formát ZIP, a prozkoumejte strukturu, která tím vznikla.
- Z asociační tabulky na Vašem počítači přečtěte, jaké aplikace jsou přiřazeny rozpoznávaným formátům cvičných souborů. Do posledního sloupce tabulky doplňte zjištěné údaje.
- Soubory, které jste ještě neotevřeli a k nimž existují v asociační tabulce aplikace, si těmito aplikacemi prohlédněte. Do sloupce „Obsah“ v tabulce stručně doplňte, co soubor obsahuje (celkem je ve 12 souborech jen 5 různých obsahů).

Shrnutí a kontrolní otázky

1. Jaký je možný postup zjištění souborového formátu v systému typu Unix?
2. Jaký je možný postup zjištění souborového formátu v systému typu Windows?
3. Vysvětlete zkratku MIME. Kde se používá formát MIME a jaké má vlastnosti?
4. Jak poznáte archiv typu ZIP?
5. Jak poznáte soubor formátu PDF?
6. Jak poznáte obrázek ve formátu PNG?
7. Lze pomocí uvedených prostředků jednoznačně rozpoznat každý binární formát? Pro své tvrzení uveďte příklady.
8. Na jakém principu jsou založeny dokumentní formáty kancelářského balíku Office 2007? Jak jsou reprezentována veškerá data v těchto formátech?
9. Jaké význačné vlastnosti má formát RTF?
10. Které datové formáty jsou založeny na principu XML?