



## NÁSTROJE PRE PRÁCU SO SÚBORMI

### NÁSTROJE NA ÚPRAVU TEXTOVÝCH SÚBOROV



#### Teoretická časť

V OS Linux existuje množstvo nástrojov na manipuláciu s textom. Všetky z nasledujúcich programov *tr*, *sort*, *uniq*, *cut*, *paste*, *diff*, *cmp* a *grep* nemenia zdrojový súbor, ale spracovaný text posielajú na štandardný výstup (obrazovka). Ak chceme, aby výsledkom bol nový súbor, je potrebné výstup presmerovať pomocou znaku `>`, alebo použiť príslušný prepínač. Všetky z uvedených programov okrem nástroja *tr* načítavajú vstup zo súboru.

Program *tr* slúži na nahradzovanie alebo mazanie znakov. Vstup mu treba načítať napr. presmerovaním vstupu **`tr < subor ...`** alebo pomocou programu *cat* a „rúry“, t. j. **`cat subor | tr ...`**. Ak chceme napr. v súbore *file01* nahradiť všetky znaky *a* znakom *b* a výsledný súbor uložiť do súboru *file02* použijeme **`tr < file01 a b > file02`**. Program *tr* dokáže nahradzovať len jednotlivé znaky. Ak by sme chceli nahradzovať celé slová, môžeme použiť napr. program *sed* (práca s ním ale nie je pre začiatočníkov).

Program *sort* slúži na triedenie obsahu súboru, ak má súbor stĺpce, tak aj podľa vybraného stĺpca. Program *cut* slúži na vystrihovanie stĺpcov. Ako štandardný oddeľovač stĺpcov pre obidva programy slúži *Tab*.

Program *uniq* odstraňuje duplicitné riadky idúce za sebou.

Program *diff* porovnáva obsah dvoch súborov. Na obrazovku vypíše riadky, ktoré sú v obidvoch súboroch rozdielne. Program *cmp* spustený s prepínačmi *-bl* vypíše presne, v ktorých bajtoch je rozdiel.

Použitie tohto učebného materiálu je určené výhradne pre Duálne vzdelávanie realizované SPŠ elektrotechnickou Košice v spolupráci s Deutsche Telekom IT Solutions Slovakia.

Autor: J. Ploščica  
Verzia 3

Predmet: ZIL, 1. ročník  
Strana 1 z 8

Program *paste* vypisuje obsah dvoch súborov vedľa seba. Môžeme ho použiť ako alternatívu k *diff* na hľadanie rozdielov v súboroch, alebo ako opak programu *cut* – spájať pomocou neho viac súborov do súboru so stĺpcami.

Program *grep* (Global Regular Expression Print) vyhľadáva v súbore riadky obsahujúce zadaný reťazec.



## Pomôcky

Virtuálny stroj s CentOS 7 vytvorený vo VMware vSphere. Úloha je určená pre jedného žiaka.



## Úlohy

1. Zapnite virtuálny stroj s CentOS 7 a prihláste sa do jeho grafického režimu ako používateľ vytvorený počas inštalácie, v tomto návode to bude bežný používateľ *kohut*.
2. Otvorte si pseudoterminál, v domovskom priečinku používateľa *kohut* vytvorte nový priečinok, presuňte sa doňho a vytvorte v ňom súbor *udaje* s nasledujúcim obsahom:

<i>Vrabec</i>	<i>Jan</i>	<i>1995</i>	<i>Kosice</i>
<i>Straka</i>	<i>Milan</i>	<i>1984</i>	<i>Presov</i>
<i>Stehlik</i>	<i>Juraj</i>	<i>1999</i>	<i>Vranov</i>
<i>Sykora</i>	<i>Peter</i>	<i>2001</i>	<i>Zvolen</i>
<i>Motyl</i>	<i>Milan</i>	<i>19733</i>	<i>Nitra</i>

Ako oddeľovač použite *Tab.* (Preklep v roku narodenia u *Motyla* je zámerný).

3. Pomocou *tr* nahradzte všetky znaky *a* v súbore znakom *b*. Výstup zobrazte na obrazovku.
4. Nahradzte znak *a* znakom *b* a znak *c* znakom *d*. Výstup zobrazte na obrazovku. Použite príkaz ***tr < udaje ac bd***.
5. Zistite aký je výsledok, ak v skupinách znakov nie je rovnaký počet znakov, t. j. napr. ***tr < udaje ac bde*** alebo ***tr < udaje ace bd***.
6. Nahradzte znaky *a b c* znakom *!*. Výstup zobrazte na obrazovku.

Použitie tohto učebného materiálu je určené výhradne pre Duálne vzdelávanie realizované SPŠ elektrotechnickou Košice v spolupráci s Deutsche Telekom IT Solutions Slovakia.

7. Nahradzte všetky malé písmena veľkými. Použite postupne príkazy: **tr < udaje a-z A-Z** a **tr < udaje [:lower:] [:upper:]**. Prezrite si manuálovú stránku programu *tr*, všimnite si, ako sa dajú vyjadriť nasledujúce množiny znakov:
- všetky číslice
  - všetky písmená
  - všetky alfanumerické znaky
  - kláves *Backspace*
  - kláves *Enter (new line)*
  - kláves *Tab (horizontal tab)*
8. Nahradzte všetky číslice znakom *?*. Výstup zobrazte na obrazovku.
9. Nahradzte všetky písmená (malé aj veľké) znakom *#*. Je možné, že budete musieť použiť úvodzovky. Výstup zobrazte na obrazovku.
10. Nahradzte všetky písmená aj číslice znakom *@*. Výstup zobrazte na obrazovku.
11. Odstráňte zo súboru všetky číslice 9. Použite príkaz **tr < udaje -d 9**. Výstup zobrazte na obrazovku.
12. Odstráňte zo súboru viacnásobný výskyt číslíc 9. Použite príkaz **tr < udaje -s 9**. Výstup zobrazte na obrazovku.
13. Opravte súbor *udaje* (odstráňte viacnásobný výskyt čísla 3 v čísle 19733). Výstup presmerujte do súboru s názvom *opraveneudaje*.
14. Nahradzte oddeľovač *Tab* v súbore *opraveneudaje* oddeľovačom *:* a výstup presmerujte do súboru *opraveneudajesdvojbodkou*. Použite príkaz **tr < opraveneudaje "\t" ":" > opraveneudajesdvojbodkou**.
15. Analogickým spôsobom vytvorte súbory *opraveneudajesmedzerou* a *opraveneudajesbodkociarkou*.
16. Zistíte, čo urobí príkaz **tr < opraveneudaje "\t" "\b"**.
17. Zistíte, čo urobí príkaz **tr < opraveneudaje "\n" -**.
18. Nahradzte v súbore *opraveneudaje* meno *Motyl* menom *Mucha*. Najskôr použite príkaz **tr < opraveneudaje Motyl Mucha**; výsledok by vás nemal uspokojiť, preto použite príkaz **sed s/Motyl/Mucha/ opraveneudaje**.
19. Zopakujte predchádzajúcu úlohu ale výstup uložte do súboru *udajesmuchom*.
20. Zobrazte si výpis bežiacich procesov príkazom **ps -ef**. Aby bol výpis kratší, výstup pošlite programu *head*, t. j. použite príkaz **ps -ef | head**. Výpis obsahuje

stĺpce; upravte ho tak, aby boli stĺpce oddelené bodkočiarkou (V pôvodnom výpise sú stĺpce oddelené viacerými medzerami; najskôr odstráňte ich viacnásobný výskyt a až potom nahradte medzeru bodkočiarkou).

21. Zoradte súbor *opraveneudaje* v abecednom poradí podľa prvého stĺpca. Môžete použiť príkaz **sort opraveneudaje** alebo **sort -k 1 opraveneudaje**.
22. Zoradte súbor *opraveneudaje* v abecednom poradí podľa druhého stĺpca. Použite príkaz **sort -k 2 opraveneudaje**.
23. Zoradte súbor *opraveneudajesdvojbodkou* v abecednom poradí podľa druhého stĺpca reverzne. Tento krát musíte zadať navyše aj to, že oddeľovačom je dvojbodka. Použite príkaz **sort -k 2 -r -t “:” opraveneudajesdvojbodkou**.
24. Zoradte súbor *opraveneudajesmedzerou* v abecednom poradí podľa miesta narodenia. Výstup presmerujte do súboru *udajepodlamiestanarodenia*. Na presmerovanie použite znak **>**.
25. Zoradte súbor *opraveneudajesbodkociarkou* podľa roku narodenia, nech výpis začína najmladším človekom. Výstup presmerujte do súboru *udajepodlaveku*. Na presmerovanie použite prepínač **-o**.
26. Zoradte pôvodný súbor *udaje* podľa roku narodenia, použite postupne príkazy **sort -k 3 udaje** a **sort -k 3 -n udaje**. Interpretujte rozdiel v jednotlivých výpisoch.
27. Vystrihnite zo súboru *opraveneudaje* prvý a tretí stĺpec. Použite príkaz **cut -f 1,3 opraveneudaje**.
28. Vystrihnite zo súboru *opraveneudajesmedzerou* druhý a štvrtý stĺpec. Keďže medzera nie je štandardný oddeľovač, musíte zadať aj oddeľovač. Použite príkaz **cut -f 2,4 -d “ ” opraveneudajesmedzerou**.
29. Pomocou programov *cut* a *sort* vystrihnite zo súboru *opraveneudajesbodkociarkou* iba krstné mená, zoradte ich podľa abecedy, pritom odstráňte aj viacnásobný výskyt mien. Správny prepínač programu *sort* nájdite v manuálových stránkach, prípadne použite program *uniq*. Výsledok musí vyzeráť nasledovne:

*Jan*

*Juraj*

*Milan*

*Peter*

30. Pomocou programu *cut* vystrihnite zo súboru *opraveneudajesdvojbodkou* v každom riadku tretí, piaty a desiaty znak. Použite príkaz **cut -c 3,5,10 opraveneudajesdvojbodkou**.
31. Pomocou programu *cut* vystrihnite zo súboru *opraveneudajesmedzerou* v každom riadku štvrtý až jedenásty znak. Použite príkaz **cut -c 4-11 opraveneudajesmedzerou**.
32. Pomocou programu *cut* vystrihnite zo súboru *opraveneudajesbodkociarkou* v každom riadku znaky od začiatku až po deviaty znak. Môžete použiť príkaz **cut -c 1-9 opraveneudajesbodkociarkou** alebo **cut -c -9 opraveneudajesbodkociarkou**.
33. Pomocou programu *cut* vystrihnite zo súboru *opraveneudaje* v každom riadku znaky od piateho až do konca riadku. Výstup zobrazte na obrazovku.
34. Pomocou programu *diff* porovnajte obsah súborov *udaje opraveneudaje*. Použite príkaz **diff udaje opraveneudaje**. Použite ten istý príkaz s prepínačom *-y*. Interpretujte výstupy týchto príkazov. Prezrite si manuálové stránky programu *diff*.
35. Obsah súborov *udaje opraveneudaje* porovnajte pomocou programu *cmp*. Najskôr použite príkaz **cmp udaje opraveneudaje**, potom príkaz **cmp -bl udaje opraveneudaje**.
36. Namiesto programov *diff* a *cmp* použite na manuálne vyhľadanie rozdielov v súboroch *udaje* a *opraveneudaje* program *paste*. Obidva súbory vypíšte na obrazovku vedľa seba príkazom **paste udaje opraveneudaje**.
37. Pomocou programu *diff* porovnajte obsah súborov *opraveneudaje* a *udajesmuchom*.
38. Prekopírujte do pracovného priečinka súbor */etc/services*. Pomocou programu *split* ho rozdeľte na 3 časti. Tieto 3 časti spojte pomocou programu *cat* do súboru s názvom *sluzby*. Pomocou programu *diff* porovnajte súbory *services* a *sluzby*. Medzi súbormi *services* a *sluzby* by nemal byť rozdiel.
39. Pomocou programu *cut* vystrihnite postupne všetky stĺpce zo súboru *opraveneudaje* a uložte ich do súborov s názvami *priezvisko*, *meno*, *roknarodenia*, *miestonarodenia*.

40. Pomocou programu *paste* vytvorte zo súborov *priezvisko*, *meno*, *roknarodenia*, *miestonarodenia* súbor *podlaroku*, ktorý obsahuje stĺpce v poradí *roknarodenia*, *miestonarodenia*, *priezvisko*, *meno* oddelené štandardným oddeľovačom *Tab*.
41. Pomocou programu *paste* vytvorte zo súborov *priezvisko*, *meno*, *roknarodenia*, *miestonarodenia* súbor *podlamiesta*, ktorý obsahuje stĺpce v poradí *miestonarodenia*, *priezvisko*, *meno*, *roknarodenia* oddelené dvojbodkou.
42. Pomocou programu *grep* vypíšte zo súboru *opraveneudaje* všetky riadky, v ktorých sa vyskytuje reťazec *Motyl*. Použite príkaz **grep Motyl opraveneudaje**.
43. Overte, že program *grep* rozlišuje veľké a malé písmená, najskôr použite príkaz **grep motyl opraveneudaje**; výpis nebude obsahovať žiadny riadok. Pomocou prepínača *-i*, t. j. príkazom **grep -i motyl opraveneudaje** potlačte v programe *grep* rozlišovanie malých a veľkých písmen.
44. Pomocou programu *grep* vypíšte zo súboru *opraveneudaje* všetky riadky, v ktorých sa nevyskytuje reťazec *Motyl*. Použite príkaz **grep -v Motyl opraveneudaje**.
45. Príkazom **grep Moty opraveneudaje** vypíšte zo súboru *opraveneudaje* všetky riadky, ktoré obsahujú reťazec *Moty*; program vypíše riadok obsahujúci meno *Motyl*. Príkazom **grep -w Moty opraveneudaje** vypíšte zo súboru *opraveneudaje* všetky riadky, ktoré obsahujú slovo *Moty*; výpis nebude obsahovať žiadny riadok. Príkazom **grep -w Motyl opraveneudaje** vypíšte zo súboru *opraveneudaje* všetky riadky, ktoré obsahujú slovo *Motyl*, výsledkom bude jeden riadok.
46. Pomocou programu *grep* vypíšte zo súboru *opraveneudaje* všetky riadky, v ktorých sa vyskytuje reťazec *Milan*. Použite príkaz **grep Milan opraveneudaje**.
47. Pomocou programu *grep* vypíšte zo súboru *opraveneudaje* všetky riadky, v ktorých sa vyskytuje reťazec *Milan* aj s číslami riadkov. Použite príkaz **grep -n Milan opraveneudaje**.
48. Pomocou programu *grep* spočítajte v súbore *opraveneudaje* všetky riadky, v ktorých sa vyskytuje reťazec *Milan*. Najskôr použite príkaz **grep Milan opraveneudaje | wc -l**, potom príkaz **grep -c Milan opraveneudaje**.
49. Pomocou programu *grep* vypíšte zo súboru *opraveneudaje* všetky riadky, v ktorých sa vyskytuje reťazec *Stehlik*. Použite príkaz **grep Stehlik opraveneudaje**.

50. Pomocou programu *grep* vypíšte zo súboru *opraveneudaje* všetky riadky, kde sa vyskytuje reťazec *Stehlik* spolu s jedným riadkom, ktorý sa nachádza nad nimi. Použite príkaz **grep -A 1 Stehlik opraveneudaje**.
51. Pomocou programu *grep* vypíšte zo súboru *opraveneudaje* všetky riadky, kde sa vyskytuje reťazec *Stehlik* spolu s dvoma riadkami, ktorý sa nachádzajú pred nimi. Použite príkaz **grep -B 2 Stehlik opraveneudaje**.
52. Pomocou programu *grep* vypíšte zo súboru *opraveneudaje* všetky riadky, kde sa vyskytuje reťazec *Stehlik* spolu s jedným riadkom pred a jedným riadkom po danom riadku. Použite príkaz **grep -C 1 Stehlik opraveneudaje**.
53. V konfiguračných súboroch sa často vyskytujú komentáre, t. j. riadky, ktoré majú na začiatku znak *#*. Presuňte sa do priečinka */etc* a potom si zobrazte obsah nejakého súboru z tohto priečinka, napr. obsah súboru *login.defs* príkazom **cat login.defs**. Vypíšte si obsah súboru *login.defs* bez komentárov; použite príkaz **grep -v “#” login.defs**.
54. Prezrite si obsah súboru */etc/services*; v tomto súbore sú okomentované aj jednotlivé sieťové služby. Vypíšte si obsah súboru */etc/services* bez riadkov, ktoré začínajú znakom *#*; použite príkaz **grep -v ^# services**, prípadne **grep -v ^# services | less**.
55. Preskúmajte riadky s výskytmi slov *http*, *HTTP*, *https* a *HTTPS* v súbore */etc/services*. Postupujte podľa nasledujúcich pokynov:
- Najskôr použite príkazy **grep http services**, **grep HTTP services** a **grep -i http services**. Potom použite príkazy **grep http services | wc -l**, **grep HTTP services | wc -l** a **grep -i http services | wc -l**. Vysvetlite prečo vznikli rozdiely v počte získaných riadkov.
  - Použite príkazy **grep http services**, **grep -w http services**. Potom použite príkazy **grep http services | wc -l**, **grep -w http services | wc -l**. Vysvetlite prečo vznikli rozdiely v počte získaných riadkov.
  - Použite príkazy **grep -w http services**, **grep http services | grep -v https**. Potom použite príkazy **grep -w http services | wc -l**, **grep http services | grep -v https | wc -l**. Vysvetlite prečo vznikli rozdiely v počte získaných riadkov. Ak sa vám rôzne výsledky nedarí vysvetliť, použite príkaz **grep http services | grep -v https | grep -vw http**.

56. Príkazom **su** - zmeníte svoju identitu na superpoužívateľa *root* a ako tento používateľ postupne vytvorte nových používateľov s loginmi aj príslušným komentárom, ktorým je skutočné meno nasledujúcimi príkazmi:

**useradd -c "Roman Hruska" hruska**

**useradd -c "Peter Slivka" slivka**

**useradd -c "Emil Malina" malina**

57. Príkazom **exit** sa v termináli vráťte k bežnému používateľovi. Ako bežný používateľ si prezrite si obsah súboru */etc/passwd*. Kombináciou programov *grep* (premýšľajte si, aký reťazec je v príslušných riadkoch rovnaký), *cut* a *sort* vytvorte v domovskom priečinku tohto používateľa súbor *pouzivatelia*, ktorý bude obsahovať iba loginy používateľov a ich skutočné mená, pričom zoznam bude zoradený abecedne podľa loginov.

58. Zmeňte opäť svoju identitu na superpoužívateľa *root* a presuňte sa do priečinka */etc*. Príkazom **grep -r nameserver \*** vyhľadajte v tomto priečinku rekurzívne všetky súbory, v ktorých sa nachádza reťazec *nameserver*.

59. Príkazom **ip address show** si zobrazte informácie o stave sieťových adaptérov; malo by medzi nimi byť rozhranie s názvom *ens192*. Príkazom **grep -r ens192 \*** vyhľadajte jeho konfiguračný súbor. Obsah nájdeného súboru si potom vypíšte.

60. Príkazom **ip address show** si znovu zobrazte informácie o stave sieťových adaptérov, pokúste sa z daného výpisu vyfiltrovať iba MAC adresu rozhrania **ens192**. Jedným z možných riešení je: **ip address show | grep -A 1 ens192: | tail -n 1 | tail -c +5 | cut -f 2 -d " "**.

61. Príkazom **exit** sa v termináli vráťte k bežnému používateľovi. Ako bežný používateľ zmažte všetky vytvorené súbory aj priečinkov, ktorý ste vytvorili na začiatku.