

Міністерство освіти і науки України
Одеський національний політехнічний університет
Інститут комп'ютерних систем
Кафедра інформаційних систем

Лабораторна робота №4
з дисципліни «Операційні системи»

Тема: «Складна обробка текстових даних засобами оболонки Unix-подібних ОС інтерфейсу командного рядка»

Виконав:

студент групи АІ-204

Євчев Д.Ф.

Перевірив:

Блажко О.А.

Одеса – 2021

Мета роботи: придбання навичок складної обробки текстових даних роботи засобами оболонки Unix-подібних ОС інтерфейсу командного рядка..

Завдання для виконання

2.1 Обробка текстових даних результатів роботи команд

Виконав завдання, знаходячись на *Linux*-сервері за адресою з IP = 91.219.60.189

- 1) Вивів на екран перші десять назв каталогів або файлів, назви яких починаються з перших трьох букв моого прізвища в латиниці

```
[evchev_denis@vpsj3IeQ ~]$ find / -name "evc*" 2> /dev/null
/var/spool/mail/evchev_denis
/home/evchev_denis
/home/evchev_denis/evchev_lab_3
```

- 2) Назва моєго домашнього каталогу містить моє прізвище та ім'я в транслітерації. Вивів на екран список назв домашніх каталогів користувачів, в назві яких присутня перша буква моого прізвища.

```
[evchev_denis@vpsj3IeQ ~]$ find /home -maxdepth 1 -type d -name "e*" 2> /dev/null
/home/ershov_oleksandr
/home/evchev_denis
/home/eliseev_danilo
```

- 3) Модифікував рішення попереднього завдання так, що на екран було виведено імена користувачів, відсортовані в порядку зростання значення, та без дублювання цих значень (використовув конвеєр з команд *cut*, *sort*, *uniq*).

```
[evchev_denis@vpsj3IeQ ~]$ find /home -maxdepth 1 -type d -name "e*" 2> /dev/null | cut -d'_' -f2 | sort | uniq
danilo
denis
oleksandr
```

- 4) В лабораторній роботі № 2 було описано структуру файлу */etc/passwd*, який використовується для зберігання даних про облікові записи користувачів в *Unix*-подібних ОС. Створив файл з назвою *accounts.csv*, який буде містити облікові записи з наступними колонками, розділеними символом «кома»:
ім'я користувача (логін), ідентифікатор користувача, початковий каталог користувача.

```
[evchev_denis@vpsj3IeQ ~]$ ls
evchev_lab_3 Lab2.pdf                               os.lab1.cpl251.html OS-LW1
Lab2.docx     Operating-System.-Laboratory-Work-1 os.lab1.utf.html    test.html
[evchev_denis@vpsj3IeQ ~]$ cut -d':' -f1,3,6 /etc/passwd | tr ':' ',' > accounts.csv
[evchev_denis@vpsj3IeQ ~]$ ls
accounts.csv  Lab2.docx   Operating-System.-Laboratory-Work-1 os.lab1.utf.html test.html
evchev_lab_3  Lab2.pdf    os.lab1.cpl251.html          OS-LW1
[evchev_denis@vpsj3IeQ ~]$ cat accounts.csv
root,0,/root
bin,1,/bin
daemon,2,/sbin
adm,3,/var/adm
lp,4,/var/spool/lpd
sync,5,/sbin
shutdown,6,/sbin
halt,7,/sbin
mail,8,/var/spool/mail
operator,11,/root
games,12,/usr/games
ftp,14,/var/ftp
nobody,99,
systemd-network,192,
dbus,81,
polkitd,999,
postfix,89,/var/spool/postfix
chrony,998,/var/lib/chrony
sshd,74,/var/empty/sshd
soft,1000,/home/soft
saslauth,997,/run/saslauthd
mailnull,47,/var/spool/mqueue
smmsp,51,/var/spool/mqueue
emps,996,/home/emps
rpc,32,/var/lib/rpcbind
rpcuser,29,/var/lib/nfs
nfsnobody,65534,/var/lib/nfs
oracle,54321,/home/oracle
apache,995,/home/apache
mysql,27,/var/lib/mysql
named,25,/var/named
exim,93,/var/spool/exim
vmail,5000,/var/local/vmail
arestov_yaroslav,54322,/home/arestov_yaroslav
martinyuk_denis,54323,/home/martinyuk_denis
bogachik_igor,54324,/home/bogachik_igor
```

5) З файлу *accounts.csv* отримав на екран перелік облікових записів, в яких четверта цифра ідентифікатора користувача співпадає з четвертою цифрою мого ідентифікатора користувача.

```
[evchev_denis@vpsj3IeQ ~]$ grep -E "[0-9]{3}8" "accounts.csv"
romanenko_oleksandr,54380,/home/romanenko_oleksandr
ahmadi_muhamad,54381,/home/ahmadi_muhamad
ahmadi_umran,54382,/home/ahmadi_umran
li_chuan,54383,/home/li_chuan
babich_artem,54384,/home/babich_artem
polovenko_illya,54385,/home/polovenko_illya
berislavskij_vladislav,54386,/home/berislavskij_vladislav
krutienko_andrij,54387,/home/krutienko_andrij
belobrov_artur,54388,/home/belobrov_artur
evchev_denis,54389,/home/evchev_denis
```

2.2 Обробка текстового файлу формату CSV

Нехай існують два файли *HTML*-формату *DOC* з табличними даними, які визначено у відповідності із номером моєї команди та номером учасника в команді, як в лабораторній роботі № 3 -

https://docs.google.com/spreadsheets/d/1ITMLdjcw7nboUo9nFw4NwkWcvjbZeZ_ZlDVb2Hg4laUQ

Створив файл формату *CSV*, який містить табличні данні, використовуючи наступні кроки:

- 1) базове створення файлу *CSV*-формату, використовуючи *Google Docs* за методичними рекомендаціями

The screenshot shows a Microsoft Excel window with the title bar 'EvchevDenis.csv – Блокнот'. The menu bar includes 'Файл', 'Правка', 'Формат', 'Вид', 'Справка'. The main content area displays a large amount of data in a single column, representing the CSV file. The data consists of pairs of names and numbers, such as 'Назва,Місто/Район, "Усього тис.грн", Підприємства Тис.грн,"Підприємства У відсотках до загального обсягу реалізованої продукції' followed by a series of numbers. The bottom of the window shows standard Excel navigation controls: 'Стр 1, стлб 1', '100%', 'Windows (CRLF)', and 'UTF-8'.

- 2) копіювання файлів на сервер за адресою з IP = 91.219.60.189 у домашній каталог учасника команди;

```
C:\Users\Администратор>pscp.exe -P 22 C:\Users\Администратор\Desktop\EvchevDenis.csv evchev_denis@91.219.60.189:/home/evchev_denis
evchev_denis@91.219.60.189's password:
EvchevDenis.csv | 2 kB | 2.9 kB/s | ETA: 00:00:00 | 100%
```

- 3) модифікація файлу *CSV*-формату:

- кодування символів – *UTF-8*
- роздільник між колонками – символ коми;
- назви файлів визначити з урахуванням транслітерації мого прізвища та номеру файлу.
- заміна в числах коми на точку з одночасним видаленням лапки;
- заміна підрядка « ... 1» на порожнє значення (якщо таке є);
- заміна символу – (дефіс) на порожнє значення (якщо таке є);
- видалення з файлу стовпчиків окрім первого та другого

```
[evchev_denis@vpsj3IeQ ~]$ sed 's/\"\\([0-9]+\\) [,]\\([0-9]+\\)\"/\\1.\\2/g' | sed 's/\\(,\\)...1\\/1/g' | sed 's/\\(,\\)\\~/\\1/g' | cut -f1,2 -d, EvchevDenis.csv > 2.csv
^C
[evchev_denis@vpsj3IeQ ~]$ cat 2.csv
Назва,Місто/Район
м.Одеса,Місто
м.Балта,Місто
м.Білгород-Дністровський,Місто
м.Біляївка,Місто
м.Ізмаїл,Місто
м.Чорноморськ,Місто
м.Помільськ,Місто
м.Теплодар,Місто
м.Южне,Місто
Ананіївський,Район
Арцизький,Район
Балтський,Район
Білгород-Дністровський,Район
Біляївський,Район
Березівський,Район
Болградський,Район
Великомихайлівський,Район
Іванівський,Район
Ізмаїльський,Район
Кілійський,Район
Кодимський,Район
Лиманський,Район
Поліський,Район
Окнянський,Район
Любашівський,Район
Миколаївський,Район
Овідіопольський,Район
Роздільнянський,Район
Ренійський,Район
Савранський,Район
Саратський,Район
Тарутинський,Район
Татарбунарський,Район
Захаріївський,Район
Ширяївський,Район
```

4) передача створеного CSV-файлу в GitHub-репозиторій;

```
[evchev_denis@vpsj3IeQ Operating-System-Laboratory-Work-4]$ mv /home/evchev_denis/2.csv .
[evchev_denis@vpsj3IeQ Operating-System-Laboratory-Work-4]$ ls
1.csv  2.csv  Operating-System.-Laboratory-Work-1
[evchev_denis@vpsj3IeQ Operating-System-Laboratory-Work-4]$ git add 2.csv
[evchev_denis@vpsj3IeQ Operating-System-Laboratory-Work-4]$ git commit -m AddingFile
[main 25a0800] AddingFile
 1 file changed, 36 insertions(+)
 create mode 100644 2.csv
[evchev_denis@vpsj3IeQ Operating-System-Laboratory-Work-4]$ git push
warning: push.default is unset; its implicit value is changing in
Git 2.0 from 'matching' to 'simple'. To squelch this message
and maintain the current behavior after the default changes, use:

  git config --global push.default matching

To squelch this message and adopt the new behavior now, use:
```

```
git config --global push.default simple
```

See 'git help config' and search for 'push.default' for further information.
(the 'simple' mode was introduced in Git 1.7.11. Use the similar mode
'current' instead of 'simple' if you sometimes use older versions of Git)

```
Username for 'https://github.com': EvchevDenis
Password for 'https://EvchevDenis@github.com':
Counting objects: 4, done.
Delta compression using up to 2 threads.
Compressing objects: 100% (3/3), done.
Writing objects: 100% (3/3), 674 bytes | 0 bytes/s, done.
Total 3 (delta 0), reused 0 (delta 0)
To https://github.com/GrEFeRFeeD/Operating-System-Laboratory-Work-4
 e8fe711..25a0800  main -> main
```

5) отримання CSV-файлу, створеного іншим учасником команди;

```
[evchev_denis@vpsj3IeQ ~]$ cp /home/evchev_denis/Operating-System-Laboratory-Work-4/2.csv /home/evchev_denis
[evchev_denis@vpsj3IeQ ~]$ ls
2.csv      evchev_lab_3  Lab2.pdf          Operating-System.-Laboratory-Work-4  os.lab1.utf.html  test.html
accounts.csv Lab2.docx   Operating-System.-Laboratory-Work-1  os.lab1.cpl251.html    OS-LW1
[evchev_denis@vpsj3IeQ ~]$ cp /home/evchev_denis/Operating-System-Laboratory-Work-4/1.csv /home/evchev_denis
[evchev_denis@vpsj3IeQ ~]$ ls
1.csv  accounts.csv  Lab2.docx  Operating-System.-Laboratory-Work-1  os.lab1.cpl251.html  OS-LW1
2.csv  evchev_lab_3  Lab2.pdf   Operating-System.-Laboratory-Work-4  os.lab1.utf.html    test.html
```

6) об'єднання двох файлів в один за першим стовпчиком;

```
[evchev_denis@vpsj3IeQ ~]$ join -t, -a 1 -a 2 -o0,1.2,2.2 -e ' -' 1.csv 2.csv > 3.csv
join: 1.csv:4: is not sorted: Ананьївський,район
join: 2.csv:3: is not sorted: м.Балта,Місто
[evchev_denis@vpsj3IeQ ~]$ cat 3.csv
Назва,місто/район,Місто/Район
м.Балта,місто, -
м.Біляївка,місто, -
Ананьївський,район, -
Арцизький,район, -
Балтський,район, -
Білгород-Дністровський,район, -
Біляївський,район, -
Верезівський,район, -
Волградський,район, -
Великомихайлівський,район, -
Іванівський,район, -
Ізмаїльський,район, -
Кілійський,район, -
Кодимський,район, -
Лиманський,район, -
м.Одеса, -,Місто
м.Балта, -,Місто
м.Білгород-Дністровський, -,Місто
м.Біляївка, -,Місто
м.Ізмаїл, -,Місто
м.Чорноморськ, -,Місто
м.Подільськ, -,Місто
м.Теплодар, -,Місто
м.Южне, -,Місто
Ананьївський, -,Район
Арцизький, -,Район
Балтський, -,Район
Білгород-Дністровський, -,Район
Біляївський, -,Район
```

7) виведення на екран змісту файлу з рішення попереднього завдання у зворотному напрямі для колонки за номером 2 - для 1-го участника команди, або за номером 3 – для 2-го участника команди.

```
[evchev_denis@vpsj3IeQ ~]$ sort --field-separator=' ' --key=3 -r 3.csv
Ширівський,район,Район
Татарбунарський,район,Район
Тарутинський,район,Район
Саратський,район,Район
Савранський,район,Район
Роздільнянський,район,Район
Ренійський,район,Район
Подільський,район,Район
Окнянський,район,Район
Овідіопольський,район,Район
Миколаївський,район,Район
Любашівський,район,Район
Лиманський, -,Район
Кодимський, -,Район
Кілійський, -,Район
Ізмаїльський, -,Район
Іванівський, -,Район
Захарівський,район,Район
Великомихайлівський, -,Район
Болградський, -,Район
Біллівський, -,Район
Білгород-Дністровський, -,Район
Березівський, -,Район
Балтський, -,Район
Арцизький, -,Район
Ананьївський, -,Район
Назва,місто/район,Місто/Район
м.Южне, -,Місто
м.Чорноморськ, -,Місто
м.Теплодар, -,Місто
```

Висновок: придбав навички роботи з оболонкою Unix-подібних ОС інтерфейсу командного рядка та утилітами обробки текстових даних. Найскладнішим завданням було 2.3.2 - робота з текстовими файлами в різних кодуваннях символів.