

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:
Ст. гр. АІ-204
Дорожкін Михайло

Перевірив(-ла):
Блажко О. А.
Дрозд М.О.

Одеса 2021

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

Завдання до виконання:

2.1 Обробка текстових даних результатів роботи команд

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

- 1) Виведіть на екран перші десять назв каталогів або файлів, назви яких починаються з перших трьох букв вашого прізвища в латиниці
- 2) Назва вашого домашнього каталогу містить ваші прізвище та ім'я в транслітерації. Виведіть на екран список назв домашніх каталогів користувачів, в назві яких присутня перша буква вашого прізвища.
- 3) Модифікуйте рішення попереднього завдання так, що на екран було виведено імена користувачів, відсортовані в порядку зростання значення, та без дублювання цих значень (використайте конвеєр з команд cut, sort, uniq).
- 4) В лабораторній роботі № 2 було описано структуру файлу /etc/passwd, який використовується для зберігання даних про облікові записи користувачів в Unix-подібних ОС. Створіть 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.
- 5) З файлу accounts.csv отримайте на екран перелік облікових записів, в яких четверта цифра ідентифікатора користувача співпадає з четвертою цифрою вашого ідентифікатора користувача.

2.2 Обробка текстового файлу формату CSV

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

<https://docs.google.com/spreadsheets/d/1ITMLdjcw7nboUo9nFw4NwkWcvjbZeZZlDVb2Hg4laUQ>

Необхідно створити файл формату CSV, який містить табличні данні, використовуючи наступні кроки:

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

<https://drive.google.com/file/d/17wTiweC-1QybKQje49VdqFahkely1Ou>:

- кодування символів – UTF-8
- роздільник між колонками – символ коми;
- назви файлів визначити з урахуванням транслітерації свого прізвища та номеру файлу.

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

3) модифікація файлу CSV-формату:

- 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 першого та другого

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

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

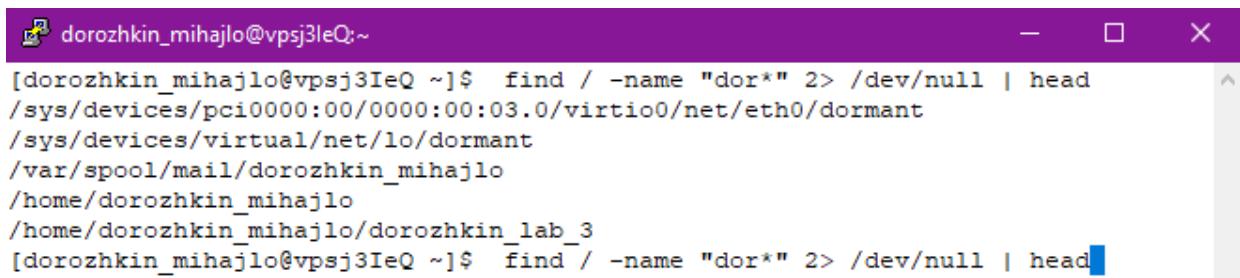
6) об'єднання двох файлів в один за першим стовпчиком;

7) виведення на екран змісту CSV-файлу з рішення попереднього завдання із сортуванням рядків файла-таблиці у зворотному напрямі значень колонки за номером 2 – для 1-го учасника команди, або за номером 3 – для 2-го учасника команди.

Виконання завдань:

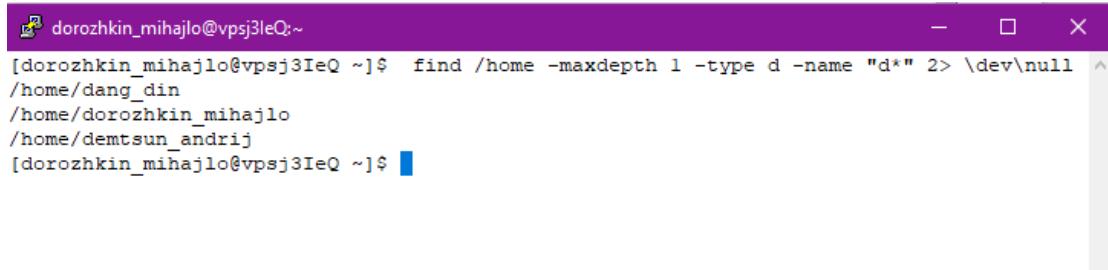
2.1 Обробка текстових даних результатів роботи команд

1) Виведіть на екран перші десять назв каталогів або файлів, назви яких починаються з перших трьох букв вашого прізвища в латиниці.



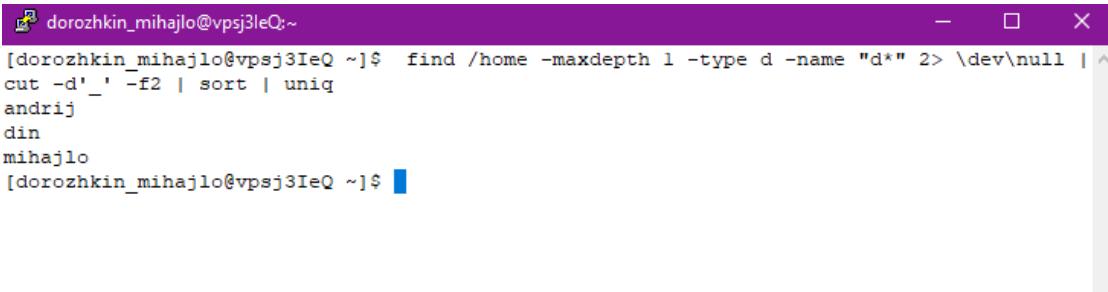
```
[dorozhkin_mihajlo@vpsj3IeQ:~]$ find / -name "dor*" 2> /dev/null | head
/sys/devices/pci0000:00/0000:00:03.0/virtio0/net/eth0/dormant
/sys/devices/virtual/net/lo/dormant
/var/spool/mail/dorozhkin_mihajlo
/home/dorozhkin_mihajlo
/home/dorozhkin_mihajlo/dorozhkin_lab_3
[dorozhkin_mihajlo@vpsj3IeQ:~]$ find / -name "dor*" 2> /dev/null | head
```

2) Назва вашого домашнього каталогу містить ваші прізвище та ім'я в транслітерації. Виведіть на екран список назв домашніх каталогів користувачів, в назві яких присутня перша буква вашого прізвища.



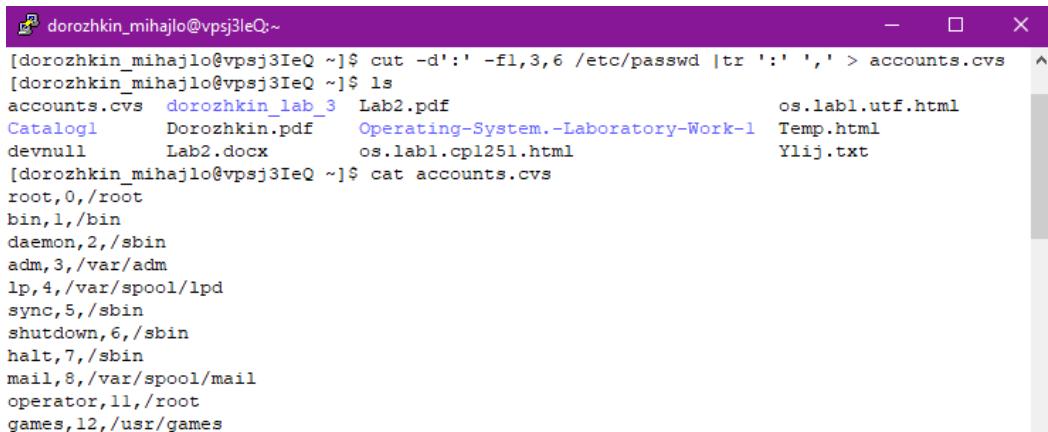
```
[dorozhkin_mihajlo@vpsj3IeQ:~]$ find /home -maxdepth 1 -type d -name "d*" 2> \dev\null
/home/dang_din
/home/dorozhkin_mihajlo
/home/demtsun_andrij
[dorozhkin_mihajlo@vpsj3IeQ:~]$
```

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



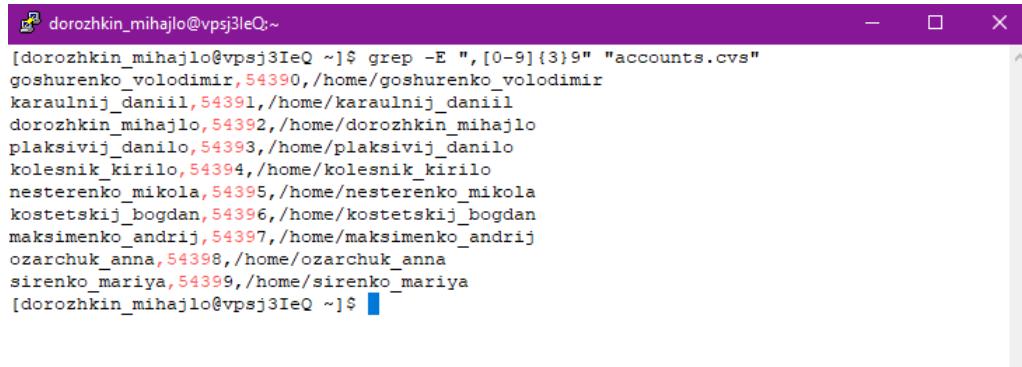
```
[dorozhkin_mihajlo@vpsj3IeQ:~]$ find /home -maxdepth 1 -type d -name "d*" 2> \dev\null | ^
cut -d'_' -f2 | sort | uniq
andrij
din
mihajlo
[dorozhkin_mihajlo@vpsj3IeQ:~]$
```

4) В лабораторній роботі № 2 було описано структуру файлу /etc/passwd, який використовується для зберігання даних про облікові записи користувачів в Unix-подібних ОС. Створіть 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.



```
[dorozhkin_mihajlo@vpsj3IeQ:~]$ cut -d':' -f1,3,6 /etc/passwd | tr ':' ',' > accounts.csv
[dorozhkin_mihajlo@vpsj3IeQ:~]$ ls
accounts.csv  dorozhkin_lab_3  Lab2.pdf          os.lab1.utf.html
Catalogl      Dorozhkin.pdf    Operating-System.-Laboratory-Work-1 Temp.html
devnull       Lab2.docx       os.lab1.cpl251.html Ylij.txt
[dorozhkin_mihajlo@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
```

5) З файлу accounts.csv отримайте на екран перелік облікових записів, в яких четверта цифра ідентифікатора користувача співпадає з четвертою цифрою вашого ідентифікатора користувача (54392)



```
[dorozhkin_mihajlo@vpsj3IeQ:~]$ grep -E "[0-9]{3}9" "accounts.csv"
goshurenko_vladimir,54390,/home/goshurenko_vladimir
karaulnij_daniil,54391,/home/karaulnij_daniil
dorozhkin_mihajlo,54392,/home/dorozhkin_mihajlo
plaksivij_danilo,54393,/home/plaksivij_danilo
kolesnik_kirilo,54394,/home/kolesnik_kirilo
nesterenko_mikola,54395,/home/nesterenko_mikola
kostetskij_bogdan,54396,/home/kostetskij_bogdan
maksimenko_andrij,54397,/home/maksimenko_andrij
ozarchuk_anna,54398,/home/ozarchuk_anna
sirenko_mariya,54399,/home/sirenko_mariya
[dorozhkin_mihajlo@vpsj3IeQ ~]$
```

2.2 Обробка текстового файлу формату CSV

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

<https://docs.google.com/spreadsheets/d/1ITMLdjcw7nboUo9nFw4NwkWcvjbZeZZlDVb2Hg4laUQ>

Необхідно створити файл формату CSV, який містить табличні данні, використовуючи наступні кроки:

1) базове створення файлу CSV-формату, використовуючи Google Docs за методичними рекомендаціями (Мал. №1)

<https://drive.google.com/file/d/17wTiweC-1QybKQje49VdqFahkely1Ou>:

- кодування символів – UTF-8
- роздільник між колонками – символ коми;
- назви файлів визначити з урахуванням транслітерації свого прізвища та номеру файлу.

```
Файл Правка Формат Вид Справка
Назва,Місто/район,"Усього, одиниць",Підприємства одиниць,"Підприємства у відсотках
до загальної кількості
підприємств",Фізичні особи-підприємці одиниць,"Фізичні особи-підприємці у відсотках
до загальної кількості
фізичних осіб-підприємців"
Одеська область,,137609,25871,100,111738,100
Одеса,Місто,76396,15576,60.2,60820,54.4
Балта,Місто,1151,84,0.3,1067,1
Білгород-Дністровський,Місто,4340,547,2.1,3793,3.4
Біляївка,Місто,637,136,0.5,501,0.4
Ізмаїл,Місто,4340,571,2.2,3769,3.4
Чорноморськ,Місто,4454,864,3.3,3590,3.2
Подільськ,Місто,1747,119,0.5,1628,1.5
Теплодар,Місто,600,85,0.3,515,0.5
Южне,Місто,1456,199,0.8,1257,1.1
Ананьївський,Район,885,101,0.4,784,0.7
Арцизький,Район,1973,315,1.2,1658,1.5
Балтський,Район,615,129,0.5,486,0.4
Білгород-Дністровський,Район,1902,309,1.2,1593,1.4
Біляївський,Район,4406,544,2.1,3862,3.5
Березівський,Район,1279,310,1.2,969,0.9
Болградський,Район,2829,394,1.5,2435,2.2
Великомихайлівський,Район,878,200,0.8,678,0.6
Іванівський,Район,1052,193,0.7,859,0.8
Ізмаїльський,Район,1693,534,2.1,1159,1
Кілійський,Район,1997,270,1,1727,1.5
Коринтський Район,960,125,0.5,825,0.7
```

Мал №1

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

```
Microsoft Windows [Version 10.0.19042.868]
(c) Корпорация Майкрософ트 (Microsoft Corporation), 2020. Все права защищены.

C:\Users\lexur>C:\Users\lexur\Desktop\pscp.exe -P 22 C:\Users\lexur\Desktop\Dorozhkin_Mihajlo1.csv dorozhkin_mihajlo@91.219.60.189:/home/dorozhkin_mihajlo
dorozhkin_mihajlo@91.219.60.189's password:
Dorozhkin_Mihajlo1.csv    | 2 kB |   2.4 kB/s | ETA: 00:00:00 | 100%
C:\Users\lexur>
```

3) модифікація файлу CSV-формату:

- заміна в числах коми на точку з одночасним видаленням лапки;
- заміна підрядка « ... 1» на порожнє значення (якщо таке ε);
- заміна символу – (дефіс) на порожнє значення (якщо таке ε);
- видалення з файлу стовпчиків окрім першого та другого

```
[dorozhkin_mihajlo@vpsj3leQ ~]$ sed 's/\\"([0-9]\+\)[,]\"([0-9]\+\)\\"/\1.\2 /g' Dorozhkin_Mihajlo1.csv > 1.csv
[dorozhkin_mihajlo@vpsj3leQ ~]$ sed 's/(\,\)\.\.1/\1/g' 1.csv > 2.csv
[dorozhkin_mihajlo@vpsj3leQ ~]$ sed 's/(\,\)\-\//\1/g' 2.csv > 3.csv
[dorozhkin_mihajlo@vpsj3leQ ~]$ cut -f1,2 -d, 3.csv > 1.csv
[dorozhkin_mihajlo@vpsj3leQ ~]$
```

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

The terminal window shows the following command and its execution:

```
dorozhkin_mihajlo@vpsj3IeQ:~/Operating-System.-Laboratory-Work-4
Password for 'https://lexur22@github.com':
No refs in common and none specified; doing nothing.
Perhaps you should specify a branch such as 'master'.
Everything up-to-date
[dorozhkin_mihajlo@vpsj3IeQ Operating-System.-Laboratory-Work-4]$ git pus -u origin --all
git: 'pus' is not a git command. See 'git --help'.

Did you mean one of these?
  push
  pull
[dorozhkin_mihajlo@vpsj3IeQ Operating-System.-Laboratory-Work-4]$ push
-bash: push: command not found
[dorozhkin_mihajlo@vpsj3IeQ Operating-System.-Laboratory-Work-4]$ git push -u origin --all
Username for 'https://github.com': lexur22
Password for 'https://lexur22@github.com':
Counting objects: 3, done.
Delta compression using up to 2 threads.
Compressing objects: 100% (2/2), done.
Writing objects: 100% (3/3), 675 bytes | 0 bytes/s, done.
Total 3 (delta 0), reused 0 (delta 0)
To https://github.com/lexur22/Operating-System.-Laboratory-Work-4
 * [new branch]      master -> master
Branch master set up to track remote branch master from origin.
[dorozhkin_mihajlo@vpsj3IeQ Operating-System.-Laboratory-Work-4]$
```

Below the terminal, a screenshot of the GitHub repository page for 'lexur22 / Operating-System.-Laboratory-Work-4' is shown. The repository has 1 branch (master) and 0 tags. A recent commit by 'lexur22' named 'NewFile' was pushed 1 minute ago, containing a file named '1.csv'.

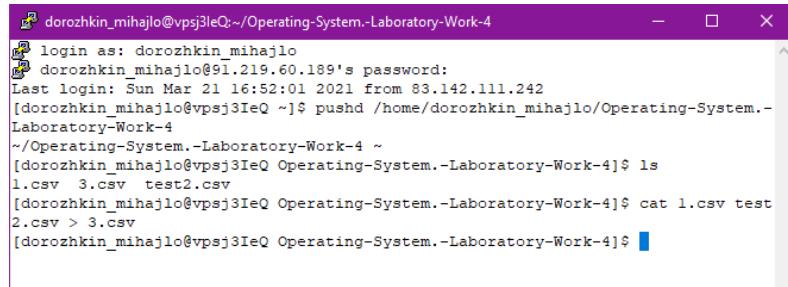
The GitHub page interface includes navigation tabs (Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, Settings), a commit summary, and a button to 'Add a README'.

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

The terminal window shows the command used to download the CSV file:

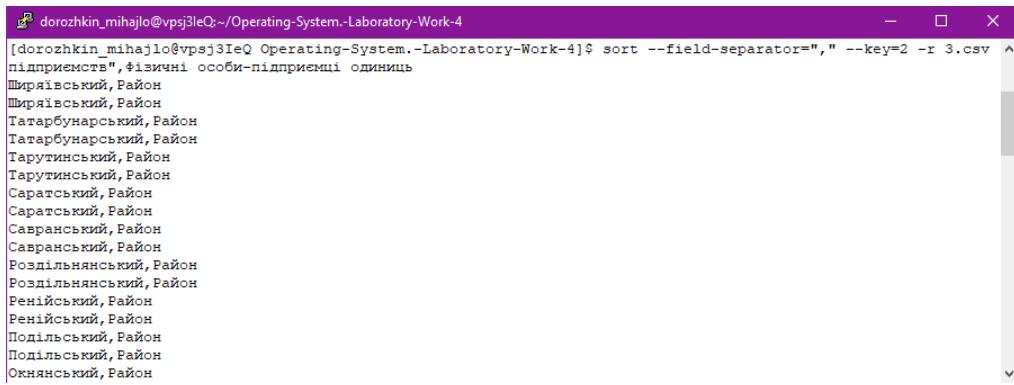
```
C:\Users\lexur>C:\Users\lexur\Desktop\pscp.exe -P 22 C:\Users\lexur\Desktop\test2.csv dorozhkin_mihajlo@91.219.60.189:/home/dorozhkin_mihajlo/Operating-System.-Laboratory-Work-4
dorozhkin_mihajlo@91.219.60.189's password:
test2.csv | 1 kB | 1.4 kB/s | ETA: 00:00:00 | 100%
C:\Users\lexur>
```

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



```
[dorozhkin_mihajlo@vpsj3IeQ:~/Operating-System.-Laboratory-Work-4]
[dorozhkin_mihajlo@vpsj3IeQ ~]$ login as: dorozhkin_mihajlo
[dorozhkin_mihajlo@vpsj3IeQ ~]$ dorozhkin_mihajlo@91.219.60.189's password:
Last login: Sun Mar 21 16:52:01 2021 from 83.142.111.242
[dorozhkin_mihajlo@vpsj3IeQ ~]$ pushd /home/dorozhkin_mihajlo/Operating-System.-Laboratory-Work-4
~/Operating-System.-Laboratory-Work-4 ~
[dorozhkin_mihajlo@vpsj3IeQ Operating-System.-Laboratory-Work-4]$ ls
1.csv 3.csv test2.csv
[dorozhkin_mihajlo@vpsj3IeQ Operating-System.-Laboratory-Work-4]$ cat 1.csv test
2.csv > 3.csv
[dorozhkin_mihajlo@vpsj3IeQ Operating-System.-Laboratory-Work-4]$
```

7) виведення на екран змісту CSV-файлу з рішення попереднього завдання із сортуванням рядків файла-таблиці у зворотному напрямі значень колонки за номером 2 – для 1-го участника команди.



```
[dorozhkin_mihajlo@vpsj3IeQ:~/Operating-System.-Laboratory-Work-4]
[dorozhkin_mihajlo@vpsj3IeQ Operating-System.-Laboratory-Work-4]$ sort --field-separator="," --key=2 -r 3.csv
підприємств", фізичні особи-підприємці одиниць
Ширяївський, Район
Ширяївський, Район
Татарбунарський, Район
Татарбунарський, Район
Тарутинський, Район
Тарутинський, Район
Саратський, Район
Саратський, Район
Савранський, Район
Савранський, Район
Роздільнянський, Район
Роздільнянський, Район
Ренійський, Район
Ренійський, Район
Подільський, Район
Подільський, Район
Окнянський, Район
```

Висновок: Під час виконання лабораторної роботи я працював із складною обробкою текстових даних засобами оболонки UNIX-подібних ОС з інтерфейсом командного рядку. Найскладнішим для мене виявилось завдання з пошуку по 3-ом своїм первих буквах прізвища інші схожі каталоги.