

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 ”**

Виконала:  
Студентка групи АІ-203  
Мягих А.М.

Перевірив:  
Блажко О.А.

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

## **План:**

### **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/1ITMLdjcw7nboUo9nFw4NwkWcvjbZeZzIDVb2Hg4laUQ>

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

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

<https://drive.google.com/file/d/17wTiwec-1QybKQje49VdjQFahkely1Ou:>

– 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 файлу-таблиці у зворотному напрямі значень колонки за номером 2 - для 1-го учасника команди, або за номером 3 – для 2-го учасника команди.

## Завдання:

### 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

```
[myagkih_arina@vpsj3IeQ ~]$ find / -name -name "mya*" 2> /dev/null | head
/var/spool/mail/myagkih_arina
/var/softaculous/myadmin
/var/softaculous/myadmin/images/myadmin.png
/usr/local/webuzo/enduser/softimages/top15/48/myadmin.png
/usr/local/webuzo/enduser/softimages/top15/70/myadmin.png
/usr/local/webuzo/enduser/softimages/top15/myadmin.png
/home/myagkih_arina
/home/myagkih_arina/myagkih_lab_3
[myagkih_arina@vpsj3IeQ ~]$
```

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.

```
[myagkih_arina@vpsj3IeQ ~]$ find /home -maxdepth 1 -type d -name "m*" 2> /dev/null
/home/muradyan_volodimir
/home/magdenko_mihajlo
/home/markovskij_danilo
/home/maksimenko_andrij
/home/myagkih_arina
/home/malofeev_denis
/home/matnenko_stanislav
/home/martinyuk_denis
/home/melnichenko_egor
[myagkih_arina@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).

```
[myagkih_arina@vpsj3IeQ ~]$ find /home -maxdepth 1 -type d -name "m*" 2>
ll | cut -d'_' -f2 | sort | uniq
andrij
arina
danilo
denis
egor
mihajlo
stanislav
volodimir
```

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.

```
[myagkih_arina@vpsj3IeQ ~]$ cut -d':' -f1,3,6 /etc/passwd | tr ':' ' ' > account
s.csv
[myagkih_arina@vpsj3IeQ ~]$ ls
??????_????_??_?_?3.docx      Myagkih_Arina.txt
accounts.csv                   myagkih_lab_3
lab_os.docx                   Operating-System.-Laboratory-Work-1
lab_os.html                   os.lab1.cpl251.html
lab_os_html_171cb5d9e9db9067.png  os.lab1.utf.html
lab_os.pdf
[myagkih_arina@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
saslauthd,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
vmmail,5000,/var/local/vmail
```

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

```
[myagkih_arina@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
[myagkih_arina@vpsj3IeQ ~]$
```

## 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

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

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

<https://drive.google.com/file/d/17wTiwec-1QybKQje49VdjQFahkely1Ou>:

– кодування символів – UTF-8

– роздільник між колонками – символ коми;

– назви файлів визначити з урахуванням транслітерації свого прізвища та номеру файлу.

Обсяг реалізованої продукції (товарів, послуг) суб'єктів господарювання по містах обласного значення та районах у 2019 році - Лист1.csv — Блокнот	
Файл	Правка
Формат	Вид
Справка	
<p>Назва ,місто / район,"Усього, тис.грн",підприємства,,фізичні особи-підприємці, ,,тис.грн,"у відсотках до загального обсягу реалізованої продукції (товарів, послуг)",тис.грн,"у відсотках до загального обсягу реалізованої продукції (товарів, послуг)"</p> <p>Одеська область,місто,444690830.2,390704866.5,100,53985963.7,100</p> <p>м.Одеса,місто,317274775.2,283787934.8,72.6,33486840.4,62</p> <p>м.Балта,місто,874682.3,582941.6,0.2,291740.7,0.5</p> <p>м.Білогір'я-Дністровський,місто,4964551.9,3463838.4,0.9,1500713.5,2.8</p> <p>м.Біляївка,місто,2651732.8,1820977.6,0.5,230755.2,0.4</p> <p>м.Ізмаїл,місто,6860810.7,5364392.2,1.4,1496418.5,2.8</p> <p>м.Чорноморськ,місто,18669600.3,17232547.3,4.4,1437053.2,7</p> <p>м.Подільськ,місто,1815201.1,1101252.1,0.3,713949.1,3</p> <p>м.Теплодар,місто,421408,172148.4,0.249259.6,0.5</p> <p>м.Южне,місто,11804189.1,11242232.2,2.9,561956.9,1</p> <p>Ананьївський,район,1043363,744720.2,0.2,298642.8,0.5</p> <p>Арцизький,район,3127999.2,2447197.3,0.6,680801.9,1.3</p> <p>Балтський,район,1083805.2,908309.6,0.2,175495.6,0.3</p> <p>Білогір'я-Дністровський,район,3959467.5,3428337.5,0.9,531130.1</p> <p>Біляївський,район,7951390.2,6253696.1,6,1697694.2,3.1</p> <p>Березівський,район,1639175.9,1322152.9,0.3,317023.0,6</p> <p>Болградський,район,2947780.9,2098218.1,0.5,849562.8,1.6</p> <p>Великомихайлівський,район,1404693.8,1172522.1,0.3,232171.7,0.4</p> <p>Іванівський,район,1343910.1,1017123.7,0.3,326786.4,0.6</p> <p>Ізмаїльський,район,1424835.6,928683.4,0.2,496152.2,0.9</p> <p>Кілійський,район,1877366.9,1289280.5,0.3,588086.4,1.1</p> <p>Кодимський,район,1025366.3,720882.1,0.2,304484.2,0.6</p> <p>Лиманський,район,15120156.7,13919025.3,3.6,1201131.4,2.2</p>	

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

```
c:\>pscp.exe -P 22 C:\myagkih_arina.csv myagkih_arina@91.219.60.189:/home/myagkih_arina
myagkih_arina@91.219.60.189's password:
myagkih_arina.csv      | 3 kB | 3.0 kB/s | ETA: 00:00:00 | 100%
```

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

```
[myagkih_arina@vps331eQ ~]$ sed 's/\("[\([0-9]\+\)\([,]\)[\([0-9]\+\)\([/]\1\2/g' | sed 's/\([,]\)\([/]\1/g' | sed 's/\([,]\)\([/]\1/g' | cut -d, -f1,2 myagkih_arina.csv > 1.csv
^C
[myagkih_arina@vps331eQ ~]$ ls
1.csv          accounts.csv   lab_os.html    lab_os.pdf     Myagkih_Arina.txt  Operating-System.-Laboratory-Work-1  os.lab1.utf.html
??????_????_??_?_?3.docx  lab_os.docx    lab_os_html_171cb5d9e9db9067.png  myagkih_arina.csv  myagkih_lab_3      os.lab1.cp1251.html
[myagkih_arina@vps331eQ ~]$ less 1.csv
```

```
Назва ,місто / район
,
обсягу реалізованої
продукції (товарів, послуг) "
обсягу реалізованої
продукції (товарів, послуг) "
Одеська область,місто
м.Одеса,місто
м.Валта,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
Березівський,район
Волградський,район
Великомихайлівський,район
Іванівський,район
Ізмаїльський,район
Кілійський,район
Кодимський,район
Лиманський,район
Подільський,район
Окнянський,район
Любашівський,район
Миколаївський,район
Овідіопольський,район
Роздільнянський,район
Ренійський,район
Савранський,район
Саратський,район
Тарутинський,район
Татарбунарський,район
Захарівський,район
Ширяївський,район
1.csv (END)
```

- 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;

```
[myagkih_arina@vpsj3IeQ ~]$ mkdir for_merge
[myagkih_arina@vpsj3IeQ ~]$ cp 1.csv for_merge/
[myagkih_arina@vpsj3IeQ ~]$ git checkout main
Switched to branch 'main'
[myagkih_arina@vpsj3IeQ ~]$ ls -l
total 4676
-rw-rw-r-- 1 myagkih_arina myagkih_arina 0 Mar 21 16:30 3.csv
-rw-rw-r-- 1 myagkih_arina myagkih_arina 180896 Mar 14 12:36 ??????_?????_??_?_?3.docx
-rw-rw-r-- 1 myagkih_arina myagkih_arina 5365 Mar 21 13:50 accounts.csv
drwxrwxr-x 2 myagkih_arina myagkih_arina 4096 Mar 21 16:48 for_merge
-rw-rw-r-- 1 myagkih_arina myagkih_arina 1031 Mar 21 16:48 grishenko_2.csv
-rw-rw-r-- 1 myagkih_arina myagkih_arina 0 Mar 21 16:48 Hryshchenko_Oleksandra.txt
-rw-rw-r-- 1 myagkih_arina myagkih_arina 0 Mar 21 16:47 joined.csv
-rw-rw-r-- 1 myagkih_arina myagkih_arina 33473 Mar 14 12:47 lab_os.docx
-rw-rw-r-- 1 myagkih_arina myagkih_arina 25167 Mar 14 13:01 lab_os.html
-rw-rw-r-- 1 myagkih_arina myagkih_arina 16181 Mar 14 13:01 lab_os_html_171cb5d9e9db9067.png
-rw-rw-r-- 1 myagkih_arina myagkih_arina 64680 Mar 14 12:49 lab_os.pdf
-rw-rw-r-- 1 myagkih_arina myagkih_arina 3076 Mar 21 14:45 myagkih_arina.csv
drwxrwxr-x 2 myagkih_arina myagkih_arina 4096 Mar 14 12:25 myagkih_lab_3
-rw-rw-r-- 1 myagkih_arina myagkih_arina 0 Mar 21 15:28 operating
drwxrwxr-x 3 myagkih_arina myagkih_arina 4096 Mar 21 15:42 Operating-System.-Laboratory-Work-1
-rw-rw-r-- 1 myagkih_arina myagkih_arina 22104 Mar 14 13:07 os.lab1.cp1251.html
-rw-rw-r-- 1 myagkih_arina myagkih_arina 25167 Mar 14 13:05 os.lab1.utf.html
-rw-rw-r-- 1 myagkih_arina myagkih_arina 689066 Mar 21 16:48 ГрищенкоОлександра.pdf
-rw-rw-r-- 1 myagkih_arina myagkih_arina 3038357 Mar 21 16:48 Лаб1 OC.pptx
-rw-rw-r-- 1 myagkih_arina myagkih_arina 642638 Mar 21 16:48 Лаб3 OC.pdf
[myagkih_arina@vpsj3IeQ ~]$ cp grishenko_2.csv for_merge/
[myagkih_arina@vpsj3IeQ for_merge]$ ls -l
total 8
-rw-rw-r-- 1 myagkih_arina myagkih_arina 1498 Mar 21 16:48 1.csv
-rw-rw-r-- 1 myagkih_arina myagkih_arina 1031 Mar 21 16:49 grishenko_2.csv
[myagkih_arina@vpsj3IeQ for_merge]$ join -t, -a 1 -a 2 -o0,1.2,2.2 -e ' ' 1.csv grishenko_2.csv > joined.csv
join: grishenko_2.csv:4: is not sorted: Ананьївський,район
join: 1.csv:5: is not sorted: обсягу реалізованої
[myagkih_arina@vpsj3IeQ for_merge]$ ls -l
total 12
-rw-rw-r-- 1 myagkih_arina myagkih_arina 1498 Mar 21 16:48 1.csv
-rw-rw-r-- 1 myagkih_arina myagkih_arina 1031 Mar 21 16:49 grishenko_2.csv
-rw-rw-r-- 1 myagkih_arina myagkih_arina 2373 Mar 21 16:49 joined.csv
[myagkih_arina@vpsj3IeQ for_merge]$ less joined.csv
[myagkih_arina@vpsj3IeQ for_merge]$ less 1.csv
[myagkih_arina@vpsj3IeQ for_merge]$ less 1.csv
[myagkih_arina@vpsj3IeQ for_merge]$ less grishenko_2.csv
[myagkih_arina@vpsj3IeQ for_merge]$ less 1.csv
[myagkih_arina@vpsj3IeQ for_merge]$ less joined.csv
[myagkih_arina@vpsj3IeQ for_merge]$
```

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



```
[myagkih_arina@vpsj3IeQ for_merge]$ sort --field-separator="," --key=3 -r joined.csv
Ширяївський, район, район
Татарбунарський, район, район
Тарутинський, район, район
Саратський, район, район
Савранський, район, район
Роздільнянський, район, район
Ренійський, район, район
Подільський, , район
Окнянський, , район
Овідіопольський, , район
Миколаївський, , район
Любашівський, , район
Лиманський, , район
Кодимський, , район
Кілійський, , район
Ізмаїльський, , район
Іванівський, , район
Захарівський, район, район
Великомихайлівський, , район
Болградський, , район
Біляївський, , район
Білгород-Дністровський, , район
Березівський, , район
Балтський, , район
Арцизький, , район
Ананьївський, , район
м.Біляївка, , місто
м.Балта, , місто
Назва, місто / район, Місто/район
, продукції (товарів, послуг) "
продукції (товарів, послуг) ",
Подільський, район,
Окнянський, район,
Одеська область1, місто,
Овідіопольський, район,
    обсягу реалізованої ,,
    обсягу реалізованої ,,
м.Южне, місто,
м.Чорноморськ, місто,
м.Теплодар, місто,
м.Подільськ, місто,
м.Одеса, місто,
м.Ізмаїл, місто,
Миколаївський, район,
м.Біляївка, місто,
м.Білгород-Дністровський, місто,
м.Балта, місто,
Любашівський, район,
Лиманський, район,
Кодимський, район,
Кілійський, район,
```

**Висновок:** під час виконання лабораторної роботи було використано складну обробку текстових даних засобами оболонки Unix-подібних ОС інтерфейсу командного рядка. Створено файл формату .csv, за допомогою команди join з'єднали файли учасника групи з власним файлом.