

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 AI-204

Озарчук А.С.

**Перевірили:** Блажко О. А

Дрозд М. О.

## 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 No 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 No 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

<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

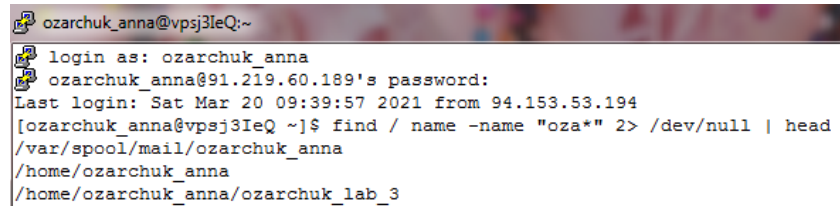
1-го учасника команди, або за номером 3 – для 2-го учасника команди.

# Результати виконання пунктів завдань

## 2.1 Обробка текстових даних результатів роботи команд

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

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



```
ozarchuk_anna@vpsj3IeQ:~  
login as: ozarchuk_anna  
ozarchuk_anna@91.219.60.189's password:  
Last login: Sat Mar 20 09:39:57 2021 from 94.153.53.194  
[ozarchuk_anna@vpsj3IeQ ~]$ find / name -name "oza*" 2> /dev/null | head  
/var/spool/mail/ozarchuk_anna  
/home/ozarchuk_anna  
/home/ozarchuk_anna/ozarchuk_lab_3
```

2) Назва мого домашнього каталогу містить моє прізвище та ім'я в транслітерації.

Вивела на екран список назв домашніх каталогів користувачів, в назві яких присутня перша буква мого прізвища.

```
[ozarchuk_anna@vpsj3IeQ ~]$ find /home -maxdepth 1 -type d -name "o*" 2> /dev/null  
/home/oracle  
/home/ozarchuk_anna
```

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

```
[ozarchuk_anna@vpsj3IeQ ~]$ find /home -maxdepth 1 -type d -name "o*" 2> /dev/null | cut -d'_' -f2 | sort | uniq  
anna  
/home/oracle
```

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

```
[ozarchuk_anna@vpsj3IeQ ~]$ ls
Catalog      Laba.html                                os.lab1.utf.html
Laba3.odt    Operating-System.Laboratory-Work-1      ozarchuk_lab_3
Laba3.pdf    os.lab1.cp1251.html                    Tree.txt
[ozarchuk_anna@vpsj3IeQ ~]$ cut -d':' -f1,3,6 /etc/passwd | tr ':' ' ' > accounts.csv
[ozarchuk_anna@vpsj3IeQ ~]$ ls
accounts.csv  Laba3.odt  Laba.html                                os.lab1.cp1251.html  ozarchuk_lab_3
Catalog      Laba3.pdf  Operating-System.Laboratory-Work-1      os.lab1.utf.html    Tree.txt
[ozarchuk_anna@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
sasauth,997,/run/sasauthd
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
```

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

```
[ozarchuk_anna@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
[ozarchuk_anna@vpsj3IeQ ~]$ █
```

## 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 No 3 -

<https://docs.google.com/spreadsheets/d/1ITMLdjcw7nboUo9nFw4NwkWcvjbZeZZIDVb2Hg4laUQ>

Я створила 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.

Назва, місто/район	Зібрана площа - га	Зібрана площа - % від 2018	Валовий збір у масі після доробки - ц	Валовий збір у масі після доробки - % від 2018	Урожайність з 1 га
м. Балта, місто	1127,91	3,23820	95,8	21,1	105
м. Білгород-Дністровський, місто	2372,79	6,19534	38,6	8,2	48,2
м. Білівка, місто	727,60	4,8689	46,4	11,9	76,8
м. Ізмаїл, місто	к, к, к, к, к, к				
м. Подільськ, місто	к, к, к, к, к, к				
Ананьївський, район	11439,105	1,285365	116,8	24,9	110,7
Арцизький, район	13387,110	5,231094	68,8	17,3	62,5
Балтський, район	14479,135	4,375914	148,4	26,109	7
Білгород-Дністровський, район	10788,91	5,118888	53,4	11,58	2
Білівський, район	8813,94	5,129581	65,5	14,7	69,3
Березівський, район	19920,86	7,432967	101,9	21,7	117,3
Болградський, район	10908,81	6,144759	59,9	13,3	73,5
Великомихайлівський, район	13490,95	295971,96	7,21	9,101	9
Іванівський, район	14051,91	6,207782	85,3	14,8	93,1
Ізмаїльський, район	8903,86	8,122653	60,9	13,8	70,4
Кілійський, район	5292,96	7,51199	48,8	9,7	50,8
Кодимський, район	5915,82	4,164342	91,1	27,8	110,8
Лиманський, район	12988,74	7,173301	63,9	13,3	85,3
Подільський, район	9490,85	7,260106	89,1	27,4	104,2
Окнянський, район	7076,89	1,210646	111,7	29,8	125,2
Любашівський, район	9927,59	8,202661	51,1	20,4	85,4
Миколатівський, район	8906,45	6,195567	36,5	22,80	
Овідіопольський, район	6214,73	68335,38	1,11	52,4	
Роздільнянський, район	14679,111	5,203111	83,3	13,8	74,6
Реніївський, район	6662,124	5,91471	61,9	13,7	65,6
Савранський, район	7229,104	1,199442	107,9	27,6	103,8
Саратський, район	12825,71	3,181236	40,4	14,1	56,6
Тарутинський, район	19682,110	1,377099	91,1	19,2	82,8
Татарбунарський, район	6213,45	9,54165	14,8	8,7	32,1
Захарівський, район	7232,89	3,134210	92,7	18,6	103,9
Ширяївський, район	15302,80	8,277174	78,5	18,1	96,8

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

```
C:\Windows\system32\cmd.exe
Microsoft Windows [Version 6.1.7601]
(c) Корпорация Майкрософт (Microsoft Corp.), 2009. Все права защищены.
C:\Users\Вікуся>cd C:\Intel
C:\Intel>pscp.exe -P 22 C:\Intel\Ozarchuk.Anna.csv ozarchuk_anna@91.219.60.189:~/home/ozarchuk_anna
Ozarchuk_anna@91.219.60.189's password:
Ozarchuk.Anna.csv      2 KB 2.5 KB/s ETA: 00:00:00 100%
C:\Intel>
```

3) модифікація файлу CSV-формату:

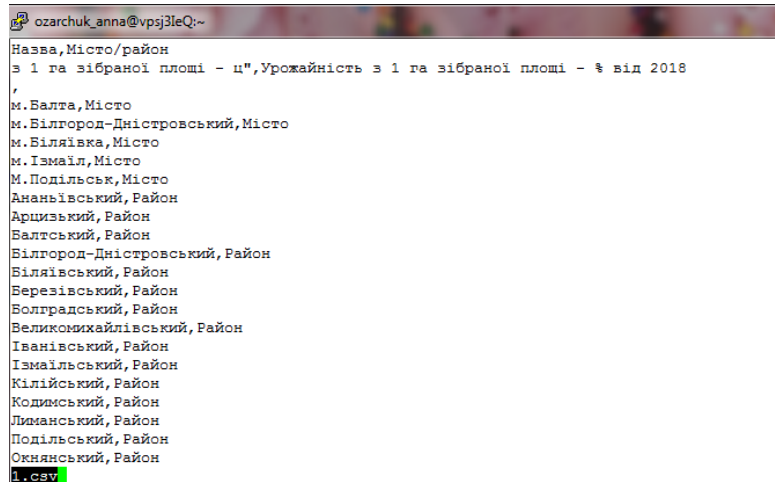
– заміна в числах коми на точку з одночасним видаленням лапки;

– 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

```
-----
[ozarchuk_anna@vpsj3IeQ ~]$ sed 's/"/([0-9]\+)[,]\([0-9]\+\)"/\1.\2 /g' | sed 's/(\,)\...1/\1/g' | sed 's/(\,)\-/\1/g' | cut -d, -f1,2 Ozarchuk
.Aнна.csv > 1.csv
^C
[ozarchuk_anna@vpsj3IeQ ~]$ ls
1.csv          Catalog       Laba3.pdf     Operating-System.Laboratory-Work-1  os.lab1.utf.html  ozarchuk_lab_3
accounts.csv   Laba3.odt     Laba.html     os.lab1.cpi251.html                 Ozarchuk.Anna.csv  Tree.txt
[ozarchuk_anna@vpsj3IeQ ~]$ less 1.csv
```



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

```
-----
[ozarchuk_anna@vpsj3IeQ Operating-System.Laboratory-Work-1]$ 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': Ozarchuk
Password for 'https://Ozarchuk@github.com':
Counting objects: 4, done.
Delta compression using up to 2 threads.
Compressing objects: 100% (3/3), done.
Writing objects: 100% (3/3), 707 bytes | 0 bytes/s, done.
Total 3 (delta 1), reused 0 (delta 0)
remote: Resolving deltas: 100% (1/1), completed with 1 local object.
To https://github.com/mariasirenko2/Operating-System.Laboratory-Work-1
20405de..a7bd805  main -> main
```

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

```
-----
[ozarchuk_anna@vpsj3IeQ Operating-System.Laboratory-Work-1]$ cp /home/ozarchuk_anna/1.csv /home/ozarchuk_anna/Operating-System.Laboratory-Work-1
[ozarchuk_anna@vpsj3IeQ Operating-System.Laboratory-Work-1]$ ls
1.csv          sirenko_mariia_2.csv  SirenkoMariia.pdf  Машина_Перейри_1_4.pptx  Машина_Перейри_full.pptx  ОС лаба 3.pdf
Ozarchuk Anna.txt  Sirenko Mariia.pdf    Sirenko Mariia.txt  Машина Перейри 2 3.pptx  ОС лаба2.pdf
```

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

```
ozarchuk_anna@vpsj3leQ:~/Operating-System.Laboratory-Work-1
[ozarchuk_anna@vpsj3leQ Operating-System.Laboratory-Work-1]$ join -t, -a 1 -a 2 -o0,1,2,2,2 -e '' 1.csv sirenko_mariia_2.csv
Назва,Місто/район,місто/район/область
, 1 га зібраної площі - ц",Урожайність з 1 га зібраної площі - % від 2018
,,
М.Балта,Місто,
М.Білгород-Дністровський,Місто,
М.Білявка,Місто,
М.Ізмаїл,Місто,
М.Подільськ,Місто,
join: 1.csv:9: is not sorted: Ананьївський, Район
Ананьївський, Район,
Арцизький, Район,
Балтський, Район,
Білгород-Дністровський, Район,
Білявський, Район,
Березівський, Район,
Болградський, Район,
Великомихайлівський, Район,
Іванівський, Район,
Ізмаїльський, Район,
Кілійський, Район,
Кодимський, Район,
Львівський, Район,
Одеська область, , область
join: sirenko_mariia_2.csv:3: is not sorted: м.Одеса,місто
м.Одеса,,місто
м.Балта (міськрада),,місто
м.Білгород-Дністровський (міськрада),,місто
м.Білявка (міськрада),,місто
м.Ізмаїл,,місто
м.Чорноморськ (Іллічівськ) (міськрада),,місто
м.Подільськ,,місто
м.Теплодар,,місто
м.Євпаторія,,місто
Ананьївський,,район
Арцизький,,район
Балтський,,район
Білгород-Дністровський,,район
Білявський,,район
Березівський,,район
Болградський,,район
Великомихайлівський,,район
Іванівський,,район
Ізмаїльський,,район
```

## 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.

```
ozarchuk_anna@vpsj3leQ:~/Operating-System.Laboratory-Work-1
[ozarchuk_anna@vpsj3leQ Operating-System.Laboratory-Work-1]$ sort --field-separator="," --key=2 -r 3.csv
, 1 га зібраної площі - ц",Урожайність з 1 га зібраної площі - % від 2018
Ширявський, 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
Захарівський, Район, район
Львівський,, район
Кодимський,, район
Кілійський,, район
Ізмаїльський,, район
Іванівський,, район
Великомихайлівський,, район
Болградський,, район
Білявський,, район
Білгород-Дністровський,, район
Березівський,, район
Балтський,, район
Арцизький,, район
Ананьївський,, район
Львівський, Район,
Кодимський, Район,
Кілійський, Район,
```



## Висновки

Під час виконання Лабораторної роботи №4 я набула практичних навичок зі складною обробкою текстових даних засобами оболонки UNIX-подібних ОС інтерфейсу командного рядка. Найскладнішим завданням було склеювання двох csv файлів.