

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ ПОЛІТЕХНИЧНИЙ УНІВЕРСИТЕТ
ІНСТИТУТ КОМП'ЮТЕРНИХ СИСТЕМ
КАФЕДРА «ІНФОРМАЦІЙНИ СИСТЕМИ»

Лабораторна робота №4

з дисципліни

«Операційні системи»

Тема

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

Виконав:

Студент групи АІ-202

Узун Михайло

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/1ITMLdjcw7nboUo9nFw4NwkWcvjbZeZZ1DVb2Hg4laUQ>

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-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 файла стовпчиків окрім першого та третього;

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

5) отримання CSV-файлу з GitHub-репозиторій, 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.

Хід роботи:

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

```
[uzun_mihajlo@vpsj3IeQ ~]$ find / -name "uzu*" 2> /dev/null | head -n10
/var/spool/mail/uzun_mihajlo
/home/uzun_mihajlo
/home/uzun_mihajlo/uzun_lab_3
[uzun_mihajlo@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.

```
[uzun_mihajlo@vpsj3IeQ ~]$ find /home/ -name "u*" 2> /dev/null
/home/uzun_mihajlo
/home/uzun_mihajlo/uzun_lab_3
/home/uzun_mihajlo/.git/hooks/update.sample
/home/uzun_mihajlo/.config/libreoffice/4/user
/home/uzun_mihajlo/.config/libreoffice/4/user/uno_packages
/home/uzun_mihajlo/.config/libreoffice/4/user/uno_packages/cache/uno_packages
/home/soft-1/.cache/masscan/src/util-malloc.h
/home/soft-1/.cache/masscan/src/unusedparm.h
/home/soft-1/.cache/masscan/src/util-bool.h
/home/soft-1/.cache/masscan/src/util-malloc.c
/home/soft-1/.cache/masscan/tmp/util-malloc.o
```

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).

```
[uzun_mihajlo@vpsj3IeQ ~]$ find /home -type d -name "u*" 2>/dev/null | cut -b 7-
24 | sort -r | uniq
uzun_mihajlo/uzun_
uzun_mihajlo/.conf
uzun_mihajlo
```

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.

```
[uzun_mihajlo@vpsj3IeQ ~]$ cut -f 1,4,6 -d: /etc/passwd > accounts.csv | sed 's/:/,/g'  
accounts.csv  
root,0,/root  
bin,1,/bin  
daemon,2,/sbin  
adm,4,/var/adm  
lp,7,/var/spool/lpd  
sync,0,/sbin  
shutdown,0,/sbin  
halt,0,/sbin  
mail,12,/var/spool/mail  
operator,0,/root  
games,100,/usr/games  
ftp,50,/var/ftp  
nobody,99,/br/>systemd-network,192,/br/>dbus,81,/br/>polkitd,997,/br/>postfix,89,/var/spool/postfix  
chrony,996,/var/lib/chrony  
sshd,74,/var/empty/sshd  
soft,1000,/home/soft  
saslauth,76,/run/saslauthd  
mailnull,47,/var/spool/mqueue  
smmsp,51,/var/spool/mqueue  
emps,1001,/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,54331,/home/arestov_yaroslav  
martinyuk_denis,54332,/home/martinyuk_denis  
bogachik_egor,54333,/home/bogachik_egor  
eliseev_danilo,54334,/home/eliseev_danilo
```

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

```
[uzun_mihajlo@vpsj3IeQ ~]$ grep [0-9][0-9][0-9][6][0-9] accounts.csv  
melnichenko_egor:54360:/home/melnichenko_egor  
perets_sergij:54361:/home/perets_sergij  
uzun_mihajlo:54362:/home/uzun_mihajlo  
polyanskij_mikola:54363:/home/polyanskij_mikola  
sherbakov_artur:54364:/home/sherbakov_artur  
ershov_oleksandr:54365:/home/ershov_oleksandr  
zagorujko_maksim:54366:/home/zagorujko_maksim  
zijodulloev_ehson:54367:/home/zijodulloev_ehson  
veselkova_anna:54368:/home/veselkova_anna  
voyakovskij_dmitro:54369:/home/voyakovskij_dmitro  
[uzun_mihajlo@vpsj3IeQ ~]$ █
```

2.2 Обробка текстового файлу формату CSV

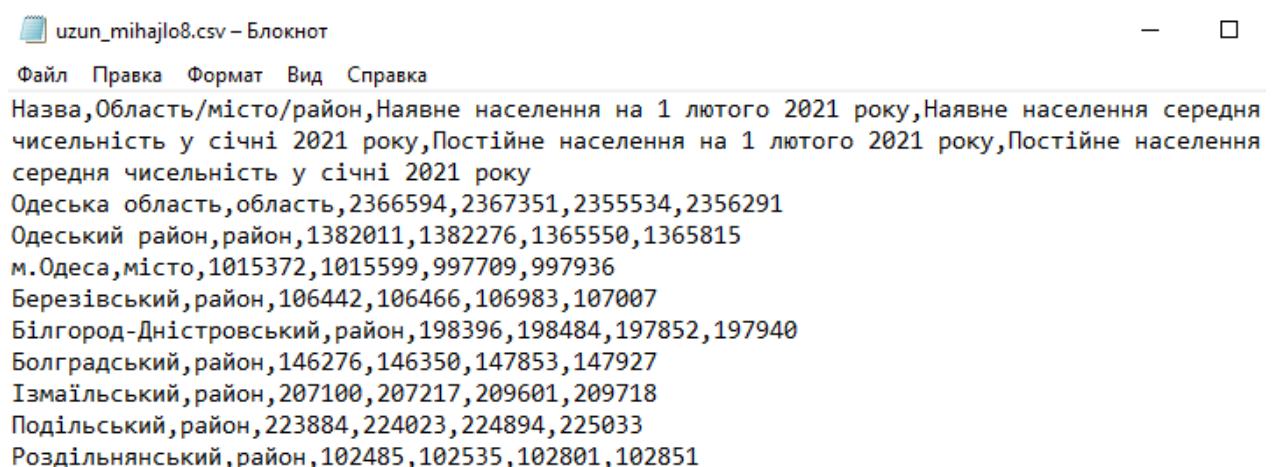
Нехай існують два файли HTML-формату DOC з табличними даними, які визначено у відповідності із номером вашої команди та номером участника в команді, як в лабораторній роботі № 3 -
<https://docs.google.com/spreadsheets/d/1ITMLdjcw7nboUo9nFw4NwkWcvjbZeZZlDVb2Hg4laUQ>

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

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

<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.



Назва,Область/місто/район,Наявне населення на 1 лютого 2021 року,Наявне населення середня чисельність у січні 2021 року,Постійне населення на 1 лютого 2021 року,Постійне населення середня чисельність у січні 2021 року
Одеська область,область,2366594,2367351,2355534,2356291
Одеський район,район,1382011,1382276,1365550,1365815
м.Одеса,місто,1015372,1015599,997709,997936
Березівський,район,106442,106466,106983,107007
Білгород-Дністровський,район,198396,198484,197852,197940
Болградський,район,146276,146350,147853,147927
Ізмаїльський,район,207100,207217,209601,209718
Подільський,район,223884,224023,224894,225033
Роздільнянський,район,102485,102535,102801,102851

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

```
C:\Users\Misha\Desktop\ОНПУ\0C\лаба4>pscp.exe -P 22 uzun_mihajlo8.csv uzun_mihajlo@91.219.60.189:/home/uzun_mihajlo
uzun_mihajlo@91.219.60.189's password:
uzun_mihajlo8.csv | 0 kB | 1.0 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 третього;

```
[uzun_mihajlo@vpsj3IeQ ~]$ cut -f1,3 -d, uzun_mihajlo8.csv > uzun_mihajlo8_1.csv
[uzun_mihajlo@vpsj3IeQ ~]$ cat uzun_mihajlo8_1.csv
Назва,Наявне населення на 1 лютого 2021 року
Одеська область,2366594
Одеський район,1382011
м.Одеса,1015372
Березівський,106442
Білгород-Дністровський,198396
Болградський,146276
Ізмаїльський,207100
Подільський,223884
Роздільнянський,102485
[uzun_mihajlo@vpsj3IeQ ~]$
```

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



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

```
1.sh          Lab2.html        MyOSParam.sh      perets_sergij8.csv
accounts.csv  Lab2.pdf         os.lab1.cpl251.html uzun_lab_3
Lab2.docx     my change file.sh os.lab1.utf.html uzun_mihajlo8_1.csv
```

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

```
[uzun_mihajlo@vpsj3IeQ ~]$ join -t, -jl 1 -j2 1 uzun_mihajlo8.csv perets_sergij8.csv > one.csv
join: perets_sergij8.csv:8: is not sorted: Ананьівський,11439
join: uzun_mihajlo8.csv:4: is not sorted: м.Одеса,1015372
[uzun_mihajlo@vpsj3IeQ ~]$ cat one.csv
Назва,Наявне населення на 1 лютого 2021 року,Зібрана площа
Одеська область,2366594,286578
Подільський,223884,9490
Роздільнянський,102485,14679
[uzun_mihajlo@vpsj3IeQ ~]$
```

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.

```
[uzun_mihajlo@vpsj3IeQ ~]$ sort -rn -t, -k3 one.csv
Одеська область,2366594,286578
Роздільнянський,102485,14679
Подільський,223884,9490
Назва,Наявне населення на 1 лютого 2021 року,Зібрана площа
[uzun_mihajlo@vpsj3IeQ ~]$
```

Висновок: під час виконання даної лабораторної роботи були отримані навички складної обробки текстових даних роботи засобами оболонки Unix-подібних ОС інтерфейсу командного рядка.