

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
Нестеренко М.О.

Перевірив:
Блажко О. А.

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

Завдання для виконання:

1. Обробка текстових даних результатів роботи команд

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

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

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

1.3) 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).

1.4) 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.

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

2. 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:

2.1) 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.

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

2.3) 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

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

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

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

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

Хід роботи:

1.1) Виводять на екран перші десять назв каталогів або файлів, назви яких починаються з перших трьох букв прізвища в латиниці – bel (рис. 1.1)

```
bash login: ssh: run: 14 17.20.11 2021 from 178.130.200.177
[nesterenko_mikola@vpsj3IeQ ~]$ find / name -name "nes*" 2> /dev/null |head
/var/spool/mail/nesterenko_mikola
/usr/lib/modules/3.10.0-1127.19.1.el7.x86_64/kernel/drivers/infiniband/hw/nest
/usr/lib/modules/3.10.0-693.el7.x86_64/kernel/drivers/infiniband/hw/nest
/usr/share/icons/Adwaita/cursors/nestw-resize
/usr/include/rdma/nest-abi.h
/usr/include/c++/4.8.2/bits/nested_exception.h
/home/nesterenko_mikola
/home/nesterenko_mikola/nesterenko_lab_3
[nesterenko_mikola@vpsj3IeQ ~]$
```

Рис. 1.1

1.2) Назва домашнього каталогу містить прізвище та ім'я в транслітерації. Виводять на екран список назв домашніх каталогів користувачів, в назві яких присутня перша буква прізвища – b (рис. 1.2)

```
[nesterenko_mikola@vpsj3IeQ ~]$ find /home -maxdepth 1 -type d -name "n*" 2> /dev/null
/home/nguen_trong
/home/nikolaenko_sergij
/home/nosov_andrij
/home/nesterenko_mikola
/home/nezhivih_mariya
[nesterenko_mikola@vpsj3IeQ ~]$
```

Рис. 1.2

1.3) Модифікують рішення попереднього завдання так, щоб на екран було виведено імена користувачів, відсортовані в порядку зростання значення, та без дублювання цих значень (використовують конвеєр з команд cut, sort, uniq).

```
[nesterenko_mikola@vpsj3IeQ ~]$ find /home -maxdepth 1 -type d -name "n*" 2> /dev/null | cut -d '_' -f2 | sort | uniq
andrij
mariya
mikola
sergij
trong
[nesterenko_mikola@vpsj3IeQ ~]$
```

Рис. 1.3

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

```
[nesterenko_mikola@vpsj3IeQ ~]$ cut -d ':' -f1,3,6 /etc/passwd | tr ':' ',' > accounts.csv
[nesterenko_mikola@vpsj3IeQ ~]$ ls
accounts.csv      nesterenko_lab_3      os.lab1.cp1251.html   testdoc.docx
index.html        Operating-System.-Laboratory-Work-1  os.lab1.utf.html     testdoc.pdf
[nesterenko_mikola@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
```

Рис 1.4

1.5) З файлу accounts.csv отримують на екран перелік облікових записів, в яких четверта цифра ідентифікатора користувача співпадає із власною четвертою цифрою ідентифікатора користувача (54395).

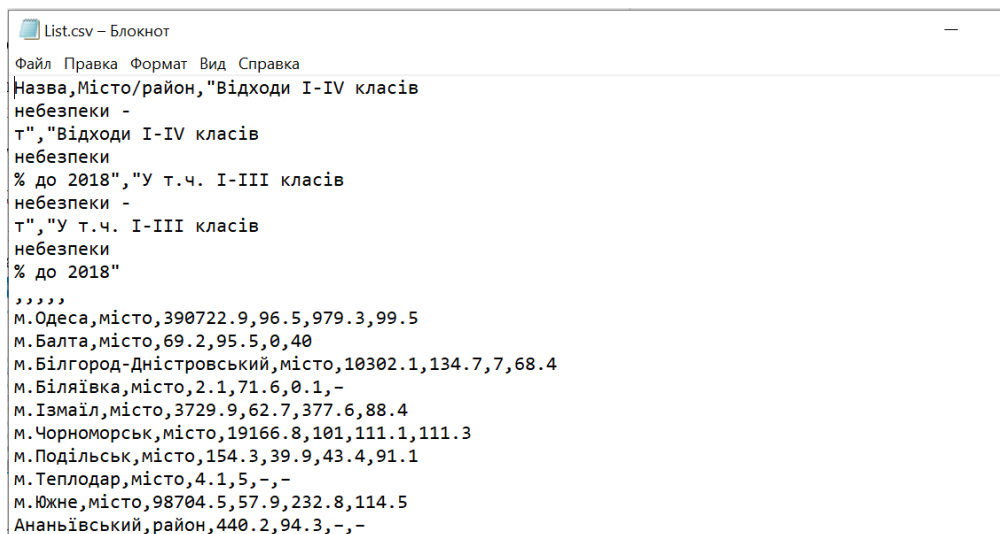
```
[nesterenko_mikola@vpsj3IeQ ~]$ grep -E ",[0-9]{3}9" "accounts.csv"
goshurenko_volodimir,54390,/home/goshurenko_volodimir
karaulnij_daniil,54391,/home/karaulnij_daniil
dorozhkin_mihajlo,54392,/home/dorozhkin_mihajlo
plaksivij_danilo,54393,/home/plaksivij_danilo
kolesnik_kirilo,54394,/home/kolesnik_kirilo
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
[nesterenko_mikola@vpsj3IeQ ~]$
```

Рис. 1.5

2. Обробка текстового файлу формату CSV

2.1) Створюють файл CSV-формату, використовуючи Google Docs за методичними рекомендаціями, із наступними характеристиками (рис. 2.1):

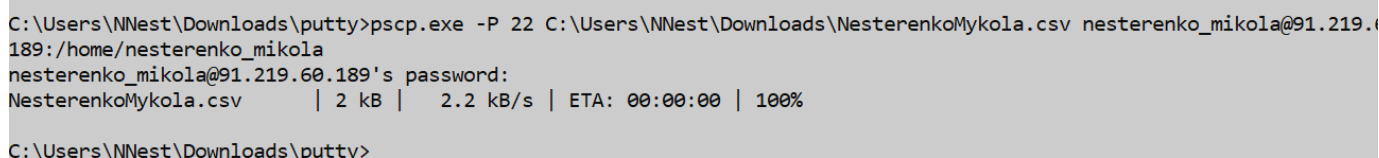
- кодування символів – UTF-8
- роздільник між колонками – символ коми;
- назви файлів визначити з урахуванням транслітерації свого прізвища та номеру файлу.



```
List.csv - Блокнот
Файл Правка Формат Вид Справка
Назва,Місто/район,"Відходи I-IV класів
небезпеки -
т","Відходи I-IV класів
небезпеки
% до 2018","У т.ч. I-III класів
небезпеки -
т","У т.ч. I-III класів
небезпеки
% до 2018"
,,,,,
м.Одеса,місто,390722.9,96.5,979.3,99.5
м.Балта,місто,69.2,95.5,0,40
м.Білгород-Дністровський,місто,10302.1,134.7,7,68.4
м.Біляївка,місто,2.1,71.6,0.1,-
м.Ізмаїл,місто,3729.9,62.7,377.6,88.4
м.Чорноморськ,місто,19166.8,101,111.1,111.3
м.Подільськ,місто,154.3,39.9,43.4,91.1
м.Теплодар,місто,4.1,5,-,-
м.Южне,місто,98704.5,57.9,232.8,114.5
Ананьївський,район,440.2,94.3,-,-
```

Рис 2.1

2.2) Копіюють файл на сервер за адресою з IP = 91.219.60.189 у домашній каталог (рисю 2.2)



```
C:\Users\NNest\Downloads\putty>pscp.exe -P 22 C:\Users\NNest\Downloads\NesterenkoMykola.csv nesterenko_mikola@91.219.60.189:/home/nesterenko_mikola
nesterenko_mikola@91.219.60.189's password:
NesterenkoMykola.csv      | 2 kB | 2.2 kB/s | ETA: 00:00:00 | 100%
C:\Users\NNest\Downloads\putty>
```

Рис. 2.2

2.3) модифікація файлу CSV-формату (рис. 2.3):

- 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

```
[nesterenko_mikola@vpsj3IeQ ~]$ sed 's/^\"([0-9]\+)[,]\\"([0-9]\+)[,]\\"^1\\.2 /g' | sed 's/^\"(.\\)...1^1/g' | sed 's/^\"(.\\)^1/g' |
t -f1,3 -d, NesterenkoMykola.csv > 2.csv
^C
[nesterenko_mikola@vpsj3IeQ ~]$ ls
2.csv          index.html      NesterenkoMykola.csv      os.lab1.cpl251.html      testdoc.docx
accounts.csv   nesterenko_lab_3 Operating-System.-Laboratory-Work-1 os.lab1.utf.html      testdoc.pdf
[nesterenko_mikola@vpsj3IeQ ~]$ less 2.csv
Назва, "Вісхоги І-ІV класів
небезпеки -
м"
небезпеки
% go 2018"
небезпеки -
м"
небезпеки
% go 2018"
/
м.Огеса,390722.9
м.Валта,69.2
м.Вілеороз-Дністровський,10302.1
м.Віляівка,2.1
м.Ізмаїл,3729.9
м.Чорноморськ,19166.8
м.Подільськ,154.3
м.Теплогар,4.1
м.Южне,98704.5
Ананьївський,440.2
Арцизький,6617.6
Валтський,1141.9
Вілеороз-Дністровський,9012.9
Віляівський,2525.9
Верезівський,1226
Волградський,1378.9
Великомихайлівський,1027.5
```

Рис 2.3

2.4) передача створеного CSV-файлу в GitHub-репозиторій (рис 2.4);

```
[nesterenko_mikola@vpsj3IeQ ~]$ git push origin master -f
Username for 'https://github.com': bezpechno
Password for 'https://bezpechno@github.com':
Counting objects: 6, done.
Delta compression using up to 2 threads.
Compressing objects: 100% (5/5), done.
Writing objects: 100% (6/6), 1.43 KiB | 0 bytes/s, done.
Total 6 (delta 0), reused 0 (delta 0)
To https://github.com/kirillkolesnyk/Operating-System.-Laboratory-Work-4
+ 497104c...61d9351 master -> master (forced update)
[nesterenko_mikola@vpsj3IeQ ~]$
```

Рис. 2.4

2.5) отримання CSV-файлу, створеного іншим учасником команди (рис 2.5)

```
[nesterenko_mikola@vpsj3IeQ ~]$ cp Operating-System.-Laboratory-Work-4/1.csv
me/nesterenko_mikola
[nesterenko_mikola@vpsj3IeQ ~]$
```

Рис 2.5

2.6) об'єднання двох файлів в один за першим стовпчиком (рис 2.6);

```
[nesterenko_mikola@vpsj3IeQ ~]$ join -t, -a 1 -a 2 -o0,1,2,2.2 -e ' ' 1.csv 2.csv
Назва,Обсягу  вуків - м,"Вісхоу I-IV класів
м.Одеса,20120.9,
м.Балта,52.6,
м.Білогород-Дністровський,449.7,
м.Віляївка,45.7,
м.Ізмаїл,831.5,
м.Чорноморськ,1605.9,
join: 1.csv:8: is not sorted: м.Погольськ,394.6
м.Погольськ,394.6,
м.Теплогар,48.5,
м.Южне,1900.7,
Ананьївський,315.8,
Арцизький,254,
Балтський,18.7,
Білогород-Дністровський,164.5,
Віляївський,151,
Верезівський,70.3,
Болградський,1150.3,
Великомихайлівський,239,
Іванівський,24.2,
```

Рис. 2.6

2.7) виведення на екран змісту файлу з рішення попереднього завдання у зворотному напрямі для колонки за номером 2 - для 1-го учасника команди, або за номером 3 – для 2-го учасника команди (рис 2.7).

```
[nesterenko_mikola@vpsj3IeQ ~]$ sort --field-separator="," --key=3 -r 3.csv
Назва,Обсягу  вуків - м,"Вісхоу I-IV класів
Погільський,,9.9
м.Южне,,98704.5
Любашівський,,944.1
Кодимський,,91.4
Розгільнянський,,903.6
Білогород-Дністровський,,9012.9
Тарутинський,202,8395.6
м.Балта,,69.2
Арцизький,,6617.6
Саратський,,49.6
Ананьївський,,440.2
Захарівський,,426
м.Теплогар,,4.1
м.Одеса,,390722.9
Окнянський,,37.5
```

Рис 2.7

Висновки: в ході виконання лабораторної роботи, були придбані навички складної обробки текстових даних роботи засобами оболонки Unix подібних ОС інтерфейсу командного рядка. Найскладнішим завданням було сортування двох csv файлу.