

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

Бєлобров А. О.

Перевірив:

Блажко О. А.

Одеса – 2021

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 ваші 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/1ITMLdjcw7nboUo9nFw4NwkWcvjbZeZZIDVb2Hg4laUQ>

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-1QybKQje49VdjFahkely1Ou>:

- 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;

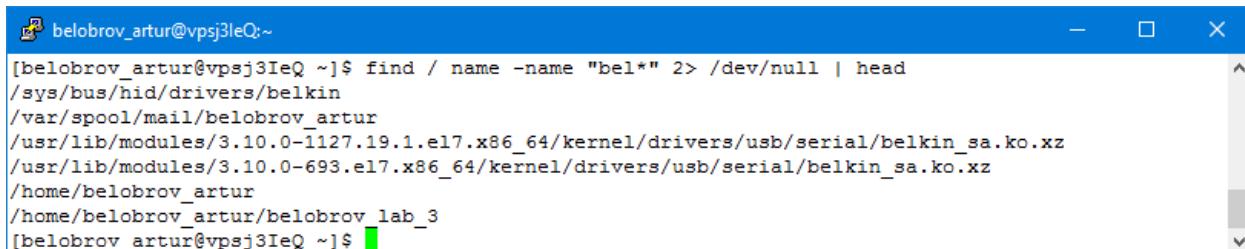
2.5) отримання CSV-файлу, створеного іншим учасником команди;

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:

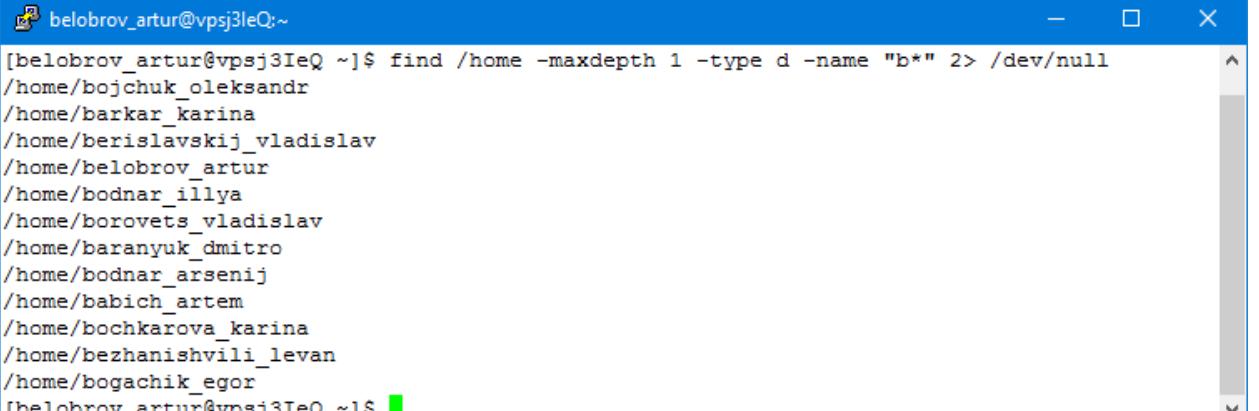
1.1) Виводять на екран перші десять назв каталогів або файлів, назви яких починаються з перших трьох букв прізвища в латиниці – bel (рис. 1.1)



```
belobrov_artur@vpsj3IeQ:~$ find / name -name "bel*" 2> /dev/null | head
/sys/bus/hid/drivers/belkin
/var/spool/mail/belobrov_artur
/usr/lib/modules/3.10.0-1127.19.1.e17.x86_64/kernel/drivers/usb/serial/belkin_sa.ko.xz
/usr/lib/modules/3.10.0-693.e17.x86_64/kernel/drivers/usb/serial/belkin_sa.ko.xz
/home/belobrov_artur
/home/belobrov_artur/belobrov_lab_3
[belobrov_artur@vpsj3IeQ ~]$
```

Рис. 1.1

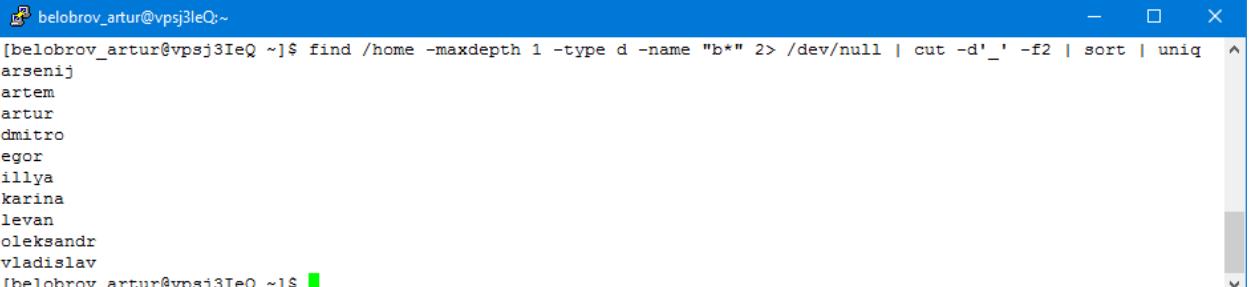
1.2) Назва домашнього каталогу містить прізвище та ім'я в транслітерації. Виводять на екран список назв домашніх каталогів користувачів, в назві яких присутня перша буква прізвища – b (рис. 1.2)



```
belobrov_artur@vpsj3IeQ:~$ find /home -maxdepth 1 -type d -name "b*" 2> /dev/null
/home/bojchuk_oleksandr
/home/barkar_karina
/home/berislavskij_vladislav
/home/belobrov_artur
/home/bodnar_illya
/home/borovets_vladislav
/home/baranyuk_dmitro
/home/bodnar_arsenij
/home/babich_artem
/home/bochkarova_karina
/home/bezhanishvili_levan
/home/bogachik_egor
[belobrov_artur@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).



```
belobrov_artur@vpsj3IeQ:~$ find /home -maxdepth 1 -type d -name "b*" 2> /dev/null | cut -d'_' -f2 | sort | uniq
arsenij
artem
artur
dmitro
egor
illya
karina
levan
oleksandr
vladislav
[belobrov_artur@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.

```

belobrov_artur@vpsj3IeQ:~$ clear
[belobrov_artur@vpsj3IeQ ~]$ ls
BelobrovArthur.docx belobrov_lab_3 Cow.txt os.lab1.cp1251.html test.html
BelobrovArthur.pdf Catalog Operating-System.-Laboratory-Work-1 os.lab1.utf.html
[belobrov_artur@vpsj3IeQ ~]$ cut -d':' -f1,3,6 /etc/passwd | tr ':' ',' > accounts.csv
[belobrov_artur@vpsj3IeQ ~]$ ls
accounts.csv BelobrovArthur.pdf Catalog Operating-System.-Laboratory-Work-1 os.lab1.utf.html
BelobrovArthur.docx belobrov_lab_3 Cow.txt os.lab1.cp1251.html test.html
[belobrov_artur@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
saslauth,997,/run/saslauthd
mailnull,47,/var/spool/mqueue

```

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 (54388).

```

belobrov_artur@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
[belobrov_artur@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.

The screenshot shows a Windows Notepad window titled "BelobrovArthur.csv — Блокнот". The menu bar includes "Файл", "Правка", "Формат", "Вид", and "Справка". The main content area displays a CSV file with the following data:

Назва,місто/район,"Велика рогата худоба усього голів",Велика рогата худоба усього у % до відповідної дати 2020р.,Вел
м.Балта,місто,,,,,
м.Біляївка,місто,,,,,
Ананьївський,район,414,107.5,157,108.3,,,
Арцизький,район,443,37.1,233,44.5,15462,99.1,,
Балтський,район,76,47.2,51,63.8,,,
Білгород-Дністровський,район,1529,98.3,741,100.3,943,45.2,9.7,68.1
Біляївський,район,900,80.8,329,69.6,1375,77.8,,
Березівський,район,2707,87,1149,95.6,,,
Болградський,район,354,40.5,199,54.4,16957,103.8,,
Великомихайлівський,район,399,81.1,145,79.7,812,69.6,,

Рис 2.1

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

The screenshot shows a Windows Command Prompt window titled "C:\Windows\system32\cmd.exe". The command entered is:

```
C:\Users\Admin>pscp.exe -P 22 C:\Users\Admin\Desktop\BelobrovArthur.csv belobrov_artur@91.219.60.189:/home/belobrov_artur  
belobrov_artur@91.219.60.189's password:  
BelobrovArthur.csv | 2 kB | 2.0 kB/s | ETA: 00:00:00 | 100%
```

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

```
[belobrov_artur@vpsj3IeQ:~]$ sed 's/"\([0-9]+\+\)[,]\([0-9]+\+\"/1.\2 /g' | sed 's/\(,\)\+1/\1/g' | sed 's/\(,\)-/\1/g' | cut -d, -f1,2 BelobrovArthur.r.csv > 1.csv
^C
[belobrov_artur@vpsj3IeQ:~]$ ls
1.csv          Catalog
2.csv          Cow.txt
accounts.csv   Operating-System.-Laboratory-Work-1
BelobrovArthur.csv os.lab1.cp1251.html
BelobrovArthur.docx os.lab1.utf.html
BelobrovArthur.pdf test.html
belobrov_lab_3
[belobrov_artur@vpsj3IeQ:~]$ less 1.csv
Назва,місто/район
худоба усього
голів",Велика рогата худоба усього у % до відповідної дати 2020р.
м.Балта,місто
м.Біляївка,місто
Ананьївський,район
Арцизький,район
Балтський,район
Білгород-Дністровський,район
Біляївський,район
Березівський,район
Болградський,район
Великомихайлівський,район
Іванівський,район
Ізмаїльський,район
Кілійський,район
Кодимський,район
Лиманський,район
Подільський,район
Окнянський,район
Любашівський,район
Миколаївський,район
Овідіопольський,район
```

Рис 2.3

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

```
[belobrov_artur@vpsj3IeQ:~]$ git push -u origin --all
Username for 'https://github.com': GrEFeRFeeD
Password for 'https://GrEFeRFeeD@github.com':
Counting objects: 75, done.
Delta compression using up to 2 threads.
Compressing objects: 100% (56/56), done.
Writing objects: 100% (75/75), 867.74 KiB | 0 bytes/s, done.
Total 75 (delta 4), reused 0 (delta 0)
remote: Resolving deltas: 100% (4/4), done.
```

Рис. 2.4

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

```
[belobrov_artur@vpsj3IeQ:~]$ cp Operating-System-Laboratory-Work-4/2.csv /home/belobrov_artur
[belobrov_artur@vpsj3IeQ:~]$ ls
1.csv      BelobrovArthur.csv  belobrov_lab_3  Operating-System-Laboratory-Work-4  test.html
2.csv      BelobrovArthur.docx  Catalog        os.lab1.cp1251.html
accounts.csv BelobrovArthur.pdf Cow.txt       os.lab1.utf.html
```

Рис 2.5

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

```
[belobrov_artur@vpsj3leQ:~]$ clear  
[belobrov_artur@vpsj3leQ:~]$ join -t, -a 1 -a 2 -o0,1.2,2.2 -e '' 1.csv 2.csv  
Назва,місто/район,Місто/Район  
м.Балта,місто,  
м.Біляївка,місто,  
join: 1.csv:4: is not sorted: Ананьївський,район  
Ананьївський,район,  
Арцизький,район,  
Балтський,район,  
Білгород-Дністровський,район,  
Біляївський,район,  
Березівський,район,  
Болградський,район,  
Великомихайлівський,район,  
Іванівський,район,  
Ізмаїльський,район,  
Кілійський,район,  
Кодимський,район,  
Лиманський,район,  
м.Одеса,,Місто  
join: 2.csv:3: is not sorted: м.Балта,Місто  
м.Балта,,Місто  
м.Білгород-Дністровський,,Місто
```

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 участника команди, або за номером 3 – для 2-го участника команди (рис 2.7).

```
[belobrov_artur@vpsj3leQ:~]$ sort --field-separator="," --key=2 -r 3.csv  
Ширяївський,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  
Лиманський,,Район  
Кодимський,,Район  
Кілійський,,Район  
Ізмаїльський,,Район  
Іванівський,,Район  
Великомихайлівський,,Район  
Болградський,,Район  
Біляївський,,Район  
Білгород-Дністровський,,Район  
Березівський,,Район  
Балтський,,Район
```

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 склеювання двох csv файлів.