

МІНІСТЕРСТВО ОСВІТИ Й НАУКИ УКРАЇНИ  
ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ ПОЛІТЕХНІЧНИЙ УНІВЕРСИТЕТ  
Інститут комп'ютерних систем  
Кафедра інформаційних систем

Лабораторна робота № 4  
За дисципліною: "Операційні системи"  
Тема: «Складна обробка текстових даних засобами оболонки  
Unix-подібних ОС інтерфейсу командного рядка»

Виконав:  
Студентка групи АІ-205  
Светашов Д.В.  
Перевірили:  
Блажко О.А.  
Дрозд М.О.

Одеса 2021

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

## 2 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

```
[svetashov_danilo@vpsj3IeQ ~]$ find / -name "sve*" 2> /dev/null | head -n 10
/var/spool/mail/svetashov_danilo
/home/svetashov_danilo
/home/svetashov_danilo/svetashov_lab_3
[svetashov_danilo@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.

```
[svetashov_danilo@vpsj3IeQ home]$ find -maxdepth 2 -name "s*" 2> /dev/null
./sherbakov_georgij
./sherbakov_artur
./strinaglyuk_ivan
./sinyakov_igor
./sultanov_artem
./sokolov_oleksij
./shostak_roman
./sivash_andrij
./stepanenko_gleb
./sirenko_mariya
./sidelnikov_mikita
./svetashov_danilo
./svetashov_danilo/svetashov_lab_3
./shapovalova_viktoriya
./soft-1
./shulyak_mikola
```

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

```
[svetashov_danilo@vpsj3IeQ home]$ find -maxdepth 1 -name "s*" 2> /dev/null |cut -d '_' -f2 | sort |uniq
andrij
artem
artur
danilo
georgij
gleb
igor
ivan
mariya
mikita
mikola
oleksij
roman
./soft-1
viktoriya
[svetashov_danilo@vpsj3IeQ home]$
```

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 з файлу /etc/passwd новий файл з назвою accounts.csv, який буде містити всі облікові записи користувачів, але лише з колонками «ім'я користувача (логін)», «ідентифікатор користувача», «початковий каталог користувача». Колонки повинні мати символом розділу символ «кома», як у стандартному CSV-форматі.

```
[svetashov_danilo@vpsj3IeQ ~]$ sed '1,33d; s/:x:/,/; s/:[0-9]*:::/,/; s/:\bin\bash/' /etc/passwd | cat >accounts.csv
[svetashov_danilo@vpsj3IeQ ~]$ cat accounts.csv
arestov_yaroslav,54322,/home/arestov_yaroslav
martinyuk_denis,54323,/home/martinyuk_denis
bogachik_egor,54324,/home/bogachik_egor
eliseev_danilo,54325,/home/eliseev_danilo
bochkarova_karina,54326,/home/bochkarova_karina
tsigankova_anastasiya,54327,/home/tsigankova_anastasiya
vyatkin_danilo,54328,/home/vyatkin_danilo
harkevich_leonid,54329,/home/harkevich_leonid
kalarashan_kostyantyn,54330,/home/kalarashan_kostyantyn
nosov_andrij,54331,/home/nosov_andrij
kebab_dmitro,54332,/home/kebab_dmitro
prihodko_egor,54333,/home/prihodko_egor
```

```
sed '1,33d; s/:x:/,/; s/:[0-9]*:::/,/; s/:\bin\bash/' /etc/passwd | cat >accounts.csv
```

\*указывает на то, что последовательность может быть любой длины  
1,33d; удаляем строки диапазона

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

```
[svetashov_danilo@vpsj3IeQ ~]$ grep -E [^0-9][0-9]{3}2 "accounts.csv"
arestov_yaroslav,54322,/home/arestov_yaroslav
martinyuk_denis,54323,/home/martinyuk_denis
bogachik_egor,54324,/home/bogachik_egor
eliseev_danilo,54325,/home/eliseev_danilo
bochkarova_karina,54326,/home/bochkarova_karina
tsigankova_anastasiya,54327,/home/tsigankova_anastasiya
vyatkin_danilo,54328,/home/vyatkin_danilo
harkevich_leonid,54329,/home/harkevich_leonid
svetashov_danilo,54420,/home/svetashov_danilo
shostak_roman,54421,/home/shostak_roman
chobotar_oleksandra,54422,/home/chobotar_oleksandra
shapovalova_viktoriya,54423,/home/shapovalova_viktoriya
alekseeva_alina,54424,/home/alekseeva_alina
guda_liliya,54425,/home/guda_liliya
topuzanov_bogdan,54426,/home/topuzanov_bogdan
[svetashov_danilo@vpsj3IeQ ~]$
```

```
grep -E [^0-9][0-9]{3}2 "accounts.csv"
```

[^0-9] так мы показываем, что первые три числа могут быть любыми.

После мы считываем любые первые 3 цифры, а после этого и 4ю, в моем случае это 2!

## 2.2 Обробка текстового файлу формату CSV

### Варіант 8-2

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:

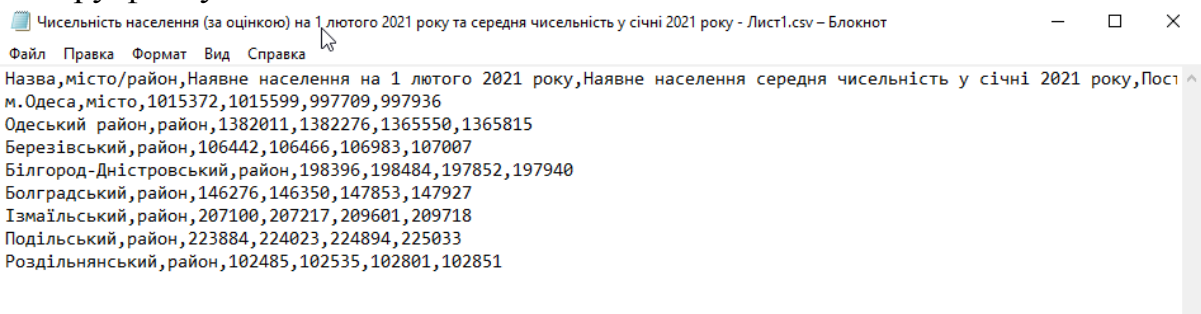
1) базове створення файлу CSV-формату, використовуючи Google Docs за методичними рекомендаціями

<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;

```
C:\Users\Даниил>pscp.exe -P 22 C:\Users\Даниил\Desktop\table.csv svetashov_danilo@91.219.60.189:/home/svetashov_danilo
svetashov_danilo@91.219.60.189's password:
table.csv | 0 kB | 0.9 kB/s | ETA: 00:00:00 | 100%
C:\Users\Даниил>
```

### 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;

```
[svetashov_danilo@vpsj3IeQ ~]$ sed 's/\("[0-9\+\\)\[,]\("[0-9\+\\)\\"/\1.\2 /g' table.csv | cut -f1,2 -d, 1>table2.csv
[svetashov_danilo@vpsj3IeQ ~]$ ls
accounts.csv                                OC_svetashov_danilo_lab_2_html_59106f44e9ae973a.png  OC_svetashov
OC_lab_work_3.doc                          OC_svetashov_danilo_lab_2_html_6458038398eda204.png  OC_svetashov
OC_lab_work_3.pdf                          OC_svetashov_danilo_lab_2_html_6bb51e6bc2fa5a0a.png  OC_svetashov
OC_svetashov_danilo_lab_2.docx              OC_svetashov_danilo_lab_2_html_729a32def4314607.png  OC_svetashov
OC_svetashov_danilo_lab_2.html              OC_svetashov_danilo_lab_2_html_7855ba8049d3ff96.png  OC_svetashov
OC_svetashov_danilo_lab_2_html_12bb55dd42931c9.png  OC_svetashov_danilo_lab_2_html_83f5b19edb26f0b6.png  OC_svetashov
OC_svetashov_danilo_lab_2_html_22ca638ce51d9ce7.png  OC_svetashov_danilo_lab_2_html_8ae0b79bb905a43b.png  Operating-Sy
OC_svetashov_danilo_lab_2_html_23e6cbf0e40cd785.png  OC_svetashov_danilo_lab_2_html_962041ec23dbc921.png  os.labi.cp1
OC_svetashov_danilo_lab_2_html_26f7662e759cb84e.png  OC_svetashov_danilo_lab_2_html_a9326364ab6c93b9.png  os.labi.utf
OC_svetashov_danilo_lab_2_html_2c792f6c140b94d0.png  OC_svetashov_danilo_lab_2_html_ae93560649e77d0.png  svetashov_la
OC_svetashov_danilo_lab_2_html_38d087a45b9a557a.png  OC_svetashov_danilo_lab_2_html_c0e3d58cf0c33b2c.png  table2.csv
OC_svetashov_danilo_lab_2_html_46fe78cc8fa592d8.png  OC_svetashov_danilo_lab_2_html_c26cf7f0886de73e.png  table.csv
[svetashov_danilo@vpsj3IeQ ~]$ cat table2.csv
Назва,місто/район
м.Одеса,місто
Одеський район,район
Верезівський,район
Білгород-Дністровський,район
Волградський,район
Ізмаїльський,район
Подільський,район
Роздільнянський,район
[svetashov_danilo@vpsj3IeQ ~]$
```

```
sed 's/\("[0-9\+\\)\[,]\("[0-9\+\\)\\"/\1.\2 /g' table.csv | cut -f1,2 -d, 1>table2.csv
```

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

```
Username for 'https://github.com': daniil-svetashov
Password for 'https://daniil-svetashov@github.com':
Counting objects: 4, done.
Delta compression using up to 2 threads.
Compressing objects: 100% (3/3), done.
Writing objects: 100% (3/3), 445 bytes | 0 bytes/s, done.
Total 3 (delta 0), reused 0 (delta 0)
To https://github.com/daniil-svetashov/svetashov_shostack
 79f0984..a4061ff main -> main
[svetashov_danilo@vpsj3IeQ svetashov_shostack]$
```

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

```
[svetashov_danilo@vpsj3IeQ svetashov_shostack]$ ls
shostak81ver2.csv start table2.csv
```

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

```

shostak81ver2.csv start table2.csv
[svetashov_danilo@vpsj3IeQ svetashov_shostack]$ join -t, -a 1 -a 2 -o0,1.2,2.2 -
e '' s^C
[svetashov_danilo@vpsj3IeQ svetashov_shostack]$ join -t, -a 1 -a 2 -o0,1.2,2.2 -e '' shostak81ver2.cs
v table2.csv | cat > table3.csv
join: shostak81ver2.csv:4: is not sorted: ,
join: table2.csv:4: is not sorted: Березівський, район
[svetashov_danilo@vpsj3IeQ svetashov_shostack]$ cat table3.csv
Назва, місто/район, місто/район
з 1 га зібраної площі - ц", "Урожайність,
,, 1 га зібраної площі - % до 2018р."
,,
м. Балта, місто,
м. Вілгород-Дністровський, місто,
м. Біляївка, місто,
м. Ізмаїл, місто,
м. Одеса, місто
М. Подільськ, місто,
Ананьївський, район,
Арцизький, район,
Балтський, район,
Вілгород-Дністровський, район,
Біляївський, район,
Березівський, район,
Болградський, район,
Великомихайлівський, район,
Іванівський, район,
Ізмаїльський, район,
Кілійський, район,
Кодимський, район,
Лиманський, район,
Одеський район, район
Березівський, район
Вілгород-Дністровський, район
Болградський, район
Ізмаїльський, район
Подільський, район, район
Окнянський, район,
Любашівський, район,
Миколаївський, район,
Овідіопольський, район,
Роздільнянський, район, район
Ренійський, район,
Савранський, район,
Саратський, район,
Тарутинський, район,
Татарбунарський, район,
Захарівський, район,
Ширяївський, район,
[svetashov_danilo@vpsj3IeQ svetashov_shostack]$ █

```

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.

```

[svetashov_danilo@vpsj3IeQ svetashov_shostack]$ sort --field-separator="," --key=3 -r table3.csv
Роздільнянський, район, район
Подільський, район, район
Одеський район, район
Ізмаїльський, район
Болградський, район
Вілгород-Дністровський, район
Березівський, район
Назва, місто/район, місто/район
м. Одеса, місто
Ширяївський, район,
Татарбунарський, район,
Тарутинський, район,
Саратський, район,
Савранський, район,
Ренійський, район,
Окнянський, район,
Овідіопольський, район,
М. Подільськ, місто,
м. Ізмаїл, місто,
Миколаївський, район,
м. Біляївка, місто,
м. Вілгород-Дністровський, місто,

```

Висновки: в ході виконання лабораторної роботи, були придбані навички складної обробки текстових даних роботи засобами оболонки Unix подібних ОС інтерфейсу командного рядка.