

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:

Студент групи AI-202

Лукашак Д. О.

Перевірив:

Блажко О. А.

Одеса 2021

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

Завдання:

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. 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 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 файлу-таблиці у зворотному напрямі значень колонки за номером 2 – для 1-го учасника команди, або за номером 3 – для 2-го учасника команди.

Виконання завдання

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

```
[lukashak_daniil@vpsj3IeQ ~]$ find / -name "luk*" 2>stderr | head -n10
/var/spool/mail/lukashak_daniil
/usr/lib/grub/i386-pc/luks.mod
/home/lukashak_daniil
/home/lukashak_daniil/lukashak_lab_3
/home/lukashak_daniil/lukashak_lab_3/lukashak_daniil
```

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.

```
[lukashak_daniil@vpsj3IeQ ~]$ find /home/ -name "l*" 2> stderr
/home/li_chuan
/home/lukashak_daniil
/home/lukashak_daniil/lukashak_lab_3
/home/lukashak_daniil/lukashak_lab_3/lukashak_daniil
/home/lukashak_daniil/Operating-System.-Laboratory-Work-1/.git/logs
/home/lukashak_daniil/Operating-System.-Laboratory-Work-1/lab.pptx
/home/lukashak_daniil/.config/libreoffice
/home/lukashak_daniil/.config/libreoffice/4/user/uno_packages/cache/log.txt
/home/lukashak_daniil/.config/libreoffice/4/user/extensions/shared/lastsynchronized
/home/lukashak_daniil/.config/libreoffice/4/user/extensions/bundled/lastsynchronized
/home/lobko_daniil
/home/loboda_oleksandra
/home/soft-1/.cache/masscan/src/logger.h
/home/soft-1/.cache/masscan/src/logger.c
/home/soft-1/.cache/masscan/tmp/logger.o
/home/li_choan
/home/lihmanov artem
```

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).

```
[lukashak_daniil@vpsj3IeQ ~]$ find /home -name "l*" 2> stderr | cut -d/ -f3 | sort | uniq
li_choan
li_chuan
lihmanov_artem
lobko_daniil
loboda_oleksandra
lukashak_daniil
soft-1
```

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.

```
[lukashak_daniil@vpsj3IeQ ~]$ cut -d: -f1,3,6 /etc/passwd | sed 's:/:/:/g'> accounts.csv
[lukashak_daniil@vpsj3IeQ ~]$ less 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
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
```

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

```
[lukashak_daniil@vpsj3IeQ ~]$ sed -n "/,[0-9]\{3\}5[0-9]/p" accounts.csv
kovach_dmitro,54350,/home/kovach_dmitro
sidelnikov_mikita,54351,/home/sidelnikov_mikita
lukashak_daniil,54352,/home/lukashak_daniil
markovskij_danilo,54353,/home/markovskij_danilo
melnichenko_egor,54354,/home/melnichenko_egor
perets_sergij,54355,/home/perets_sergij
uzun_mihajlo,54356,/home/uzun_mihajlo
polyanskij_mikola,54357,/home/polyanskij_mikola
sherbakov_artur,54358,/home/sherbakov_artur
ershov_oleksandr,54359,/home/ershov_oleksandr
```

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.

	A	B	C	D	E	F
1	Назва	Місто/район	Відходи I-IV класів небезпеки - т	Відходи I-IV класів небезпеки - % до 2018р.	У т.ч. I-III класів небезпеки - т	У т.ч. I-III класів небезпеки - % до 2018р.
2	м.Одеса	місто	390722.9	96.5	979.3	99.5
3	м.Балта	місто	69.2	95.5	0	40
4	м.Білгород-Дністровський	місто	10302.1	134.7	7	68.4
5	м.Біляївка	місто	2.1	71.6	0.1	
6	м.Ізмаїл	місто	3729.9	62.7	377.6	88.4
7	м.Чорноморськ	місто	19166.8	101	111.1	111.3
8	м.Подільськ	місто	154.3	39.9	43.4	91.1
9	м.Теплодар	місто	4.1	5		
10	м.Южне	місто	98704.5	57.9	232.8	114.5
11	Ананівський	район	440.2	94.3		
12	Арцизький	район	6617.6	97.5	1	23.4
13	Балтський	район	1141.9	130.4	1	153.1
14	Білгород-Дністровський	район	9012.9	85.6	11.6	129.9
15	Біляївський	район	2525.9	111	24.9	307.5
16	Березівський	район	1226	100.2		
17	Болградський	район	1378.9	95.1	0.3	
18	Великомихайлівський	район	1027.5	136.2	0.9	56.2
19	Іванівський	район	1194.3	95.6	0.1	9.9
20	Ізмаїльський	район	2845.2	101.7	5.5	126.2
21	Кілійський	район	10429.4	97.1	7.1	108
22	Кодимський	район	91.4	104.6	0.5	100
23	Лиманський	район	27768.4	115.9	90.3	514.6
24	Подільський	район	9.9	87.3	0.9	68.7
25	Окнянський	район	37.5	99.5	0	100
26	Любашівський	район	944.1	66.9	3.2	470.4
27	Миколаївський	район				
28	Овідіопольський	район	35464.8	117.5	44.9	120.9
29	Роздільнянський	район	903.6	7.4	0.7	87.8
30	Ренійський	район	3469.2	87.8	1.8	48.1
31	Савранський	район	166.8	90.1	6.3	161.7
32	Саратський	район	49.6	94.3	0.3	100.4
33	Тарутинський	район	8395.6	100.9	2	91.3

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

```
PS C:\Users\Даниил Лукашак> scp D:\lukashak1.csv lukashak_daniil@91.219.60.189:/home/lukashak_daniil/lukashak1.csv
lukashak_daniil@91.219.60.189's password:
lukashak1.csv 100% 2217 153.1KB/s 00:00
```

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;

```
[lukashak_daniil@vpsj3IeQ ~]$ sed 's/\("[0-9]\+\)[,]\("[0-9]\+\)"\/\1\.\2 /g' lukashak1.csv > lukashak2.csv
[lukashak_daniil@vpsj3IeQ ~]$ sed 's/\(,\)\...1\/\1/g' lukashak2.csv > lukashak3.csv
[lukashak_daniil@vpsj3IeQ ~]$ sed 's/\(,\)\)\-\/\1/g' lukashak3.csv > lukashak4.csv
[lukashak_daniil@vpsj3IeQ ~]$ cut -f1,3 -d, lukashak4.csv > lukashak5.csv
[lukashak_daniil@vpsj3IeQ ~]$ less lukashak5.csv
```

Назва,Відходи I-IV класів небезпеки - т

м.Одеса,390722.9

м.Балта,69.2

м.Білгород-Дністровський,10302.1

м.Біляївка,2.1

м.Ізмаїл,3729.9

м.Чорноморськ,19166.8

м.Подільськ,154.3

м.Теплодар,4.1

м.Южне,98704.5

Ананьївський,440.2

Арцизький,6617.6

Балтський, .9

Білгород-Дністровський,9012.9

Біляївський,2525.9

Березівський,1226

Болградський,1378.9

Великомихайлівський,1027.5

Іванівський,1194.3

Ізмаїльський,2845.2

Кілійський,10429.4

Кодимський,91.4

Лиманський,27768.4

Подільський,9.9

Окнянський,37.5

Любашівський,944.1

Миколаївський,

Овідіопольський,35464.8

Роздільнянський,903.6

Ренійський,3469.2

Савранський,166.8

Саратський,49.6

Тарутинський,8395.6

Татарбунарський,250

Захарівський,426

Ширяївський,125

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;

```
[lukashak_daniil@vpsj3IeQ ~]$ join -t, -j1 1 -j2 1 lukashak5.csv sidelnikov_5.csv > common.csv
join: lukashak5.csv:2: is not sorted: м.Одеса,390722.9
join: sidelnikov_5.csv:3: is not sorted: м.Одеса,109.90
[lukashak_daniil@vpsj3IeQ ~]$ cat common.csv
Роздільнянський,903.6,10.60
Ренійський,3469.2,20.40
Савранський,166.8,116.00
Саратський,49.6,45.50
Тарутинський,8395.6,0
Татарбунарський,250,35.70
Захарівський,426,
Ширяївський,125,101.60
```

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.

```
[lukashak_daniil@vpsj3IeQ ~]$ sort -rn -t, -k3 common.csv
Савранський,166.8,116.00
Ширяївський,125,101.60
Саратський,49.6,45.50
Татарбунарський,250,35.70
Ренійський,3469.2,20.40
Роздільнянський,903.6,10.60
Тарутинський,8395.6,0
Захарівський,426,
```

Висновки

Виконуючи цю лабораторну роботу, ми придбали навички роботи з CSV файлами в UNIX подібних ОС, та навчилися використовувати регулярні вирази для пошуку. Однією з складних команд виявилася sed, так як не дуже зрозуміле її застосування через велику кількість роздільників у параметрі регулярного виразу.