Цевки, филтри и регуларни изрази

## Цев­ки

Опе­ра­то­рот "|", кој се на­ре­ку­ва цев­ка (pipe), се ко­рис­ти за пов­рзу­ва­ње­то на ко­ман­ди­те. При­тоа, из­ле­зот од прва­та ко­ман­да се ко­рис­ти ка­ко влез на вто­ра­та. Мно­гу чес­та ком­би­наци­ја е след­на­та:

|  |
| --- |
| student@os:~$ ls –al | less |

Име­но, ако во име­ни­кот има мно­гу да­то­те­ки, из­ле­зот од ls "ќе избе­га". Ко­ман­да­та less ќе го пре­зе­ме тој из­лез и ќе го при­ка­же стра­на по стра­на.  
Со "|" мо­же да се пов­рзу­ва­а­т про­из­во­лен број ко­ман­ди. Так­ва ко­ман­дна ли­ни­ја се на­ре­ку­ва pipeline. На при­мер со:

|  |
| --- |
| student@os:~$ ls –al Lab2 | sort | cat > lista.txt |

лис­та­та на да­то­те­ки во име­ни­кот Lab2 ќе се сорти­ра и ќе се пре­на­со­чи во да­то­те­ка­та lista.txt.

## Филтри

Фил­три­те се ко­ман­ди ко­и обич­но вр­ша­т тран­сфор­маци­и на не­ко­ја да­то­те­ка, на из­ле­зот од дру­га ко­ман­да или да­ва­а­т ин­фор­маци­и за нив. При­мер на так­ва ко­ман­да е ко­ман­да­та grep. Та­а мо­же да се ко­рис­ти и ка­ко са­мос­тој­на ко­ман­да или во ком­би­наци­ја со дру­га. Неј­зи­ни­от оп­шт об­ли­к е:

**$grep** *ре­гу­ла­рен\_из­раз да­то­те­ка(и)*

grep чи­та од да­то­те­ка­та(и) или од стан­дар­д­ни­от влез и на стан­дар­д­ни­от из­лез ги при­ка­жу­ва ре­до­ви­те ко­и сод­ржа­т ни­за од знаци ко­ја од­го­ва­ра на ре­гу­лар­ни­от из­раз. На­једнос­тав­на­та фор­ма е ко­га ка­ко ре­гу­ла­рен из­раз се ко­рис­ти ни­за од знаци. На при­мер:

|  |
| --- |
| student@os:~$ grep zdravo poraka |

ќе ги при­ка­же си­те ре­до­ви од да­то­те­ка­та poraka ко­и ја сод­ржа­т ни­за­та zdravo. Мно­гу по­чес­то grep се ко­рис­ти во ком­би­наци­ја со дру­ги ко­ман­ди пов­рза­ни со цев­ка. Ко­ман­да­та:

|  |
| --- |
| student@os:~$ cat poraka | grep '^....97' |

ќе ги при­ка­же си­те ре­до­ви од poraka ко­и за­поч­ну­ва­а­т со ни­за од 4 знаци по ко­ја сле­ди '97'. Ако на­мес­то ^ се упот­ре­би $ (на кра­јот) , то­гаш ќе се ба­ра­а­т они­е ре­до­ви ко­и

зав­ршу­ва­а­т на за­да­де­на­та ни­за од знаци. Со оп­ци­ја­та -v се про­на­о­ѓа­а­т си­те ре­до­ви од да­то­те­ка­та ко­и не сод­ржа­т од­ре­де­на ни­за од знаци. Со:

|  |
| --- |
| student@os:~$ grep –v zdravo poraka |

ќе се при­ка­жа­т си­те ре­до­ви од по­ра­ка ко­и не сод­ржа­т zdravo.

Еве еден по­с­ло­жен при­мер на ко­ман­да со фил­тер ко­ја мно­гу чес­то се ко­рис­ти:

|  |
| --- |
| student@os:~$ ls –l | grep '\.txt' | less |

Ре­зул­та­тот ќе би­де лис­та на си­те да­то­те­ки во те­ков­ни­от име­ни­к ко­и зав­ршу­ва­а­т на нас­тав­ка .txt, из­лис­та­ни стра­на по стра­на (со зна­кот " \ " се от­стра­ну­ва спе­ци­јал­но­то зна­че­ње на точ­ка­та, ис­то ва­жи и за ос­та­на­ти­те спе­ци­јал­ни знаци).

Ко­ман­да­та sort ис­то та­ка спа­ѓа во гру­па­та на фил­три­те. Се упот­ре­бу­ва во вид:

**$sort** *оп­ци­и да­то­те­ка*

или пов­рза­на со цев­ка со дру­га ко­ман­да.

Не­ко­и од оп­ци­и­те на sort се:

-f - сорти­ра­ње по абе­це­ден ре­дос­ле­д без раз­ли­ку­ва­ње на го­ле­ми­те и ма­ли­те бук­ви

-n - сорти­ра­ње по ну­ме­рич­ки вреднос­ти

-r - сорти­ра­ње по опа­ѓач­ки ре­дос­лед

-k x (x е цел број) - сорти­ра­ње спо­ред x-ти­от збор во ре­дот

При­ме­ри:

|  |
| --- |
| student@os:~$ ls | sort -f |

ги сорти­ра да­то­те­ки­те од те­ков­ни­от име­ни­к по абе­це­ден ре­дос­лед

|  |
| --- |
| student@os:~$ ls –l | sort –rn –k 5 |

ги сорти­ра да­то­те­ки­те по го­ле­ми­на по опа­ѓач­ки ре­дос­лед

|  |
| --- |
| student@os:~$ who | sort –k 5 |

ги сорти­ра ко­рис­ни­ци­те на сис­те­мот спо­ред нив­но­то вре­ме на на­ја­ву­ва­ње.

Не­ко­гаш са­ка­ме да за­да­де­ме ни­за од по­ве­ќе ко­ман­ди одед­наш на ко­ман­дна­та ли­ни­ја. То­а мо­же да се ре­а­ли­зи­ра со по­мош на опе­ра­то­рот ";" кој се ко­рис­ти на сли­чен на­чин ка­ко и "|", но не ги пов­рзу­ва ко­ман­ди­те ме­ѓу се­бе.

## Ре­гу­лар­ни из­ра­зи

Ре­гу­лар­ни­те из­ра­зи се упот­ре­бу­ва­а­т при раз­ни ма­ни­пу­лаци­и со текст. Ти­е прет­ста­ву­ва­а­т за­ме­на за ни­зи од знаци ко­и за­до­во­лу­ва­а­т од­ре­де­ни ус­ло­ви.

Еден ре­гу­ла­рен из­раз се сос­то­и од опе­ра­то­ри ко­и опи­шу­ва­а­т ни­зи од знаци (слич­но ка­ко \* (за­ме­ну­ва би­ло ко­и ка­рак­те­ри) и ? (за­ме­ну­ва еден ка­рак­тер)). Следни­те ком­би­наци­и од опе­ра­то­ри мо­жа­т да се ко­рис­та­т за кон­струк­ци­ја на ре­гу­лар­ни из­ра­зи:

. - за­ме­ну­ва би­ло кој знак  
^ - оз­на­чу­ва ба­ра­ње од по­че­ток на ред, но не­ко­гаш во за­вис­нос­т од кон­тек­стот го не­ги­ра зна­че­ње­то на мно­жес­тво ка­рак­те­ри  
$ - оз­на­чу­ва ба­ра­ње од крај на ред  
\c - ако c е спе­ци­ја­лен знак, се тре­ти­ра ка­ко оби­чен  
[c1-c2] - би­ло кој знак од ин­тер­ва­лот c1-c2  
[c1c2...cn] - би­ло кој знак од ни­за­та c1c2...cn  
[^c1-c2] - би­ло кој знак кој не при­па­ѓа на ин­тер­ва­лот c1-c2

[^c1c2...cn] - би­ло кој знак кој не при­па­ѓа на ни­за­та c1c2...cn

\< - оз­на­чу­ва ба­ра­ње од по­че­ток на збор

\> - оз­на­чу­ва ба­ра­ње од крај на збор

\b - оз­на­чу­ва ба­ра­ње од по­че­ток или крај на збор

\B - оз­на­чу­ва ба­ра­ње би­ло кој ка­рак­тер но не на по­че­ток или крај на збор

За сек­венци од ис­т ка­рак­тер мо­же и да ги ко­рис­ти­ме сле­дни­ве ре­гу­лар­ни из­ра­зи:

\{n\} - точ­но n па­ти пов­то­ру­ва­ње на по­сле­дни­от ка­рак­тер  
\{n,\} - n или по­ве­ќе па­ти пов­то­ру­ва­ње на по­сле­дни­от ка­рак­тер  
\{n,m\} - ми­ни­мум n а мак­си­мум m па­ти пов­то­ру­ва­ње на по­сле­дни­от ка­рак­тер

При­ме­ри:

a.c - би­ло ко­ја ни­за ко­ја поч­ну­ва на а, зав­ршу­ва на c и има еден знак ме­ѓу нив   
[a-zA-Z] - си­те афа­бет­ски знаци

[0-9].....[aA][^a-z] - би­ло ко­ја ни­за ко­ја поч­ну­ва со ци­фра, по­то­а има уш­те 5 знаци, сле­ди 'а' или 'А', а по­сле­дни­от знак не е ма­ла бук­ва.

ca\{5,10\}t - би­ло кој збор што поч­ну­ва со c а зав­ршу­ва со t а из­ме­ѓу има по­ме­ѓу 5 и 10 ка­рак­те­ри 'а'.

\<c.\*h\> - би­ло кој збор што поч­ну­ва на c и зав­ршу­ва на h

След­на­та ко­ман­да ги лис­та си­те да­то­те­ки ко­и има­а­т нас­тав­ка .cc

|  |
| --- |
| student@os:~$ ls \*.cc |

до­де­ка пак со ко­ман­да­та

|  |
| --- |
| student@os:~$ ls –l [abcxyz]\* | less |

ќе се из­лис­та­а­т си­те име­ни­ци и да­то­те­ки чи­и ими­ња поч­ну­ва­а­т на бук­ви­те 'а', 'b', 'c', 'x', 'y' или 'z'.

За да ги из­бри­ше­те си­те да­то­те­ки со име prog и нас­тав­ка ко­ја има еден знак, за­да­де­те:

|  |
| --- |
| student@os:~$ rm prog? |

до­де­ка пак со ко­ман­да­та

|  |
| --- |
| student@os:~$ rm \*[^1-9] |

ќе се из­бри­ша­т си­те да­то­те­ки чи­е име не зав­ршу­ва на број

Со ко­ман­да­та:

|  |
| --- |
| student@os:~$ mkdir newdir  student@os:~$ cp [a-z]\* newdir/ |

ќе ги ис­ко­пи­ра си­те да­то­те­ки, чи­е име за­поч­ну­ва со  бук­ва во по­ди­ме­ни­кот newdir од те­ков­ни­от име­ник.

|  |
| --- |
| student@os:~$ cp [pP]art[0-9] ../book |

ќе ги ис­ко­пи­ра си­те да­то­те­ки, чи­е име за­поч­ну­ва со p или P, по­то­а сле­ди ни­за­та art и зав­ршу­ва со ци­фра, во име­ни­к book кој се на­о­ѓа едно ни­во по­го­ре од она на те­ков­ни­от име­ник. Вни­ма­вај­те име­ни­кот book тре­ба да се кре­и­ра до­кол­ку не по­стои!

|  |
| --- |
| student@os:~$ mv {ii,[0-9]}\*.{txt,doc,jpg} /home/student/lab2 |

ќе се пре­мес­та­т си­те да­то­те­ки чи­е име за­поч­ну­ва на ii или не­ко­ја ци­фра, па сле­ду­ва ни­за од ка­рак­те­ри и зав­ршу­ва со .txt или .doc или .jpg во име­ни­кот student/lab2

Ко­га се ко­рис­та­т ре­гу­лар­ни из­ра­зи во ком­би­наци­ја со grep, нај­доб­ро е ре­гу­лар­ни­от из­раз се­ко­гаш да се заг­ра­ди со ' ', што ќе го спре­чи ин­тер­пре­те­рот да го проце­си­ра из­ра­зот пред да го пре­да­де на grep.

Со:

|  |
| --- |
| student@os:~$ grep –n '^\$' money |

ќе се из­лис­та­а­т си­те ли­ни­и во да­то­те­ка­та money ко­и за­поч­ну­ва­а­т со $ за­едно со нив­ни­от ре­ден број.

|  |
| --- |
| student@os:~$ ls –l | grep '^d.\* ' |

ќе ги из­лис­та де­тал­но си­те име­ни­ци во те­ков­ни­от име­ник.

Со:

|  |
| --- |
| student@os:~$ grep 'main' \*.c |

ќе се из­лис­та­а­т си­те ли­ни­и од си­те да­то­те­ки со нас­тав­ка .c во те­ков­ни­от име­ни­к ко­и сод­ржа­т ни­за main, а до­де­ка пак со

|  |
| --- |
| student@os:~$ grep \<main\>' \*.c |

ќе се из­лис­та­а­т си­те ли­ни­и од си­те да­то­те­ки со нас­тав­ка .c во те­ков­ни­от име­ни­к ко­и го сод­ржа­т збо­рот main.

Со:

|  |
| --- |
| student@os:~$ ls –l | grep '\.txt$' | cat > poraka |

ќе се да­де де­та­лен из­веш­тај на си­те да­то­те­ки ко­и ја има­а­т нас­тав­ка­та .txt и при­то­а из­ле­зот ќе се пре­на­со­чи во да­то­те­ка­та poraka.

## Ко­ман­да­та cut

Ко­ман­да­та cut врши из­дво­ју­ва­ње ко­ло­ни од по­да­тоци. Од­нос­но ко­ман­да­та ко­рис­ти IFS (Input Field Separator) кој препоз­на­ва по што да се врши од­де­лу­ва­ње­то. Се ко­рис­ти за­едно со сле­дни­ве оп­ции:

-d ':' - де­ли­ми­тер

-f x (ка­де x е цел број) - врши од­де­лу­ва­ње на x-та ко­ло­на

-c x-y (ка­де x и y се це­ли бро­е­ви) - врши од­де­лу­ва­ње на ка­рак­те­ри­те од x до y во по­да­то­кот

При­мер:

|  |
| --- |
| student@os:~$ ls –l Lab2 | grep ' ^-' | tr -s ' ' | cut –f 5 –d ' ' |

Во овој при­мер се врши лис­та­ње на си­те да­то­те­ки од ди­рек­то­ри­у­мот Lab2, а по­то­а со на­ред­ба­та tr –s ' ' се врши спо­ју­ва­ње на по­ве­ќе праз­ни мес­та во едно за­то­а што мо­же да се слу­чи да се ја­ва­т по­ве­ќе праз­ни мес­та по­ме­ѓу ко­ло­ни­те при лис­та­ње­то со ls –l и при­то­а на­ред­ба­та cut да **не** ни ги да­де по­са­ку­ва­ни­те ре­зул­та­ти. Овој из­лез со по­мош на цев­ка се пре­на­со­чу­ва на на­ред­ба­та cut ко­ја што ќе ја из­дво­и 5-та­та ко­ло­на (од­нос­но го­ле­ми­на­та на да­то­те­ки­те) од ос­та­на­ти­те ко­ло­ни ко­и се од­де­ле­ни со по едно праз­но мес­то.

## Ко­ман­да­та awk

Ко­га ја ко­рис­ти­ме awk во ком­би­наци­ја со print, то­а зна­чи да се ис­пе­ча­та­т са­мо спе­ци­фи­ци­ра­ни по­ли­ња. Се ко­рис­ти на следни­ов на­чин:

**$awk** '{print $бр\_*ко­ло­на*, $ бр\_*ко­ло­на*, ... , $ бр\_*ко­ло­на*; }'

При­мер:

|  |
| --- |
| student@os:~$ ls –l | grep '^-' | awk '{print $1, $5, $8;}' |

ќе се ис­пе­ча­та­т на ек­ран за се­ко­ја да­то­те­ка од ва­ши­от те­ко­вен име­ник: при­ви­ле­ги­и­те, неј­зи­на­та го­ле­ми­на и име­то.

## Опе­ра­то­рот grawe ` `

Во слу­ча­и ко­га е пот­реб­но ре­зул­та­тот од из­вршу­ва­ње на ко­ман­да да се смес­ти во сис­тем­ска про­мен­ли­ва мо­же да се ко­рис­ти *grave* опе­ра­то­рот (` ко­ман­да ` ка­де на­вод­ни­ци­те се они­е под коп­че­то Escape кај ~). Тој пре­диз­ви­ку­ва из­вршу­ва­ње на ко­ман­да­та и пре­на­со­чу­ва­ње на стан­дар­д­ни­от из­лез во про­мен­ли­ва­та. Во следни­от при­мер, сод­ржи­на­та на те­ков­ни­от име­ни­к се смес­ту­ва во про­мен­ли­ва­та лис­та:

|  |
| --- |
| student@os:~$ lista=`ls`  student@os:~$ echo "Sodrzinata na tekovniot imenik**:** $lista" |

до­де­ка со:

|  |
| --- |
| student@os:~$ echo "Ima `ls $HOME | wc -l` datoteki vo $HOME"  Ima 51 datoteki vo /home/student |

се при­ка­жу­ва бро­јот на да­то­те­ки во име­ни­кот $HOME. (На­по­ме­на уш­те ед­наш: да се вни­ма­ва на на­вод­ни­ци­те кај grave опе­ра­то­рот!)

# 

# За на час

1. За­пи­ше­те го из­ле­зот од ко­ман­ди­те во да­то­те­ка­та zad1.txt:

echo "Najaveni se:"  
who  
echo "Vkupno:"   
who | wc –l

{ echo “Najaveni se:” ;who ;echo “Vkupno:” ;who | wc -l ;} >zad1.txt

1. Про­нај­ди ги си­те да­то­те­ки во те­ков­ни­от име­ни­к со ек­стен­зи­ја .txt и резултатот за­пи­ше­те ги во да­то­те­ка­т zad2.txt.   
     
   ls | grep ‘\.txt’ | cat > zad2.txt
2. Из­број­те кол­ку да­то­те­ки и по­ди­ме­ни­ци (за­едно) има­те во те­ков­ни­от име­ни­к и из­ле­зот за­пи­ше­те го во да­то­те­ка­та zad3.txt.  
     
   ls -l | grep ‘^[-d]’ | wc -l > zad3.txt
3. Кре­и­рај да­то­те­ка zad4.txt со след­на­ва сод­ржи­на:

Zz .. ss  
Ll . k L ... io  
  
nano zad4.txt

1. Из­лис­тај­те ги да­то­те­ки­те ко­и во сво­ја­та сод­ржи­на го сод­ржа­т ка­рак­те­рот .(точ­ка) и ре­зул­та­тот за­пи­ше­те го во да­то­те­ка­та zad5.txt.  
     
   grep -l ‘\.’ \* > zad5.txt
2. Из­број­те кол­ку ко­рис­ни­ци чи­и ин­дек­си за­поч­ну­ва­а­т на **11** се на­ја­ве­ни мо­мен­тал­но на сис­те­мот и из­ле­зот за­пи­ше­те го во да­то­те­ка­та zad6.txt.   
     
   who | grep ‘^11’ | wc -l > zad6.txt
3. Из­број­те кол­ку име­ни­ци има во ва­ши­от до­ма­шен име­ни­к и из­ле­зот за­пи­ше­те го во да­то­те­ка­та zad7.txt.  
     
   ls -l | grep ‘^d’ | wc -l > zad7.txt
4. Да се ис­пи­ша­т са­мо ими­ња­та на си­те да­то­те­ки (без име­ни­ци) во те­ков­ни­от име­ни­к и из­ле­зот за­пи­ше­те го во да­то­те­ка­та zad8.txt. (**Упат­ство:** Ко­рис­те­те цев­ки, grep и awk. Да се иг­но­ри­ра­а­т име­ни­ци­те и да­то­те­ки­те ко­и сод­ржа­т праз­но мес­то во нив­но­то име).  
     
   ls -p | grep -v ‘[\s/]’ | cat > zad8.txt
5. Со ко­рис­те­ње на sed ко­ман­да­та, да се ис­пи­ша­т са­мо ими­ња­та на си­те име­ни­ци (без да­то­те­ки) во те­ков­ни­от име­ник, а по­то­а да се сорти­ра­а­т по аз­бу­чен ре­дос­ле­д и ре­зул­та­тот да се пре­на­со­чи во да­то­те­ка zad9.txt.  
     
   ls -1p | sed -n ‘/\/$/p’ | sort > zad9.txt
6. Да се ис­пи­ша­т си­те да­то­те­ки и име­ни­ци ко­и има­а­т при­ви­ле­ги­ја за за­пи­шу­ва­ње од стра­на на ко­рис­ни­кот и ре­зул­та­тот да се пре­на­со­чи во да­то­те­ка zad10.txt  
     
   ls -l | grep ‘^..w’ | awk -f script.sh | cat > zad10.txt  
     
   script.sh  
   {  
    x = 10  
    while(x<=NF) {  
    printf “%s ”, $x;  
    x++;  
    }  
   printf “\n”;  
   }
7. Да се из­лис­та­а­т си­те на­ја­ве­ни ко­рис­ни­ци чи­е ко­рис­нич­ко име зав­ршу­ва на 5 по аз­бу­чен ре­дос­ле­д и вре­ме­то ко­га ти­е се на­ја­ви­ле и ре­зул­та­тот да се пре­на­со­чи во да­то­те­ка zad11.txt   
     
   who | awk ‘$1 ~ /5$/ { print $1, $4 ; }’ | sort -f -k -1 > zad11.txt
8. Да се нап­ра­ви ко­пи­ра­ње на си­те да­то­те­ки ко­и за­поч­ну­ва­а­т на zad и се со ек­стен­зи­ја .txt во име­ни­кот zadaci. Да се кре­и­ра име­ни­кот zadaci во до­маш­ни­от име­ник.  
     
   mkdir zadaca  
   cp zad\*.txt zadaci/

# 

# За­да­чи за веж­ба­ње

На­пи­ше­те ре­гу­ла­рен из­раз кој :

* Има ба­ре­м едно i по што сле­да­т про­из­во­лен број на бро­е­ви: i[0-9]\*
* Би­ло ко­и 5 ка­рак­те­ри: .\{5\}
* Збо­ро­ви од 3 бук­ви прва­та е c крај­на­та е t а по­ме­ѓу мо­же да е би­ло ко­ја са­мог­лас­ка: c[aeiou]t
* Збо­ро­ви од 3 бук­ви прва­та е c крај­на­та е t а из­ме­ѓу мо­же да е би­ло ко­ја сог­лас­ка: c[^aeiou]t
* Збо­ро­ви од 5 бук­ви ка­де крај­на­та е c а сод­ржа­т и точ­ка: [a-zA-z\.]\{4\}c
* Збо­ро­ви што поч­ну­ва­а­т со с а зав­ршу­ва со t а из­ме­ѓу има 5 до 10 ка­рак­те­ри "a": ca\{5,10\}t
* Збо­ро­ви што поч­ну­ва­а­т со ii а зав­ршу­ва со student а из­ме­ѓу има 4 до 6 ци­ф­ри: ii[0-9]\{4,6\}student
* Збо­ро­ви што поч­ну­ва­а­т на "p" и зав­ршу­ва­а­т на са­мог­лас­ка: p.\*[aeiou]
* Збо­ро­ви што поч­ну­ва­а­т со точ­ка ".": \<\.
* Збо­ро­ви што за­поч­ну­ва­а­т со а, а зав­ршу­ва­а­т со z а из­ме­ѓу има не­оп­ре­дел­но број бук­ви: \<a.\*z\>