







LPIC-1. Examen 101. Objectius 103.2

LPI 103.3 Processar fluxos de text utilitzant filtres

http://acacha.org/mediawiki/index.php/LPI_103.2



Objectius

103.2. Processar fluxos de text utilitzant filtres	
	<ul style="list-style-type: none"> ▪ Objectiu: Els candidats han de ser capaços d'aplicar filtres a fluxos de text. ▪ Pes: 3
	<p>Àrees Clau de Coneixement:</p> <ul style="list-style-type: none"> ▪ Enviar fitxers de text i fluxos de sortida a través de filtres processadors de textos per a modificar la sortida utilitzant les instruccions estàndard UNIX incloses al paquet textutils de GNU.
	<p>La següent és una llista parcial de fitxers, termes i utilitats utilitzades:</p> <ul style="list-style-type: none"> ▪ cat ▪ cut ▪ expand ▪ fmt ▪ head ▪ od ▪ join ▪ nl ▪ paste ▪ pr ▪ sed ▪ sort ▪ split ▪ tail ▪ tr ▪ unexpand ▪ uniq ▪ wc
	<p>Apunts: LPI 103.2. Processar fluxos de text utilitzant filtres</p>



Paquet textutils

- ♦ Ara és un tercera part del paquet coreutils:

```
$ sudo apt-get install textutils
```

```
S'està llegint la llista de paquets... Fet
```

```
S'està construint l'arbre de dependències
```

```
S'està llegint la informació de l'estat... Fet
```

Nota: s'està seleccionant coreutils en comptes de textutils

coreutils ja es troba en la versió més recent.

0 actualitzats, 0 nous a instal·lar, 0 a suprimir i 0 no actualitzats.

- ♦ Shellutils (LPI 103.1)
- ♦ Textutils (LPI 103.2)
- ♦ Fileutils (LPI 103.3)



man versus info

♦ Coreutils és un paquet de GNU

- ♦ Les ordres de coreutils tenen tant man com info.
- ♦ Al tractar-se però d'eines de GNU, trobareu més informació a les pàgines info
- ♦ Cal tenir en compte que per consultar les pàgines info cal
 - \$ info coreutils cat
- ♦ **man** a la wiki del curs
- ♦ **info** a la wiki del curs



cat

♦ cat (concatenate)

- ♦ Concatena fitxers
- ♦ Resultat a la sortida estàndard

```
$ cat numeros personas
10
9
15
ana
pep
Josep
$ cat primer.txt
segon.txt > combinat.txt
```

- ♦ Suporta un nombre arbitrari de fitxers:

```
$ cat fitxer1.txt fitxer2.txt fitxer3.txt ...
```

- ♦ Pot concatenar només un fitxer --> Consultar el fitxer:

```
$ cat /etc/passwd
```

- Per a fitxers grans és més recomanable un "paginador" com les ordres **more** i **less**.



cat

◆ Opcions

- ◆ Mostrar on s'acaben les línies (--show-ends) amb un dolar:

```
$ cat -E fitxer1 fitxer2
```

- ◆ Mostrar números de línia:

```
$ cat -n fitxer2
```

- ◆ Comprimir múltiples línies en blanc (-s o--squeeze-blank):

```
$ cat -s Fitxer amb espais
```



cat

- ◆ Mostrar caràcters especials. Per exemple el tabulador:

```
$ cat -T fitxer3
```

```
1^Iun
```

```
2^Idos
```

```
3^Itres
```

```
4^Iquatre
```

Els tabuladors es mostren com ^I. El fitxer original és:

```
$ cat fitxer3
```

```
1      un
```

```
2      dos
```

```
3      tres
```

```
4      quatre
```

- ◆ Copiar l'entrada estàndard a la sortida estàndard

```
$ cat
```

- ◆ L'ordre **tac** concatena al revés

```
$ tac fitxer2.txt fitxer1.txt > combinat.txt
```



cat

♦ Caràcter -

```
$ echo "-----" | cat numeros1 - numeros2
1
2
3
-----
un
dos
tres
```

```
$ cat numeros1 - numeros2 < numeros3
1
2
3
4
1      un
2      dos
3      tres
4      quatre
un
dos
tres
quatre
```




cut

♦ cut (tallar)

- ♦ Extreu porcions de text de les línies de l'entrada estàndard i les mostra per la sortida estàndard.
- ♦ 3 opcions:
 - **Per bytes:** opció **-b** (**--bytes=list**) extreu la llista de bytes especificats del fitxer d'entrada.
 - **Per caràcters:** opció **-c** (**--characters=list**) extreu la llista de caràcters especificats. Bytes i caràcters acostumen a ser el mateix a no ser que s'utilitzi una codificació de caràcters multibyte.
 - **Per camps:** opció **-f** (**--fields=list**) mostra la llista de camps especificats del fitxer d'entrada. **Per defecte el separador de camps és el tabulador.** Es poden especificar altres separadors de camp amb l'opció **-d char** o **--delim=char** o **--delimiter=char**.



cut

♦ Exemples de llistes:

- ♦ **2**: Especifica el segon camp, caràcter o byte
- ♦ **2-6**: Un rang (de 2 a 6)
- ♦ **-4 o 4- o 4+**: un rang obert
- ♦ Extreure els camps 3 i 5 d'un fitxer separat per :

```
$ cut -d: -f3,5 /etc/passwd
```

- ♦ Altres exemples:
- ♦ Opcions:

```
$ date | cut -c12-19
```

```
6:48:25
```

```
$ ifconfig eth0 | grep HWaddr | \  
cut -d " " -f 11
```

```
00:0C:76:96:A3:73
```

- ♦ **-s o --only-delimited**: cut no mostra les línies que no contenen el delimitador en cas d'extracció de camps.



Expand i unexpand

- ♦ **Converteixen tabuladors en espais i viceversa:**
 - ♦ Permet agrupar una sèrie d'espais en tabuladors
 - ♦ Els tabuladors són més útils que els espais per tal de navegar còmodament pels fitxers
 - ♦ **unexpand**: espais a tabuladors
 - ♦ **expand**: tabuladors a espais

```
$ echo "          asdf sdf" | unexpand | cat -T -
^I^I asdf sdf
```



fmt

◆ Permet formatar fitxers

- ◆ Formata paràgrafs, amb línies massa llargues
- ◆ La mida màxima d'una línia és 75 caràcters.
- ◆ Es pot canviar la mida màxima de línia amb l'opció -w
- ◆ Consulteu [fmt](#) a la wiki del curs

```
$ fmt -w 30 tirant
Molt excel·lent, virtuós
e gloriós Príncep, Rei
expectant: Jatsia per vulgada
fama fos informat de vostres
virtuts, molt majorment ara
he hagut notícia d'aquelles,
per vostra senyoria voler-me
...
```



head i tail

♦ Mostren l'inici (head=cap) i el final (tail=cua) d'un fitxer

```
$ sudo head /var/log/syslog  
$ sudo tail /var/log/syslog
```

- ♦ 10 línies per defecte.
- ♦ Es pot modificar la quantitat d'informació mostrada per número de línies o per mida:
 - **Per mida en bytes.** -c num o --bytes=num: Mostra els primers num bytes
 - **Per número de línies.** -n num o --lines=num: Mostra les primeres num línies.
- ♦ **Fitxers de log:**

```
$ tail -f /var/log/syslog  
La opció --pid=pid permet finalitzar  
-f quan un procés acaba
```



od

♦ Permet consultar fitxers binaris

- ♦ Els fitxer de tipus text es poden mostrar de forma senzilla per la línia de comandes amb comandes com **cat**, **more** o **less**.
- ♦ En canvi els fitxers binaris mostraran caràcters estranys (garbage characters) si els mostreu amb aquest tipus de comandes (recordeu l'ordre reset per fer neteja)

```
$ cat /bin/ls
$ od /bin/ls
0000000 042577 043114 000401 000001 000000 000000 000000 000000
0000020 000002 000003 000001 000000 115440 004004 000064 000000
...
```



od

♦ Mostra fitxers binaris en octal i altres formats

- ♦ Els fitxers de text també es poden mostrar amb od
- ♦ Hi ha una codificació de caràcters:
- ♦ La primera columna és la posició
- ♦ La resta de columnes és el contingut

```
$ cat numeros
1
2
3
4
5
6
7
```

```
$ od numeros
0000000 005061 005062 005063 005064 005065 005066 005067
0000016
```

➤ Codificació ascii en octal

```
$ sudo atp-get install ascii
```

```
$ ascii -o
```

0 NUL	20 DLE	40	60 0	100 @	120 P	140 `	160 p
1 SOH	21 DC1	41 !	61 1	101 A	121 Q	141 a	161 q
2 STX	22 DC2	42 "	62 2	102 B	122 R	142 b	162 r
3 ETX	23 DC3	43 #	63 3	103 C	123 S	143 c	163 s
4 EOT	24 DC4	44 \$	64 4	104 D	124 T	144 d	164 t
5 ENQ	25 NAK	45 %	65 5	105 E	125 U	145 e	165 u
6 ACK	26 SYN	46 &	66 6	106 F	126 V	146 f	166 v
7 BEL	27 ETB	47 '	67 7	107 G	127 W	147 g	167 w
10 BS	30 CAN	50 (70 8	110 H	130 X	150 h	170 x
11 HT	31 EM	51)	71 9	111 I	131 Y	151 i	171 y
12 LF	32 SUB	52 *	72 :	112 J	132 Z	152 j	172 z
13 VT	33 ESC	53 +	73 ;	113 K	133 [153 k	173 {
14 FF	34 FS	54 ,	74 <	114 L	134 \	154 l	174
15 CR	35 GS	55 -	75 =	115 M	135]	155 m	175 }
16 SO	36 RS	56 .	76 >	116 N	136 ^	156 n	176 ~
17 SI	37 US	57 /	77 ?	117 O	137 _	157 o	177 DEL

```
$ locale
```

```
$ man utf-8
```

Part de utf-8 és compatible amb ASCII

od

◆ Altres formats

◆ Hexadecimal

```
$ od -t x numeros
0000000 0a320a31 0a340a33 0a360a35 00000a37
0000016
$ ascii -x
...
 1 SOH      11 DC1      21 !        31 1        41 A        51 Q        61 a        71 q
 2 STX      12 DC2      22 "        32 2        42 B        52 R        62 b        72 r
 3 ETX      13 DC3      23 #        33 3        43 C        53 S        63 c        73 s
 4 EOT      14 DC4      24 $        34 4        44 D        54 T        64 d        74 t
 5 ENQ      15 NAK      25 %        35 5        45 E        55 U        65 e        75 u
 6 ACK      16 SYN      26 &        36 6        46 F        56 V        66 f        76 v
 7 BEL      17 ETB      27 '        37 7        47 G        57 W        67 g        77 w
```

◆ decimal

```
$ od -t d numeros
0000000 171051569 171182643 171313717 2615
0000016
$ od -t a numeros
0000000 1 nl 2 nl 3 nl 4 nl 5 nl 6 nl 7 nl
0000016
```



join

♦ Combina fitxers segons un camp

- ♦ El camp per defecte és el primer
- ♦ Camps separats per espais

```
$ cat telefons.txt
93-894-56-45   Linus Torvalds
977-44-45-34   Richard Stallman
93-144-56-51   Pepe Pardo Jeans
93-345-53-21   George Orwell
$ cat grups_telefons.txt
93-894-56-45   Grup1
977-44-45-34   Grup1
93-144-56-51   Grup3
93-345-53-21   Grup2
Podeu combinar els dos fitxers amb:
$ join telefons.txt grups_telefons.txt
93-894-56-45 Linus Torvalds Grup1
977-44-45-34 Richard Stallman Grup1
93-144-56-51 Pepe Pardo Jeans Grup3
93-345-53-21 George Orwell Grup2
```



join

- ◆ Per defecte join utilitza el primer camp com a camp coincident. Es poden especificar altres camps de la següent forma (un exemple):
 - **-1 3 -2 2**: El camp 3 del primer fitxer amb el camp 2 del primer fitxer
 - **-1 1 -2 1**: Aquesta és la opció per defecte: el camp 1 del primer i segon fitxer.
- ◆ Podeu fer que join ignori majúscules i minúscules (per defecte és case sensitive) amb l'opció -i.
- ◆ Amb l'opció -o FORMAT' es poden especificar opcions més complexes per a la sortida (resultat de la comanda)
- ◆ La comanda join és la comanda principal per manipular fitxers com a base de dades.

IMPORTANT: Cal que els dos fitxers tinguin els camps clau o camps a coincidir en el mateix ordre. Sovint s'utilitza prèviament la comanda sort per tal de complir amb aquest requisit.



nl

- ♦ **Mostra números de línia (number lines)**
 - ♦ També hi ha opcions amb cat, però ni són tant completes

```
$ nl /etc/group
1  root:x:0:
2  daemon:x:1:
3  bin:x:2:
4  sys:x:3:
5  adm:x:4:sergi,bego
6  tty:x:5:
7  disk:x:6:
8  lp:x:7:
9  mail:x:8:
10 news:x:9:
11 uucp:x:10:
```

```
$ cat -n /etc/group
$ cat -b /etc/group
```



nl

♦ Estil de la numeració:

- ♦ Es pot canviar amb l'opció -b STYLE (--body-numbering=STYLE). Els estils són:
 - **a**: números a totes les línies
 - **t**: números a totes les línies no buides
 - **n**: sense números (per defecte a les capçaleres i als peus de pàgina)
 - **pBRE**: números només a les línies que coincideixen amb la expressió regular BRE
- ♦ Suporta capçaleres i peus de pàgina
- ♦ Consulteu el manual per a més opcions

paste

- ♦ **Fusiona els continguts de dos fitxers en una sola línia:**

línia:

```
$ paste telefons.txt grups_telefons.txt
93-894-56-45   Linus Torvalds      93-894-56-45   Grup1
977-44-45-34   Richard Stallman    977-44-45-34   Grup1
93-144-56-51   Pepe Pardo Jeans    93-144-56-51   Grup3
93-345-53-21   George Orwell       93-345-53-21   Grup2
```

- ♦ Crear una sortida de dos columnes o dos files:

```
$ paste numeros numeros1
1      un
2      dos
3      tres
4      quatre
5      cinc
6      sis
$ paste -s numeros numeros1
1 2 3      4      5      6      7
un dostres quatre cinc sis set
```



pr

♦ **pr (print)**

- ♦ Prepara un text per a ser imprimit.
- ♦ Afegir pàgines, capçaleres i peus de pàgina

```
$ pr tirant | lpr
```

- ♦ Opcions:
 - **Columnes múltiples.** -numcols o --columns=numcols: permet crear un text de múltiples columnes
 - **Doble espai.** -d o --double-space: doble espai entre línies.



sed

♦ **sed (stream editor)**

- ♦ Editor de flux de dades que ens permet dur a terme transformacions bàsiques de text en un flux d'entrada (un fitxer o una entrada des d'una canonada).
- ♦ Accions:
 - Suprimir una línia determinada d'un fitxer
 - Suprimir un conjunt de línies determinades.
 - Suprimir línies que coincideixin amb un patró determinat.
 - Reemplaçar una cadena de text.
 - Afegir una línia després de cada línia de l'arxiu.

♦ **Sintaxi:**

```
$ sed [opcions] -f script-file [fitxer-entrada]
```




sed

♦ Sintaxi 2:

```
$ sed [opcions] script-text [fitxer-entrada]
```

- ♦ En els dos casos anteriors, [fitxer-entrada] és el camí (path) del fitxer que volem modificar. La comanda sed modifica el contingut d'un fitxer i envia els canvis per la sortida estàndard.
- ♦ Sed funciona mitjançant un conjunt d'ordres. Les ordres les podem especificar directament amb **script-text** o, mitjançant el paràmetre -f, podem utilitzar un fitxer que contingui les ordres. Normalment el conjunt d'ordres especificades directament es posen entre cometes simples.
- ♦ Algunes de les ordres més importants que podeu utilitzar amb sed són:



sed

♦ Ordres de sed

Ordre	Adreces	Descripció
=	O o 1	Mostra el número de línia actual
a\text or 1	O o 1	Afegeix text al fitxer.
i\text	O o 1	Insereix text al fitxer.
r fitxer1	O o 1	Afegeix el text del fitxer1 al fitxer
c\text	Rang	Reemplaça el rang de línies seleccionades pel text proporcionat
s/regexp/text	Rang	Reemplaça el text que coincideix amb l' expressió regular (regexp)rang de línies seleccionades pel text proporcionat (text)
w fitxer	Rang	Escriu l'espai del patró actual al fitxers indicat.
q	Rang	Acaba l'script immediatament, però mostra l'espai del patró actual.
q	Rang	Acaba l'script immediatament, però mostra l'espai del patró actual.
q	Rang	Acaba l'script immediatament, però mostra l'espai del patró actual.
Q	Rang	Acaba l'script immediatament.



sed

♦ Exemples

- ♦ Cercar i reemplaçar només la primera coincidència de cada línia

```
$ sed 's/2008/2009/' calendari-2008.txt > calendari-2009.txt
```

- ♦ Cercar i reemplaçar totes les coincidències (g:global)

```
$ sed 's/2008/2009/g' calendari-2008.txt > calendari-2009.txt
```

- ♦ Esborrar la línia 0 d'un fitxer: `$ sed -i '10d' file`

- ♦ Esborrar els tres últims caràcters d'un fitxer:

```
$ sed 's/...$//' exemple.txt >  
exemple_sense_tres_ultims_caracters.txt
```



sed

- ◆ Esborrar els 2 primers caràcters d'un fitxer

```
$ sed 's/^..$//' exemple.txt >  
exemple_sense_tres Ultims_caracters.txt
```

- ◆ Convertir de DOS a Unix

- Amb bash, per convertir de DOS a UNIX, escriuiu:
 - \$ sed 's/^M\$//' input.txt > output.txt
- Premeu Ctrl-V seguit de Ctrl-M per obtenir el patró del símbol especial.
- Per convertir de Unix a DOS:
 - \$ sed 's/\$"/`echo \\r`/' input.txt > output.txt
- També podeu utilitzar **tofrodoS**



sort

♦ **sort=ordenar en anglès**

- ♦ Ordena les línies d'un fitxer d'entrada segons el criteri que indiquem.
- ♦ Podem ordenar alfabèticament o numèricament, ascendentment o descendentment, definir la part de la línia que s'ha de considerar per ordenar, etc...
- ♦ Ordenar un fitxer segons els contingut de tota la línia:

```
$ sort /etc/passwd | head
avahi-autoipd:x:104:110:Avahi autoip daemon,,,:/var/lib/avahi-
autoipd:/bin/false
avahi:x:110:119:Avahi mDNS daemon,,,:/var/run/avahi-daemon:/bin/false
bin:x:2:2:bin:/bin:/bin/sh
backup:x:34:34:backup:/var/backups:/bin/sh
bego:x:1001:1001:Bego Solé,,,,:/home/bego:/bin/bash
...
```



sort

- ♦ Ordena numèricament (-n) pel tercer camp (paràmetre -k3), considerant que el caràcter separador de camps és : (paràmetre -t:):

```
$ sort -t: -k3 -n /etc/passwd
```

- ♦ Ordena el resultat de ls -l /etc pel cinquè camp (la mida del fitxer) descendentment (paràmetre -r). Fem servir el separador de camps per defecte (un o més espais en blanc).

```
$ ls -l /etc | sort -k5 -n -r | more
```



sort

- ♦ **Per ordenar un fitxer numèric:**

```
$ cat numeros
10
23
3
4
54
8
12
11
34
Podem executar:
$ sort -g numeros
o
$ sort -n numeros
```



split





tr

♦ tr (translate)

- ♦ Modifica o elimina caràcters de l'entrada estàndard.

- ♦ La sintaxi és la següent:

```
$ tr [options] SET1 [SET2]
```

- **SET1**: El conjunt de caràcters a buscar per tal de ser reemplaçats.
- **SET2**: El conjunt de caràcters que reemplaçaran al SET1.
- Passar a minúscules:

```
$ cat noms  
Linux Torvalds  
Richard Stallman  
Ian Murdock
```

```
$ tr LTRSIM ltrsim < noms  
linux torvalds  
richard stallman  
ian murdock
```

tr

- ◆ Eliminar caràcters

```
$ tr -d a < noms
Linux Torvlds
Richrd Stllmn
In Murdock
```

- ◆ Guardar el resultat en un fitxer:

```
$ tr -d a < noms > noms_sense_a
```

- ◆ Passar un fitxer CSV a tabuladors

```
$ tr , \\t < fitxerCSV.txt > fitxerTabuladors.txt
fitxerCSV.txt:
1,"2TESA","02","CA","29",2,5,,
1,"2TESA","02","CA",,4,1,,
2,"2ECOHU","02","CA","25",1,3,,
2,"2ECOHU","02","CA","25",2,3,,
6,"3ESC","01","CA","E17",1,8,,
6,"3ESC","01","CA","E17",3,4,,
6,"3ESC","01","CA","E17",5,5,,
```



tr

◆ Resultat:

1	"2TESA"	"02"	"CA"	"29"	2	5	
1	"2TESA"	"02"	"CA"		4	1	
2	"2ECOHU"		"02"	"CA"	"25"	1	3
2	"2ECOHU"		"02"	"CA"	"25"	2	3
6	"3ESC"	"01"	"CA"	"E17"	1	8	
6	"3ESC"	"01"	"CA"	"E17"	3	4	
6	"3ESC"	"01"	"CA"	"E17"	5	5	



uniq

♦ Elimina línies duplicades

- ♦ Només té en compte línies successives

```
$ cat shakespeare  
To  
be  
or  
not  
to  
be  
that  
is  
the  
question
```

```
$ sort shakespeare.txt | uniq  
be  
is  
not  
or  
question  
that  
the  
to  
To
```

```
$ uniq shakespeare  
To  
be  
or  
not  
to  
be  
that  
is  
the  
question
```



WC

♦ **wc (word count)**

- ♦ wc compte les paraules, així com també pot comptar línies, caràcters i bytes.

```
$ wc fitxer.txt  
200 2100 14965 fitxer.txt
```

- ♦ **Format:**

- 200 línies (es compten la quantitat de caràcters de nova línia)
- 2100 paraules
- 14965 bytes

```
Longitud de la línia més llarga  
$ wc -L /etc/profile  
70 /etc/profile
```

- ♦ Es pot limitar la informació de la sortida amb:

- -l (--lines): mostra el número de línies.
- -w (--words): mostra el número de paraules.
- -c (--bytes): Mostra el número de bytes.
- -m (--chars): Mostra només els caràcters



Reconeixement 3.0 Unported

Sou lliure de:



copiar, distribuir i comunicar públicament l'obra



fer-ne obres derivades

Amb les condicions següents:



Reconeixement. Heu de reconèixer els crèdits de l'obra de la manera especificada per l'autor o el llicenciador (però no d'una manera que suggereixi que us donen suport o rebeu suport per l'ús que feu l'obra).

- Quan reutilitzeu o distribuïu l'obra, heu de deixar ben clar els termes de la llicència de l'obra.
- Alguna d'aquestes condicions pot no aplicar-se si obteniu el permís del titular dels drets d'autor.
- No hi ha res en aquesta llicència que menyscabi o restringeixi els drets morals de l'autor.

Advertiment

Els drets derivats d'usos legítims o altres limitacions reconegudes per llei no queden afectats per l'anterior
Això és un resum fàcilment llegible del text legal (la llicència completa).

<http://creativecommons.org/licenses/by/3.0/deed.ca>

LPIC-1. Examen 101. Objectiu 103.2

ICE-UPC



Autor: Sergi Tur Badenas