

## 92 - Text manipulation Programs

### Extracting specific columns from columns delimited text files

**cut** *-ddelimiter -ffield[s] -cn filename* Extracts columns of text from a file

- dd** Sets Delimiter char *d* (can only be ONE char)
- fn** Extract the field list/range *n,n2, ...*  
Fields numbers should be separated by comas(,)
- cn** Extract the character *n* or range of characters
- s, --only-delimited**  
Do not print lines not containing delimiters
- output-delimiter=STRING**  
Use STRING as the output delimiter the default is to use the input delimiter.

When using **-f** or **-c** Ranges/lists can be of the following formats:

<b>n</b>	<i>n</i> 'th byte, character or field, counted from 1
<b>n-</b>	from <i>n</i> 'th byte, character or field, to end of line
<b>n-m</b>	from <i>n</i> 'th to <i>m</i> 'th (included) byte, character or field
<b>-m</b>	from first to <i>m</i> 'th (included) byte, character or field

eg.

```
grep :5.: /etc/passwd | cut -d: -f 1,3
Show a list of non-system users (UID > 500)
```

### Converting TABs to Spaces

**expand** *-tn -i filename* Converts TABs to spaces from a file

- tn** Sets the number of spaces per TAB  
Default = 8 spaces
- i** Convert only start of lines

### Styled Text lines numbering

**nl** *-h style -b style -f style filename* Number lines of a file

<b>-h</b>	Header of text ,	starts with \:\:\:
<b>-b</b>	Body of text,	starts with \:\:
<b>-f</b>	Footer of text,	starts with \:

#### Styles:

<b>a</b>	Number all lines
<b>t</b>	Number only test lines
<b>n</b>	No numbering for this part

**Dumping a file in Named char, ASCII, OCTAL or Hexadecimal format**

```
od [options] filename
```

## Options

```
-j, --skip-bytes=BYTES      skip BYTES input bytes first on each file
-N, --read-bytes=BYTES      limit dump to BYTES input bytes per file
-s, --strings[=BYTES]      output strings of at least BYTES graphic chars

-v, --output-duplicates     do not use * to mark line suppression
-w, --width[=BYTES]         output BYTES bytes per output line

-t, --format=TYPE           select output format or formats as follows:
```

Short options	Std option	Formats
-a	same as -t a,	select named characters
-b	same as -t oC,	select octal bytes
-c	same as -t c,	select ASCII characters or backslash escapes
-d	same as -t u2,	select unsigned decimal shorts
-f	same as -t fF,	select floats
-h	same as -t x2,	select hexadecimal shorts
-i	same as -t d2,	select decimal shorts
-l	same as -t d4,	select decimal longs
-o	same as -t o2,	select octal shorts
-x	same as -t x2,	select hexadecimal shorts

**Paste columns from one file to another**

```
paste -dDelimiter file1 file2 - Paste content of file1 to content of file2 vertically
```

```
-d Sets the delimiter between the groups of columns the output
( One Char only, Default is TAB)
```

**Format a raw text before printing**

```
pr -h header -l LinesPerPage -o width filename
-l Number of lines per page (default is 66)
-o Offset: Page Left margin
-h What for header
```

**Translating certain types of characters to another**

**tr "Chars set 1" "Char set 2" < filename**

Translate all the corresponding "Chars set 1" into "Char set 2"

**tr -d "Chars to delete" < filename** Delete all of these chars from file

**Character sets format:**

<code>\NNN</code>	character with octal value NNN (1 to 3 octal digits)
<code>\\</code>	backslash
<code>\a</code>	audible BEL
<code>\b</code>	backspace
<code>\f</code>	form feed
<code>\n</code>	new line
<code>\r</code>	return
<code>\t</code>	horizontal tab
<code>\v</code>	vertical tab
<code>CHAR1-CHAR2</code>	all characters from CHAR1 to CHAR2 in ascending order
<code>[CHAR*]</code>	in SET2, copies of CHAR until length of SET1
<code>[CHAR*REPEAT]</code>	REPEAT copies of CHAR, REPEAT octal if starting with 0
<code>[:alnum:]</code>	all letters and digits
<code>[:alpha:]</code>	all letters
<code>[:blank:]</code>	all horizontal whitespace
<code>[:cntrl:]</code>	all control characters

**Counting the Characters, words, lines of a file**

**wc -c -w -l filename** Count the number of Characters, Words and Lines in file  
 The output is in the same order.  
 Default (without options) is Chars, Words, and Lines.

<b>-c</b>	Count only the characters in the file
<b>-w</b>	Count only the words in the file
<b>-l</b>	Count only the lines in the file