Useful shell commands

head/tail, cut, sort, uniq

Virginie Orgogozo

March 2011

cat = shows the content of a file
less = shows the content of a file

Head and tail

cd pcfb/examples/

$head ctd.txt
shows the first 10 lines

$head -n 2 *.pdb
shows the first 2 lines

$history | tail -n 15
shows the 15 most recent items in your command history

$tail -n +2 Thalas*.txt
shows from the second line to the end

$head -n -1 Thalas*.txt
shows from the second line to the 10th line

Cut

cd pcfb/examples/

$cut -f 1,3 Thal*.txt
returns columns 1 and 3 delimited by tabs

$cut -f 1-3 Thal*.txt
returns columns 1 to 3 delimited by tabs

$cut -c 16-20,30 Thal*.txt
returns characters 16 to 20 and 30 from each line

$grep ">" FPexamples.fta | cut -c 2-11
prints out the gene names

$head ctd.txt | cut -f 5,7 -d ","
returns columns 5 and 7. These are delimited by , in the original
file and in the output.

Be careful when space is used between columns because there
are sometimes two spaces instead of one.

Sort

$grep ">" FPexcerpt.fta | sort
lines are sorted by alphabetical order

Options

-n
sorts by numerical value rather than alphabetically

-f
Make all lines uppercase before sorting

-r
sorts in reverse order

-k 3
sorts lines based on column 3 , with columns delimited by space or tab
$head Thal.txt | sort -k 2

-t ","
uses commas for delimiters

-u
returns a unique representative of repeated items

The ASCII list

32	SP	64	@	96	`	128	Ç	160	á	192	└	224	Ó
33	!	65	A	97	a	129	ü	161	í	193	┴	225	ß
34	"	66	B	98	b	130	é	162	ó	194	┬	226	Ô
35	#	67	C	99	c	131	â	163	ú	195	├	227	Ò
36	$	68	D	100	d	132	ä	164	ñ	196	─	228	õ
37	%	69	E	101	e	133	à	165	Ñ	197	┼	229	Õ
38	&	70	F	102	f	134	å	166	ª	198	ã	230	µ
39	'	71	G	103	g	135	ç	167	º	199	Ã	231	þ
40	(	72	H	104	h	136	ê	168	¿	200	╚	232	Þ
41	)	73	I	105	i	137	ë	169	®	201	╔	233	Ú
42	*	74	J	106	j	138	è	170	¬	202	╩	234	Û
43	+	75	K	107	k	139	ï	171	½	203	╦	235	Ù
44	,	76	L	108	l	140	î	172	¼	204	╠	236	ý
45	-	77	M	109	m	141	ì	173	¡	205	=	237	Ý
46	.	78	N	110	n	142	Ä	174	«	206	╬	238	¯
47	/	79	O	111	o	143	Å	175	»	207	¤	239	´
48	0	80	P	112	p	144	É	176	░	208	ð	240	-
49	1	81	Q	113	q	145	æ	177	▒	209	Ð	241	±
50	2	82	R	114	r	146	Æ	178	▓	210	Ê	242	_
51	3	83	S	115	s	147	ô	179	│	211	Ë	243	¾
52	4	84	T	116	t	148	ö	180	┤	212	È	244	¶
53	5	85	U	117	u	149	ò	181	Á	213	ı	245	§
54	6	86	V	118	v	150	û	182	Â	214	Í	246	÷
55	7	87	W	119	w	151	ù	183	À	215	Î	247	¸
56	8	88	X	120	x	152	ÿ	184	©	216	Ï	248	°
57	9	89	Y	121	y	153	Ö	185	╣	217	┘	249	¨
58	:	90	Z	122	z	154	Ü	186	║	218	┌	250	·
59	;	91	[	123	{	155	ø	187	╗	219	█	251	¹
60	<	92	\	124			156	£	188	╝	220	▄	252	³
61	=	93	]	125	}	157	Ø	189	¢	221	¦	253	²
62	>	94	^	126	~	158	×	190	¥	222	Ì	254	■
63	?	95	_	127	DEL	159	ƒ	191	┐	223	▀	255	

Uniq

Removes identical lines that are in immediate succession and keeps a single line.

Options

-c
counts the number of occurrence of each unique line and write it before each unique line
$cut -c 12-21 ctd.txt | uniq -c

-f 4
ignores the first 4 fields (columns delimited by any number of spaces) in determining uniqueness

-i
ignore case when determining uniqueness