

Scientific Computing 1
Handout 2
October 14, 2014

Regular Expressions

Basic rules for regular expressions

.	matches any single character except linebreaks
^	matches the beginning of the string/line
\$	matches the end of the string/line
[list]	any one character from list. Here list can be a single character, a number of characters, or a character range given with -
[^list]	any one character that is NOT in list.
()	guarantees precedence of the enclosed expression. (optional)
(re)	matches the expression re
re1 re2	matches either the expression re1 or re2
re?	matches at most one appearance of re. Note that in <code>sed</code> you need to either write <code>\?</code> or use the <code>-r</code> commandline switch when using this.
re+	matches one or more subsequent appearances of re
re*	matches none or arbitrarily many subsequent appearances of re
re{n,m}	matches at least n and at most m subsequent appearances of re. Both n and m can be omitted either with or without the comma. Then n means exactly n matches. n, stands for at least n matches and ,m for at most m matches.
(re1)(re2)	matches re1 followed by re2—in search and replace operations the corresponding matches can be referred to by <code>\1</code> and <code>\2</code>
\	escapes, i.e., removes the special meaning of, the following special character.

Some enlightening examples

a?b	matches a string of one or two characters eventually starting with a but necessarily ending on b
^From	matches a line/string beginning with From
^\$	matches an empty line/string
^X*YZ	matches any line/string starting with arbitrarily many X characters followed by YZ
linux	matches the string linux
[a-z]+	matches any string consisting of at least one but also more lower case letters
^[aA]	any line/string that does not start with an a or A.