# Regular Expressions

Grep[Mode] [Regular Expression] [Filename]

## Modes:

* -h (Do not display filenames)
* -i (Ignore case)
* -l (List only filenames containing matching lines
* -n (Precede each matching line with its line number
* -v (Negate matches)
* -x Match whole line only (fgrep only)
* -e (expression - specify expression as option)
* -f filename

## Regular Expressions:

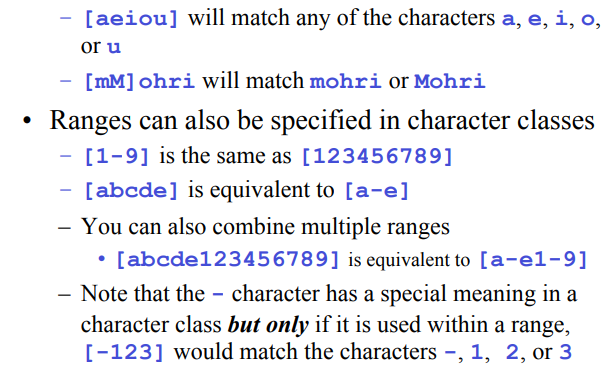
|  |  |  |
| --- | --- | --- |
| Regular Expression | Meaning | Example |
| . | The dot expression matches any character. | “o.o” matches any line with a character in between two o’s |
| [] | Character classes can be used to match any specific set of characters. | b[eor]at matches beat, brat, boat |
| [^] | Character classes can be negated. | b[^eor]at matches all lines not containing beat, brat, boat |
| ^ | Is an anchor, and means beginning of the line | ^b[eor]at would only match if beat,breat, or boat was in the start of the line. |
| $ | Is an anchor, means end of the line. | b[eor]at$ would only match if the 3 words where in the end of the line. |
| ^$ | Empty lines |  |
| ^word$ | Only matches if the word is alone on a line |  |
| \* | The \* is used to define zero or more occurrences. | Ya\*y matches yaaaaaaay, and yy. |
| + | The plus (+) means “one or more”.  Equivalent to {1,} | abc+d will match ‘abcd’ , ‘abccd’ , or ‘abccccccd’ but will not match ‘abd’. |
| ? | The ‘?’ (question mark) specifies an optional character, the single character that immediately precedes it  Equivalent to {0,1} | July? will match ‘Jul’ or ‘July’ |
| \n | Backreference specifier, where n is a number. Looks for nth subexpression | For example, to find if the first word of a line is the same as the last:  ^\([[:alpha:]]\{1,\}\) .\* \1$ |

\*, ?, and + are known as quantifiers

Quantifiers can be used with subexpressions.

* (a\*c)+ will match ‘c’ , ‘ac’ , ‘aac’ or ‘aacaacac’ but will not match ‘a’ or a blank line

# More about character classes:



Named character classes are: alpha, lower, upper, alnum, digit, punct, cntrl.

|  |  |
| --- | --- |
| [[:alpha:]] - Named character classes | [a-zA-Z] |
| [[:alnum:]] | [a-zA-Z0-9 |
| [45[:lower:]] | [45a-z] |

# Repetition ranges

|  |  |
| --- | --- |
| {} | notation can specify a range of repetitions for the immediately preceding regex |
| {n} | Means exactly n occurrences |
| {n,} | Means at least n occurrences |
| {n,m} | means at least n occurrences but no more than m occurrences |

.{0,} same as .\*

a{2,} same as aaa\*

# Subexpressions

If you want to group part of an expression so that \* or { } applies to more than just the previous character, use ( ) notation

Subexpresssions are treated like a single character

a\* matches 0 or more occurrences of a

abc\* matches ab, abc, abcc, abccc, …

(abc)\* matches abc, abcabc, abcabcabc, …

(abc){2,3} matches abcabc or abcabcabc