

# Basic Regular Expressions

## Quantifiers

- ? zero or one
- \* zero or more
- + one or more
- {n} exactly n times
- {n,} n or more times
- {n1,n2} n1 to n2 times

For example:

“a+” matches one or more "a"s, like "a", "aaa" or "aaaaaaaaa"

“cats?” matches two strings: "cat" and "cats"

## Boundaries

- \b word boundary
- \B not a word boundary
- \< word boundary on the left
- \> word boundary on the right
- ^ the start of a string or a line
- \$ the end of a string or a line
- \A the start of a string (i.e. the whole document)
- \Z the end of a string (i.e. the whole document)

For example:

“\bb\b” matches the string "back" in " back "

“\bb\b” matches the string "back" in " background "

## Character Classes

- \d one digit
- \D one non-digital character
- \w a letter, a digit or an underscore
- \W one character that is not a letter, a digit or an underscore
- \s a whitespace character: a space, a tab, a newline, etc.
- \S one character that is not a whitespace character
- \t a tab
- \n a newline character

One symbol is used without a slash, though:

- . any character except a newline

For example:

“\d{3}” matches any three numbers in a row, like "364" or "111"

“\W” matches spaces or punctuation marks, like ", " or "!"

## Character Ranges

[ ... ] one of the characters in the brackets

[^ ... ] any character except the characters in the brackets. `^` is an indicator of negation

For example:

“[aeuoi]” matches one letter that represents a vowel sound, like "u"

“[^aeuoi]” matches one character that is not "a", "e", "u", "o" or "i", like "b"

“[3-7]” matches a digit between 3 and 7, like "6"

## Logical OR

(string1|string2) string1 or string2

For example:

“cat|dog|rat” matches strings "cat", "dog" and "rat"

## Match groups

(...) a match group that can be referenced

(?:...) a match group that cannot be referenced (a passive group)

\number a reference to the match group using its position in the regexp

For example:

“(d+)-(d+)-\1-\2” matches "11-22-11-22", etc.

“(?:d+)-(d+)-\1” matches "11-22-22", etc.

## Matching Literal Characters

\. matches a period (".")

\] matches a right square bracket ("]")

\\$ matches a dollar sign ("\$")

\\ matches a backslash ("\")

## Lookarounds

(?=...) positive lookahead

(?<=...) positive lookbehind

(?!...) negative lookahead

(?<!...) negative lookbehind

For example:

“(?<=iphone)d” matches "7" in "iphone7"

“VBG(?: JJ[RS]?)” matches "VBG" in "reading VBG NN" and does not match "VBG" in "lasting VBG JJ" or "VBG JJokey"

## **References**

1. A nice reference about regular expressions: <http://www.rexegg.com/regex-quickstart.html>
2. A regexp cheat sheet: <http://www.cheatography.com/davechild/cheat-sheets/regular-expressions/>
3. A nice place to practice your regex-writing skills: <http://www.regexr.com/>