


REGEX 

Python Regular Expressions

character classes

- - any character except new line
- \d - digit (0-9)
- \D - not a digit (0-9)
- \w - word character (a-z, A-Z, 0-9, _)
- \W - not a word character
- \s - whitespace, tab, newline
- \S - not a whitespace

Assertions

- \b - word boundary
- \B - not a word boundary
- ^ - beginning of a string
- \$ - end of a string

Anchors

[]

- matches characters in brackets

[^]

- matches characters not in brackets

|

- either or

()

- group

Quantifiers

*

- 0 or more

+

- 1 or more

{

- 0 or 1

{3}

- exact number

{3,4}

- range of numbers (min, max)

"re" package

1. `match` : checks for a match only at the beginning of the string
2. `search` : locates the pattern in the string
3. `findall` : find all occurrences of the strings
4. `finditer` : return an iterator yielding match objects over all non-overlapping matches