

Let's Practice **Regex**

Open regex.txt file from Moodle,
open it with Notepad++

Go to search, enable regular expression and match case

Query:
Search for “abc”

Regex:

Query:
Search for “abc”

Regex:
abc

Query:
Search for “.”

Regex:

. is a MetaCharacters

•
Any character except new line

MetaCharacters need to be escaped using \

Query:
Search for “.”

Regex:

\.

MetaCharacters need to be escaped using \

MetaCharacters:

. [{ () \ ^ \$ | ? * +

Query:
Search for “+”

Regex:

Query:
Search for “+”

Regex:

\+

Query:
Search for “.com”

Regex:

Query:
Search for “.com”

Regex:
\.com

Query:

Search for any digit (0-9)

Regex:

Query:

Search for any digit (0-9)

Regex:

\d

Query:

Search for anything except a digit

Regex:

Query:

Search for anything except a digit

Regex:

`\D`

Query:

Search for any word character (a-z, A-Z, 0-9, _)

Regex:

Query:

Search for any word character (a-z, A-Z, 0-9, _)

Regex:

\w

Query:

Search for anything except a word character

Regex:

Query:

Search for anything except a word character

Regex:

`\W`

Query:

Search for a whitespace (space, tab, newline)

Regex:

Query:

Search for a whitespace (space, tab, newline)

Regex:

\s

Query:

Search for a **not** whitespace (space, tab, newline)

Regex:

Query:

Search for a whitespace (space, tab, newline)

Regex:

\s

- . - 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 (space, tab, newline)
- \S - Not Whitespace (space, tab, newline)

Regex Anchors

`\b`

`\B`

`^`

`$`

Query:

Search for “Ha” at the start of a word (not in the middle)

Regex:

Query:

Search for “Ha” at the start of a word (not in the middle)

Regex:

\bHa

`\b`

word boundary

Query:

Search for “Ha” at middle

Regex:

Query:

Search for “Ha” at the start of a word (not in the middle)

Regex:

\BHa

\B

Not a word boundary

Query:

Search for line started with "Mr"

Regex:

Query:

Search for line started with "Mr"

Regex:

`^Mr`

^

Beginning of a String

Query:

Search for line end with "com"

Regex:

Query:

Search for line end with “com”

Regex:

com\$

\$

End of a String

Regex Anchors

`\b` - Word boundary

`\B` - Not a word boundary

`^` - Beginning of a String

`$` - End of a String

Find Patterns

Phone Numbers

Query:

Search for phone numbers like

321-555-3332

222.444.6698

444 555 6666

Regex:

Query:

Search for phone numbers like

321-555-3332

222.444.6698

444 555 6666

Regex:

```
\d\d\d.\d\d\d.\d\d\d\d
```

Query:

Search for phone numbers only with - as separators

321-555-3332

Regex:

Query:

Search for phone numbers only with - as separators

321-555-3332

Regex:

\d\d\d[-]\d\d\d[-]\d\d\d\d

[]

Matches Characters in brackets

Query:

Search for phone numbers only with - or . as separators

321-555-3332

321.555.3332

Regex:

Query:

Search for phone numbers only with - or . as separators

321-555-3332

321.555.3332

Regex:

```
\d\d\d[-.]\d\d\d[-.]\d\d\d\d
```

[1-7]

numbers between 1 to 7

[a-z]

letters between a to z

[a-zA-Z]
all letters

[^a-z]

everything that is not a lowercase letter

Quantifiers

*

0 or more

+

1 or more

?

0 or 1

{3}

exact 3 of something

Phone number

\d{3}.\d{3}.\d{4}

$\{3,7\}$

range of numbers (min, max)

3 to 7 of something

Emails

Emails

`[a-zA-Z][a-zA-Z0-9_]*@[a-zA-Z]+\.[a-zA-Z]`

words started with Mr.

Mr

words started with Mr.

. character

Mr\.

words started with Mr.

could have 0 or 1 . character

Mr\.

words started with Mr.

then one space

Mr\.\s

words started with Mr.

then one or more characters

Mr\.\?\s\w+

words started with Mr. or Ms.

words started with Mr.

then one or more characters

`M(r|s)\.?\s\w+`

- [] - Matches Characters in brackets
- [^] - Matches Characters NOT in brackets
- | - Either Or
- () - Group

files with the **pdf** extension

files with the **pdf** extension

`^\w+\.pdf$`

- .
- \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 (space, tab, newline)
- \S - Not Whitespace (space, tab, newline)

- \b - Word Boundary
- \B - Not a Word Boundary
- ^ - Beginning of a String
- \$ - End of a String

- [] - Matches Characters in brackets
- [^] - Matches Characters NOT in brackets
- | - Either Or
- () - Group

- * - 0 or More
- + - 1 or More
- ? - 0 or One
- {3} - Exact Number
- {3,7} - Range of Numbers (Minimum, Maximum)

Practice more using
<https://regexlearn.com/learn/regex101>