DM Lab: 9 Regular Expressions

→ At its most basic level, a regex can be considered a method of pattern matching or matching patterns within a string. It is also useful in extracting useful information from various types of inputs.

→ It may be simple as the common letters on following line

match: abcdefg
match: abcde
match: abC
ans: abc

metacharacter

→ \d : any digit from 0-9
\D : any non-digit character
* : zero or more repetitions
+ : One or more repetitions.

match: abc123xyz
match: define "123"
match: var g=123
ans: 123
ans: \D+123\D+
ans: \D+\d+\D+

. (dot) ← Wildcard metacharacter
↑ can match any single character

\. ← matching of period character

Match: cat.
Match: 896.
Match: ?=+.
Skip: abc1
ans: .*\.

matching specific characters
↓
[ ]
[abc] : match a single a, b or c

Match : can
Match : man
Match : fan
Skip : dan
Skip : ran
Skip : pan
ans: [cmf]an

↓ ——— Excludes specific ~~ope~~ characters.
[^]
[^abc] : match any single character except for letters a, b or c

match: tog
match: dog
skip: bog
ans: [^b]og
[hd]og

character range

[-]

[0-9] match any single digit characters from 0 to 9.

[a-z] characters a to z

[A-Za-z0-9_]: \w : alphanumeric characters

\W : non-alphanumeric character.

match: Ana
match: Bob
match: Cpc
skip: aax
skip: bby
skip: ccz
ans: [A-C] [n-p][a-c]
[^a-c][^a-c][^x-z]
[^a-c]+[^x-z]

{ } ← how many repetitions

a{3}: match the character 'a' exactly 3 times
.{2,6}: between 2 and six of any character

Match: waaazzzup
Match: waazzzup
Skip: wazup
ans: waz{2,4}up
ans: \D{2}z{2,4}\D{2}

Match: aaaabcc
Match: aabbbbbc
Match: aacc
Skip: a
ans: aa+b*c+
ans: a{2,4}b{0,4}c{1,2}

? ← optionality

ab?c ← abc, ac

Match: 1 file found?
Match: 2 files found?
Match: 24 files found?
Skip: No files found
ans: \d+ files? found\?

\s ← match any of specific whitespaces
\S ← non whitespace characters

Match: 1. abc
Match: 2. abc
Match: 3. abc
Skip 4. abc

solution: ~~\d+~~ \d\.\s+abc
\S+\s+\S+

^.... $ ← starts & ends

Match: Mission : successful
Skip: Last Mission: unsuccessful
Skip: Next Mission: successful upon capture

ans: ^Mission:successful$

( ) ← Match groups | Capture groups

^(IMG\d+\.png)$ ← Capture & extract full filename
^(IMG\d+)\.png $ ← Captures the part before the period

capture: file_record_transcript.pdf

capture: file_07241999.pdf

skip: testfile_fake.pdf.tmp

ans: (file.+)\.pdf$

(( )) : nested groups

Capture: Jan 1987
Capture: May 1969
Capture: Aug 2011

ans: (\w + (\d+))

Capture: 1280x720
Capture: 1920x1600
Capture: 1024x768

ans: (\d{4}) x (\d{3,4})
(\d+) x (\d+)

(|) └ logical OR : different possible characters

match: I love dogs
matan: I love cats
skip: I love logs
skip: I love oogs

ans: I love (cats|dogs)