I. Bảng regex:

Note: Chỉ match 1 ký tự

. : Khớp (match) với bất kỳ ký tự nào

^ : Khớp tại điểm bắt đầu

$ : Khớp ở cuối dòng

[abc] : Khớp với a hoặc b hoặc c

[abc][vz] : Khớp với a hoặc b hoặc c theo sau là v hoặc z

[^abc]: Khi dấu ^ xuất hiện như là nhân vật đầu tiên trong dấu ngoặc vuông, nó phủ nhận mô hình

Điều này có thể khớp với bất kỳ ký tự nào ngoại trừ a hoặc b hoặc c

[a-d1-7] : Phù hợp với một chuỗi giữa a và điểm d và con số từ 1 đến 7

X|Z : Tìm X hoặc Z

XZ Tìm X và theo sau là Z.

$ : Kiểm tra kết thúc dòng.

\d : Số bất kỳ, viết ngắn gọn cho [0-9]

\D : Ký tự không phải là số, viết ngắn gon cho [^0-9]

\s : Ký tự khoảng trắng, viết ngắn gọn cho [ \t\n\x0b\r\f]

\S : Ký tự không phải khoản trắng, viết ngắn gọn cho [^\s]

\w : Ký tự chữ, viết ngắn gọn cho [a-zA-Z\_0-9]

\W : Ký tự không phải chữ, viết ngắn gọn cho [^\w]

\S+ : Một số ký tự không phải khoảng trắng (Một hoặc nhiều)

\b : Ký tự thuộc a-z hoặc A-Z hoặc 0-9 hoặc \_, viết ngắn gọn cho [a-zA-Z0-9\_]

\* : Xuất hiện 0 hoặc nhiều lần, viết ngắn gọn cho {0,}

+ : Xuất hiện 1 hoặc nhiều lần, viết ngắn gọn cho {1,}

? : Xuất hiện 0 hoặc 1 lần, ? viết ngắn gọn cho {0,1}

{X} : Xuất hiện X lần, {}

{X,Y} : Xuất hiện trong khoảng X tới Y lần

\*? : \* có nghĩa là xuất hiện 0 hoặc nhiều lần, thêm ? phía sau nghĩa là tìm kiếm khớp nhỏ nhất

.\* : Khớp với bất kỳ ký tự nào 0 hoặc nhiều lần

- Ex1: chuoi = "abc11111111111111"; match = chuoi.matches(".\*") => return true

- Ex2: chuoi = "mnp"; match = chuoi.matches("^m.+"); => return true

- Ex3: Kiểm tra toàn bộ s3 Ký tự bất kỳ xuất hiện 1 lần: . - tiếp theo là n, xuất hiện 1 hoặc tối đa 3 lần. - Kết thúc bởi p: p$

Kết hợp các quy tắc: . , {X,Y}, $=> true

s3 = "2nnp";

match = s3.matches(".n{1,3}p$"); => return true

II. Các ký tự đặc biệt trong Java Regex (Special characters):

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

III. Common Regex:

. lấy 1 ký tự bất kỳ

\D+: lấy ký tự đầu là chữ và ký tự đặc biệt, các ký tự còn lại bất kỳ

\d+: lấy ký tự đầu là số, các ký tự còn lại bất kỳ

123: match với 123

...\D: 3 ký tự đầu bất kỳ, ký tự cuối là chữ và ký tự đặc biệt( không phải là số)

[cmf]an: lấy c or m or f và 2 ký tự cuối là an

[^drp]an: lấy ký tự đầu không phải là 1 trong 4 ký tự: c or m or f và 2 ký tự cuối là an

[A-C][n-p][a-c]: lấy ký tự theo thứ tự alphabet a - z

[^a-c][^a-c][^x-z]: lấy ký tự không theo thứ tự alphabet a - c và x - z

waz{3,5}up: lấy ký tự w hoặc a hoặc z xuất hiện 3 đến 5 lần sau đó là up

a{2,4}b{0,4}c{1,2}: lấy ký tự a xuất hiện 2 đến 4 lần, b xuất hiện 0 đến 4 lần, c xuất hiện 1 đến 2 lần

aa+b\*c+: lấy ký tự aa, sau đó + : sau đó a xuất hiện 1 hoặc nhiều lần, sau đó đến b xuất hiện 0 hoặc nhiều lần, sau đó

đến c xuất hiện 1 hoặc nhiều lần

\d+ files\* found\?: lấy ký tự đầu là số xuất hiện 1 hoặc nhiều lần, sau đó đến file, s xuất hiện 0 hoặc nhiều lần, found và cuối cùng là dấu chấm hỏi

\d\.\s+abc: ký tự đầu là số sau đó đến dấu chấm, tiếp là khoảng trắng xuất hiện 1 hoặc nhiều lần s+, cuối cùng là abc

^Mission: successful$: bắt đầu là Mission và kết thúc là successful

^(file.+).pdf$: bắt đầu là file, sau đó đến dấu chấm xuất hiện 1 hoặc nhiều lần, đuôi là .pdf

(\w+ (\d+)) Jan 1987 và Jan 1987 1987

(\d+)x(\d+): 1280x720 và 1280 720

I love (cats|dogs): lấy I love cats or dogs

.\*: lấy ký tự đầu tiên bất kỳ xuất hiện 0 hoặc nhiều lần

Exercise 1: Matching Numbers

^-?\d+(,\d+)\*(\.\d+(e\d+)?)?$

|  |  |  |
| --- | --- | --- |
| Match | 3.14529 | Success |
| Match | -255.34 | Success |
| Match | 128 | Success |
| Match | 1.9e10 | Success |
| Match | 123,340.00 | Success |
| Skip | 720p |  |

Exercise 2: Matching Phone Numbers

|  |  |  |  |
| --- | --- | --- | --- |
| Capture | 415-555-1234 |  | Success |
| Capture | 650-555-2345 |  | Success |
| Capture | (416)555-3456 |  | Success |
| Capture | 202 555 4567 |  | Success |
| Capture | 4035555678 |  | Success |
| Capture | 1 416 555 9292 |  |  |

Exercise 3: Matching Emails

^([\w\.]\*)

|  |  |  |  |
| --- | --- | --- | --- |
| Capture | tom@hogwarts.com |  | Success |
| Capture | tom.riddle@hogwarts.com |  | Success |
| Capture | tom.riddle+regexone@hogwarts.com |  | Success |
| Capture | tom@hogwarts.eu.com |  | Success |
| Capture | potter@hogwarts.com |  | Success |
| Capture | harry@hogwarts.com |  | Success |
| Capture | hermione+regexone@hogwarts.com |  |  |

Exercise 4: Matching HTML

1. Thẻ tag

<(\w+)

|  |  |  |  |
| --- | --- | --- | --- |
| Capture | <a>This is a link</a> |  | Success |
| Capture | <a href='https://regexone.com'>Link</a> |  | Success |
| Capture | <div class='test\_style'>Test</div> |  | Success |
| Capture | <div>Hello <span>world</span></div> |  |  |

2. >([\w\s]\*)<

|  |  |  |  |
| --- | --- | --- | --- |
| Capture | <a>This is a link</a> |  | Failed |
| Capture | <a href='https://regexone.com'>Link</a> |  | Failed |
| Capture | <div class='test\_style'>Test</div> |  | Failed |
| Capture | <div>Hello <span>world</span></div> |  |  |

3. ='([\w://.]\*)'

|  |  |  |  |
| --- | --- | --- | --- |
| Capture | <a href='https://regexone.com'>Link</a> |  | Failed |
| Capture | <div class='test\_style'>Test</div> |  |  |

Exercise 5: Matching specific filenames:

(\w+)\.(jpg|png|gif)$

|  |  |  |  |
| --- | --- | --- | --- |
| Capture | img0912.jpg |  | Success |
| Capture | updated\_img0912.png |  | Success |
| Skip | documentation.html |  | To be completed |
| Capture  Exercise 6: ^\s\*(.\*)\s\*$ | favicon.gif  Trimming Space from start and end of line |  |  |
| Capture | The quick brown fox... |  | Success |
| Capture | jumps over the lazy dog. |  |  |

Exercise 7: Extracting Data From Log Entries

(\w+)\(([\w\.]+):(\d+)\)

|  |  |  |  |
| --- | --- | --- | --- |
| Capture | E/( 1553): at widget.List.makeView(ListView.java:1727) |  | Success |
| Capture | E/( 1553): at widget.List.fillDown(ListView.java:652) |  | Success |
| Capture | E/( 1553): at widget.List.fillFrom(ListView.java:709) |  |  |

Problem 8: Parsing and extracting data from a URL

(\w+)://([\w\-\.]+)(:(\d+))?

|  |  |  |  |
| --- | --- | --- | --- |
| Capture | ftp://file\_server.com:21/top\_secret/life\_changing\_plans.pdf |  | Success |
| Capture | https://regexone.com/lesson/introduction#section |  | Success |
| Capture | file://localhost:4040/zip\_file |  | Success |
| Capture | https://s3cur3-server.com:9999/ |  | Success |
| Capture | market://search/angry%20birds |  |  |