1. What is the name of the feature responsible for generating Regex objects?

Compile () function returns the regex object.

1. Why do raw strings often appear in Regex objects?

Raw string ignores the meaning of the meta character.

1. What is the return value of the search() method?

Search() method either returns none (if the pattern doesn’t gets matched), or re.matchObject contains information about the matching part of the string.

1. From a Match item, how do you get the actual strings that match the pattern?

For example:

import re

match = '\d\d\d-\d\d\d-\d\d\d\d'

phone\_match = re.compile(match) # here regex is compiled into pattern object.

a = phone\_match.search("Amardeep's mobile number is 222-777-7522")

print(a.group())

1. In the regex which created from the r'(\d\d\d)-(\d\d\d-\d\d\d\d)', what does group zero cover? Group 2? Group 1?

group(0) will give (\d\d\d)-(\d\d\d-\d\d\d\d)

group(1) will give (\d\d\d) (i.e the first 3 digit)

group(2) will give (\d\d\d-\d\d\d\d)

1. In standard expression syntax, parentheses and intervals have distinct meanings. How can you tell a regex that you want it to fit real parentheses and periods?

Period and Parenthesis can be escaped with backslash.

1. The findall() method returns a string list or a list of string tuples. What causes it to return one of the two options?

If one or more capturing groups are present in the pattern, return a list of groups; this will be a list of tuples if the pattern has more than one group.

For Example:

import re

match = '(\d\d\d)-(\d\d\d-\d\d\d\d)'

phone\_match = re.compile(match) # here regular expression are compiled into pattern object.

phone\_match.findall("Amardeep's mobile number is 222-777-7522 797-944-4264")

Output:

[('222', '777-7522'), ('797', '944-4264')]

1. In standard expressions, what does the | character mean?

The above symbol represents the “OR” symbol. It checks whether the pattern before or after the “OR” symbol is present in the string or not.

For example:

a|b means that it will match any string that contains a or b

1. In regular expressions, what does the character stand for?

All characters, except those having special meaning in regex, matches themselves.

10.In regular expressions, what is the difference between the + and \* characters?

Plus(+) symbol matches one or more occurrence of the regex preceding the + symbol

For Example:

ab+c will match abbc, abc, abbbbbbbbbbc. But this pattern will not match “ac” because ‘a’ is not followed by “b” .

Also,

Star(\*) symbol matches zero or more occurrence of the regex preceding the \* symbol.

For Example:

ab\*c will match ac, abc, abbbbbc. But it will not match aacc, acccc,c,a .

1. What is the difference between {4} and {4,5} in regular expression?

{4} means exactly four times repetition of a particular pattern in the given string.

\d{4,5} means if we need to find numbers from 4 to 5 digits.

1. What do you mean by the \d, \w, and \s shorthand character classes signify in regular expressions?

\d : matches any decimal digit , this is equivalent to the set class [0-9]

\w: matches any alpha numeric character , this is equivalent to the class [a-zA-Z0-9]

\s: matches any white space character.

1. What do means by \D, \W, and \S shorthand character classes signify in regular expressions?

\D: matches any non-digit character , this is equivalent to the set class [^0-9]

\W: matches any non-alpha numeric character.

\S: matches any non- white space character.

1. What is the difference between (.\*?) and (.\*) ?

Dot(.) means any character can come in place of Dot(.)

Star(\*) means zero or more occurrence of character after which this symbol is placed

Question mark(?) means zero or one occurrence of character after which this symbol is placed.

Now,

1) (.\*)? captures the group of zero or one time(?). And that group will consisit of characters of any lenght because of(\*) , and that character can be any character beacuse of (.)

2) (.\*) capture the one or more because of(.) of any charcter because of (\*)

3) (.\*)? matches any character (.) any number of times (\*) and that too as few times as possible because to make the regex match(?)

1. What is the syntax for matching both numbers and lowercase letters with a character class?

[0-9a-z]

1. What is the procedure for making a normal expression in regex case insensitive?

Make the parameter, flags=re.IGNORECASE

1. What does the . character normally match? What does it match if re.DOTALL is passed as 2nd argument in re.compile()?

a)Dot(.) character normally matches any character except then new line character.

b)Re.DOTALL is used to tune the behaviour of DOT(.). As we know that DOT(.) will identify any character except the newline character. But when write re.DOTALL we can tune this behaviour and now it will be able to identify the newline character.

18. If numReg = re.compile(r'\d+'), what will numRegex.sub('X', '11 drummers, 10 pipers, five rings, 4 hen') return?

'X drummers, X pipers, five rings, X hen

Explanation:

Wherever we find the number then just replace with “**X**”

19. What does passing re.VERBOSE as the 2nd argument to re.compile() allow to do?

re.VERBOSE allows representing the regular expression in a more readable way .

For Example:

import re

match = re.search(r"""(?P<first\_two>[\d]{2}) # The first two digit

- # A literal

(?P<last\_three>[\d]{3})# The last three digit

""", '25-555', re.VERBOSE)

print(match)

20. How would you write a regex that match a number with comma for every three digits? It must match the given following:

'42'

'1,234'

'6,368,745'

but not the following:

'12,34,567' (which has only two digits between the commas)

'1234' (which lacks commas)

Solution:

r"^\d{1,3}(,\d{3})\*$”

21. How would you write a regex that matches the full name of someone whose last name is Watanabe? You can assume that the first name that comes before it will always be one word that begins with a capital letter. The regex must match the following:

'Haruto Watanabe'

'Alice Watanabe'

'RoboCop Watanabe'

but not the following:

'haruto Watanabe' (where the first name is not capitalized)

'Mr. Watanabe' (where the preceding word has a nonletter character)

'Watanabe' (which has no first name)

'Haruto watanabe' (where Watanabe is not capitalized)

**Solution**:

import re

pattern = r"[A-Z][A-za-z]+\sWatanabe$"

text = 'Haruto Watanabe'

re.findall(pattern, text)

22. How would you write a regex that matches a sentence where the first word is either Alice, Bob, or Carol; the second word is either eats, pets, or throws; the third word is apples, cats, or baseballs; and the sentence ends with a period? This regex should be case-insensitive. It must match the following:

'Alice eats apples.'

'Bob pets cats.'

'Carol throws baseballs.'

'Alice throws Apples.'

'BOB EATS CATS.'

but not the following:

'RoboCop eats apples.'

'ALICE THROWS FOOTBALLS.'

'Carol eats 7 cats.'

**Solution**:

import re

pattern = r"^(\bAlice\b|\bBob\b|\bCarol\b)(\s)(\beats\b|\bpets\b|\bthrows\b)(\s)(\bapples\b|\bcats\b|\bbaseballs\b)(.$)"

text = 'Carol pets baseballs.'

re.findall(pattern, text)