Practice Questions

1. What is the function that creates Regex objects?

Re.compile()

2. Why are raw strings often used when creating Regex objects?

So that no characters are escaped

3. What does the search() method return?

It will return *None* if the pattern is not found, and it will return a *Match Object* if a match is found.

4. How do you get the actual strings that match the pattern from a Match object?

Use the .group() method.

5. In the regex created from r'(\d\d\d)-(\d\d\d-\d\d\d\d)', what does group 0 cover? Group 1? Group 2?

Group 0 is the whole thing -> (\d\d\d)-(\d\d\d-\d\d\d\d)

Group 1 is (\d\d\d)

Group 2 is (\d\d\d-\d\d\d\d)

6. Parentheses and periods have specific meanings in regular expression syntax. How would you specify that you want a regex to match actual parentheses and period characters?

Use backslash then the parenthesis or period. For ex: \( or \) and \.

7. The findall() method returns a list of strings or a list of tuples of strings. What makes it return one or the other?

If there are no groups then it returns a list of strings. If there are groups then it returns a list of tuples of strings.

8. What does the | character signify in regular expressions?

It’s called the pipe and it will match one thing or another.

9. What two things does the ? character signify in regular expressions?

It could mean that part of the pattern is optional, or it could mean that it’s nongreedy.

10. What is the difference between the + and \* characters in regular expressions?

The + means one or more

The \* means zero or more

11. What is the difference between {3} and {3,5} in regular expressions?

{3} means match the previous thing three times

{3,5} means match the previous thing three, four, or five times and it will match the longest possible by default.

12. What do the \d, \w, and \s shorthand character classes signify in regular expressions?

\d -> decimal

\w -> word

\s -> whitespace

13. What do the \D, \W, and \S shorthand character classes signify in regular expressions?

\D -> anything NOT decimal

\W -> anything NOT word

\S -> anything NOT whitespace

14. How do you make a regular expression case-insensitive?

Pass re.IGNORECASE or re.I to re.compile() as the second argument.

15. What does the . character normally match? What does it match if

re.DOTALL is passed as the second argument to re.compile()?

16. What is the difference between .\* and .\*??

.\* match any character zero or more times (greedy)

.\*? match any character zero or more times (nongreedy)

17. What is the character class syntax to match all numbers and lowercase letters?

Re.compile(r’[0-9a-z]’)

18. If

numRegex = re.compile(r'\d+')

what will

numRegex.sub('X', '12 drummers, 11 pipers, five rings, 3 hens')

return?

X drummers, X pipers, five rings, X hens

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

Put any amount of whitespace and comments in the argument

20. How would you write a regex that matches a number with commas for every three digits? It must match the 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)

Re.compile(r’^\d{1,3}(,\d{3})\*$)

21. How would you write a regex that matches the full name of someone

whose last name is Nakamoto? 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:

• 'Satoshi Nakamoto'

• 'Alice Nakamoto'

• 'RoboCop Nakamoto'

but not the following:

• 'satoshi Nakamoto' (where the first name is not capitalized)

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

• 'Nakamoto' (which has no first name)

• 'Satoshi nakamoto' (where Nakamoto is not capitalized)

Re.compile(r’^[A-Z]\w+\sNakamoto’)

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.'

Re.compile(r’^Alice|Bob|Carol\s(eats|pets|throws)\s(apples|cats|baseballs)\.$’, re.I)