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

Sol: re.compile(pattern)

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

Sol. Raw string (r') treats backslash as literal character instead of escape characters. Backslash are often used in regex objects.

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

Sol. It returns all the matching pattern from the string instead of match function which returns only first string

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

Sol: .group() function can be used

5. 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?

Sol. group(0) is the entire match

Group (1) is first parenthesis

Group(2) is second parenthesis

6. 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?

Sol. You can use backslash to escape it. Ex. \. \(\)

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

Sol. When we use group() funtion with findall() we will get tuple of list otherwise we will get only list.

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

Sol. | Represents either or condition

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

Sol. Means zero or one

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

Sol. + Matches one or more

* Match zero or more

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

Sol. {4} matches exactly 4 instances of the preceding group

{4,5} matches instances between 4 and 5 with 4 and 5 inclusive

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

Sol. \d match single digit

\w match single word

\s match single space

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

Sol. Match any character that is not single digit , single word or single space respectively.

14. What is the difference between .*? and .*?

Sol. .* Means greedy : match as many reps aa possible

.*? Means non greedy: match as few rep as possible

Ex. eeeAiiZuuuuAoooZeeee

A.*Z :AiiZuuuuAoooZ

A.*?Z: AiiZ and AoooZ

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

Sol:

Class[a-z0-9]

16. What is the procedure for making a normal expression in regax case insensitive?

Sol. re.compile(r'xyz', re.IGNORECASE)

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

Sol. . Matches all characters except newline

Re.DOTALL matches all characters and also newline

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

Sol. (11 Drummers, 10 Piper's, five rings, X hen)

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

Sol. Allows to add whitespace and comment to string passed in re.compile

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

Sol: re.compile(r'^\d{1,3}(,\d{3})*$')

'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)

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)

Sol. re.compile(r'^[A-Z][a-z]*.\sWatanabe')

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.'

Sol. re.compile(r'\s(Alice|Bob|Carol)\s(eats|pets|throws)\s(apples|cats|baseballs)\.', re.IGNORECASE)