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

The re. compile() function returns Regex objects.

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

Raw strings are used so that backslashes do not have to be escaped.

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

The search() method returns Match objects.

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

The group() method returns strings of the matched text.

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?

Group 0 ⇒ entire match

Group 2 ⇒ covers the second set of parentheses.

Group 1 ⇒ covers the first set of parentheses.

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?

Periods and parentheses can be escaped with a **backslash** like **\., \( and \)**.

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

If the regex has no groups, a list of strings is returned. If the regex has groups, a list of tuples of strings is returned.

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

The | character signifies matching either, or between two groups.

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

The “?” character is a quantifier that mean "match zero or one of the preceding group".

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

Both are Quantifiers which account for:

\* - 0 or More matches

+ - 1 or More matches

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

{4} - Exact Number

{4,5} - Range of Numbers (Minimum, Maximum)

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

\d - Digit (0-9)

\w - Word Character (a-z, A-Z, 0-9, \_)

\s - Whitespace (space, tab, newline)

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

\D - Not a Digit (0-9)

\W - Not a Word Character

\S - Not Whitespace (space, tab, newline)

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

The .\* performs a greedy match, and the .\*? performs a nongreedy match.

. signifies every character except new line. \* signifies zero or more occurrences. so .\* matches every line of text.

? signifies zero or one occurrence so .\*? matches every character in text except newline with empty spaces detected between each character signifies for zero characters.

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

Either [0-9a-z] or [a-z0-9]

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

Passing **re.I** or **re.IGNORECASE** as the second argument to re.compile() will make the matching

case insensitive. This parameter is often called as Flag.

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

The . character normally matches any character except the newline character.

If re.DOTALL is passed as the second argument to re.compile(), then the dot will also match newline characters.

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’

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

re.VERBOSE flag allows you to write regular expressions that look nicer and are more readable by allowing you to visually separate logical sections of the pattern and add comments. Whitespace within the pattern is ignored, except when in a character class, or when preceded by an unescaped backslash, or within tokens like \*?, (?: or (?P. When a line contains a # that is not in a character class and is not preceded by an unescaped backslash, all characters from the leftmost such # through the end of the line are ignored.

**# Without Using VERBOSE**

regex\_email = re.compile(r'^([a-z0-9\_\.-]+)@([0-9a-z\.-]+)\.([a-z\.]{2, 6})$',

re.IGNORECASE)

**# Using VERBOSE**

regex\_email = re.compile(r"""

^([a-z0-9\_\.-]+) # local Part

@ # single @ sign

([0-9a-z\.-]+) # Domain name

\. # single Dot .

([a-z]{2,6})$ # Top level Domain

""",re.VERBOSE | re.IGNORECASE)

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)

regex : **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)

regex :  **r’^[A-Z]\w+ Watanabe’**

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.'

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