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

re. compile() -- generates Regex objects

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

Since regular expressions frequently use backslashes, so to avoid typing extra backslashes raw re.compile() function used in raw strings.

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

The search() method will return None if the regex pattern is not found in the string. If the pattern is found, the search() method returns a match object.

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

we can call group() on the match object to return the match result

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: (\d\d\d)-(\d\d\d-\d\d\d\d) Group 1: (\d\d\d) Group 2: (\d\d\d-\d\d\d\d)

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?

Using r'.()'

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

When called on a regex with no groups, the method findall() returns a list of string matches

When called on a regex that has groups, the method findall() returns a list of tuples of strings

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

The | character is called a pipe. You can use it anywhere you want to match one of many expressions.

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

The ? is saying "Match zero or one of the group preceding this question mark."

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

\* means "match zero or more"

+ means "match one or more."

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

{4} is saying "match this pattern three times." {4,5} means "match this pattern three to five times."

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

\d in a regex stands for a digit character - that is, any single number 0 to 9.

\w means any letter, numeric digit, or the underscore character.

\s means any space, tab, or newline character.

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

\D represents any character that is not a numeric digit from 0 to 9.

\W represents any character that is not a letter, numeric digit, or the underscore character.

\S represents any character that is not a space, tab or newline.

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

\* is a greedy mode, which tends to get the longest string that meets the conditions.

\*? is a non-greedy mode, which tends to get the shortest string that can meet the conditions.

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

[0-9a-z]

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

re.IGNORECASE **:** This flag allows for case-insensitive matching of the Regular Expression with the given string

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

The . character in a regular expression is called a wildcard and will match any character except for a newline.

The dot-star will match everything except a newline. By passing re.DOTALL as the second argument to re.compile(), you can make the dot character match all characters, including 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 hens

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

To ignore whitespace and comments inside the regular expression string.

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)

numCommas = re.compile(r'(^\d{1,3})(,\d{3})\*$') numCommas.search('12,34,567').group()

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)

re.compile(r'[A-Z][a-z]\*\s 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.'

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