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

re.compile()

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 is the entire match, group 1 covers the first set of parentheses, and group 2 covers the second 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: \., \(, and \).

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?

It can signify that part of a regular expression is optional (i.e., it occurs 0 or 1 times), or that nongreedy matching should be used.

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

+ means match one or more times.

\* means match 0 or more times.

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

{3} means match exactly 3 times.

{3,5} means match 3, 4, or 5 times.

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

\d stands for a numeric digit.

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

\s stands for any space, tab, or newline character.

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

\D stands for any character that is not a numeric digit.

\W stands for any character that is not a letter, digit, or underscore.

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

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

use the re.IGNORECASE (or re.I) argument in re.compile().

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

. normally matches any one character except the newline.

If the re.DOTALL argument is used, a . matches any one character including the newline.

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

.\* uses greedy matching.

.\*? is the nongreedy version.

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

[a-z0-9]

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?

It allows you to ignore white space and comments in 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)

r'((^\d{1,3},)(\d{3},)\*\d{3}$)|(^\d{,3}$)'

Much, much better way: 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)

r'[A-Z][a-z]\* Nakamoto'

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) (eats|pets|throws) (apples|cats|baseballs)\.', re.IGNORECASE)