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

Regex objects can be imported using the 're' module.

Import re

re.compile() returns regex objects

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

Raw strings inform the Python interpreter to interpret the escape character (\) as a string. Hence, the string '\nHey' would be literally interpreted as \nHey. '\n' will not be considered as a new line.

Regex objects' search criteria are based on many METACHARACTERS like '\d' or '\w' that represent numerical or text data respectively. If raw strings are not used, there are chances of Python interpreter skipping the '\' escape character and resulting in incorrect searches.

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

A typical search() i.e. re.search(<regex>,<string>) returns the matched string and its starting and ending indices. If we slice the input string with the start and end value, it will give us the matched string.

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

group() method can be used to return the matched string. If multiple groups have been created in the search criteria, group() method can be used to search for a specific group number also using group(n), n represents the group number, 1 being the outermost group.

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 indexing is 1 based and not zero based. Hence, **group(0)** has a special meaning, it returns the complete matched string.

Group1 is the first group from left, i.e. **(\d\d\d)**

Group2 is **(\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?

Escape character (\) can be used to achieve this. Periods and parenthesis can be escaped by suing backslash like \., \(, \).

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

findall() method returns the complete string if no grouping is done. In case of grouping, multiple search results are returned, hence, tuple of the searches are retuned.

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

The | character represents OR operation. It allows matching of either of the groups mentioned on the either side of the operator.

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

Question unclear, is it about '?' character

'?' is a meta-character that means match either 0 or 1 of the instance of the preceding group. It can also be used to indicate non-greedy matching where the smallest matching is targeted.

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

+ indicates search for 1 or more instances of the preceding group.

* indicates search for 0 or more instances of the preceding group.

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

{4} will search for exactly 4 repetitions of the preceding instance.

{4,5} will search for min 4 to max 5 repetitions of the preceding instance

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

\d Single Numerical value [0-9]

\w Single alpha numeric value including underscore `[a-zA-Z0-9_]`

\s Single white space [ ]

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

Capital versions of \d, \w and \S are inverse search criteria. They represent exactly opposite to what we discussed in the previous question.

\D Single NON-Numerical value [^0-9]

\W Single NON alpha numeric value including underscore `[^a-zA-Z0-9_]`

\S Single NON white space [^]

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

Question unclear

* Searches for the longest possible match of the preceding instance. This is greedy search

*? Searches for the shortest possible match of the preceding instance. This is non-greedy search

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

This can be performed using the character class [0-9a-z] or [a-z0-9]

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

This can be obtained by using the IGNORECASE flag.
Example: re.search('a+', 'aaaAAA', re.IGNORECASE)

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

The ’ .’ (period) character matches a single character except newline.
If re.DOTALL flag is provided as an input, it matches for the newline also.

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

numRegex.sub() replaces the occurrences of a particular sub-string with other sub-string. Here the compiled substring would return 11, 10 and 4 if matched (‘\d+’ corresponds to 1 or more numerical values). These are replaced by ‘X’ using the numRegex.sub()

Return value: ‘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 inclusion of whitespace and comments within a regex search.

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)

Required expression

r’\d{1-3},(,\d{3})*$’

This will match any string starting with 1-3 numerals, followed by a comma and 3 numerical group. * can match zero or more instances, hence a number with 1-3 digits will also match.

20. 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 is

[A-Z][a-z]*\sWatanabe

21. 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.serach(r'(Alice|Bob|Carol)\s(eats|pets|thows)\s(apples|cats|baseballs)\.', re.IGNORECASE)