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

Solution1:-

The re.compile() function returns Regex objects.

Regular Expressions, also known as “regex” or “regexp”, are used to match strings of text such as particular characters, words, or patterns of characters. It means that we can match and extract any string pattern from the text with the help of regular expressions.

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

Solution2:-

According to Python docs, raw string notation (r"text") keeps regular expressions meaningful and confusion-free. Without it, every backslash ('\') in a regular expression would have to be prefixed with another one to escape it.

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

Solution3:-

The re.search() method takes a regular expression pattern and a string and searches for that pattern within the string. If the search is successful, search() returns a match object or None otherwise.

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

Solution4:-

print(re.match(pattern,string).group())

1. 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?

Solution5:-

Group 0 is the entire match, group 1 covers the first set of parentheses, and group 2 covers the second set of parentheses.

Code:-

pattern=r'(d(e))f'

m = re.search(pattern, 'abcdef')

print(m.group(0))

print(m.group(1))

print(m.group(2))

output:-

def

de

e

1. 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?

Solution6:-

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

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

Solution7:-

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.

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

Solution8:-

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

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

Solution9:-

A character class matches any one of a set of characters. It is used to match the most basic element of a language like a letter, a digit, space, a symbol etc.

/s : matches any whitespace characters such as space and tab

/S : matches any non-whitespace characters

/d : matches any digit character

/D : matches any non-digit characters

/w : matches any word character (basically alpha-numeric)

/W : matches any non-word character

/b : matches any word boundary (this would include spaces, dashes, commas, semi-colons, etc)

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

Solution10:-

The asterisk symbol ( \* ):

It tells the computer to match the preceding character (or set of characters) for 0 or more times (upto infinite).

Example : The regular expression ab\*c will give ac, abc, abbc,

abbbc….ans so on.

The Plus symbol ( + ):

It tells the computer to repeat the preceding character (or set of characters) for atleast one or more times(upto infinite).

Example : The regular expression ab+c will give abc, abbc,

abbc, … and so on.

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

Solution11:-

It tells the computer to repeat the preceding character (or set of characters) for as many times as the value inside this bracket.

Example : {4} means that the preceding character is to be repeated 4

times, {min,} means the preceding character is matches min or more

times. {min,max} means that the preceding character is repeated at

least min & at most max times. Hence {4,5} means that that the preceding character is repeated at least 4 & at most 5 times

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

Solution12:-

/d : matches any digit character

/w : matches any word character (basically alpha-numeric)

/s : matches any whitespace characters such as space and tab

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

Solution13:-

/D : matches any non-digit characters

/W : matches any non-word character

/S : matches any non-whitespace characters

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

Solution14:-

Input:-

eeeAiiZuuuuAoooZeeee

then A.\*Z yields 1 match: AiiZuuuuAoooZ

and A.\*?Z yield 2 match AiiZ and AoooZ

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

Solution15:-

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

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

Solution16:-

USE re.IGNORECASE TO DO CASE INSENSITIVE REGULAR EXPRESSIONS

Assign re.IGNORECASE as an optional parameter in regular expression calls to do case insensitive regular expressions.

regex\_search = re.search('test', 'TeSt', re.IGNORECASE)

print(regex\_search)

OUTPUT

<re.Match object; span=(0, 4), match='TeSt'>

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

Solution17:-

The dot symbol can take place of any other symbol, that is why it is called the wildcard character.

Example :

The Regular expression .\* will tell the computer that any character can be used any number of times.

Code:-

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?

Solution18:-It will replace numerical values as x

Code:-

numReg = re.compile(r'\d+')

numReg.sub('X', '11 drummers, 10 pipers, five rings, 4 hen')

output:-

'X drummers, X pipers, five rings, X hen'

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

Solution19:-

The re. VERBOSE argument allows you to add whitespace and comments to the string passed to re.

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)

Solution20:-

re.compile(r'^\d{1,3}(,\d{3})\*$') will create this regex, but other regex strings can produce a similar regular expression.

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)

Solution21:-

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

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.'

Solution22:-

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