ASSIGNMENT NO 7

Nirbhay Ahir

[Nirbhay.ahir36@gmail.com](mailto:Nirbhay.ahir36@gmail.com)

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

Ans: re.compile() is the feature responsible for generating Regex objects.

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

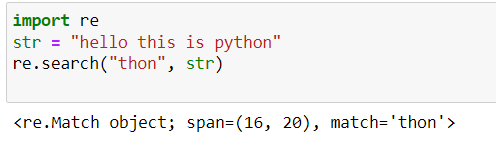
Ans: Python raw string is created by prefixing a string literal with 'r' or 'R'.

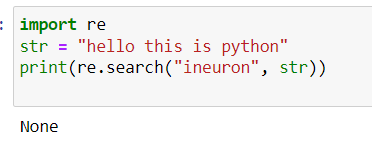
Python raw string treats backslash (\) as a literal character.

This is useful when we want to have a string that contains backslash and don't want it to be treated as an escape character.

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

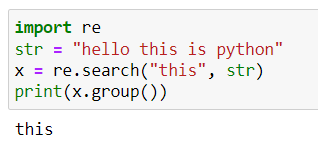
Ans: The re.search(pattern,string) function searches for some substring in a string and returns a match object if found, else it returns none.





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

Ans: 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?

Ans: Group 0 covers the entire pattern , group 1 covers the first set of parentheses, and group 2 covers the second set of parentheses.

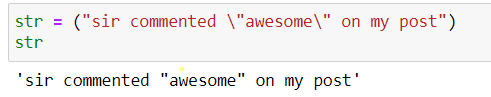
Group 0 - r'(\d\d\d)-(\d\d\d-\d\d\d\d)'

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?

Ans: The \.  \(  and  \)  escape characters in the raw string passed to re.compile() will match actual parenthesis characters.



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

Ans: 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?

Ans: . In standard expressions ‘|’ means OR operator.

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

Ans: The character and what does it specify is given below

^ : Starts with $ : Ends with : Zero or more occurrences : One or more occurrences

? : Zero or one occurrences

{}: Exactly the specified number of occurrences

| : Either or

(): Capture and group

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

Ans: In regular Expressions, ‘\*’ Represents Zero or more occurances of the preceeding group, whereas  ‘+’  represents One or more occurances of the preceeding group.

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

Ans: {4} means that its preceeding group should repeat 4 times. where as {4,5} means that its preceeding group should repeat mininum 4 times and maximum 5 times inclusively.

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

Ans: The shorthand characters signifies following expressions

\w – Matches a word character equivalent to [a-zA-Z 0-9]

\d – Matches digit character equivalent to [0-9]

\s – Matches whitespace character

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

Ans: The shorthand characters signifies following expressions

\W – Matches any non-alphanumeric character

\D – Matches any non-digit character

\S – Matches any non-whitespace character

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

Ans: \*? is a Greedy mode, which returns the longest string that meets the condition. Whereas .\*? is a non greedy mode which returns the shortest string that meets the condition.

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

Ans: The Synatax is Either [a-z0-9] or [0-9a-z]

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

Ans: We can pass re.IGNORECASE as a flag to make a normal expression case insensitive

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

Ans: Dot . character matches everything in input except newline character. By passing re.DOTALL as a flag 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?

Ans: the output will be 'X drummers, X pipers, five rings, X hen'

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

Ans: re.VERBOSE will allow to add whitespace and comments to string passed to re.compile().

20. How would you write a regex that match a number with comma for every three digits? It must match the given following:

'42'

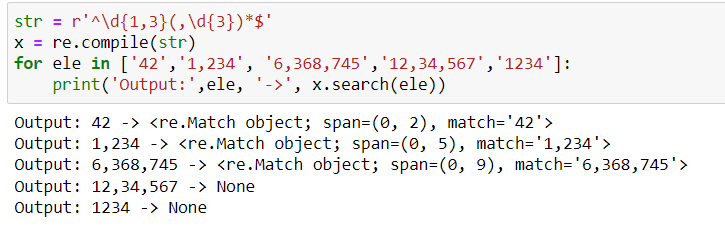
'1,234'

'6,368,745'

but not the following:

'12,34,567' (which has only two digits between the commas)

'1234' (which lacks commas)

Ans: 

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'

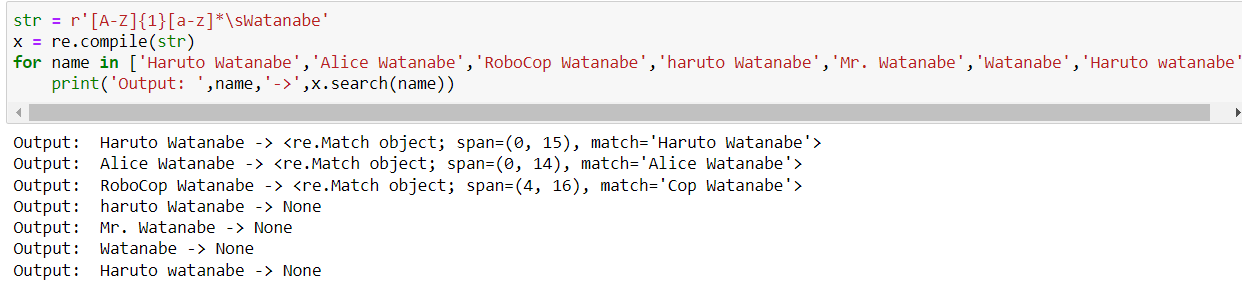
but not the following:

'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)

Ans: 

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.'

Ans:

