[REGULAR EXPRESSION] [FLIP ROBO ASSIGNMENT] **BIJAY PAREEK** BATCH NO: DS2308 import regex as re Question 1- Write a Python program to replace all occurrences of a space, comma, or dot with a colon. Sample Text- 'Python Exercises, PHP exercises.' Expected Output: Python:Exercises::PHP:exercises: In [2]: Sample_text = 'Python Exercises, PHP exercises.' print(re.sub("[,.]", ":", Sample_text)) Python:Exercises::PHP:exercises: Question 2- Write a Python program to find all words starting with 'a' or 'e' in a given string. In [3]: string = "early bird catches the worm" pattern = $("[ae]\w+")$ matches = re.findall(pattern, string) print(matches) ['early', 'atches'] Question 3- Create a function in python to find all words that are at least 4 characters long in a string. The use of the re.compile() method is mandatory. In [4]: string = "early bird catches the worm" pattern = $(r'' b w{4,} b'')$ match = re.findall(pattern, string) print(match) ['early', 'bird', 'catches', 'worm'] Question 4- Create a function in python to find all three, four, and five character words in a string. The use of the re.compile() method is mandatory. In [5]: string = "early bird catches the worm" pattern = $(r'' b w{3,} b'')$ match = re.findall(pattern, string) print(match) ['early', 'bird', 'catches', 'the', 'worm'] Question 5- Create a function in Python to remove the parenthesis in a list of strings. The use of the re.compile() method is mandatory. Sample Text: ["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data Science World)", "Data (Scientist)"] **Expected Output:** example.com hr@fliprobo.com github.com Hello Data Science World Data Scientist text = ["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data Science World)", "Data (Scientist)"] list = re.compile(r"[()]") for text in text: print(re.sub(list,"",text)) example .com hr@fliprobo .com github .com Hello Data Science World Data Scientist Question 6- Write a python program to remove the parenthesis area from the text stored in the text file using Regular Expression. Sample Text: ["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data Science World)", "Data (Scientist)"] Expected Output: ["example", "hr@fliprobo", "github", "Hello", "Data"] Note- Store given sample text in the text file and then to remove the parenthesis area from the text. file=open("text_file.txt",'r') text_files = file.read() print(re.sub(r"\([^)]+\)", "", text_files)) Question 7- Write a regular expression in Python to split a string into uppercase letters. Sample text: "ImportanceOfRegularExpressionsInPython" **Expected Output:** ['Importance', 'Of', 'Regular', 'Expression', 'In', 'Python'] sample_text = "ImportanceOfRegularExpressionsInPython" pattern = $'[A-Z][^A-Z]*'$ matches = re.findall(pattern, sample_text) print(matches) ['Importance', 'Of', 'Regular', 'Expressions', 'In', 'Python'] Question 8- Create a function in python to insert spaces between words starting with numbers. Sample Text: "RegularExpression1IsAn2ImportantTopic3InPython" **Expected Output:** Regular Expression 1IsAn 2Important Topic 3In Python def word_space(str1): In [8]: **return** re.sub('\w\\s([a-z])','\1\2', str1) print(word_space("RegularExpression 1IsAn 2 ImportantTopic 3InPython")) RegularExpression 1IsAn 2 ImportantTopic 3InPython Question 9- Create a function in python to insert spaces between words starting with capital letters or with numbers. Sample Text: "RegularExpression1IsAn2ImportantTopic3InPython" **Expected Output:** RegularExpression 1 IsAn 2 ImportantTopic 3 InPython In [9]: def capital_words_spaces(str1): return re.sub(r"(\w)([A-Z])", r'\1\2', str1) print(capital_words_spaces("RegularExpression1IsAn2ImportantTopic3InPython")) RegularExpression1IsAn2ImportantTopic3InPython Question 10- Write a python program to extract email address from the text stored in the text file using Regular Expression. Sample Text- Hello my name is Data Science and my email address is xyz@domain.com and alternate email address is xyz.abc@sdomain.domain.com. Please contact us at hr@fliprobo.com for further information. **Expected Output:** ['xyz@domain.com', 'xyz.abc@sdomain.domain.com'] ['hr@fliprobo.com'] Note- Store given sample text in the text file and then extract email addresses. In [10]: sample_text = 'Sample Text- Hello my name is Data Science and my email address is xyz@domain.com and alternate email address is xyz.abc@sdomain.domain.com. Ple pattern = $(r''[a-z0-9\.\-+_]+@[a-z0-9\.\-+_]+\.[a-z]+")$ matches = re.findall(pattern, sample_text) print(matches) ['xyz@domain.com', 'xyz.abc@sdomain.domain.com', 'hr@fliprobo.com'] Question 11- Write a Python program to match a string that contains only upper and lowercase letters, numbers, and underscores. def text_match(text): patterns = $'^[a-zA-Z0-9]*$ if re.search(patterns, text): return 'Found a match!' else: return('Not matched!') print(text_match("BIJAY")) print(text_match("_pareek")) print(text_match('pareek')) print(text_match('1998')) Found a match! Found a match! Found a match! Found a match! Question 12- Write a Python program where a string will start with a specific number. In [12]: string = "5_no is my car" pattern = $("^5")$ matches = re.findall(pattern, string) print(matches) ['5'] Question 13- Write a Python program to remove leading zeros from an IP address In [13]: | ip_address = "216.08.008.094.196" string = re.sub($' \setminus [0]^*'$, '.', ip_address) print(string) 216.8.8.94.196 Question 14- Write a regular expression in python to match a date string in the form of Month name followed by day number and year stored in a text file. Sample text: 'On August 15th 1947 that India was declared independent from British colonialism, and the reins of control were handed over to the leaders of the Country'. **Expected Output-**August 15th 1947 Note- Store given sample text in the text file and then extract the date string asked format. Question 15- Write a Python program to search some literals strings in a string. Sample text: 'The quick brown fox jumps over the lazy dog.' Searched words: 'fox', 'dog', 'horse' In [14]: patterns = ['fox', 'dog', 'horse'] text = 'The quick brown fox jumps over the lazy dog.' **for** pattern **in** patterns: print('Searching for "%s" in "%s" ->' % (pattern, text),) if re.search(pattern, text): print('Matched!') print('Not Matched!') Searching for "fox" in "The quick brown fox jumps over the lazy dog." -> Matched! Searching for "dog" in "The quick brown fox jumps over the lazy dog." -> Matched! Searching for "horse" in "The quick brown fox jumps over the lazy dog." -> Not Matched! Question 16- Write a Python program to search a literals string in a string and also find the location within the original string where the pattern occurs Sample text: 'The quick brown fox jumps over the lazy dog.' Searched words: 'fox' sample_text = 'Thequickbrownfoxjumpsoverthelazydog' In [15]: result = re.search('fox', sample_text) print(result) <regex.Match object; span=(13, 16), match='fox'> Question 17- Write a Python program to find the substrings within a string. Sample text: 'Python exercises, PHP exercises, C# exercises' Pattern: 'exercises'. text = 'Python exercises, PHP exercises, C# exercises' pattern = 'exercises' match = re.findall(pattern, text) print(match) ['exercises', 'exercises', 'exercises'] Question 18- Write a Python program to find the occurrence and position of the substrings within a string. In [17]: string = 'Python exercises, PHP exercises, C# exercises' pattern = 'exercises' for match in re.finditer(pattern, string): s = match.start() e = match.end() print('Found "%s" at %d:%d' % (string[s:e], s, e)) Found "exercises" at 7:16 Found "exercises" at 22:31 Found "exercises" at 36:45 Question 19- Write a Python program to convert a date of yyyy-mm-dd format to dd-mm-yyyy format. def convert_date_format(dt): **return** re.sub(r'($\d{4}$)-($\d{1,2}$)-($\d{1,2}$)', ' $\d{3-\d{1,2}}$ ', dt) dt1 = "2023-08-26" print("date in YYY-MM-DD Format: ",dt1) print("date in DD-MM-YYYY Format: ",(dt1)) date in YYY-MM-DD Format: 2023-08-26 date in DD-MM-YYYY Format: 2023-08-26 Question 20- Create a function in python to find all decimal numbers with a precision of 1 or 2 in a string. The use of the re.compile() method is mandatory. Sample Text: "01.12 0132.123 2.31875 145.8 3.01 27.25 0.25" Expected Output: ['01.12', '145.8', '3.01', '27.25', '0.25'] In [19]: def find_decimal_numbers(text): pattern = $re.compile(r'\b\d+\.\d\{1,2\}\b')$ decimal_numbers = pattern.findall(text) return decimal_numbers text = "01.12 0132.123 2.31875 145.8 3.01 27.25 0.25" decimal_numbers = find_decimal_numbers(text) print(decimal_numbers) ['01.12', '145.8', '3.01', '27.25', '0.25'] Question 21- Write a Python program to separate and print the numbers and their position of a given string. In [20]: string = "Bijay Pareek has won by 43 runs in his cricket match" pattern = $re.compile(r''(\d+)'')$ for match in pattern.finditer(string): print(match) print(match.group()) print('Index position', match) <regex.Match object; span=(24, 26), match='43'> Index position <regex.Match object; span=(24, 26), match='43'> Question 22- Write a regular expression in python program to extract maximum/largest numeric value from a string. Sample Text: 'My marks in each semester are: 947, 896, 926, 524, 734, 950, 642' Expected Output: 950 sample_text ='My marks in each semester are: 947, 896, 926, 524, 734, 950, 642' maximun_numeric_value = re.findall('\d+', sample_text) maximun_numeric_value = map(int,maximun_numeric_value) print(maximun_numeric_value) <map object at 0x0000019D7086E6A0> Question 23- Create a function in python to insert spaces between words starting with capital letters. Sample Text: "RegularExpressionIsAnImportantTopicInPython" Expected Output: Regular Expression Is An Important Topic In Python def capital_letters(str1): In [22]: return re.sub(r"(\w)([A-Z])", r"\1 \2", str1) print(capital_letters("RegularExpressionIsAnImportantTopicInPython")) Regular Expression Is An Important Topic In Python Question 24- Python regex to find sequences of one upper case letter followed by lower case letters In [23]: string = "Python" pattern = ('[A-Z]+[a-z]+\$')match = re.search(pattern, string) print(match) <regex.Match object; span=(0, 6), match='Python'> Question 25- Write a Python program to remove continuous duplicate words from Sentence using Regular Expression. Sample Text: "Hello hello world world" Expected Output: Hello hello world string = "Hello hello world world" matches = $re.sub(r'\b(\w+)(\1\b)+',r'\1',string)$ print(matches) Hello hello world Question 26- Write a python program using RegEx to accept string ending with alphanumeric character. regex = '[a-zA-Z0-9]\$' In [25]: def check_string(my_string): if(re.search(regex, my_string)): print('The string ends with alphanumberic character') print('The string is not alphanumeric character') my_string_1 = "MYSQL" print('The string is :') print(my_string_1) check_string(my_string_1) my_string_2 = "MYSQL003" print('\nThe string is :') print(my_string_2) check_string(my_string_2) The string is : MYSQL The string ends with alphanumberic character The string is : MYSQL003 The string ends with alphanumberic character Question 27-Write a python program using RegEx to extract the hashtags. Sample Text: """RT @kapil kausik: #Doltiwal I mean #xyzabc is "hurt" by #Demonetization as the same has rendered USELESS <U+00A0><U+00BD><U+00B1><U+0089> "acquired funds" No wo""" Expected Output: ['#Doltiwal', '#xyzabc', '#Demonetization'] sample_text = """RT @kapil_kausik: #Doltiwal I mean #xyzabc is "hurt" by #Demonetization as the same has rendered USELESS <U+00A0><U+00BD><U+00B1><U+0089> "acc In [26]: regex = "(#\w+)" matches = re.findall(regex, sample_text) print(matches) ['#Doltiwal', '#xyzabc', '#Demonetization'] Question 28- Write a python program using RegEx to remove <U+..> like symbols Check the below sample text, there are strange symbols something of the sort <U+..> all over the place. You need to come up with a general Regex expression that will cover all such symbols. Sample Text: "@Jags123456 Bharat band on 28??<U+00A0><U+00BD><U+00B8><U+0082>Those who are protesting #demonetization are all different party leaders" Expected Output: @Jags123456 Bharat band on 28??Those who are protesting #demonetization are all different party leaders sample_text = "@Jags123456 Bharat band on 28??<U+00A0><U+00BD><U+00BB><U+00B8>>Those who are protesting #demonetization are all different party leaders" pattern = r''<U'+[A-Za-z0-9]*>"matches = re.sub(pattern, "", sample_text) print(matches) @Jags123456 Bharat band on 28??Those who are protesting #demonetization are all different party leaders Question 29- Write a python program to extract dates from the text stored in the text file. Sample Text: Ron was born on 12-09-1992 and he was admitted to school 15-12-1999. Note- Store this sample text in the file and then extract dates. string =' Ron was born on 12-09-1992 and he was admitted to school 15-12-1999.' pattern = $\sqrt{4}$ matches = re.findall(pattern, string) print(matches) ['12-09-1992', '15-12-1999'] Question 30- Create a function in python to remove all words from a string of length between 2 and 4. The use of the re.compile() method is mandatory. Sample Text: "The following example creates an ArrayList with a capacity of 50 elements. 4 elements are then added to the ArrayList and the ArrayList is trimmed accordingly." Expected Output: following example creates ArrayList a capacity elements. 4 elements added ArrayList ArrayList trimmed accordingly. sample_text ="The following example creates an ArrayList with a capacity of 50 elements. 4 elements are then added to the ArrayList and the ArrayList is trimme pattern = re.compile($r'\W^*\b\w\{2,4\}\b'$) print(pattern.sub('', sample_text)) following example creates ArrayList a capacity elements. 4 elements added ArrayList ArrayList trimmed accordingly. [<<<<<<<<<<<<<<THANK YOU>>>>>>>>>>>>>>>