

Regex Assignment

Question 1- Write a Python program to replace all occurrences of a space, comma, or dot with a colon.

Ans:

```
SampleText="Python Exercises, PHP exercises."
```

```
pattern=r'\s|\.|,'
```

```
output=re.sub(pattern,':',SampleText)
```

output:

```
'Python:Exercises::PHP:exercises.'
```

Question 2- Create a dataframe using the dictionary below and remove everything (commas (,), !, XXXX, ;, etc.) from the columns except words.

Ans:

```
Dictionary= {'SUMMARY': ['hello, world!', 'XXXXX test', '123four, five;; six...']}
```

```
new_dict=pd.DataFrame(Dictionary)
```

```
pattern1=r'\,|\!|\:|\;|\.|\\d+|[X]
```

```
result=new_dict['SUMMARY'].str.replace(pattern1,"",case=None)
```

output:

```
0      hello world
1           test
2    four five six
Name: SUMMARY, dtype: object
```

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.

Ans:

```
def match_string(text):
```

```
    pattern=re.compile(r'\b\w{4}\b')
```

```
    result=pattern.findall(text)
```

```
    return result
```

```
string = "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."
```

```
result=match_string(string)
```

output:

```
['find', 'that', 'long']
```

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.

Ans:

```
def match_strings(text):
    pattern=re.compile(r'\b(\w{3,5})\b')
    result=pattern.findall(text)
    return result
strings = "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.."
result=match_strings(strings)
```

output:

```
['find', 'all', 'three', 'four', 'and', 'five', 'words', 'The', 'use',
 'the']
```

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.

Ans:

```
sampletext=["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data Science World)", "Da
ta (Scientist)"]
def remove_parentheses(strings):
    pattern = re.compile(r"s\(|\)")
    modified_strings = []
    for string in strings:
        modified_string = pattern.sub("", string)
        modified_strings.append(modified_string)
    return modified_strings
sample_text = ["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data Science World)", "
Data (Scientist)"]
result = remove_parentheses(sample_text)
```

output:

```
['example.com', 'hr@fliprobo.com', 'github.com', 'HelloData Science Wo
rld', 'DataScientist']
```

Question 6- Write a python program to remove the parenthesis area from the text stored in the text file using Regular Expression.

Ans:

```
pattern = re.compile(r"\([^)]*\)")
def remove_parentheses_area(strings):

    new_text = []
    for string in strings:
        new_string = pattern.sub("", string)
        new_text.append(new_string)
```

```

    return new_text
sample_text = ["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data Science World)", "Data (Scientist)"]
result = remove_parentheses_area(sample_text)
output:
['example ', 'hr@fliprobo ', 'github ', 'Hello ', 'Data ']

```

Question 7- Write a regular expression in Python to split a string into uppercase letters.

Ans:

```

Sampletext= "ImportanceOfRegularExpressionsInPython"
pattern=r'[IOREIP]|[A-Z]|[^a-z][a-z]+'
result=re.findall(pattern,Sampletext)
result
['Importance', 'Of', 'Regular', 'Expressions', 'In', 'Python']

```

Question 8- Create a function in python to insert spaces between words starting with numbers.

Ans:

```

SampleText= "RegularExpression1IsAn2ImportantTopic3InPython"
def space_create(sample):
    pattern=r"(\d+)([a-zA-Z]+)"
    result=re.sub(pattern,r' \1\2',sample)
    return result
txt=space_create(SampleText)
txt

```

output:

```
'RegularExpression 1IsAn 2ImportantTopic 3InPython'
```

Question 9- Create a function in python to insert spaces between words starting with capital letters or with numbers.

Ans:

```

SampleText="RegularExpression1IsAn2ImportantTopic3InPython"
def space(string):
    pattern=r"(\d+)([a-z]*)"
    result=re.sub(pattern,r' \1\2',string)
    return result
txt=space(SampleText)
txt

```

output:

```
'RegularExpression 1 IsAn 2 ImportantTopic 3 InPython'
```

Question 10- Use the github link below to read the data and create a dataframe. After creating the dataframe extract the first 6 letters of each country and store in the dataframe under a new column called first_five_letters.

Ans:

```
df=pd.read_csv("https://raw.githubusercontent.com/dsrs Scientist/DSData/master/happiness_score_data
set.csv")
pattern=r'(\w{6})'
df['first_five_letters']=df['Country'].str.extract(pattern)
df.sample(15)
```

Question 11- Write a Python program to match a string that contains only upper and lowercase letters , numbers, and underscores.

Ans:

```
string="ImpartofbatchDSC_0523"
pattern=r'\w+$'
result=re.findall(pattern,string)
print(result)
```

output:

```
[ 'ImpartofbatchDSC_0523' ]
```

Question 12- Write a Python program where a string will start with a specific number.

Ans:

```
string1="123xyz a Python program where a string will start with a 256specific number."
pattern1=r'^[0-9]+'
result=re.findall(pattern1,string1)
print(result)
```

output:

```
[ '123' ]
```

Question 13- Write a Python program to remove leading zeros from an IP address

Ans:

1.

```
ip=re.sub(r'[0]', '', 'abc00123xyz456_0')
ip
```

output:

```
'abc123xyz456_'
```

2.

```
ip1=re.sub(r'[0]', '', 'fe80::9ae0:1cb6:614a:522b 2402:8100:27b0:4907:cbc5:50b2:6d81:5039 2402:8100:2
7b0:4907:a8f4:acee:e282:7ba2')
ip1
```

output:

```
'fe8::9ae:1cb6:614a:522b 242:81:27b:497:cbc5:5b2:6d81:539 242:81:27b:4
97:a8f4:acee:e282:7ba2'
```

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.

Ans:

Sampletext="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."

```
pattern=r'(\b[Aa-z])+([0-9a-z]).+(\d+)'
```

```
output=re.search(pattern,Sampletext)
```

output

output:

```
<regex.Match object; span=(3, 19), match='August 15th 1947'>
```

Question 15- Write a Python program to search some literals strings in a string.

Ans:

Sampletext='The quick brown fox jumps over the lazy dog.'

```
pattern=['fox','dog','horse']
```

```
for i in pattern:
```

```
    match_string=re.findall(i,Sampletext)
```

```
    print(match_string)
```

output:

```
[ 'fox' ]
```

```
[ 'dog' ]
```

```
[ ]
```

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

Ans:

Sampletext='The quick brown fox jumps over the lazy dog.'

```
pattern=r'\b(fox)'
```

```
output=re.search(pattern,Sampletext)
```

output

output:

```
regex.Match object; span=(16, 19), match='fox'>
```

Question 17- Write a Python program to find the substrings within a string.

Ans:

Sampletext='Python exercises, PHP exercises, C# exercises'

```
pattern='exercises'
```

```
for i in re.findall(pattern,Sampletext):
```

```
print('detect : ', i)
```

output:

```
detect : exercises
detect : exercises
detect : exercises
```

Question 18- Write a Python program to find the occurrence and position of the substrings within a string.

Ans:

```
Sampletext='Python exercises, PHP exercises, C# exercises'
pattern='exercises'
for i in re.finditer(pattern,Sampletext):
```

```
    print(i)
```

output:

```
<regex.Match object; span=(7, 16), match='exercises'>
<regex.Match object; span=(22, 31), match='exercises'>
<regex.Match object; span=(36, 45), match='exercises'>
```

Question 19- Write a Python program to convert a date of yyyy-mm-dd format to dd-mm-yyyy format.

Ans:

```
def change_date(ch_date):
    pattern=r'(\d{4})-(\d{1,2})-(\d{1,2})'
    convert= re.sub(pattern, '\\3-\\2-\\1', ch_date)
    return convert
date = "2023-12-16"
change=change_date_format(date)
change
```

Output:

```
"16-12-2023"
```

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.

Ans:

```
SampleText='01.12 0132.123 2.31875 145.8 3.01 27.25 0.25'
def find_decimal(string):
    pattern=re.compile(r'\b(\d*\.\d{1,2})\b')
    output=pattern.findall(string)
    return output
txt=find_decimal(SampleText)
txt
```

output:

```
['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.

Ans:

```
text='Suryakumar Yadav scores 100 , 56, 29 in this series'
pattern=r'(\d+)'
check=re.finditer(pattern,text)
for i in check:
    print(i.group())
    print(i)
```

output:

```
100
<regex.Match object; span=(24, 27), match='100'>
56
<regex.Match object; span=(30, 32), match='56'>
29
<regex.Match object; span=(34, 36), match='29'>
```

Question 22- Write a regular expression in python program to extract maximum/largest numeric value from a string.

Ans:

```
sampletext='My marks in each semester are: 947, 896, 926, 524, 734, 950, 642'
pattern=r'\d+'
check=re.findall(pattern,sampletext)
check=np.array(check,dtype=int)
check=max(check)
check
```

output:

```
950
```

Question 23- Create a function in python to insert spaces between words starting with capital letters.

Ans:

```
SampleTex='RegularExpressionIsAnImportantTopicInPython'
```

```
def add_space(string):
```

```
    pattern=r'([A-Z])'
```

```
    space=re.sub(pattern,r' \1',string)
```

```
    return space
```

```
txt=add_space(SampleTex)
```

```
txt
```

output:

```
' 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

Ans:

```
Text="Python regex to Find sequences Of One upper case letter followed by lower case letters"
```

```
pattern=r'[A-Z][a-z]+'
```

```
find=re.findall(pattern,Text)
```

```
find
```

output:

```
['Python', 'Find', 'Of', 'One']
```

Question 25- Write a Python program to remove continuous duplicate words from Sentence using Regular Expression.

Ans:

```
sampletext='Hello hello world world'
```

```
pattern=r'\b(\w+)(\s\1\b)+'
```

```
result=re.sub(pattern,r' \1',sampletext)
```

```
result
```

output:

```
'Hello hello world'
```

Question 26- Write a python program using RegEx to accept string ending with alphanumeric character.

Ans:

```
string="Write a python program using RegEx to accept string ending with alphanumeric character123."
```

```
pattern=r'\w+\.$'
```

```
find=re.findall(pattern,string)
```

```
find
```

output:


```
['character123.']
```

Question 27-Write a python program using RegEx to extract the hashtags.

Ans:

```
SampleText='RT @kapil_kausik: #Doltiwal I mean #xyzabc is "hurt" by #Demonetization as the same has rendered USELESS <ed><U+00A0><U+00BD><ed><U+00B1><U+0089> "acquired funds" No wo'
pattern=r'(\#\w+)'
result=re.findall(pattern,SampleText)
result
```

output:

```
['#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.

Ans:

```
SampleText="@Jags123456 Bharat band on 28??<ed><U+00A0><U+00BD><ed><U+00B8><U+0082>Those who are protesting #demonetization are all different party leaders"
pattern=r'<U\+(\w{4}>)'
result=re.sub(pattern,"",SampleText)
```

result

output:

```
'@Jags123456 Bharat band on 28??<ed><ed>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.

Ans:

```
SampleText="Ron was born on 12-09-1992 and he was admitted to school 15-12-1999."
pattern=r'\b\d{1,2}-\d{1,2}-\d{4}\b|\b\d{4}-\d{1,2}-\d{1,2}\b'
check=re.findall(pattern,SampleText)
check
```

output:

```
['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. b. The use of the re.compile() method is mandatory.

Ans:

SampleText='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.'

```
patt=re.compile(r'\b(\w{2,4})\b')
```

```
text=patt.sub("",SampleText)
```

```
text
```

output:

```
' following example creates ArrayList a capacity elements. 4 elements added ArrayList ArrayList trimmed accordingly.'
```