

Regular Expressions-Assignment

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 [73]: import regex as re
import pandas as pd
```

```
In [74]: sample_text = 'Python Exercises, PHP exercises.'

result = re.sub(r"[\s|,|.]", ":", sample_text)
print(result)
```

Python:Exercises::PHP:exercises:

Question 2- Write a Python program to find all words starting with 'a' or 'e' in a given string.

```
In [75]: string="Data science is the study of data to extract meaningful insights fo

pattern = '\sa\w+\s|\se\w+\s'

result=re.findall(pattern,string)
print(result)
```

[' extract ', ' approach ', ' and ', ' artificial ', ' and ', ' engineerin
g ', ' analyze ', ' amounts ']

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 [76]: text = "APJ Abdul Kalam is a very kind-hearted individual who generously per
pattern_str = r"\b\w{4}\b"
pattern = re.compile(pattern_str)

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

['very', 'kind', 'many']

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 [77]: string = "SQL is a powerful programming language that is used for communicat
pattern = r"\b\w{3}\b|\b\w{4}\b|\b\w{5}\b"
string_pattern = re.compile(pattern)

result = string_pattern.findall(string)
print(result)
```

['SQL', 'that', 'used', 'for', 'with', 'and', 'data', 'types', 'from']

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)"]

```
In [78]: sample = ["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (
pattern_rule = re.compile(r"[A-Za-z]+\s(?:^|)")
for i in sample:

    result = pattern_rule.findall(i)
    print(result)
```

['Data S', 'cience W']

In []:

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)"]

```
In [79]: sample_text = ["example (.com)", "hr@fliprobo (.com)", "github (.com)", "He
for item in sample_text:
    result = re.sub(r" ?\[^\]+\)", "", item)
    print(result)
```

```
example
hr@fliprobo
github
Hello
Data
```

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

Sample text: "ImportanceOfRegularExpressionsInPython"

```
In [80]: text = "ImportanceOfRegularExpressionsInPython"
x = re.findall("[A-Z][^A-Z]*", text)
print(x)
```

```
['Importance', 'Of', 'Regular', 'Expressions', 'In', 'Python']
```

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

Sample Text: "RegularExpression1IsAn2ImportantTopic3InPython"

```
In [81]: x = "RegularExpression1IsAn2ImportantTopic3InPython"

y = re.sub(r"([a-zA-Z]+)(\d+)", r"\1 \2", x)
print(y)
```

```
RegularExpression 1IsAn 2ImportantTopic 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 [82]: x = "RegularExpression1IsAn2ImportantTopic3InPython"

y = re.sub(r"([a-zA-Z]+)(\d+)", r"\1 \2 ", x)
print(y)
```

RegularExpression 1 IsAn 2 ImportantTopic 3 InPython

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 (<mailto:xyz@domain.com>) and alternate email address is xyz.abc@sdomain.domain.com (<mailto:xyz.abc@sdomain.domain.com>). Please contact us at hr@fliprobo.com (<mailto:hr@fliprobo.com>) for further information.

```
In [83]: Sample = "Hello my name is Data Science and my email address is xyz@domain
pattern = r"\b[A-Za-z0-9._]+@[A-Za-z0-9.-]+\.[A-Za-z]{2,}\b"

email = re.findall(pattern, Sample)
print(email)
```

['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.

```
In [84]: target_string = "XyZab569_ is your password for login into the account"
result = re.match(r"[A-Za-z0-9_]+", target_string)

print(result)
```

<regex.Match object; span=(0, 9), match='XyZab569_'>

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

```
In [85]: string = ["8-4546512", "4-5623245", "9-456789"]
pattern = re.compile(r"^4-\d+")
for i in string:
    text = pattern.findall(i)
    print(text)
```

```
[]
['4-5623245']
[]
```

```
In [86]: string = ["8-4546512", "4-5623245", "9-456789"]
pattern = re.compile(r"^9-\d+")
for i in string:
    text = pattern.findall(i)
    print(text)
```

```
[]
[]
['9-456789']
```

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

```
In [87]: ip_address = "11.075.09.36"
text = re.sub(r"[0]", "", ip_address)
print(text)
```

```
11.75.9.36
```

```
In [88]: ip= ["11.075.09.36", "110.120.369.250", "78.520.360.307", "20.140.021.05"]
for i in ip:
    address = re.sub(r"[0]", "", i)
    print(address)
```

```
11.75.9.36
11.12.369.25
78.52.36.37
2.14.21.5
```

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

```
In [89]: text = "On August 15th 1947 that India was declared independent from British
rule = r"\b([A-Z][a-z]+ \d{1,2}(th|nd|st)? \d{4})\b"
date = re.findall(rule, text)
print(date)
```

```
[('August 15th 1947', 'th')]
```

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.'

```
In [90]: text = "The quick brown fox jumps over the lazy dog."
pattern = [ 'fox', 'dog', 'horse' ]
for items in pattern:
    print( 'Searching for "%s"' % (items))
    match = re.search(items,text)
    print(match)

    # if re.search(items,text):
    #     print("Matched !")
    # else:
    #     print("Not Matched !")
```

```
Searching for "fox"
<regex.Match object; span=(16, 19), match='fox'>
Searching for "dog"
<regex.Match object; span=(40, 43), match='dog'>
Searching for "horse"
None
```

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.'

```
In [91]: text = "The quick brown fox jumps over the lazy dog."
pattern = 'fox'

match = re.search(pattern,text)
print(match)
```

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

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

Sample text : 'Python exercises, PHP exercises, C# exercises'

```
In [92]: sample = "Python exercises, PHP exercises, C# exercises"
pattern = "exercises"

match = re.findall(pattern,sample)
print(match)

['exercises', 'exercises', 'exercises']
```

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

```
In [93]: sample = ["Python exercises", "PHP exercises", "C# exercises"]
pattern = "exercises"
for i in sample:

    match = re.search(pattern,i)
    print(match)

<regex.Match object; span=(7, 16), match='exercises'>
<regex.Match object; span=(4, 13), match='exercises'>
<regex.Match object; span=(3, 12), match='exercises'>
```

Its showing the position and occurrence of the substrings within a string

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

```
In [94]: date = "2022-12-23"
pattern = r"(\d{4})-(\d{1,2})-(\d{1,2})"

format = re.sub(pattern,"\\3-\\2-\\1",date)
print(format)

23-12-2022
```

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.

```
In [95]: sample = "01.12 0132.123 2.31875 145.8 3.01 27.25 0.25"
rule = r"\b(\d+\.\d{1,2})\b"
string = re.compile(rule)

match = string.findall(sample)
print(match)

['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 [96]: text = "13 is my lucky number"
pattern = r"\d+"

element = re.search(pattern, text)
print(element)

<regex.Match object; span=(0, 2), match='13'>
```

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

```
In [97]: Sample_text = "My marks in each semester are: 947, 896, 926, 524, 734, 950,"

Maxnum = re.findall("\d+", Sample_text)

Maxnum = map(int, Maxnum)

print("Largest Number: ", max(Maxnum))

Largest Number: 950
```

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

```
In [98]: string = "RegularExpressionIsAnImportantTopicInPython"
pattern = r"(\w)([A-Z])"

match = re.sub(pattern, r"\1 \2", string)
print(match)

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 [99]: list = ['Python', 'PYTHON', 'aB', 'Letters']

pattern = r"[A-Z]+[a-z]+$"

for i in list:
    match = re.search(pattern,i)
    print(match)

<regex.Match object; span=(0, 6), match='Python'>
None
None
<regex.Match object; span=(0, 7), match='Letters'>
```

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

```
In [100]: sample = "Hello hello world world"
pattern = r'\b(\w+)(\s+\1\b)+'

result = re.sub(pattern,r"\1",sample)
print(result)

Hello hello world
```

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

```
In [101]: String = input("Enter the String: ")
pattern = r"\w$"

if(re.search(pattern,String)):
    print("Accept")
else:
    print("Reject")
```

Enter the String: Sumit987888
Accept

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

```
In [102]: text = """RT @kapil_kausik: #Doltiwal I mean #xyzabc is "hurt" by #Demoneti
rule = r"#[A-Za-z]+"

match = re.findall(rule,text)
print(match)
```

['#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.

```
In [103]: sample = "@Jags123456 Bharat band on 28??<ed><U+00A0><U+00BD><ed><U+00B8><U
pattern = r"<U\+\w{4}>"

remove_item = re.sub(pattern,"",sample)
print(remove_item)
```

@Jags123456 Bharat band on 28??<ed><ed>Those who are protesting #demoneti
zation are all different party leaders

Question 29- Write a python program to extract dates from the text stored in the text file.

```
In [104]: with open("C:/Users/DELL/Downloads/Extract_dates") as file:
for line in file:
    pattern = "\d{1,2}-\d{1,2}-\d{4}"
    dates = re.findall(pattern,line)
    print(dates)
```

['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.

```
In [105]: text = "The following example creates an ArrayList with a capacity of 50 ele  
pattern = re.compile(r"\w*\b\w{2,4}\b")  
  
result = pattern.sub("",text)  
print(result)
```

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

In []: