

Solution - Regular Expressions

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:

Solution :

```
import re

def replace_characters(text):
    pattern = r'[ ,.]'
    replacement = ':'
    result = re.sub(pattern, replacement, text)
    return result

given_text = 'Python Exercises, PHP exercises.'
output = replace_characters(given_text)
print(output)
```

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

Solution :

```
string = 'Python Exercises, PHP exercises.'
pattern = r'\b[aeAE]\w+\b'

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

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.

Solution :

```
def four_long_words(text):
    pattern = re.compile(r'\b\w{4,}\b')
    matches = pattern.findall(text)
```

```
sample = 'Python Exercises, PHP exercises.'  
four_long_words(sample)
```

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.

Solution :

```
def words(text):  
    pattern = re.compile(r'\b\w{3,5}\b')  
    matches = pattern.findall(text)  
    return matches  
  
sample_text = 'Python Exercises, PHP exercises.'  
variable = words(sample_text)  
print(variable)
```

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

Solution :

```
def words(strings):  
    pattern = re.compile(r'\((.*?)\)')  
    strings = [pattern.sub('', s) for s in strings]  
    return strings  
  
sample_text = ['example (.com)', 'hr@fliprobo (.com)', 'github (.com)', 'Hello (Data  
Science World)', 'Data (Scientist)']  
list = words(sample_text)  
for strings in list:  
    print(list)
```

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.

Solution :

```
def words(strings):
    pattern = re.compile(r'\((.*?)\)')
    strings = [pattern.sub('', s) for s in strings]
    return strings

sample_text = ['example (.com)', 'hr@fliprobo (.com)', 'github (.com)', 'Hello (Data Science World)', 'Data (Scientist)']
list = words(sample_text)
for strings in list:
    print(list)
```

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']

Solution :

```
string = 'ImportanceOfRegularExpressionsInPython'
pattern = '[A-Z][a-z]*'

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

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

Sample Text: "RegularExpression1IsAn2ImportantTopic3InPython"

Expected Output: RegularExpression 1IsAn 2ImportantTopic 3InPython

Solution :

```
string = 'ImportanceOfRegularExpressionsInPython'
pattern = '([A-Za-z]+)(\d)'
```

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

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

Solution :

```
string = 'RegularExpression1IsAn2ImportantTopic3InPython'
pattern = r'(?<=[a-zA-Z0-9]) (?=[A-Z\d]) '

result = re.compile(pattern)
result1 = result.split(string)
print(result1)
```

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.

Solution :

```
string = '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'
pattern = r'\b[\w\.-]+@[ \w\.-]+\b'

result = re.compile(pattern)
result1= result.findall(string)
print(result1)
```

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

Solution :

```
string = ['WritePythone', 'Code_username', 'Code_username_2023', 'OhitsPythone!!']
pattern = '^[a-zA-Z0-9_]+$'

result = re.compile(pattern)
result1= result.findall(string)
print(result1)
```

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

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

Solution :

```
string = '178.191.001.002'
pattern = r'\b0+(\d+)\b'

result = re.compile(r'\b0+(\d+)\b')
print(result)
```

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.

Solution :

```
string = '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(?:August)\s+\d{1,2}(?:st|nd|rd|th)?\s+\d{4}\b'

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

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'

Solution :

```
string = 'The quick brown fox jumps over the lazy dog.'
match = re.search('fox',string)
match1 = re.search('dog',string)
match2 = re.search('horse',string)
print(match)
print(match1)
print(match2)
```

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'

Solution :

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

match = re.search('fox',string)
print(match)

# Span(16,19) is location for given word.
```

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

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

Pattern : 'exercises'.

Solution :

```
string = 'Python exercises, PHP exercises, C# exercises'

pattern = 'exercises'

replace = 'C#'

sub_and_count = re.subn(pattern,replace,string)

print(sub_and_count)
```

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

Solution :

```
string = 'Python exercises, PHP exercises,exercises'

match = re.search('Python',string)
match1 = re.search('exercises',string)
match2 = re.search('PHP',string)

print(match)
print(match1)
print(match2)
```

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

Solution :

```
string = '2021-09-06'
pattern = r'(\d{4})-(\d{2})-(\d{2})'

result = re.compile(pattern)
print(result)
```

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.

Solution :

```
def decimalNumbers(text):
    pattern = re.compile(r'\b\d+(\.\d{1,2})?\b')
    string = pattern.findall(text)
    return string

text = 'I have rupees of 0.25 0.50 0.75'
decimalNumbers = decimalNumbers(text)
print(decimalNumbers)
```

Question 21- Write a Python program to separate and print the numbers and their position of a given string.

Solution :

```
def string(text):
    pattern = re.compile(r'\b\d+\b')
    find = pattern.finditer(text)
    for match in find:
```

```

        number = match.group()
        position = match.start()
        print(f"{number},{position}")

text = 'I have rupees of 0.25 0.50 0.75'
string(text)

```

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

Solution :

```

string = 'My marks in each semester are: 947, 896, 926, 524, 734, 950, 642'
pattern = r'\b\d+\b'

result = re.compile(pattern)
matches = result.findall(string)
max_num = max(map(int, matches)) #map for converting string to int
print(max_num)

```

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

Solution :

```

string = 'RegularExpressionIsAnImportantTopicInPython'
pattern = r'(?<=[a-z])(?=[A-Z])'

result = re.compile(pattern)
new = result.sub(' ', string)
print(new)

```

Question 24- Python regex to find sequences of one upper case letter followed by lower case letters

Solution :

```

string = 'Regular Expression Is An Important Topic In Python'
pattern = r'[A-Z][a-z]+'

```



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

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

Solution :

```
string = 'Hello hello world world'
pattern = r'\b(\w+) (?:\s+\1)+\b'

result = re.compile(pattern)
new = result.sub(r'\1', string)
print(new)
```

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

Solution :

```
string = 'WritePythone', 'Code_username', 'Code_username_2023', 'OhitsPythone!!'
pattern = r'^.*[a-zA-Z0-9]$$'

result = re.compile(pattern)
for word in string:
    if result.match(word):
        print({word}, 'ends with an alphanumeric character.')
    else:
        print({word}, 'does not end with an alphanumeric 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 <ed><U+00A0><U+00BD><ed><U+00B1><U+0089> "acquired funds" No wo""

Expected Output: ['#Doltiwal', '#xyzabc', '#Demonetization']

Solution :

```
string = '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'
pattern = r'#\w+'

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

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??<ed><U+00A0><U+00BD><ed><U+00B8><U+0082>Those who are protesting #demonetization are all different party leaders"

Expected Output: @Jags123456 Bharat band on 28??<ed><ed>Those who are protesting #demonetization are all different party leaders

Solution :

```
string = '@Jags123456 Bharat band on 28??<U+00A0><U+00BD><U+00B8><U+0082>Those who are protesting #demonetization are all different party leaders'
pattern = r'<U\[A-F0-9\]>+'

result = re.compile(pattern,string)
words = max(map(int, result)) #map for coverting string to int
print(words)
```

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.

Solution :

```
def dates(files):

    with open(files, 'r') as file:
        text = file.read()
        pattern = re.compile(r'\d{2}-\d{2}-\d{4}')
        dates1 = pattern.findall(text)
        return dates1

filename = 'sample_text.txt'
extractedDates = dates(filename)
print(extractedDates)
```

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.

Solution :

```
string = '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.'  
pattern = r'\b\w{2,4}\b'  
  
result = re.compile(pattern)  
new = result.sub(r' ', string)  
print(new)
```