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:

Ans: import re

text = 'Python Exercises, PHP exercises.'

print(re.sub("[,.]", ":", text))

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

Ans import re

Input.

text = "The following example creates an ArrayList with a capacity of 50 elements. Four elements are then added to the ArrayList and the ArrayList is trimmed accordingly."

#find all the words starting with 'a' or 'e'

list = re.findall("[ae]\w+", text)

Print result.

print(list)

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 import re

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

print(re.findall(r"\b\w{4,}\b", text))

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 import re

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

print(re.findall(r"\b\w{3,5}\b", text))

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

Ans. import re

items = ["example.com", "hr@fliprobo.com", "github.com)", "Hello Data Science World (.com)","Data Scientist"]

for item in items:

    print(re.sub(r" ?\([^)]+\)", "", item))

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.

Ans. import re

```
items = ["example (.com)", "hr@fliprobo (.com)", "github (.com)", "Hello (Data Science World)", "Data (Scientist)"]

for item in items:
    print(re.sub(r" ?\([^)]+\)", "", item))
```

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']

Ans. import re

```
text = "ImportanceOfRegularExpressionsInPython"

print(re.findall('[A-Z][^A-Z]*', text))
```

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

Sample Text: “RegularExpression1IsAn2ImportantTopic3InPython"

Expected Output: RegularExpression 1IsAn 2ImportantTopic 3InPython

Ans. import re

```
def capital_words_spaces(str1):
  return re.sub(r"(\w)([A-Z])", r"\1 \2", str1)
print(capital_words_spaces("RegularExpression1"))
print(capital_words_spaces("IsAn2ImportantTopic"))
print(capital_words_spaces("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

Ans. Import re

```
def capital_words_spaces(str1):
      return re.sub(r”(\w) ([A-Z])”, r”\1 \2”,  str1)
print(capital _words_spaces(“Python”))
print(capital_words_spaces(“PythonExcercises”))
print(capital_words_spaces(“PythonExcercisesPracticeSolution”))
```

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.

Ans. import re

```
# Example string

s = """Hello my name is Data Science and my email address is xyz@domain.com and alternate email address is xyz.abc@sdomain.domain.com and hr@fliprobo.com"""


# \S matches any non-whitespace character

# @ for as in the Email

# + for Repeats a character one or more times

lst = re.findall('\S+@\S+', s)


# Printing of List

print(lst)
```

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

Ans.
```
import re
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("The quick brown fox jumps over the lazy dog."))
print(text_match("Python_Exercises_1"))
```

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

Ans.
```
import re
def match_num(string):
   text = re.compile(r"^5")
   if text.match(string):
      return True
   else:
      return False
print(match_num('4-2345861'))
print(match_num('5-2345861'))
```

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

Ans.
```
import re
ip = "218.05.084.186"
string = re.sub('\.[0]*', '.', ip)
print(string)
```

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.

Ans. import re

```
#storing the value of datestring in a variable

datestring = '15-08-1947'

#use re.match() functions to match the datestring

str =re.match('(\d{2})[/.-](\d{2})[/.-](\d{4})$', datestring)

#again declaring the datestring variable with different date format

#printing the str.group()

print ("August 15th 1947", str.group())

datestring = '1947-08-15'

#matching the datestring with re.match() functions.

str=re.match('(\d{2})[/.-](\d{2})[/.-](\d{4})$', datestring)

#printing the str
```

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'

Ans. import re

```
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!')
   else:
      print('Not Matched!')                print (‘Matched!’)
        else:
        print ('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'

Ans import re

```
pattern = 'fox'
text = 'The quick brown fox jumps over the lazy dog.'
match = re.search(pattern, text)
s = match.start()
e = match.end()
print('Found "%s" in "%s" from %d to %d ' % \
   (match.re.pattern, match.string, s, e))
```

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

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

Pattern : 'exercises'.

Ans. import re

```
text = 'Python exercises, PHP exercises, C# exercises'
pattern = 'exercises'
for match in re.findall(pattern, text):
    print('"%s"' % match)
```

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

Ans. import re

```
text = 'Python exercises, PHP exercises, C# exercises'
pattern = 'exercises'
for match in re.finditer(pattern, text):
    s = match.start()
    e = match.end()
    print('Found "%s" at %d:%d' % (text[s:e], s, e))
```

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

Ans. import re

```
def change_date_format(dt):
        return re.sub(r'(\d{4})-(\d{1,2})-(\d{1,2})', '\\3-\\2-\\1', dt)
dt1 = "2026-01-02"
print("Original date in YYY-MM-DD Format: ",dt1)
print("New date in DD-MM-YYYY Format: ",change_date_format(dt1))
```

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']

Ans. import re

```
def find_decimal_numbers(string):
  pattern = re.compile(r'\d+\.\d{1,2}')
  decimal_numbers = re.findall(pattern, string)
  return decimal_numbers
sample_text = "01.12 0132.123 2.31875 145.8 3.01 27.25 0.25"
output = find_decimal_numbers(sample_text)
print(output)
```

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

Ans. import re

```
# Input.
text = "The following example creates an ArrayList with a capacity of 50 elements. Four elements are then added to the ArrayList and the ArrayList is trimmed accordingly."

for m in re.finditer("\d+", text):
    print(m.group(0))
    print("Index position:", m.start())
```

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

Ans. import re

```
#input
string='947, 896, 926, 524, 734, 950, 642'
#seperate number from string
number = re.findall('\d+', string)
#convert it into integer
number = map(int, number)
print("Max_value:",max(number))
```

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

Ans.

```
import re

def putSpace(input):

    # regex [A-Z][a-z]* means any string starting
    # with capital character followed by many
    # lowercase letters
    words = re.findall('[A-Z][a-z]*', input)

    # Change first letter of each word into lower
    # case
    for i in range(0,len(words)):
      words[i]=words[i][0].lower()+words[i][1:]
    print(' '.join(words))


# Driver program
if __name__ == "__main__":
    input = 'RegularExpressionIsAnImportantTopicInPython'
    putSpace(input)
     print(‘ ‘.join(words))
#Driver program
If _name_==”_main_”:
    input = ‘BruceWayneIsBatman’
   putSpace(input)
```

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

Ans. import re

```
def text_match(text):
        patterns = '[A-Z]+[a-z]+$'
        if re.search(patterns, text):
                return 'Found a match!'
        else:
                return('Not matched!')
print(text_match("AaBbGg"))
print(text_match("Python"))
print(text_match("python"))
print(text_match("PYTHON"))
print(text_match("aA"))
print(text_match("Aa"))
```

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

Ans def unique_list(text_str):

```
   l = text_str.split()
   temp = []
   for x in l:
      if x not in temp:
```

```
        temp.append(x)
    return ' '.join(temp)

text_str = "Hello hello world world"
print("Original String:")
print(text_str)
print("\nAfter removing duplicate words from the said string:")
print(unique_list(text_str))
```

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

```
Ans import re
def check(ip_str):
    re_exp = '[a-zA-z0-9]$'
    if(re.search(re_exp, ip_str)):
        return "The string is ending with alphanumeric char!"

    else:
        return "The string does not ends with alphanumeric char!"

ip_str = input("Enter the string: ")
print(check(ip_str))
```

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']

Ans. # program to print all the hashtags in a text

```
text1 = """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"""

textList = text1.split()

for i in textList:

  if(i.startswith("#")):

    x = i.replace("#", '')

    print(x)
```

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

Ans. import re

```
input_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"

pattern = r"<U\+\w{4}>"
```

```
output_text = re.sub(pattern, "", input_text)

print(output_text)
```

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.

Ans.

```
# Python3 code to demonstrate working of
# Detect date in String
# Using python-dateutil()
from dateutil import parser

# initializing string
test_str = "Ron was born on 12-09-1992 and he was admitted to school 15-12-1999"

# printing original string
print("The original string is : " + str(test_str))

# extracting date using inbuilt func.
```

```
res = parser.parse(test_str, fuzzy=True)


# printing result

print("Computed date : " + str(res)[:10])
```

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.

Ans.

```
import re

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."

# remove words between 2 and 4

shortword = re.compile(r'\W*\b\w{1,3}\b')

print(shortword.sub('', text))
```