

# 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:**

```
text= "Python Exercises, PHP exercises."  
x=re.sub("\s|,|",":",text)  
print (x)
```

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

**Ans:**

```
re.findall("[a,e]\w+",text)
```

**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:**

```
target_string= "Rohit Sharma has scored 43 CENTURIES and 91 HALFCENTURIES in his cricket career "  
pattern=re.compile(r"\b\w{4,}\b")  
for match in pattern.finditer(target_string):  
    print (match.group())
```

**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:**

```
target_string= "A rose by any other name would smell as sweet. "  
pattern=re.compile(r"\b\w{3,5}\b")
```

```
for match in pattern.finditer(target_string):
```

```
    print (match.group())
```

**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:**

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

```
for i in text:
```

```
    print (re.sub(r" ?\(", " ", i))
```

**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:**

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

```
for i in text
print (re.sub(r" ?\([^\)]+\)" , " " , i))
```

**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:**

```
text2 = "ImportanceOfRegularExpressionsInPython"
match=re.findall('[A-Z][^A-Z]*',text2)
print (match)
```

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

**Sample Text:** "RegularExpression1IsAn2ImportantTopic3InPython"

**Expected Output:** 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

**Ans:**

```
text="RegularExpression1IsAn2ImportantTopic3InPython"
pattern = r'(\d{1})([A-Z]+)'
result =re.sub (pattern, r'\1 \2' , text)
print (result)
```

**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.

**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:**

```
file=open ('data.txt')
for line in file:
    pattern= '[a-zA-Z0-9._%~]+@[a-zA-Z0-9]+\.[a-zA-Z]{2,}'
    email=re.findall(pattern,line)
    print (email)
```

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

**Ans:**

```
Pattern= "^[A-Za-z0-9_]*"
```

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

**Ans:**

```
Pattern = r"^5"
```

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

**Ans:**

```
ip="245.07.019.196"
```

```
string=re.sub(r"\.[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.

**Ans:**

```
pattern= r"(\d{2})([.-])(\d{2})([.-])(\d{4})$"
```

**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:**

```
file=open ('data 2.txt')
```

```
for i in file:
```

```
pattern= r"\b([A-Z][a-z]+ \d{1,2}(:st|nd|rd|th)? \d{4})\b"
```

```
match=re.match (pattern,i)
```

```
print (match)
```

**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'

**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:**

```
pattern = 'fox'
text = 'The quick brown fox jumps over the lazy dog.'
match = re.search(pattern, text)
A=match.start()
B=match.end()
print (A,B)
```

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

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

**Pattern :** 'exercises'.

**Ans:**

```
text='Python exercises, PHP exercises, C# exercises'
Pattern ='exercises'
for i in re.findall (Pattern,text):
    print (i)
```

**Question 18-** Write a Python program to find the occurrence and position 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.

**Ans:**

```
re.sub('(\d{4})-(\d{1,2})-(\d{1,2})', '(\d{1,2})-(\d{1,2})-(\d{4})',date_text)
```

**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']

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

**Ans:**

```
text="Every day i used to take 4 bananas & 2 apples"
for i in re.finditer("\d+",text):
    print (i.group(0))
    print ("Index position:", i.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:**

```
text="My marks in each semester are: 947, 896, 926, 524, 734, 950, 642"
number=re.findall("\d+",text)
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:**

```
text="RegularExpressionIsAnImportantTopicInPython"
pattern = r'(\w)([A-Z]+)'
result =re.sub (pattern, r'\1 \2', text)
print (result)
```

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

**Ans:**

```
Pattern= [A-Z]+[a-z].*
```

**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:**

```
sample="Hello hello world world"
pattern= r'(\w+) (?:\w+\1\b)+'
replace=re.sub(pattern,r'\1',sample)
print (replace)
```

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

**Ans:**

```
r'[A-Za-z0-9]$\'
```

**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:**

```
def extract_text(text):
    sample = re.findall(r'#\w+',text)
    print (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'
print (extract_text(text))
```



**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:**

```
file = open ('data3.txt')
pattern= r'<U\[A-Z0-9]{2}>'
for i in file:
    replace= re.sub (pattern, "",i)
    print (i)
```

**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:**

```
pattern= r'(\d{1,2})(\.-)(\d{1,2})(\.-)(\d{4})'
```

**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:**

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

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"

```
print (pattern.sub(",text))
```