

Regular Expressions

September 9, 2019

Problem 1: Return the first word of a given string

```
In [1]: import re
        demo_string = "always learning CSE1001 is really fun"
```

```
In [2]: result=re.findall(r'.',demo_string)
        print(result)
```

```
['a', 'l', 'w', 'a', 'y', 's', ' ', 'l', 'e', 'a', 'r', 'n', 'i', 'n', 'g', ' ', 'C', 'S', 'E', '1', '0', '0', '1', ' ', 'i', 's', ' ', 'r', 'e', 'a', 'l', 'l', 'y', ' ', 'f', 'u', 'n', '']
```

```
In [3]: result=re.findall(r'\w',demo_string)
        print(result)
```

```
['a', 'l', 'w', 'a', 'y', 's', 'l', 'e', 'a', 'r', 'n', 'i', 'n', 'g', 'C', 'S', 'E', '1', '0', '0', '1', 'i', 's', 'r', 'e', 'a', 'l', 'l', 'y', 'f', 'u', 'n']
```

```
In [4]: result=re.findall(r'\w*',demo_string)
        print(result)
```

```
['always', ' ', 'learning', ' ', 'CSE1001', ' ', 'is', ' ', 'really', ' ', 'fun', '']
```

```
In [5]: result=re.findall(r'\w+',demo_string)
        print(result)
```

```
['always', 'learning', 'CSE1001', 'is', 'really', 'fun']
```

```
In [6]: result=re.findall(r'^\w+',demo_string)
        print(result)
```

```
['always']
```

Problem 2: Return the last word of a given string

```
In [7]: result=re.findall(r'\w+$',demo_string)
        print(result)
```

```
['fun']
```

Problem 3: Return the first two character of each word

```
In [8]: result=re.findall(r'\w\w',demo_string)
        print(result)
```

```
['al', 'wa', 'ys', 'le', 'ar', 'ni', 'ng', 'CS', 'E1', '00', 'is', 're', 'al', 'ly', 'fu']
```

```
In [9]: result=re.findall(r'\b\w\w',demo_string)
        print(result)
```

```
['al', 'le', 'CS', 'is', 're', 'fu']
```

Problem 4: Return the domain type of given email-ids

```
In [10]: demo_string = 'abcd@gmail.com, xyx@ac.in, pqrs@yahoo.biz.in'
```

```
In [11]: result=re.findall(r'@\w+',demo_string)
         print(result)
```

```
['@gmail', '@ac', '@yahoo']
```

```
In [12]: result=re.findall(r'@\w+.\w+',demo_string)
         print(result)
```

```
['@gmail.com', '@ac.in', '@yahoo.biz']
```

```
In [13]: result=re.findall(r'@\w+.(\\w+)',demo_string)
         print(result)
```

```
['com', 'in', 'biz']
```

```
In [14]: result=re.findall(r'@\\w+.(\\w+)',demo_string)
         print(result)
```

```
['com', 'in', 'biz']
```

Problem 5: extract the date from the given string

```
In [15]: demo_string = 'PQR 34-3456 12-05-2007, XYZ 56-4532 11-11-2011, ABC 67-8945 31-01-2009'
```

```
In [16]: result=re.findall(r'\d{2}-\d{2}-\d{4}',demo_string)
         print(result)
```

```
['12-05-2007', '11-11-2011', '31-01-2009']
```

to extract only day from the date

```
In [17]: result=re.findall(r'(\d{2})-\d{2}-\d{4}',demo_string)
         print(result)
```

```
['12', '11', '31']
```

to extract only month from the date

```
In [18]: result=re.findall(r'\d{2}-\d{2}-\d{4}',demo_string)
         print(result)
```

```
['05', '11', '01']
```

to extract only the year from the date

```
In [19]: result=re.findall(r'\d{2}-\d{2}-\d{4}',demo_string)
         print(result)
```

```
['2007', '2011', '2009']
```

Problem 6: extract all the words of a string which starts with vowel

```
In [20]: demo_string = "always learning CSE1001 is really fun"
```

```
In [21]: result=re.findall(r'[aeiouAEIOU]\w+',demo_string)
         print(result)
```

```
['always', 'earning', 'E1001', 'is', 'eally', 'un']
```

```
In [22]: result=re.findall(r'\b[aeiouAEIOU]\w+',demo_string)
         print(result)
```

```
['always', 'is']
```

Problem 7: extract all the words of a string which starts with consonant

```
In [23]: result=re.findall(r'\b[^aeiouAEIOU ]\w+',demo_string)
         print(result)
```

```
['learning', 'CSE1001', 'really', 'fun']
```

Problem 8: Validate a phone number (phone number must be of 10 digits and starts with 6 or 7 or 8 or 9)

```
In [24]: phone_no='9999999999, 8675724402, 5434215789'  
        result = re.findall(r'[6-9]{1}[0-9]{9}', phone_no)  
        print(result)
```

```
['9999999999', '8675724402']
```

Problem 9: Split a string with multiple delimiters

```
In [25]: demo_string = line = 'tulasi,prasad;sariki,vit chennai'  
        result= re.split(r'[;,\s]', demo_string)  
        print(result)
```

```
['tulasi', 'prasad', 'sariki', 'vit', 'chennai']
```