

# Regular Expression - RE

18-04-2023

```
In [1]: 1 # regular expression = pattern and validation = use to data validate and insert in database after validat
2 # first you have to import RE(inbuild module) in your program
3 # metacharacter = to create pattern = / ? * ^
4 # \n = new line
5 # \t = tab
6 # r = row string
7 # web scrapping is the use of regular expression
8 #
```

```
In [2]: 1 import re
2 s1="Data science is the study of data to extract meaningful insights for business. It is a multidisciplin
3 print(s1)
```

Data science is the study of data to extract meaningful insights for business. It is a multidisciplinary approach that combines principles and practices from the fields of mathematics, statistics, artificial intelligence, and computer engineering to analyze large amounts of data. This analysis helps data scientists to ask and answer questions like what happened, why it happened, what will happen, and what can be done with the results.

```
In [6]: 1 #re.search()
2 pattern="business"
3 m1=re.search(pattern,s1)
4 print(m1)
5 print(m1.group()) # for finding the value which we have to find
```

```
<re.Match object; span=(69, 77), match='business'>
business
```

```
In [8]: 1 #re.match() # match starting word
        2 p1="business"
        3 m1=re.match(p1,s1)
        4 print(m1)
```

None

```
In [9]: 1 p1="Data science"
        2 m1=re.match(p1,s1)
        3 print(m1)
```

<re.Match object; span=(0, 12), match='Data science'>

```
In [10]: 1 s2="python python tableau"
         2 p1="python"
         3 p2="tableau"
         4 m1=re.search(p1,s2)
         5 print(m1)
         6 m2=re.search(p2,s2)
         7 print(m2)
```

<re.Match object; span=(0, 6), match='python'>  
<re.Match object; span=(14, 21), match='tableau'>

```
In [11]: 1 m1=re.match(p1,s2)
         2 print(m1)
         3 m2=re.match(p2,s2)
         4 print(m2)
```

<re.Match object; span=(0, 6), match='python'>  
None

```
In [13]: 1 #re.findall() # always return the values in the form of list
         2 m3=re.findall(p1,s2)
         3 print(m3)
```

['python', 'python']

```
In [15]: 1 s3="My roll no is 12"
          2 p1="\d" #metacharacter = \d = for digit
          3 match=re.findall(p1,s3)
          4 print(match)
```

```
['1', '2']
```

```
In [16]: 1 s3="My roll no is 12"
          2 p1="\d\d" #metacharacter = \d = for single digit
          3 match=re.findall(p1,s3)
          4 print(match)
```

```
['12']
```

```
In [17]: 1 s4="Product prices are 10 20 30 250 120 10"
          2 p1="\d\d"
          3 m4=re.findall(p1,s4)
          4 print(m4)
```

```
['10', '20', '30', '25', '12', '10']
```

```
In [18]: 1 s4="Product prices are 10 20 30 250 120 10"
          2 p1="\d\d\d"
          3 m4=re.findall(p1,s4)
          4 print(m4)
```

```
['250', '120']
```

```
In [22]: 1 s5="My name is Ankita and My Roll no is 12"
          2 p1="[A-Z][a-z]*" # metacharacter = * = rest of all a-z
          3 m5=re.findall(p1,s5)
          4 print(m5)
```

```
['My', 'Ankita', 'My', 'Roll']
```

```
In [24]: 1 s6="9898989899 1234567890 7766554890"
2 p1="\d{10}"
3 m6=re.findall(p1,s6)
4 print(m6)

['9898989899', '1234567890', '7766554890']
```

```
In [25]: 1 s6="9898989899 1234567890 77665"
2 p1="\d{10}"
3 m6=re.findall(p1,s6)
4 print(m6)

['9898989899', '1234567890']
```

```
In [27]: 1 text="today is Apr 18, 2023 and tommarow call me at 234 567-8763 or 234-567-8763"
2 p1="\d{3}-\d{3}-\d{4}"
3 m6=re.findall(p1,text)
4 print(m6)

['234-567-8763']
```

```
In [29]: 1 text="today is Apr 18, 2023 and tommarow call me at 234 567-8763 or 234-567-8763"
2 p1="\d{3} \d{3}-\d{4}"
3 m6=re.findall(p1,text)
4 print(m6)

['234 567-8763']
```

```
In [31]: 1 p2="\d{3}[\s-]\d{3}-\d{4}" #metacharacter = \s= space ,
2 m7=re.findall(p2,text) # for or condition we use [\s-]
3 print(m7)

['234 567-8763', '234-567-8763']
```

```
In [37]: 1 msg="Available numbers are +91 8765453280 and +1 (821)-654-9876"
          2 a1="[+]\d{2}[\s]\d{10}"
          3 x1=re.findall(a1,msg)
          4
          5 a2="[+]\d[\s][()]\d{3}[]]-\d{3}-\d{4}"
          6 x2=re.findall(a2,msg)
          7 print(x1)
          8 print(x2)
```

```
['+91 8765453280']
```

```
['+1 (821)-654-9876']
```

```
In [38]: 1 msg="Available numbers are +91 8765453280 and +1 (821)-654-9876"
          2 a1="[+]\d{2}[\s]\d{10}|[+]\d[\s][()]\d{3}[]]-\d{3}-\d{4}" # for this we use |
          3 x1=re.findall(a1,msg)
          4
          5 print(x1)
          6
```

```
['+91 8765453280', '+1 (821)-654-9876']
```

```
1 # 19-04-2023
```

```
In [1]: 1 import re
          2 print("cs\txyz") # \t = tab
```

```
cs      xyz
```

```
In [2]: 1 import re
          2 print(r"cs\txyz") # r = row string
```

```
cs\txyz
```

In [4]:

```
1 import re
2 text="05/3/2017 3/01/2017 1/6/17 34/11/937 may 21,2017 21st mar 2017"
3 day=r"0?[0-9]|1?[0-9]|2?[0-9]|3?[01]" # / for multiple or conditions
4 m1=re.findall(day,text) # this program only for date
5 print(m1)
```

```
['05', '3', '2', '01', '7', '3', '01', '2', '01', '7', '1', '6', '1', '7', '3', '4', '1', '1', '9', '3', '7', '2', '1', '2', '01', '7', '2', '1', '2', '01', '7']
```

In [5]:

```
1 text="05/3/2017 3/01/2017 1/6/17 34/11/937 may 21,2017 21st mar 2017"
2 month=r"0?[0-9]|1?[0-2]"
3 m1=re.findall(month,text)
4 print(m1)
```

```
['05', '3', '2', '01', '7', '3', '01', '2', '01', '7', '1', '6', '1', '7', '3', '4', '1', '1', '9', '3', '7', '2', '1', '2', '01', '7', '2', '1', '2', '01', '7']
```

In [7]:

```
1 text="05/3/2017 3/01/2017 1/6/17 34/11/1937 may 21,2015 21st mar 1997"
2 year=r"2?[0-9][0-9][0-9]"
3 m2=re.findall(year,text)
4 print(m2)
```

```
['2017', '2017', '1937', '2015', '1997']
```

In [9]:

```
1 text="05/3/2017 3/01/2017 1/6/17 34/11/1937 may 21,2015 21st mar 1997"
2 year="[12]\\d\\d\\d"
3 m3=re.findall(year,text)
4 print(m3)
```

```
['2017', '2017', '1937', '2015', '1997']
```

```
In [12]: 1 text='05/3/2017 3/01/2017 1/6/17 34/11/937 may 21,2017 21st mar 2017'
2 day=r'(\d{1,2}|\d{1,2}/\d{1,2})'
3 month=r'(\d{1,2}|\d{1,2}/\d{1,2})'
4 year=r'(\d{4}|\d{4}/\d{2}/\d{2})'
5 sep=r'/'
6 pattern=r'((\d{1,2}|\d{1,2}/\d{1,2})/(\d{1,2}|\d{1,2}/\d{1,2})/(\d{4}|\d{4}/\d{2}/\d{2}))'
7 m4=re.findall(pattern,text)
8 print(m4)
```

```
[('05/3/2017', '05', '3', '2017'), ('3/01/2017', '3', '01', '2017')]
```

```
In [23]: 1 x=(input("enter number"))
2 phno="[7-9]\d{9}"
3 m1=re.fullmatch(phno,x)
4 if m1!=None:
5     print("valid mobile number")
6 else:
7     print("Invalid mibile number")
```

```
enter number9764565251
```

```
valid mobile number
```

```
In [22]: 1 n=(input("enter number"))
2 p1="[7-9]\d{9}"
3 m1=re.fullmatch(p1,n)
4 if m1!=None:
5     print("valid mobile number")
6 else:
7     print("Invalid mibile number")
```

```
enter number7066738205
```

```
valid mobile number
```

```
In [ ]: 1
```