In [ ]:
```
# Regex ------> Regular expresssion


Regex is a sequence of characters that form a search patthern

regular expression is a set of characters which we can use to find and match of a group of the charatcers
```

In [1]:
```
import re
```

In [2]:
```
text= "I love my country"

patteren = "country"

patteren_value = re.compile(patteren)

s = patteren_value.search(text)
print(s)
```

```
<re.Match object; span=(10, 17), match='country'>
```

In [3]:
```
patteren_value = re.compile("country")
s1 = patteren_value.search(text)
print(s1)
print(s1.start())
print(s1.end())
print(s1.span())
```

```
<re.Match object; span=(10, 17), match='country'>
10
17
(10, 17)
```

In [4]:
```
print(re.compile("country").search(text))
```

```
<re.Match object; span=(10, 17), match='country'>
```

In [5]:
```
e produced over 800 million mobile phones to date.[21] The company grouped them together under Samsung Electronics in the 1980s
```

```
['Samsung', 'Samsung', 'Samsung']
```

In [6]:
```
f = re.finditer("Samsung",sentence)
for i in f:
    print(i)
```

```
<re.Match object; span=(9, 16), match='Samsung'>
<re.Match object; span=(243, 250), match='Samsung'>
<re.Match object; span=(384, 391), match='Samsung'>
```

In [7]:
```
f = re.split(" ",sentence)
print(f)
```

```
['In', '1980,', 'Samsung', 'acquired', 'the', 'Gumi-based', 'Hanguk', 'Jeonja', 'Tongsin', 'and', 'entered', 'telecommunicatio
ns', 'hardware.', 'Its', 'early', 'products', 'were', 'switchboards.', 'The', 'facility', 'was', 'developed', 'into', 'the',
'telephone', 'and', 'fax', 'manufacturing', 'systems', 'and', 'became', 'the', 'center', 'of', "Samsung's", 'mobile', 'phone',
'manufacturing.', 'They', 'have', 'produced', 'over', '800', 'million', 'mobile', 'phones', 'to', 'date.[21]', 'The', 'compan
y', 'grouped', 'them', 'together', 'under', 'Samsung', 'Electronics', 'in', 'the', '1980s.']
```

In [8]:
```
f = re.sub("Samsung","CNC Web World",sentence)
print(f)
```

```
In 1980, CNC Web World acquired the Gumi-based Hanguk Jeonja Tongsin and entered telecommunications hardware. Its early produc
ts were switchboards. The facility was developed into the telephone and fax manufacturing systems and became the center of CNC
Web World's mobile phone manufacturing. They have produced over 800 million mobile phones to date.[21] The company grouped the
m together under CNC Web World Electronics in the 1980s.
```

In [9]:
```
# Secial Characyters
\d--Digit (0-9)
\D-- Not a digit
\w
\W
\s
\S
^--Begining
$ -- ENd of the string
```

```
  Cell In[9], line 2
    \d--Digit (0-9)
                   ^
SyntaxError: unexpected character after line continuation character
```

In [11]:
```
sentence
```

Out[11]:
```
"In 1980, Samsung acquired the Gumi-based Hanguk Jeonja Tongsin and entered telecommunications hardware. Its early products we
re switchboards. The facility was developed into the telephone and fax manufacturing systems and became the center of Samsun
g's mobile phone manufacturing. They have produced over 800 million mobile phones to date.[21] The company grouped them togeth
er under Samsung Electronics in the 1980s."
```

In [12]:
```
d = re.findall('\d',sentence)
print(d)
```

```
['1', '9', '8', '0', '8', '0', '0', '2', '1', '1', '9', '8', '0']
```

In [13]:
```
d = re.findall('\d+',sentence)
print(d)
```

```
['1980', '800', '21', '1980']
```

In [15]:
```
d = re.findall('\w+',sentence)
print(d)
```

```
['In', '1980', 'Samsung', 'acquired', 'the', 'Gumi', 'based', 'Hanguk', 'Jeonja', 'Tongsin', 'and', 'entered', 'telecommunicat
ions', 'hardware', 'Its', 'early', 'products', 'were', 'switchboards', 'The', 'facility', 'was', 'developed', 'into', 'the',
'telephone', 'and', 'fax', 'manufacturing', 'systems', 'and', 'became', 'the', 'center', 'of', 'Samsung', 's', 'mobile', 'phon
e', 'manufacturing', 'They', 'have', 'produced', 'over', '800', 'million', 'mobile', 'phones', 'to', 'date', '21', 'The', 'com
pany', 'grouped', 'them', 'together', 'under', 'Samsung', 'Electronics', 'in', 'the', '1980s']
```

In [17]:
```
d = re.findall('\W+',sentence)
print(d)
```

```
[' ', ', ', ' ', ' ', ' ', '-', ' ', ' ', ' ', ' ', ' ', ' ', ' ', '. ', ' ', ' ', ' ', ' ', '. ', ' ', ' ', ' ', ' ', ' ', '
', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', "'", ' ', ' ', ' ', ' ', '. ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', '.
[', '] ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', '.']
```

In [ ]:
```
### Metacharacters

[] --->  a collection of characters

\  ----> A specific sequence is signaled

. ---->  any charcter

^ ---> begin with

$---> End With

* ---->zero or more occurances

+ -- one or more occurances

{} --> Extacly th especified number of occuracne
```

In [22]:
```
sent = "Mumbai is the financial capital of India, and its pin code started from 232022"

x = re.findall("[a-zA-Z]+",sent)
print(x)
```

```
['Mumbai', 'is', 'the', 'financial', 'capital', 'of', 'India', 'and', 'its', 'pin', 'code', 'started', 'from']
```

In [23]:
```
sent = "Mumbai is the financial capital of India, and its pin code started from 232022"

x = re.findall("[0-9]+",sent)
print(x)
```

```
['232022']
```

In [26]:
```
sent = "Mumbai is the financial capital of India, and its pin code started from 232022"

x = re.findall(".",sent)
print(x)
```

```
['M', 'u', 'm', 'b', 'a', 'i', ' ', 'i', 's', ' ', 't', 'h', 'e', ' ', 'f', 'i', 'n', 'a', 'n', 'c', 'i', 'a', 'l', ' ', 'c',
'a', 'p', 'i', 't', 'a', 'l', ' ', 'o', 'f', ' ', 'I', 'n', 'd', 'i', 'a', ',', ' ', 'a', 'n', 'd', ' ', 'i', 't', 's', ' ',
'p', 'i', 'n', ' ', 'c', 'o', 'd', 'e', ' ', 's', 't', 'a', 'r', 't', 'e', 'd', ' ', 'f', 'r', 'o', 'm', ' ', '2', '3', '2',
'0', '2', '2']
```

In [30]:
```
sent = "Mumbai is the financial capital of India, and its pin code started from 232022"

x = re.findall("^Mumbai",sent)
print(x)
```

```
['Mumbai']
```

In [33]:
```
sent = "Mumbai is the financial capital of India, and its pin code started from 232022"

x = re.findall("232022$",sent)
print(x)
```

```
['232022']
```

In [45]:
```
sent = "Mumbai is the financial capital of India, and its pin code started from 232022"

x = re.findall("India.{30}",sent)
print(x)
```

```
['India, and its pin code started fro']
```

In [47]:
```
sent = "Mumbai is the financial capital of India, and its pin code started from 232022"

x = re.findall("Indi*a",sent)
print(x)
```

```
['India']
```

In [58]:
```
sent = "Mumbai is the financial capital of India, and its pin code started from 232022, @gmail.com"

x = re.findall("[@]+\w+.\w+",sent)
print(x)
```

```
['@gmail.com']
```

In [ ]:
```

```