Regular Expressions

- For checking available data valid for us or not
- To extract the data

In [51]:
```
import re

```

re.search('str_pattern',str_data)

In [55]:
```
name = 'python program'
print(re.search('a',name))

```

```
<re.Match object; span=(12, 13), match='a'>
```

In [54]:
```
name = 'python program'
print(re.search('p',name))

```

```
<re.Match object; span=(0, 1), match='p'>
```

In [56]:
```
name = 'python program'
print(re.search('z',name))

```

```
None
```

print the name which contains 'u' in their name by using regular expression

In [60]:
```
name = ['raja','lakshmi','raju']
for i in name:
    if re.search('u',i):
        print(i)
```

```
raju
```

re.match(str_pattern,str_data)

In [61]:
```
name = 'python program'
print(re.match('p',name))
```

```
<re.Match object; span=(0, 1), match='p'>
```

In [62]:
```
name = 'python program'
print(re.match('r',name))
```

```
None
```

In [63]:
```
1 name = 'python program'
2 print(re.match('py',name))
```

```
<re.Match object; span=(0, 2), match='py'>
```

In [64]:
```
1 name = 'python program'
2 print(re.match('pr',name))
```

```
None
```

re.findall(str_pattern,str_data)

In [65]:
```
1 name = 'python program'
2 print(re.findall('p',name))
```

```
['p', 'p']
```

In [66]:
```
1 name = 'python program'
2 print(re.findall('z',name))
```

```
[]
```

Special characters with special meaning in regular in regular expression

. dot

    any single character

In [67]:
```
1 print(re.search('.','raja'))
```

```
<re.Match object; span=(0, 1), match='r'>
```

In [68]:
```
1 print(re.search('.',''))
```

```
None
```

In [69]:
```
1 print(re.search('r.','raja'))
```

```
<re.Match object; span=(0, 2), match='ra'>
```

In [70]:
```
1 print(re.search('j.','raja'))
```

```
<re.Match object; span=(2, 4), match='ja'>
```

In [71]:
```
1 print(re.search('m.','rajam'))
```

```
None
```

In [73]:
```
print(re.match('r.','rajam'))
```

```
<re.Match object; span=(0, 2), match='ra'>
```

In [75]:
```
print(re.match('a.','rajam'))
```

```
None
```

In [76]:
```
print(re.search('r...','rajam'))
```

```
<re.Match object; span=(0, 4), match='raja'>
```

In [78]:
```
numbers = ['12','123','1234','12345']
for i in numbers:
    if re.search('...',i):
        print(i)
```

```
123
1234
12345
```

^

Srarting with

In [80]:
```
re.search('^R','Raja')
```

Out[80]:
```
<re.Match object; span=(0, 1), match='R'>
```

In [81]:
```
re.search('^r','Raja')
```

$

ENding with

In [82]:
```
re.search('s$','Rajas')
```

Out[82]:
```
<re.Match object; span=(4, 5), match='s'>
```

In [83]:
```
re.search('a$','Rajas')
```

In [85]:
```
numbers = ['12','123','1234','12345']
for i in numbers:
    if re.search('^...$',i):
        print(i)
```

```
123
```

In [88]:
```
numbers = ['12','123','1234','12345','163','124']
for i in numbers:
    if re.search('^1.3$',i):
        print(i)
```

```
123
163
```

\d

any single digit

In [89]:
```
re.search('\d','Rajas3')
```

Out[89]:
```
<re.Match object; span=(5, 6), match='3'>
```

In [91]:
```
re.search('\d','Raja1s2')
```

Out[91]:
```
<re.Match object; span=(4, 5), match='1'>
```

In [92]:
```
re.search('\d','Rajas')
```

In [95]:
```
m = 'every 1 idgc 100 djch'
n = re.findall('\d',m)
n
```

Out[95]:
```
['1', '1', '0', '0']
```

In [96]:
```
map(int,n)
```

Out[96]:
```
<map at 0x23b5b1d4ec8>
```

In [100]:
```
r = list(map(int,n))
r
```

Out[100]:
```
[1, 1, 0, 0]
```

In [99]:
```
sum(r)
```

Out[99]:
```
2
```

In [102]:
```
# or
s = 'every 1 idgc 100 djch'
sum(list(map(int,re.findall('\d',s))))
```

Out[102]:
```
2
```

In [103]:
```
m = 'every 300 299 idgc 100 djch'
re.findall('\d\d\d',m)
```

Out[103]:
```
['300', '299', '100']
```

{min,max}

In [104]:
```
1 re.search('.{2}','Raja')
```

Out[104]:
```
<re.Match object; span=(0, 2), match='Ra'>
```

In [106]:
```
1 re.search('.{5}','Raja')
```

In [107]:
```
1 re.search('.{2,4}','Raja')
```

Out[107]:
```
<re.Match object; span=(0, 4), match='Raja'>
```

In [108]:
```
1 re.search('.{2,4}','a')
```

In [109]:
```
1 re.search('.{2,4}','Rajar')
```

Out[109]:
```
<re.Match object; span=(0, 4), match='Raja'>
```

In [110]:
```
1 m = 'hiedh 1 jncnd 23 kjnkvnf 554 jfvifhv 3333'
```

In [112]:
```
1 re.findall('\d{1,10}',m)
```

Out[112]:
```
['1', '23', '554', '333']
```

accept names which starts with either 'a','c','h'

[each]

either a or c or h

In [114]:
```
1 re.search('[ach]',"Raja")
```

Out[114]:
```
<re.Match object; span=(1, 2), match='a'>
```

In [115]:
```
1 re.search('[ch]',"Raja")
```

In [116]:
```
1 re.search('[a-k]',"Raja")
```

Out[116]:
```
<re.Match object; span=(1, 2), match='a'>
```

In [117]:
```
1 re.search('[a-h]',"lmkm")
```

[^a-z]

neither a nor b nor c . . . . .

In [118]:
```
re.search('[^acd]','acdacdhai')
```

Out[118]:
```
<re.Match object; span=(6, 7), match='h'>
```

In [125]:
```
n = ['9383682753','63772927927929','5357268267']
for i in n:
    if re.search('^[6-9][0-9]{9}$',i):
        print(i)
```

```
9383682753
```

In [124]:
```
n = ['9383682753','63772927927929','5357268267']
for i in n:
    if re.search('^[6-9]\d{9}$',i):
        print(i)
```

```
9383682753
```

In [3]:
```
import re
```

\D

oher than \d

means any single character other than digit

In [5]:
```
re.search('\D','123 raja')
```

Out[5]:
```
<re.Match object; span=(3, 4), match=' '>
```

In [6]:
```
re.search('\D','123raja')
```

Out[6]:
```
<re.Match object; span=(3, 4), match='r'>
```

In [7]:
```
re.search('\D','raja567')
```

Out[7]:
```
<re.Match object; span=(0, 1), match='r'>
```

\s

any spacing single character

In [10]:
```
re.search('\s',"Raja s")
```

Out[10]:
```
<re.Match object; span=(4, 5), match=' '>
```

In [11]:
```
re.search('\s',"Rajas")
```

In [12]:
```
1 re.search('\s'," Raja s")
```

Out[12]: <re.Match object; span=(0, 1), match=' '>

\S

otherthan \s

In [13]:
```
1 re.search('\S','Raja s')
```

Out[13]: <re.Match object; span=(0, 1), match='R'>

In [14]:
```
1 re.search('\S','   Raja s')
```

Out[14]: <re.Match object; span=(3, 4), match='R'>

\w

single identifier character

a-z A-Z 0-9

In [15]:
```
1 re.search('\w',' _9az')
```

Out[15]: <re.Match object; span=(1, 2), match='_'>

In [16]:
```
1 re.search('\w','9az')
```

Out[16]: <re.Match object; span=(0, 1), match='9'>

\W

other than \w

other than identifier characters

In [17]:
```
1 re.search('\W','raja s')
```

Out[17]: <re.Match object; span=(4, 5), match=' '>

In [18]:
```
1 n = 'hello 1 Hello'
2 len(re.findall('[a-z]',n))
```

Out[18]: 9

* star

zero or more occurances

In [20]:
```
print(re.search('a*','python'))
```

```
<re.Match object; span=(0, 0), match=''>
```

In [21]:
```
print(re.search('1*','1111222233111'))
```

```
<re.Match object; span=(0, 4), match='1111'>
```

In [22]:
```
print(re.search('1','1111222233111'))
```

```
<re.Match object; span=(0, 1), match='1'>
```

In [23]:
```
print(re.search('^s.*m$','srjam'))
```

```
<re.Match object; span=(0, 5), match='srjam'>
```

In [24]:
```
print(re.search('^s.*m$','sm'))
```

```
<re.Match object; span=(0, 2), match='sm'>
```

In [25]:
```
print(re.search('^s.*m$','ahghm'))
```

```
None
```

+

one or more

In [26]:
```
re.search('1+','222112211')
```

Out[26]:
```
<re.Match object; span=(3, 5), match='11'>
```

In [27]:
```
re.search('3+','222112211')
```

In [28]:
```
re.search('^r.+s$','rajas')
```

Out[28]:
```
<re.Match object; span=(0, 5), match='rajas'>
```

write pattern for accepting gmail

In [33]:
```
m = input()
if re.search('^\w{3,10}[@][a-zA-Z]{3,8}[.][a-zA-Z]{2,3}',m):
    print(True)
else:
    print(False)
```

```
r@gmail.com
False
```

In [47]:
```
content = 'raja got 100 for maths 150 for physics and  total marks from scho
s = re.findall('\d{1,}',content)
s
```

Out[47]:
```
['100', '150', '4000']
```

In [45]:
```
sum(list(map(int,s)))
```

Out[45]:
```
4250
```

In [48]:
```
content = 'raja got 100 for maths 150 for physics and  total marks from scho
s = re.findall('\d{1,}',content)
print(s)
m = list(filter(lambda x:len(x)==3,s))
print(m)
sum(list(map(int,m)))
```

```
['100', '150', '4000']
['100', '150']
```

Out[48]:
```
250
```

os

In [49]:
```
import os
```

In [51]:
```
# current working directory
os.getcwd()
```

Out[51]:
```
"C:\\Users\\Raj's\\Downloads\\DataScince&ML\\NewTrainer\\pythonAdvance"
```

In [52]:
```
os.listdir()
```

Out[52]:
```
['.ipynb_checkpoints',
 '23-11-2019(Functions, Modules, Packages).ipynb',
 '24-11-2019(InBulitMadules).ipynb',
 'myOwnMadule.py',
 '__pycache__']
```

In [53]:
```
os.listdir('E:')
```

Out[53]:
```
['$RECYCLE.BIN', 'SoftWare', 'Study', 'System Volume Information']
```

In [54]:
```
1 os.listdir('d:')
```

Out[54]:
```
['$RECYCLE.BIN', 'Movies', 'System Volume Information']
```

In [55]:
```
1 os.mkdir('tempfolder')
```

In [56]:
```
1 os.chdir('d:')
```

In [57]:
```
1 dir(os)
```

...

In [58]:
```
1 help(os)
```

...

In [59]:
```
1 import statistics as st
```

In [60]:
```
1 st.mean([1,2,3,4,5])
```

Out[60]:
```
3
```

In [63]:
```
1 st.median([30,20,33,32,25,33,19,90])
```

Out[63]:
```
31.0
```

In [68]:
```
1 st.median_high([30,20,33,25,32,19,90,40])
```

Out[68]:
```
32
```

In [69]:
```
1 st.median_low([30,20,33,25,3,19,90,40])
```

Out[69]:
```
25
```

In [71]:
```
1 st.mode([10,10,20,10,20])
```

Out[71]:
```
10
```

In [73]:
```
1 from collections import Counter
```

In [74]:
```
1 cnt = Counter([10,20,30,40,40,30,30,30])
2 print(list(cnt.elements()))
3 print(cnt.most_common())
```

```
[10, 20, 30, 30, 30, 30, 40, 40]
[(30, 4), (40, 2), (10, 1), (20, 1)]
```

In [75]:
```
1 cnt1 = Counter('abcabcbcbsbcbgd')
2 print(cnt1.most_common(1))
3 print(cnt1.most_common(3))
```

```
[('b', 6)]
[('b', 6), ('c', 4), ('a', 2)]
```

In [77]:
```
1 print(dir(__builtins__))
```

```
['ArithmeticError', 'AssertionError', 'AttributeError', 'BaseException', 'Block
ingIOError', 'BrokenPipeError', 'BufferError', 'BytesWarning', 'ChildProcessErr
or', 'ConnectionAbortedError', 'ConnectionError', 'ConnectionRefusedError', 'Co
nnectionResetError', 'DeprecationWarning', 'EOFError', 'Ellipsis', 'Environment
Error', 'Exception', 'False', 'FileExistsError', 'FileNotFoundError', 'Floating
PointError', 'FutureWarning', 'GeneratorExit', 'IOError', 'ImportError', 'Impor
tWarning', 'IndentationError', 'IndexError', 'InterruptedError', 'IsADirectoryE
rror', 'KeyError', 'KeyboardInterrupt', 'LookupError', 'MemoryError', 'ModuleNo
tFoundError', 'NameError', 'None', 'NotADirectoryError', 'NotImplemented', 'Not
ImplementedError', 'OSError', 'OverflowError', 'PendingDeprecationWarning', 'Pe
rmissionError', 'ProcessLookupError', 'RecursionError', 'ReferenceError', 'Reso
urceWarning', 'RuntimeError', 'RuntimeWarning', 'StopAsyncIteration', 'StopIter
ation', 'SyntaxError', 'SyntaxWarning', 'SystemError', 'SystemExit', 'TabErro
r', 'TimeoutError', 'True', 'TypeError', 'UnboundLocalError', 'UnicodeDecodeErr
or', 'UnicodeEncodeError', 'UnicodeError', 'UnicodeTranslateError', 'UnicodeWar
ning', 'UserWarning', 'ValueError', 'Warning', 'WindowsError', 'ZeroDivisionErr
or', '__IPYTHON__', '__build_class__', '__debug__', '__doc__', '__import__', '_
_loader__', '__name__', '__package__', '__spec__', 'abs', 'all', 'any', 'asci
i', 'bin', 'bool', 'breakpoint', 'bytearray', 'bytes', 'callable', 'chr', 'clas
smethod', 'compile', 'complex', 'copyright', 'credits', 'delattr', 'dict', 'di
r', 'display', 'divmod', 'enumerate', 'eval', 'exec', 'filter', 'float', 'forma
t', 'frozenset', 'get_ipython', 'getattr', 'globals', 'hasattr', 'hash', 'hel
p', 'hex', 'id', 'input', 'int', 'isinstance', 'issubclass', 'iter', 'len', 'li
cense', 'list', 'locals', 'map', 'max', 'memoryview', 'min', 'next', 'object',
'oct', 'open', 'ord', 'pow', 'print', 'property', 'range', 'repr', 'reversed',
'round', 'set', 'setattr', 'slice', 'sorted', 'staticmethod', 'str', 'sum', 'su
per', 'tuple', 'type', 'vars', 'zip']
```

In [80]:
```
1 help('modules')
```

```
Please wait a moment while I gather a list of all available modules...


C:\ProgramData\Anaconda3\lib\site-packages\IPython\kernel\__init__.py:13: Shi
mWarning: The `IPython.kernel` package has been deprecated since IPython 4.0.
You should import from ipykernel or jupyter_client instead.
  "You should import from ipykernel or jupyter_client instead.", ShimWarning)
WARNING: AstropyDeprecationWarning: astropy.utils.compat.futures is now depre
cated - use concurrent.futures instead [astropy.utils.compat.futures]
C:\ProgramData\Anaconda3\lib\site-packages\nltk\twitter\__init__.py:22: UserW
arning: The twython library has not been installed. Some functionality from t
he twitter package will not be available.
  "The twython library has not been installed. "

DEBUG:pip._internal.vcs.versioncontrol:Registered VCS backend: bzr
DEBUG:pip._internal.vcs.versioncontrol:Registered VCS backend: git
DEBUG:pip._internal.vcs.versioncontrol:Registered VCS backend: hg
DEBUG:pip._internal.vcs.versioncontrol:Registered VCS backend: svn
```

In [81]:
```
1 import sys
```

In [82]:
```
1 sys.path
```

Out[82]:
```
['C:\\ProgramData\\Anaconda3\\lib\\site-packages\\spyder\\utils\\help',
 "C:\\Users\\Raj's\\Downloads\\DataScince&ML\\NewTrainer\\pythonAdvance",
 'C:\\ProgramData\\Anaconda3\\python37.zip',
 'C:\\ProgramData\\Anaconda3\\DLLs',
 'C:\\ProgramData\\Anaconda3\\lib',
 'C:\\ProgramData\\Anaconda3',
 '',
 'C:\\ProgramData\\Anaconda3\\lib\\site-packages',
 'C:\\ProgramData\\Anaconda3\\lib\\site-packages\\win32',
 'C:\\ProgramData\\Anaconda3\\lib\\site-packages\\win32\\lib',
 'C:\\ProgramData\\Anaconda3\\lib\\site-packages\\Pythonwin',
 'C:\\ProgramData\\Anaconda3\\lib\\site-packages\\IPython\\extensions',
 "C:\\Users\\Raj's\\.ipython",
 'C:\\ProgramData\\Anaconda3\\lib\\site-packages\\astroid\\brain']
```

In [90]:
```
1 sys.version
```

Out[90]:
```
'3.7.4 (default, Aug  9 2019, 18:34:13) [MSC v.1915 64 bit (AMD64)]'
```

In [92]:
```
1 import hello
2
```

In [ ]:
```
1
```