Date :22 June 2019

Day Objective

- File Handling
  - Basic File data processing
    - Accessing and Modifying File Data
      - Character count
      - Line Count
      - File Size
      - word count
      - Unique Word count

In [12]:
```
# Read a file - File should exists(Read Mode)
#write into a file -File can exists(append) or new file can be created(Write

def readfile(filepath):

    with open(filepath,'r') as f:
        filedata=f.read()
    print(filedata)
    return filedata
filepath="DataFiles/data.txt"
readfile(filepath)
```

```
new data
Line2
Line3
Line3
Line4
data in line1
```

Out[12]:
```
'new data\nLine2\nLine3\nLine3\nLine4\ndata in line1'
```

In [5]:
```
def readfile(filepath):
    count=0
    count1=0
    char_count=0
    with open(filepath,'r') as f:
        filedata=f.read()
        for i in filedata:
            if i==" ":
                count=count+1
            elif i=="\n":
                count1=count1+1
            else:
                char_count=char_count+1
        print(count,"spaces")
        print("char_count",char_count)
        print(count1,"lines count")
        print(count1+count+char_count,"filesize")


filepath="DataFiles/data.txt"
readfile(filepath)


```

```
1 spaces
char_count 17
3 lines count
21 filesize
```

In [22]:
```
#file size
import os
os.path.getsize(filepath)
```

Out[22]:
```
20
```

In [10]:
```
#Each line into a list
def readfileintolist(filepath):

    with open(filepath,'r') as f:
        filedata=f.read()
        lines=filedata.split("\n")
        #s=len(lines)
    #return s
    return lines
filepath="DataFiles/data.txt"
readfileintolist(filepath)
```

Out[10]:
```
['new data', 'Line2', 'Line3', '']
```

In [11]:
```
#Function to count lines
def countlines(filepath):
    count=len(readfileintolist(filepath))
    return count
countlines(filepath)
```

Out[11]: 4

In [22]:
```
#Function to count the number of words in a file
import re

def numberofwords(filepath):
    with open(filepath,'r') as f:
        pattern='[ \n]'
        filedata=f.read()
        count=len(re.split(pattern,filedata))
        return count
filepath="DataFiles/data.txt"
numberofwords(filepath)
```

Out[22]: 5

In [23]:
```
#Function to get unique elements in a list
#First create a empty unique list
def uniquedata(li):
    #create an empty unique list
    unique=[]
    #for every element in the main list,check if it exists in the unique lis
    #if it does not exist,add it to unique list
    #else if it already exists,move on to the next element in the main list
    for element in li:
        if element not in unique:
            unique.append(element)
    return unique
li=[1,2,3,1,2,4]
uniquedata(li)
```

Out[23]: [1, 2, 3, 4]

In [35]:
```
def uniquewordcount(filepath):
    u=[]
    with open(filepath,'r') as f:
        filedata=f.read()
        filedata=filedata.split()
        for element in filedata:
            if element not in u:
                u.append(element)
    return u
filepath="DataFiles/data.txt"
uniquewordcount(filepath)



```

Out[35]: ['new', 'data', 'Line2', 'Line3', 'Line4', 'in', 'line1']

In [11]:
```
#frequency distribution of words
def uniquewordcount(filepath):

    f1={}
    uniquedata=readfile(filepath)
    word=set(uniquedata.split())
    for element in word:
            f1[element]=uniquedata.count(element)
    return f1
filepath="DataFiles/data.txt"
uniquewordcount(filepath)
```

```
new data
Line2
Line3
Line3
Line4
data in line1
```

Out[11]:
```
{'data': 2, 'Line3': 2, 'line1': 1, 'new': 1, 'Line4': 1, 'Line2': 1, 'in': 6}
```

In [110]:
```
l=[1,2,3,1,2,2]
f={}
count=0
for element in l:
    if element in f:
        f[element]+=1

    else:
        f[element]=1
print(f)


```

```
{1: 2, 2: 3, 3: 1}
```

In [14]:
```
import re
def wordsfromfile(filepath):
    pattern="[ \n]"
    filedata=readfile(filepath)
    allwordslist=re.split(pattern,filedata)
    return allwordslist
wordsfromfile(filepath)
```

```
new data
Line2
Line3
Line3
Line4
data in line1
```

Out[14]:
```
['new', 'data', 'Line2', 'Line3', 'Line3', 'Line4', 'data', 'in', 'line1']
```

In [ ]:
```

```