Module 2:

Count the number of words in a given file

We hope that you have learnt about what is a string, taking input string and printing strings in the module 1 and also you have used split() function to count number of words in the given string. Here we are going to learn how to open a file ,read and count number of words, spaces, lines.. etc in a given file.

Before we are going to read a file you need to know about ***split()***and ***count()*** functions. The standard ***split()*** can use only one delimiter. To split a text file into words you need multiple delimiters like blank, punctuation, math signs (+-*/), parentheses and so on.

Here's a quick some example to understand

Spliting a sentence:

```
sent = "Jack ate the apple." # Assign a stament to a variable
splitsent = sent.split(' ')      # spliting a stament
print splitsent                          # printing the data in after spliting
```

Output: ['Jack', 'ate', 'the', 'apple.']

when we split a statement, that will be converted into a list and every word is stored as an element of list.

['Jack', 'ate', 'the', 'apple.']

So if we find the length of the list that length will be equal to number words in the statement. This is one way of counting words in the data. You can use another way.

Counting number of substrings:

Syntax: count(sub[,start[,end]])

Is used to count the number of occurrences of the given item in the list.

```
l = [‘a’,’b’,’a’,’c’,’d’,’e’]
l.count(‘a’)
```

```
=> 3
l.count(‘d’)
=> 1
```

Counting No of spaces in a statement:

When we use count function to count No spaces it is very is to find out the No of words in the given data.

```
sent = "Jack ate the apple." # Assign a stament to a variable

spaces = sent.count(' ')      # spliting a stament
print “No of spces”,spaces   # printing the data in after spliting
print “No of words,” spaces+1
```

This will produce: No of spaces 3

No of words 4

Reading a file:

Files in a programming sense are really not very different from files that you use in a word processor or other application: you *open* them, do some work and then *close* them again.

You can open files with open function, which has the following syntax

Open(name[,mode])

Open() takes two arguments. The first is the filename (which may be passed as a variable or a literal string and mandatory). The second is the *mode* which is optional. The mode determines whether we are opening the file for reading(r) or writing(w).

Ex: file = open(“text1.txt”, ‘r’) # This is in reading mode

We close the file at the end with the close() method.

Ex: file.close()

Here’s a quick example to understand. See that there is a txt file which contains “Hello world! ”statement in it.

```
f = open('somefile.txt', 'r') # open file in reading mode


print f.read()                        # reading the data
'Hello, World!'
```

We can assign data to a variable and it will be consider as a string.

```
f = open('somefile.txt', 'r') # open file in reading mode
f.read()                      # reading the data at a time
a = f.read()                  # assigning data to a variable
print a                       # printing data
This will produce:
Hello, World!
```

read()   : function is used to read file at once
readline() : function is used to *reads in just a single line from a file at a time*
readlines() : function is used to reads ALL lines, and splits them by line delimiter.

Worked Example 1 :
Take a input statement from the user and count how many words are there in it .

```
a = raw_input(“Enter your statement \n”)
b = a.split()                # splitting the sentences in words
print len(b)
```

Here b is list contains words.
Input is : “This is my first program.”
Output :   5

Worked Example 2:
Open a txt file called weeks.txt which consists of week days and then read that file.

```
f = open("weeks.txt",'r') # open file in reading mode
data = f.read()          # assigning data to a variable
print data
```

This will produce:

```
        Sunday
        Monday
        Tuesday
        Wednesday
        Thursday
        Friday
        Saturday
```

Note: you need to create a text file called weeks.txt where the program file exists.

Worked Example 3 : count how many lines of data in the above txt file.

```
f = open("weeks.txt",'r')
b = f.readlines()           # reading data line by line
count = 0                   # assing 0 to a varible to count lines
for i in b:      # this for loop is used to read data line by line
    count = count+1
print count
```

This will produce :   7

Worked Example 4: Read a file name from the user and count the number of spaces in the file.

```
filename = raw_input("Enter a file name which is already exits \n")
file = open(filename,'r')
data = file.read()
count = data.count(" ")
print count
```

This will produce : How many spaces are there in a given file.
(Note: when taking a file name as input you need to give extension of that file. Ex: text1.txt)

Exercise problems :

1. Count the number of words in a given filename.
2. Count the No of characters with space and without spaces separately in the given file name.

ALL THE BEST