Lecture 6 - Lists / Dictionaries

First let's read in a file

```
 1:
 2: def readNameList(fn):
 3:
 4:     f = open(fn,"r")
 5:     if f == None:
 6:         print ( f"Invalid file {fn} — failed to open" )
 7:         return None
 8:     dt = f.readlines()
 9:     f.close()
10:     for i in range (len(dt)):
11:         s = dt[i].rstrip()
12:         dt[i] = s
13:
14:     return dt
15:
16:
17: # Automated Test
18: if __name__ == "__main__":
19:     n_err = 0
20:
21:     got = readNameList("2names.txt")
22:     expect = [
23:         "\"Gunter, Dolly R\",(072) 123—4760",
24:         "\"Polk, Hattie S\",(563) 404—0792"
25:         ]
26:     if got[0] != expect[0]:
27:         n_err = n_err + 1
28:         print ( "Error: Test 1: file read error expected {} got {}".
29:                 format (  expect[0], got[0] ) )
30:     if got[1] != expect[1]:
31:         n_err = n_err + 1
32:         print ( "Error: Test 2: file read error expected {} got {}".
33:                 format (  expect[1], got[1] ) )
34:
35:     if n_err == 0 :
36:         print ( "PASS" )
37:     else:
38:         print ( "FAILED" )
39:
```

Let's remove the punctuation

```
 1:
 2: import string
 3:
 4: def removePunctuation(txt):
 5:     for c in string.punctuation:
 6:         txt = txt.replace(c,"")
 7:     return txt
 8:
 9:
10: # Automated Test
11: if __name__ == "__main__":
12:     n_err = 0
13:
14:     got = removePunctuation("this, and: that")
15:     expect = "this and that"
16:     if got != expect:
17:         n_err = n_err + 1
18:         print ( "Error: Test 1: file read error expected {} got {}".
19:                 format (  expect, got ) )
20:
21:     if n_err == 0 :
22:         print ( "PASS" )
23:     else:
24:         print ( "FAILED" )
25:
```

List Example

```
 1:
 2: def readNameList(fn):
 3:     f = open(fn,"r")
 4:     if f == None:
 5:         print ( f"Invalid file {fn} — failed to open" )
 6:         return None
 7:     dt = f.readlines()
 8:     f.close()
 9:     for i in range (len(dt)):
10:         s = dt[i].rstrip()
11:         dt[i] = s
12:     return dt
13:
14: phone_list = readNameList("50000phone.csv")
15:
16: print ( "Enter a Name to Lookup" )
17: lookFor = input()
18: found = False
19: for i in range(len(phone_list)):
20:     if lookFor in phone_list[i]:
21:         found = True
22:         print ( "Found {}".format(phone_list[i]) )
23: if not found:
24:     print ( "no names found" )
25:
```

Dictionary

A dictionary is an association between a "key" value and a set of data that is efficient for looking thins up by the key.

let's do an "age" one...

```
>>> dd = {}
>>> dd["bob"] = 22
>>> dd["jane"] = 31
>>> dd["marry"] = 18
>>>
>>> dd["bob"]
>>> dd["jane"]
```

Now changing our phone search to use a dictionary:

```
 1: #!/Users/philip/opt/anaconda3/bin/python
 2:
 3: from readNameListCSV import readNameListCSV
 4:
 5: phone_list = readNameListCSV("50000phone.csv")
 6:
 7: print ( "Enter a Name to Lookup\n=> ", end="" )
 8: lookFor = input()
 9:
10: if lookFor in phone_list:
11:     print ( "Found {}".format(phone_list[lookFor]) )
12: else:
13:     print ( "{} not found".format(lookFor) )
14:
```

Requirements

Prompt for a file name. Then read in a file of text.
Split the file up into words. Remove any punctuation and convert each word to lower case. Count
how many times each word occurs. Print out a sorted list of the words in the file with the number of
occurrences of each word. Sort from the most frequent word to the least.

Requirements Broken Down

1. Prompt for a file name.

2. Read in a file of text.

3. Split the file up into words.

4. Remove any punctuation and convert each word to lower case.

5. Count how many times each word occurs.

6. Print out a sorted list
   - Print out a sorted list of the words in the file with the number of occurrences of each word.
   - Sort from the most frequent word to the least.

Let's count some words

```
 1: #!/Users/philip/opt/anaconda3/bin/python
 2:
 3: from readNameList import readNameList
 4: from removePunctuation import removePunctuation
 5:
 6: def main():
 7:
 8:      print ( "Enter File Name\n=> ", end="" )
```

```
 8:        print ( "Enter File Name\n=>", end="" )
 9:        fn = input()
10:        # read in file to a list of lines
11:        data = readNameList(fn)
12:
13:        # create an empty dictionary, the key will be the word
14:        #the value will be the count.
15:        freq = {}
16:
17:        for line in data:
18:
19:            # Split line up into a set of words
20:            words = line.split()
21:
22:            for word in words:
23:
24:                word = word.lower()
25:                word = removePunctuation(word)
26:
27:                if word in freq:
28:                    freq[word] += 1
29:                else:
30:                    freq[word] = 1
31:
32:        # Prints that woudl just print out in word sorted order.
33:        #for key in sorted(freq.keys()):
34:        #    cnt = freq[key]
35:        #    print ( f"{key} = {cnt}" )
36:
37:        print ( "{name:5s} : {word}".format(name="Count",word="Word") )
38:        print ( "{name:5s} : {word}".format(name="-----",word="--------------------")
39:        x = sorted(freq.items(), key=lambda x: x[1], reverse=True)
40:        for p in x:
41:            print ( f"{p[1]:5d} : {p[0]}" )
42:
43:
44:
45:
46: if __name__ == "__main__":
47:     main()
48:
```

Copyright