

## Assignment 10: Dictionaries, Lists, and Tuples

[1] **Objectives:** This assignment is the finale of the course. We will use many structures learned recently, including Lists, Tuples, and Dictionaries. The assignment will highlight the similarity and differences between these structures. In particular, we will show that dictionaries are not as convenient as lists and tuples regarding casting and sorting. Since this is the last assignment, it is relatively short in scope.

[2] **Description:** In some sense, this is somewhat related to the last assignment. However, we designed it to be independent of the previous assignment. The similarity is in counting the frequency of words used in a document. However, we put the words in a file, one word per line. There is not much “cleaning” to do.

In a list, elements play an equal role in the list. None of them is more important than the others. A dictionary is different. It has a key-value pair. Keys were designed for accessing the values. The assignment will show the difficulties of using a dictionary and how to get around them.

Sorting a dictionary is limited. We can sort by the keys without the values. Furthermore, it does not sort the dictionary directly; `sorted()` returns a list. For example, if “names” is a Python dictionary, the following function call returns a list of tuples and then casts back into a dict.

```
names = dict(sorted(names.items(), key=lambda item: item[1]))
```

The goal is to count the frequencies of the words in the text file. Print them in decreasing order of frequency.

[3] **Requirements:** The code for this assignment is relatively short. You will manipulate tuples, lists, and dictionaries in six steps to get the same result as the last assignment. In the end, print a table showing the result. Each step should consist of one line to computing followed by one print statement to show the result.

1. Read the text file (function call) and produce a list of words.
2. Build a dict to count the frequencies (make a function call).
3. Sort the dict.
4. Cast the dict to a list of tuples.
5. Sort the list in decreasing order of frequency.
6. Case the list to a dictionary.

In the end, write a “pretty” print for the dictionary.

[4] **Output:** A sample output is given below. For the six steps, the result is truncated.

[5] **Deadline:** 11:59 pm, Monday, May 1, 2023

1. List of Words: ['new', 'spyware', 'firm', 'said', 'to', 'have', 'helped', 'hack', 'iphones', 'around']
2. Dict of Words: {'new': 2, 'spyware': 3, 'firm': 1, 'said': 2, 'to': 5, 'have': 3, 'helped': 1, 'hack': 1, 'iphones': 1, 'around': 1, 'a': 1, 'according': 1, 'and': 1, 'apple': 1, 'are': 1, 'at': 1, 'been': 1, 'break': 1, 'by': 1, 'citizen': 1, 'continued': 1, 'corp': 1, 'edited': 1, 'group': 1, 'highlight': 1, 'into': 1, 'israel': 1, 'israeli': 1, 'journalists': 1, 'lab': 1, 'largely': 1, 'makers': 1, 'market': 1, 'marketed': 1, 'microsoft': 1, 'name': 1, 'newspapers': 1, 'operated': 1, 'operating': 1, 'political': 1, 'published': 1, 'researchers': 1, 'secret': 1, 'software': 1, 'sponsored': 1, 'tuesday': 1, 'university': 1, 'used': 1, 'which': 1}
3. Sort the Words: ['a', 'according', 'and', 'apple', 'are', 'around', 'at', 'been', 'break', 'by', 'citizen', 'continued', 'corp', 'edited', 'group', 'highlight', 'into', 'israel', 'israeli', 'journalists', 'lab', 'largely', 'makers', 'market', 'marketed', 'microsoft', 'name', 'newspapers', 'operated', 'operating', 'political', 'published', 'researchers', 'secret', 'software', 'sponsored', 'tuesday', 'university', 'used', 'which']
4. Dict to List: [('new', 2), ('spyware', 3), ('firm', 1), ('said', 2), ('to', 5), ('have', 3), ('helped', 1), ('hack', 1), ('iphones', 1), ('around', 1), ('a', 1), ('according', 1), ('and', 1), ('apple', 1), ('are', 1), ('at', 1), ('been', 1), ('break', 1), ('by', 1), ('citizen', 1), ('continued', 1), ('corp', 1), ('edited', 1), ('group', 1), ('highlight', 1), ('into', 1), ('israel', 1), ('israeli', 1), ('journalists', 1), ('lab', 1), ('largely', 1), ('makers', 1), ('market', 1), ('marketed', 1), ('microsoft', 1), ('name', 1), ('newspapers', 1), ('operated', 1), ('operating', 1), ('political', 1), ('published', 1), ('researchers', 1), ('secret', 1), ('software', 1), ('sponsored', 1), ('tuesday', 1), ('university', 1), ('used', 1), ('which', 1)]
5. List sorted: [('the', 6), ('to', 5), ('of', 4), ('spyware', 3), ('have', 3), ('by', 3), ('new', 2), ('said', 2), ('firm', 1), ('helped', 1), ('hack', 1), ('iphones', 1), ('around', 1), ('a', 1), ('according', 1), ('and', 1), ('apple', 1), ('are', 1), ('at', 1), ('been', 1), ('break', 1), ('citizen', 1), ('continued', 1), ('corp', 1), ('edited', 1), ('group', 1), ('highlight', 1), ('into', 1), ('israel', 1), ('israeli', 1), ('journalists', 1), ('lab', 1), ('largely', 1), ('makers', 1), ('market', 1), ('marketed', 1), ('microsoft', 1), ('name', 1), ('newspapers', 1), ('operated', 1), ('operating', 1), ('political', 1), ('published', 1), ('researchers', 1), ('secret', 1), ('software', 1), ('sponsored', 1), ('tuesday', 1), ('university', 1), ('used', 1), ('which', 1)]
6. List to Dict: {'the': 6, 'to': 5, 'of': 4, 'spyware': 3, 'have': 3, 'by': 3, 'new': 2, 'said': 2, 'firm': 1, 'helped': 1, 'hack': 1, 'iphones': 1, 'around': 1, 'a': 1, 'according': 1, 'and': 1, 'apple': 1, 'are': 1, 'at': 1, 'been': 1, 'break': 1, 'citizen': 1, 'continued': 1, 'corp': 1, 'edited': 1, 'group': 1, 'highlight': 1, 'into': 1, 'israel': 1, 'israeli': 1, 'journalists': 1, 'lab': 1, 'largely': 1, 'makers': 1, 'market': 1, 'marketed': 1, 'microsoft': 1, 'name': 1, 'newspapers': 1, 'operated': 1, 'operating': 1, 'political': 1, 'published': 1, 'researchers': 1, 'secret': 1, 'software': 1, 'sponsored': 1, 'tuesday': 1, 'university': 1, 'used': 1, 'which': 1}

Length of the list: 71

Count | Word

-----|-----

6 | the  
 5 | to  
 4 | of  
 3 | spyware  
 3 | have  
 3 | by  
 2 | new  
 2 | said  
 2 | iphones  
 2 | tools  
 2 | linked  
 2 | quadream  
 2 | software  
 2 | a  
 2 | and  
 2 | under  
 2 | research  
 1 | firm  
 1 | helped  
 1 | hack  
 1 | around  
 1 | globe  
 1 | slightly  
 1 | edited  
 1 | israel  
 1 | highlight  
 1 | continued  
 1 | use  
 1 | secret  
 1 | hacking  
 1 | sold  
 1 | little  
 1 | known  
 1 | israeli  
 1 | vendor  
 1 | been  
 1 | used  
 1 | break  
 1 | into

1 | journalists  
 1 | political  
 1 | opposition  
 1 | figures  
 1 | silently  
 1 | exploiting  
 1 | apple  
 1 | inc  
 1 | iphone  
 1 | cybersecurity  
 1 | researchers  
 1 | intrusions  
 1 | are  
 1 | ltd  
 1 | which  
 1 | markets  
 1 | name  
 1 | reign  
 1 | according  
 1 | published  
 1 | tuesday  
 1 | citizen  
 1 | lab  
 1 | group  
 1 | at  
 1 | university  
 1 | toronto  
 1 | microsoft  
 1 | corp  
 1 | makers  
 1 | operated  
 1 | largely