Introduction to Computer Programming I

Dr. Sharon Persinger

December 3, 2018

Topics

- Dictionaries, dictionary operations
- Counting words in a file

Python dictionaries

- A dictionary is a built-in Python data type for storing key-value ordered pairs. Dictionaries are also called mappings or function tables.
- A dictionary is a function given by a table of key-value pairs. What does that mean?

- Read section 11.7 – non-sequential collections
- Why is a dictionary non-sequential?

Dictionaries in Python

- Create a dictionary
  - birds = {“duck”:”bird that swims”, “hawk”:”bird that hunts”, “eagle”:”big bird that hunts”}
  - 3 key-value pairs, key and value separated by :, enclosed in { }

- Find the value associated with a key
  birds[“eagle”]
  - Evaluating a function birds

Dictionaries in Python

- Change the value associated with a key
  - birds[“eagle”] = “big bird that hunts fish”
- Add new key-value pairs
  - birds[“cardinal”] = “redbird”
- Test whether a key is in the dictionary
  - “eagle” in birds
- Return the keys as a list
  - birds.keys()

Dictionaries in Python

- Return the values as a list
  - birds.values()
- Return the set of key-value pairs as a list of tuples
  - birds.items()
- Delete the pair with specified key
  - del birds[“eagle”]
- Delete all pairs
  - birds.clear()

Dictionaries in Python

- Return the value associated with a key or some default value if key is not present

  birds.get("blue jay", "not present")

Word frequency program

- wordfreq.py

  Program that counts how many times each word occurs in a document.

- Basically, create a dictionary, a table of key-value pairs. The key is a word in the document, and the value is the number of times the word appears.

Analysis of the program wordfreq.py

```
# wordfreq.py

def byFreq(pair):
    return pair[1]

def main():
    print("This program analyzes word frequency in a file")
    print("and prints a report on the n most frequent words.\n")

    # get the sequence of words from the file
    fname = input("File to analyze: ")
    text = open(fname,'r').read()       #read the entire document into the string text
    text = text.lower()                 #change it to lowercase – why?
    for ch in '!"#$%&()*+,-./:;<=>?@[\\]^_`{|}~':
        text = text.replace(ch, ' ')    #replace function for the string text
                                        #what does this do?
    words = text.split()                #what does this do?
```

Analysis of the program wordfreq.py

```
# construct a dictionary of word counts
   counts = {}                    #make an empty dictionary
   for w in words:
      counts[w] = counts.get(w,0) + 1   #check the get function
                                        #the dictionary is complete


   # output analysis of n most frequent words.
   n = int(input("Output analysis of how many words? "))
   items = list(counts.items())   #items as list of pairs
   items.sort()                   #first sort the words alphabetically
   items.sort(key=byFreq, reverse=True)   #then sort by frequency, reverse
   for i in range(n):
      word, count = items[i]
      print("{0:<15}{1:>5}".format(word, count))  #print out n most frequent

if __name__ == '__main__':  main()
```

key parameter in the sort method

- The value of the key parameter must be a function. The function is applied to each item in the list and then the sort comparison is performed on the function value for that item.

- For the word frequency program, the key takes the function byFreq as its value. byFreq return the item with index 1 in its argument pair.

- So the result of items.sort(key=byFreq, reverse=True) is to sort the list items by the frequency number, which has index 1 in the item.
- Since the other parameter reverse is given the value True, the items are sorted in reverse or decreasing order.