CHAPTER 5 – DICTIONARIES AND STRUCTURING DATA



Support the Author: Buy the book on Amazon or the book/ebook bundle directly from No Starch Press.



Read the author's other free Python books:



DICTIONARIES AND STRUCTURING DATA

In this chapter, I will cover the dictionary data type, which provides a flexible way to access and organize data. Then, combining dictionaries with your knowledge of lists from the previous chapter, you'll learn how to create a data structure to model a tic-tac-toe board.

THE DICTIONARY DATA TYPE

Like a list, a *dictionary* is a collection of many values. But unlike indexes for lists, indexes for dictionaries can use many different data types, not just integers. Indexes for dictionaries are called *keys*, and a key with its associated value is called a *key-value pair*.

In code, a dictionary is typed with braces, {}. Enter the following into the interactive shell:

```
>>> myCat = {'size': 'fat', 'color': 'gray', 'disposition': 'loud'}
```

This assigns a dictionary to the myCat variable. This dictionary's keys are 'size', 'color', and 'disposition'. The values for these keys are 'fat', 'gray', and 'loud', respectively. You can access these values through their keys:

```
>>> myCat['size']
'fat'
>>> 'My cat has ' + myCat['color'] + ' fur.'
'My cat has gray fur.'
```

Dictionaries can still use integer values as keys, just like lists use integers for indexes, but they do not have to start at 0 and can be any number.

```
>>> spam = {12345: 'Luggage Combination', 42: 'The Answer'}
```

DICTIONARIES VS. LISTS

Unlike lists, items in dictionaries are unordered. The first item in a list named spam would be spam[0]. But there is no "first" item in a dictionary. While the order of items matters for determining whether two lists are the same, it does not matter in what order the key-value pairs are typed in a dictionary. Enter the following into the interactive shell:

```
>>> spam = ['cats', 'dogs', 'moose']
>>> bacon = ['dogs', 'moose', 'cats']
>>> spam == bacon
False
>>> eggs = {'name': 'Zophie', 'species': 'cat', 'age': '8'}
>>> ham = {'species': 'cat', 'age': '8', 'name': 'Zophie'}
>>> eggs == ham
True
```

Because dictionaries are not ordered, they can't be sliced like lists.

Trying to access a key that does not exist in a dictionary will result in a KeyError error message, much like a list's "out-of-range" IndexError error message. Enter the following into the interactive shell, and notice the error message that shows up because there is no 'color' key:

```
>>> spam = {'name': 'Zophie', 'age': 7}
>>> spam['color']
Traceback (most recent call last):
   File "<pyshell#1>", line 1, in <module>
        spam['color']
KeyError: 'color'
```

Though dictionaries are not ordered, the fact that you can have arbitrary values for the keys allows you to organize your data in powerful ways. Say you wanted your program to store data about your friends' birthdays. You can use a dictionary with the names as keys and the birthdays as values. Open a new file editor window and enter the following code. Save it as *birthdays.py*.

```
while True:
    print('Enter a name: (blank to quit)')
    name = input()
    if name == '':
        break

if name in birthdays:
    print(birthdays[name] + ' is the birthday of ' + name)
    else:
        print('I do not have birthday information for ' + name)
        print('What is their birthday?')
        bday = input()
    birthdays[name] = bday
        print('Birthday database updated.')
```

You create an initial dictionary and store it in birthdays ①. You can see if the entered name exists as a key in the dictionary with the in keyword ②, just as you did for lists. If the name is in the dictionary, you access the associated value using square brackets ③; if not, you can add it using the same square bracket syntax combined with the assignment operator ④.

When you run this program, it will look like this:

```
Enter a name: (blank to quit)

Alice

Apr 1 is the birthday of Alice

Enter a name: (blank to quit)

Eve

I do not have birthday information for Eve

What is their birthday?

Dec 5

Birthday database updated.

Enter a name: (blank to quit)

Eve
```

```
Dec 5 is the birthday of Eve Enter a name: (blank to quit)
```

Of course, all the data you enter in this program is forgotten when the program terminates. You'll learn how to save data to files on the hard drive in Chapter 8.

THE KEYS(), VALUES(), AND ITEMS() METHODS

There are three dictionary methods that will return list-like values of the dictionary's keys, values, or both keys and values: keys(), values(), and items(). The values returned by these methods are not true lists: They cannot be modified and do not have an append() method. But these data types (dict_keys, dict_values, and dict_items, respectively) can be used in for loops. To see how these methods work, enter the following into the interactive shell:

Here, a for loop iterates over each of the values in the spam dictionary. A for loop can also iterate over the keys or both keys and values:

Using the keys(), values(), and items() methods, a for loop can iterate over the keys, values, or key-value pairs in a dictionary, respectively. Notice that the values in the dict_items value returned by the items() method are tuples of the key and value.

If you want a true list from one of these methods, pass its list-like return value to the list() function. Enter the following into the interactive shell:

```
>>> spam = {'color': 'red', 'age': 42}
>>> spam.keys()
dict_keys(['color', 'age'])
>>> list(spam.keys())
['color', 'age']
```

The list(spam.keys()) line takes the dict_keys value returned from keys() and passes it to list(), which then returns a list value of ['color', 'age'].

You can also use the multiple assignment trick in a for loop to assign the key and value to separate variables. Enter the following into the interactive shell:

CHECKING WHETHER A KEY OR VALUE EXISTS IN A DICTIONARY

Recall from the previous chapter that the in and not in operators can check whether a value exists in a list. You can also use these operators to see whether a certain key or value exists in a dictionary. Enter the following into the interactive shell:

```
>>> spam = {'name': 'Zophie', 'age': 7}
>>> 'name' in spam.keys()
True
>>> 'Zophie' in spam.values()
True
>>> 'color' in spam.keys()
False
>>> 'color' not in spam.keys()
True
>>> 'color' in spam.keys()
```

In the previous example, notice that 'color' in spam is essentially a shorter version of writing 'color' in spam.keys(). This is always the case: If you ever want to check whether a value is (or isn't) a key in the dictionary, you can simply use the in (or not in) keyword with the dictionary value itself.

THE GET() METHOD

It's tedious to check whether a key exists in a dictionary before accessing that key's value. Fortunately, dictionaries have a get() method that takes two arguments: the key of the value to retrieve and a fallback value to return if that key does not exist.

Enter the following into the interactive shell:

```
>>> picnicItems = {'apples': 5, 'cups': 2}
>>> 'I am bringing ' + str(picnicItems.get('cups', 0)) + ' cups.'
'I am bringing 2 cups.'
>>> 'I am bringing ' + str(picnicItems.get('eggs', 0)) + ' eggs.'
'I am bringing 0 eggs.'
```

Because there is no 'eggs' key in the picnicItems dictionary, the default value 0 is returned by the get() method. Without using get(), the code would have caused an error message, such as in the following example:

```
>>> picnicItems = {'apples': 5, 'cups': 2}
>>> 'I am bringing ' + str(picnicItems['eggs']) + ' eggs.'
Traceback (most recent call last):
   File "<pyshell#34>", line 1, in <module>
        'I am bringing ' + str(picnicItems['eggs']) + ' eggs.'
KeyError: 'eggs'
```

THE SETDEFAULT() METHOD

You'll often have to set a value in a dictionary for a certain key only if that key does not already have a value. The code looks something like this:

```
spam = {'name': 'Pooka', 'age': 5}
if 'color' not in spam:
    spam['color'] = 'black'
```

The setdefault() method offers a way to do this in one line of code. The first argument passed to the method is the key to check for, and the second argument is the value to set at that key if the key does not exist. If the key does exist, the setdefault() method returns the key's value. Enter the following into the interactive shell:

```
>>> spam = {'name': 'Pooka', 'age': 5}
>>> spam.setdefault('color', 'black')
'black'
>>> spam
```

```
{'color': 'black', 'age': 5, 'name': 'Pooka'}
>>> spam.setdefault('color', 'white')
'black'
>>> spam
{'color': 'black', 'age': 5, 'name': 'Pooka'}
```

The first time setdefault() is called, the dictionary in spam changes to {'color': 'black', 'age': 5, 'name': 'Pooka'}. The method returns the value 'black' because this is now the value set for the key 'color'. When spam.setdefault('color', 'white') is called next, the value for that key is *not* changed to 'white' because spam already has a key named 'color'.

The setdefault() method is a nice shortcut to ensure that a key exists. Here is a short program that counts the number of occurrences of each letter in a string. Open the file editor window and enter the following code, saving it as *characterCount.py*:

```
message = 'It was a bright cold day in April, and the clocks were striking thirteen.'
count = {}

for character in message:
    count.setdefault(character, 0)
    count[character] = count[character] + 1

print(count)
```

The program loops over each character in the message variable's string, counting how often each character appears. The setdefault() method call ensures that the key is in the count dictionary (with a default value of 0) so the program doesn't throw a KeyError error when count[character] = count[character] + 1 is executed. When you run this program, the output will look like this:

```
{' ': 13, ',': 1, '.': 1, 'A': 1, 'I': 1, 'a': 4, 'c': 3, 'b': 1, 'e': 5, 'd': 3, 'g': 2, 6, 'h': 3, 'k': 2, 'l': 3, 'o': 2, 'n': 4, 'p': 1, 's': 3, 'r': 5, 't': 6, 'w': 2, 'y': 1
```

From the output, you can see that the lowercase letter c appears 3 times, the space character appears 13 times, and the uppercase letter A appears 1 time. This program will work no matter what string is inside the message variable, even if the string is millions of characters long!

PRETTY PRINTING

If you import the pprint module into your programs, you'll have access to the pprint() and pformat() functions that will "pretty print" a dictionary's values. This is helpful when you want a cleaner display of the items in a dictionary than what print() provides. Modify the previous *characterCount.py* program and save it as *prettyCharacterCount.py*.

import pprint

```
message = 'It was a bright cold day in April, and the clocks were striking
thirteen.'
count = {}

for character in message:
    count.setdefault(character, 0)
    count[character] = count[character] + 1
```

pprint.pprint(count)

This time, when the program is run, the output looks much cleaner, with the keys sorted.

```
{' ': 13,
 ',': 1,
 '.': 1,
 'A': 1,
 'I': 1,
 'a': 4,
 'b': 1,
 'c': 3,
 'd': 3,
 'e': 5,
 'g': 2,
 'h': 3,
 'i': 6,
 'k': 2,
 '1': 3,
 'n': 4,
 'o': 2,
 'p': 1,
 'r': 5,
 's': 3,
 't': 6,
 'w': 2,
```

'y': 1}

The pprint.pprint() function is especially helpful when the dictionary itself contains nested lists or dictionaries.

If you want to obtain the prettified text as a string value instead of displaying it on the screen, call pprint.pformat() instead. These two lines are equivalent to each other:

```
pprint.pprint(someDictionaryValue)
print(pprint.pformat(someDictionaryValue))
```

Using Data Structures to Model Real-World Things

Even before the Internet, it was possible to play a game of chess with someone on the other side of the world. Each player would set up a chessboard at their home and then take turns mailing a postcard to each other describing each move. To do this, the players needed a way to unambiguously describe the state of the board and their moves.

In *algebraic chess notation*, the spaces on the chessboard are identified by a number and letter coordinate, as in Figure 5-1.

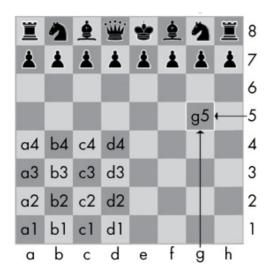


Figure 5-1. The coordinates of a chessboard in algebraic chess notation

The chess pieces are identified by letters: *K* for king, *Q* for queen, *R* for rook, *B* for bishop, and *N* for knight. Describing a move uses the letter of the piece and the coordinates of its destination. A pair of these moves describes what happens in a single turn (with white going first); for instance, the notation *2*. *Nf3 Nc6* indicates that white moved a knight to f3 and black moved a knight to c6 on the second turn of the game.

There's a bit more to algebraic notation than this, but the point is that you can use it to unambiguously describe a game of chess without needing to be in front of a

chessboard. Your opponent can even be on the other side of the world! In fact, you don't even need a physical chess set if you have a good memory: You can just read the mailed chess moves and update boards you have in your imagination.

Computers have good memories. A program on a modern computer can easily store billions of strings like '2. Nf3 Nc6'. This is how computers can play chess without having a physical chessboard. They model data to represent a chessboard, and you can write code to work with this model.

This is where lists and dictionaries can come in. You can use them to model real-world things, like chessboards. For the first example, you'll use a game that's a little simpler than chess: tic-tac-toe.

A TIC-TAC-TOE BOARD

A tic-tac-toe board looks like a large hash symbol (#) with nine slots that can each contain an X, an O, or a blank. To represent the board with a dictionary, you can assign each slot a string-value key, as shown in Figure 5-2.

You can use string values to represent what's in each slot on the board: 'x', '0', or ' ' (a space character). Thus, you'll need to store nine strings. You can use a dictionary of values for this. The string value with the key 'top-R' can represent the top-right corner, the string value with the key 'low-L' can represent the bottom-left corner, the string value with the key 'mid-M' can represent the middle, and so on.

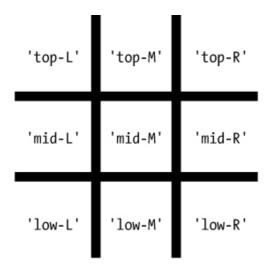


Figure 5-2. The slots of a tic-tactoe board with their corresponding keys

This dictionary is a data structure that represents a tic-tac-toe board. Store this board-as-a-dictionary in a variable named theBoard. Open a new file editor window, and enter the following source code, saving it as *ticTacToe.py*:

The data structure stored in the theBoard variable represents the tic-tactoe board in Figure 5-3.

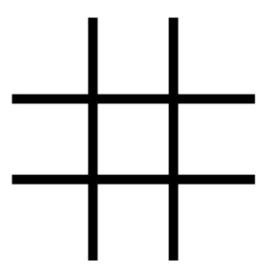


Figure 5-3. An empty tic-tac-toe board

Since the value for every key in theBoard is a single-space string, this dictionary represents a completely clear board. If player X went first and chose the middle space, you could represent that board with this dictionary:

The data structure in the Board now represents the tic-tac-toe board in Figure 5-4.

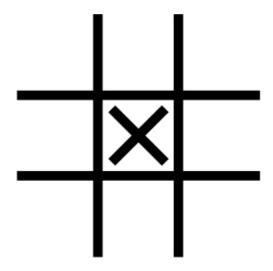


Figure 5-4. The first move

A board where player O has won by placing Os across the top might look like this:

The data structure in the Board now represents the tic-tac-toe board in Figure 5-5.

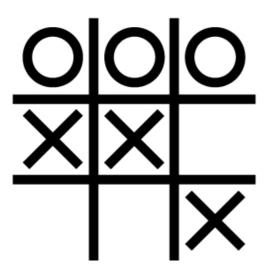


Figure 5-5. Player O wins.

Of course, the player sees only what is printed to the screen, not the contents of variables. Let's create a function to print the board dictionary onto the screen. Make the following addition to *ticTacToe.py* (new code is in bold):

When you run this program, printBoard() will print out a blank tic-tactoe board.

```
| |
-+-+-
| |
-+-+-
```

The printBoard() function can handle any tic-tac-toe data structure you pass it. Try changing the code to the following:

```
theBoard = {'top-L': 'O', 'top-M': 'O', 'top-R': 'O', 'mid-L': 'X', 'mid-M':
'X', 'mid-R': ' ', 'low-L': ' ', 'low-M': ' ', 'low-R': 'X'}

def printBoard(board):
    print(board['top-L'] + '|' + board['top-M'] + '|' + board['top-R'])
    print('-+-+-')
    print(board['mid-L'] + '|' + board['mid-M'] + '|' + board['mid-R'])
    print('-+-+-')
    print(board['low-L'] + '|' + board['low-M'] + '|' + board['low-R'])
printBoard(theBoard)
```

Now when you run this program, the new board will be printed to the screen.

```
0|0|0
-+-+-
X|X|
-+-+-
| |X
```

Because you created a data structure to represent a tic-tac-toe board and wrote code in printBoard() to interpret that data structure, you now have a program that "models" the tic-tac-toe board. You could have organized your data structure differently (for example, using keys like 'TOP-LEFT' instead of 'top-L'), but as long as the code works with your data structures, you will have a correctly working program.

For example, the printBoard() function expects the tic-tac-toe data structure to be a dictionary with keys for all nine slots. If the dictionary you passed was missing, say, the 'mid-L' key, your program would no longer work.

```
0|0|0
-+-+-
Traceback (most recent call last):
    File "ticTacToe.py", line 10, in <module>
        printBoard(theBoard)
    File "ticTacToe.py", line 6, in printBoard
        print(board['mid-L'] + '|' + board['mid-M'] + '|' + board['mid-R'])
KeyError: 'mid-L'
```

Now let's add code that allows the players to enter their moves. Modify the *ticTacToe.py* program to look like this:

```
theBoard = {'top-L': ' ', 'top-M': ' ', 'top-R': ' ', 'mid-L': ' ', 'mid-M': '
', 'mid-R': ' ', 'low-L': ' ', 'low-M': ' ', 'low-R': ' '}
```

```
def printBoard(board):
       print(board['top-L'] + '|' + board['top-M'] + '|' + board['top-R'])
       print('-+-+-')
       print(board['mid-L'] + '|' + board['mid-M'] + '|' + board['mid-R'])
       print('-+-+-')
       print(board['low-L'] + '|' + board['low-M'] + '|' + board['low-R'])
   turn = 'X'
   for i in range(9):
       printBoard(theBoard)
        print('Turn for ' + turn + '. Move on which space?')
       move = input()
       theBoard[move] = turn
       if turn == 'X':
4
            turn = '0'
        else:
            turn = 'X'
   printBoard(theBoard)
```

The new code prints out the board at the start of each new turn ①, gets the active player's move ②, updates the game board accordingly ③, and then swaps the active player ④ before moving on to the next turn.

When you run this program, it will look something like this:

```
-+-+-
 | \cdot |
-+-+-
 III
Turn for X. Move on which space?
mid-M
 -+-+-
 |X|
-+-+-
 | \cdot |
Turn for O. Move on which space?
low-L
 | \cdot |
-+-+-
 |X|
-+-+-
```

```
0 | |
```

```
--snip--

0|0|X
-+-+-

X|X|0
-+-+-
0| |X

Turn for X. Move on which space?

1ow-M

0|0|X
-+-+-

X|X|0
-+-+-

0|X|X
```

This isn't a complete tic-tac-toe game—for instance, it doesn't ever check whether a player has won—but it's enough to see how data structures can be used in programs.

NOTE

If you are curious, the source code for a complete tic-tac-toe program is described in the resources available from http://nostarch.com/automatestuff/.

NESTED DICTIONARIES AND LISTS

Modeling a tic-tac-toe board was fairly simple: The board needed only a single dictionary value with nine key-value pairs. As you model more complicated things, you may find you need dictionaries and lists that contain other dictionaries and lists. Lists are useful to contain an ordered series of values, and dictionaries are useful for associating keys with values. For example, here's a program that uses a dictionary that contains other dictionaries in order to see who is bringing what to a picnic. The totalBrought() function can read this data structure and calculate the total number of an item being brought by all the guests.

Inside the totalBrought() function, the for loop iterates over the key-value pairs in guests ①. Inside the loop, the string of the guest's name is assigned to k, and the dictionary of picnic items they're bringing is assigned to v. If the item parameter exists as a key in this dictionary, it's value (the quantity) is added to numBrought ②. If it does not exist as a key, the get() method returns 0 to be added to numBrought.

The output of this program looks like this:

```
Number of things being brought:
- Apples 7
- Cups 3
- Cakes 0
- Ham Sandwiches 3
- Apple Pies 1
```

This may seem like such a simple thing to model that you wouldn't need to bother with writing a program to do it. But realize that this same totalBrought() function could easily handle a dictionary that contains thousands of guests, each bringing thousands of different picnic items. Then having this information in a data structure along with the totalBrought() function would save you a lot of time!

You can model things with data structures in whatever way you like, as long as the rest of the code in your program can work with the data model correctly. When you first begin programming, don't worry so much about the "right" way to model data. As you gain more experience, you may come up with more efficient models, but the important thing is that the data model works for your program's needs.

SUMMARY

You learned all about dictionaries in this chapter. Lists and dictionaries are values that can contain multiple values, including other lists and dictionaries. Dictionaries are useful because you can map one item (the key) to another (the

value), as opposed to lists, which simply contain a series of values in order. Values inside a dictionary are accessed using square brackets just as with lists. Instead of an integer index, dictionaries can have keys of a variety of data types: integers, floats, strings, or tuples. By organizing a program's values into data structures, you can create representations of real-world objects. You saw an example of this with a tic-tac-toe board.

PRACTICE QUESTIONS

- Q: 1. What does the code for an empty dictionary look like?
- Q: 2. What does a dictionary value with a key 'foo' and a value 42 look like?
- Q: 3. What is the main difference between a dictionary and a list?
- Q: 4. What happens if you try to access spam['foo'] if spam is {'bar': 100}?
- Q: 5. If a dictionary is stored in spam, what is the difference between the expressions 'cat' in spam and 'cat' in spam.keys()?
- Q: 6. If a dictionary is stored in spam, what is the difference between the expressions 'cat' in spam and 'cat' in spam.values()?
- Q: 7. What is a shortcut for the following code?

```
if 'color' not in spam:
    spam['color'] = 'black'
```

Q: 8. What module and function can be used to "pretty print" dictionary values?

PRACTICE PROJECTS

For practice, write programs to do the following tasks.

FANTASY GAME INVENTORY

You are creating a fantasy video game. The data structure to model the player's inventory will be a dictionary where the keys are string values describing the item in the inventory and the value is an integer value detailing how many of that item the player has. For example, the dictionary value {'rope': 1, 'torch': 6, 'gold coin': 42, 'dagger': 1, 'arrow': 12} means the player has 1 rope, 6 torches, 42 gold coins, and so on.

Write a function named displayInventory() that would take any possible "inventory" and display it like the following:

```
Inventory:
12 arrow
42 gold coin
1 rope
6 torch
1 dagger
Total number of items: 62
Hint: You can use a for loop to loop through all the keys in a dictionary.
# inventory.py
stuff = {'rope': 1, 'torch': 6, 'gold coin': 42, 'dagger': 1, 'arrow': 12}
def displayInventory(inventory):
    print("Inventory:")
    item_total = 0
    for k, v in inventory.items():
        # FILL IN THE CODE HERE
    print("Total number of items: " + str(item_total))
displayInventory(stuff)
```

LIST TO DICTIONARY FUNCTION FOR FANTASY GAME INVENTORY

Imagine that a vanquished dragon's loot is represented as a list of strings like this:

```
dragonLoot = ['gold coin', 'dagger', 'gold coin', 'gold coin', 'ruby']
```

Write a function named addToInventory(inventory, addedItems), where the inventory parameter is a dictionary representing the player's inventory (like in the previous project) and the addedItems parameter is a list like dragonLoot. The addToInventory() function should return a dictionary that represents the updated inventory. Note

that the addedItems list can contain multiples of the same item. Your code could look something like this:

```
def addToInventory(inventory, addedItems):
    # your code goes here

inv = {'gold coin': 42, 'rope': 1}
dragonLoot = ['gold coin', 'dagger', 'gold coin', 'gold coin', 'ruby']
inv = addToInventory(inv, dragonLoot)
displayInventory(inv)
```

The previous program (with your displayInventory() function from the previous project) would output the following:

Inventory:

45 gold coin

1 rope

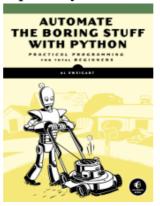
1 ruby

1 dagger

Total number of items: 48



Support the author by purchasing the print & ebook bundle from No Starch Press or separately on Amazon.



Read the author's other Creative Commons licensed Python books.





