



Content Copyright by Pierian Data

Collections Module

The collections module is a built-in module that implements specialized container data types providing alternatives to Python's general purpose built-in containers. We've already gone over the basics: dict, list, set, and tuple.

Now we'll learn about the alternatives that the collections module provides.

Counter

Counter is a *dict* subclass which helps count hashable objects. Inside of it elements are stored as dictionary keys and the counts of the objects are stored as the value.

Let's see how it can be used:

```
In [1]: from collections import Counter
```

Counter() with lists

```
In [2]: lst = [1,2,2,2,2,3,3,3,1,2,1,12,3,2,32,1,21,1,223,1]
Counter(lst)
```

```
Out[2]: Counter({1: 6, 2: 6, 3: 4, 12: 1, 21: 1, 32: 1, 223: 1})
```

Counter with strings

```
In [3]: Counter('aabsbsbsbshhbsbs')
```

```
Out[3]: Counter({'a': 2, 'b': 7, 'h': 3, 's': 6})
```

Counter with words in a sentence

```
In [4]: s = 'How many times does each word show up in this sentence word times each each word'
words = s.split()
Counter(words)
```

```
Out[4]: Counter({'How': 1,
                'does': 1,
                'each': 3,
                'in': 1,
                'many': 1,
                'sentence': 1,
                'show': 1,
                'this': 1,
                'times': 2,
                'up': 1,
                'word': 3})
```

```
In [5]: # Methods with Counter()
c = Counter(words)

c.most_common(2)
```

```
Out[5]: [('each', 3), ('word', 3)]
```

Common patterns when using the Counter() object

<code>sum(c.values())</code>	<code># total of all counts</code>
<code>c.clear()</code>	<code># reset all counts</code>
<code>list(c)</code>	<code># list unique elements</code>
<code>set(c)</code>	<code># convert to a set</code>
<code>dict(c)</code>	<code># convert to a regular dictionary</code>
<code>c.items()</code>	<code># convert to a list of (elem, cnt) pairs</code>
<code>Counter(dict(list_of_pairs))</code>	<code># convert from a list of (elem, cnt) pairs</code>
<code>c.most_common()[::-n-1:-1]</code>	<code># n least common elements</code>
<code>c += Counter()</code>	<code># remove zero and negative counts</code>

defaultdict

`defaultdict` is a dictionary-like object which provides all methods provided by a dictionary but takes a first argument (`default_factory`) as a default data type for the dictionary. Using `defaultdict` is faster than doing the same using `dict.setdefault` method.

A `defaultdict` will never raise a `KeyError`. Any key that does not exist gets the value returned by the default factory.

```
In [6]: from collections import defaultdict
```

```
In [7]: d = {}
```

```
In [8]: d['one']
```

```
-----
KeyError                                Traceback (most recent call last)
<ipython-input-8-07706fc5dc20> in <module>()
----> 1 d['one']

KeyError: 'one'
```

```
In [9]: d = defaultdict(object)
```

```
In [10]: d['one']
```

```
Out[10]: <object at 0x216de27bcf0>
```

```
In [11]: for item in d:  
         print(item)
```

one

Can also initialize with default values:

```
In [12]: d = defaultdict(lambda: 0)
```

```
In [13]: d['one']
```

```
Out[13]: 0
```

namedtuple

The standard tuple uses numerical indexes to access its members, for example:

```
In [14]: t = (12,13,14)
```

```
In [15]: t[0]
```

```
Out[15]: 12
```

For simple use cases, this is usually enough. On the other hand, remembering which index should be used for each value can lead to errors, especially if the tuple has a lot of fields and is constructed far from where it is used. A namedtuple assigns names, as well as the numerical index, to each member.

Each kind of namedtuple is represented by its own class, created by using the namedtuple() factory function. The arguments are the name of the new class and a string containing the names of the elements.

You can basically think of namedtuples as a very quick way of creating a new object/class type with some attribute fields. For example:

```
In [22]: from collections import namedtuple
```

```
In [23]: Dog = namedtuple('Dog', ['age', 'breed', 'name'])  
  
         sam = Dog(age=2, breed='Lab', name='Sammy')  
  
         frank = Dog(age=2, breed='Shepard', name="Frankie")
```

We construct the namedtuple by first passing the object type name (Dog) and then passing a string with the variety of fields as a string with spaces between the field names. We can then call on the various attributes:

```
In [24]: sam
```

```
Out[24]: Dog(age=2, breed='Lab', name='Sammy')
```

```
In [25]: sam.age
```

```
Out[25]: 2
```

```
In [26]: sam.breed
```

```
Out[26]: 'Lab'
```

```
In [27]: sam[0]
```

```
Out[27]: 2
```

Conclusion

Hopefully you now see how incredibly useful the collections module is in Python and it should be your go-to module for a variety of common tasks!