CS1210 Lecture 17 Oct. 1, 2021

- Quiz 2 next Wednesday, Oct. 6, in class
- HW4 due Tuesday
- There was some clear cheating on HW3 - uses of code that's been on the Internet for several years with idiosyncratic bug. So your initial score *might* not be your final score. I haven't addressed the Academic Honesty Policy violations yet …

Last time

- dictionaries, "a few little exercises", HW4

Today

- Tuples (10.26)
- Default and optional arguments to functions
- HW4
  - Sorting for HW4
- zip, generators and iterators

Ch 10.26 - Tuples and tuple assignment

- We've covered several Python **sequence** types: string, list, range. We've been "quietly" using another - **tuple** - when returning multiple values from a function.
- Tuples are just like lists except they are *immutable*!
- Create by typing a comma-separated list of values
  - Standard practice is to enclose the list in parentheses (but that's not required)

```
>>> myTuple = (1,2,3)
>>> myList = [1,2,3]
>>> len(myTuple)
3
>>> myTuple[1]
2

>>> myList[0] = 4
>>> myList
[4, 2, 3]
>>> myTuple[0] = 4
...
TypeError: 'tuple' object does not support item assignment
>>> newList = list(myTuple)
>>> newList
[1, 2, 3]
```

Tuples

```
>>> myTuple = 10, 11, 12
>>> myTuple
(10, 11, 12)
>>> myTuple = (3)
>>> myTuple
3
>>> myTuple = (3,)
>>> myTuple
(3,)
>>> myTuple = ()
```

parentheses not *required* to create a tuple

NO! This is just 3

Tuple of length 1

Empty tuple – length 0

Adding tuples works:

```
>>> myTuple = (1,2,3)
>>> myTuple + myTuple
(1,2,3,1,2,3)
```

But append and other operations that mutate lists do not

Tuples

You don't *need* tuples. You could use lists anywhere instead. But the immutability is sometimes desirable.

E.g. sometimes good to pass tuples as arguments to functions. The function can't then (accidentally or purposely) modify your data!

Easy to create a tuple from a list:

```
>>> importantList = [10, 11, 12]
>>> safeTuple = tuple(importantList)
>>> safeTuple
(10, 11, 12)
>>> result = dangerousFunction(safeTuple)
```

:)

Tuple assignment

We saw in HW1 that you can write things like:

```
>>> a, b = 1, 2                                  (or even (a,b) = (1,2))
>>> a
1
>>> b
2
>>> myMin, myMax = getMinAndMax(2,3,4)
```

presuming getMinAndMax returns two values

This is called tuple assignment (even though you don't really need to think about tuples here)

Tuple assignment

Useful for swapping values

```
>>> a, b = 1, 2
>>> a = b
>>> b = a
```

Does not correctly swap!

```
>>> temp = a
>>> a = b
>>> b = temp
```

Does correctly swap.

But so does

```
>>> a, b = b, a
```

```
          2, 1
    a, b = 2, 1
```

Why does this work?

Remember our old rule:

1. evaluate right hand side
2. update values var names refer t

```
>>> a
2
>>> b
1
```

Tuple assignment

Also useful for “unpacking” results from functions that return list or tuple of results

```
def twiceAndThriceN(n):
    return (2*n, 3*n)                    [2*n, 3*n] also ok

>>> twoN, threeN = twiceAndThriceN(5)
>>> twoN
10
>>> threeN
15
>>> result = twiceAndThriceN(5)          also works, of course
>>> result
(10, 15)
>>> result[0]
10
>>> result[1]
15
```

Default/optional argument values and keyword arguments to functions

Note: although it is used a few places in the book (e.g. 9.18), it's not discussed much. See official Python docs for more info.

With sort/sorted, there are optional arguments that control how the sort works: E.g.

```
>>> myL  = [1,2,3,4,5]
>>> sorted(myL, reverse = True)
[5, 4, 3, 2, 1]
>>> myL = [ ["abc", 100], ["mno", 60] , ["xyz", 80] ]
>>> sorted(myL)
[['abc', 100], ['mno', 60], ['xyz', 80]]
>>> sorted(myL, key = lambda pair: pair[1])
[['mno', 60], ['xyz', 80], ['abc', 100]]
>>> sorted(myL, key = lambda pair: pair[1], reverse = True)
[['abc', 100], ['xyz', 80], ['mno', 60]]
```

The 'key = …' and 'reverse = …' are called keyword arguments and are very convenient.

Default argument values and keyword arguments to functions

```
def foo(a, b = 0, c = True):
    if c == True:
        return a + b
    else:
        return a – b
```

Parameters b and c are optional! They have default values so you don't need to provide more than one argument when calling foo.

```
>>> foo(3)
 3
```

Default argument values and keyword arguments to functions

```
def foo(a, b = 0, c = True):
    if c == True:
        return a + b
    else:
        return a – b
```

```
>>> foo(3,1)
4
>>> foo(3,1,False)
2
```

Default argument values and keyword arguments to functions

```
def foo(a, b = 0, c = True):
    if c == True:
        return a + b
    else:
        return a – 1
```

It sometimes makes code easier to read to use the parameter name when making the function call.

```
>>> foo(3, b=1)
4
>>> foo(3, b=1,c=False)
2
>>> foo(3, c=False, b=1)
2
>>> foo(b=1, c=False, 3)
Error!
>>> foo(c=False, b=1, a = 3)
2
```

Unnamed arguments must come first, before keyword arguments!

And BE CAREFUL. Mutable default arguments are evaluated once, at function definition (different than some other languages). See bar in lec17.py

HW4

It is interesting, and not hard if you do a little bit at a time. Get it working bit by bit.

1. Read the file, storing all the messages and their labels (spam/ham). E.g.
   - Two separate lists: ham list [[‘text’, ‘me’, ‘later!’], [‘…’, …], ...] and spam list [[‘call’, ‘1412’, ‘to’, ‘win’], …] *(I recommend this option)*
   - Or one list [[‘spam’, [‘call’, ‘1412’, ‘to’, ‘win’]], [‘ham’, [‘text’, ‘me’, ‘later!’]], […], …]
   - Note: don’t keep ham/spam label/tag as part of message. I’ve seen people do this and then write special case code to ignore ‘ham’/’spam’ when processing message words in step 2 below – this can yield errors.
2. Create a ham and a spam dictionary. For each message, extract its words, and update spam or ham dictionary of word counts accordingly
   - for ‘text me later!’ increment ‘text’, ‘me’, ‘later’ entries in ham dict
3. Use the two dictionaries to compute and print some statistics
   - get total spam/ham word counts and unique word counts
   - extract most common words from dictionaries
   - print stats

HW4

1. File has some non-Ascii characters.

   use: open(fileName, encoding = 'utf-8')

2. To break line into tokens – individual elements of a line, learn how to use string **split**

   for line in fileStream:

   lineAsList = line.split()

   lec16split.py

3. get rid of extra stuff "…cool!?" learn how to use string **strip** (and/or lstrip, rstrip)
   - I *strongly recommend against* using replace() method
   - *Don't put "" (empty string) as a word in your dictionaries*

for HW4, sorting is very helpful

Why?

You'll have two dictionaries of the form:
{'free': 23, 'you': 50, 'go': 10, 'zoo': 1}

You'll need to extract words in order from most to least frequent?

sort/sorted "work" on dictionaries but do they do what we want?

```
>>> d = {'free': 23, 'you': 50, 'go': 10, 'zoo': 1}
>>> sorted(d)
['free', 'go', 'you', 'zoo']
```

helpful???

For HW4, sorting is helpful

But suppose we have a list of tuples instead of a dictionary.

tl = [('free', 23), ('you', 50), ('go', 10), ('zoo', 1)]

```
>>> sorted(tl)
[('free', 23), ('go', 10), ('you', 50), ('zoo', 1)]
```

Now helpful? Not very.
sorts based on whole tuple

sorted (and sort) have two useful optional arguments:

key: you provide a little function that to apply to item to generate key to use to sort

reverse: provide True if you want list from largest to smallest instead of default of smallest to largest

For HW4, sorting is helpful

sorted (and sort) have two useful optional arguments:

1) key: a little function that is applied to item to generate key to use to sort

```
>>> tl = [('free', 23), ('you', 50), ('go', 10), ('zoo', 1)]
>>>sorted(tl, key = item1)
>>>sorted(tl, key = lambda item: item[1])
```

if function item1 exists

But don't need to write a separate function. 'lambda' allows you to define an (anonymous) function anywhere

```
[('zoo', 1), ('go', 10), ('free', 23), ('you', 50)]
```

yes - better!

So … now also use the other optional argument - reverse

2) reverse: True if you want list from largest to smallest instead of default of smallest to largest

```
>>> sorted(tl, key = lambda item: item[1], reverse = True)
[('you', 50), ('free', 23), ('go', 10), ('zoo', 1)]
```

That's what we want!

lec17.py

For HW4, sorting is very helpful

How do you use this stuff in HW4?

You will create two dictionaries of word counts - one for ham, and one for spam

And you'll want to extract items with highest counts.

1. Saw (three slides back) that

    sort(dict)

didn't quite give us what we needed

2. Saw (in last two slides) that we can usefully sort list of tuples

So … can you make a list of tuples [… (word, count) …] from a dictionary?

- use list(d.items())
- use list(zip(list(d.keys(), list(d.values()))) what is zip?

It's perfectly fine to do things that way in HW4. It turns out you can also sort the dictionary directly using sorted and keywords key and reverse. Experiment and see if you can figure out how ..

Zip function (not in our text)

zip is a convenient function for “re-packaging” sequences

Given two (or more) sequences, zip returns an **iterator** of tuples, where the i-th tuple contains the i-th element from each of the argument sequences.

zip([1,2,3], [‘a’, ‘b’, ‘c’]) is *like* the tuple of tuples:

((1, ‘a’), (2, ‘b’), (3, ‘c’))

but isn’t exactly that…

```
>>> zippedStuff = zip([1,2,3], [‘a’, ‘b’, ‘c’])
>>> zippedStuff
<zip object at 0x1045c2c48>
```

??? It’s an **iterator**

lec17.py

Quick look at iterators (not required)

zip returns a zip object, which is a kind of **iterator.** Iterators (and generators) are important more advanced Python concepts that you are not required to know. Look up at python.org or sites like w3schools.com

```
>>> zippedStuff = zip([1,2,3], [‘a’, ‘b’, ‘c’])
>>> zippedStuff
<zip object at 0x1045c2c48>
```

What can you do with that zip object?

1) Turn it into a list:

```
>>> zippedStuffList = list(zippedStuff)
>>> zippedStuffList
[(1, 'a'), (2, 'b'), (3, 'c')]
```

common/useful

What can you do with iterator object like result of zip?

```
>>> zippedStuff = zip([1,2,3], [‘a’, ‘b’, ‘c’])
>>> zippedStuff
<zip object at 0x1045c2c48>
```

2) use it with for loops

```
>>> for element in zippedStuff:
        print(element)
(1, ‘a’)
(2, ‘b’)
(3, ‘c’)
```

What can you do with iterator object like result of zip?

```
>>> zippedStuff = zip([1,2,3], [‘a’, ‘b’, ‘c’])
>>> zippedStuff
<zip object at 0x1045c2c48>
```

3) ask for the items one by one

```
>>> next(ZippedStuff)
(1, ‘a’)
>>> next(zippedStuff)
(2, ‘b’)
>>> next(zippedStuff)
(3, ‘c’)
>>> next(zippedStuff)
Traceback (most recent call last):
  File "<pyshell#7>", line 1, in <module>
    next(zippedStuff)
StopIteration
```

next is a special built-in func. for working with iterators

What can you do with iterator object like result of zip?

```
>>> zippedStuff = zip([1,2,3], [‘a’, ‘b’, ‘c’])
```

BUT BE CAREFUL! You can only “use” iterator once!

```
>>> for element in zippedStuff:
        print(element)
(1, ‘a’)
(2, ‘b’)
(3, ‘c’)
>>> list(zippedStuff)
[]
>>> next(zippedStuff)
Traceback (most recent call last):
  File "<pyshell#15>", line 1, in <module>
    next(zippedStuff)
StopIteration
>>>
```

Nothing’s left!

Nothing’s left!

zip

Typically, we'll use list(zip(…)) or tuple(zip(…))

zipObj = zip([1, 2, 3], ['uno', 'dos', 'tres'], ['mot', 'hai', 'ba'])

zippedList = list(zipObj) (or zippedTuple = tuple(zipObj))

[(1, 'uno', 'mot'), (2, 'dos', 'hai'), (3, 'tres', 'ba')]

Notice: in example above three lists were zipped (rather than two in earlier examples – can zip any number)

What happens when lists/things being zipped aren't the same length? *Try it* …

Next Time

- Review/last minute advice for HW4
- Image editing/manipulation
- Information about Quiz 2