Performance and Special Issues

Learning CS with Python Series - Day 4

Considering import

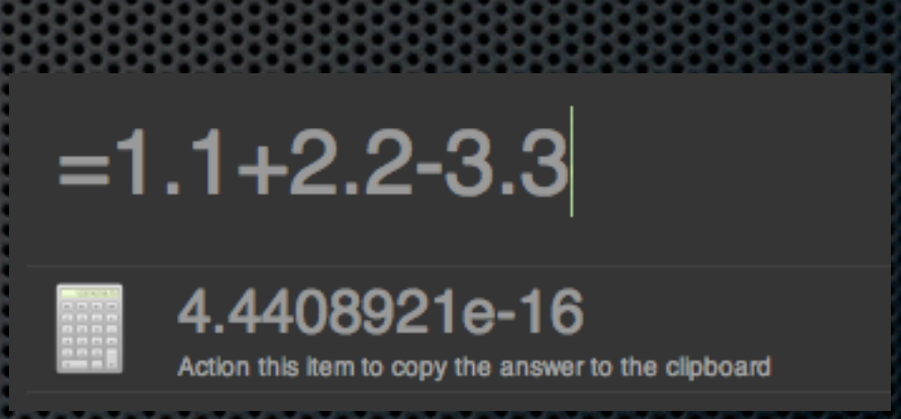
- Depending on the class you are importing, you should consider how you are pulling it into your code
  - import X: makes a reference to the class in the current namespace
  - from X import *: imports every single method and ties it into the current namespace (this could be slow)
  - from X import Z: only imports Z into the current namespace, if you are only using Z, this is what you should do

Removing variables

```
def MyFunc(x):
    somevar = ‘Hello’
    return x+somevar
y = MyFunc(x)
del x
...
```

Decimal Data Type

- Just like 1/3 can't be perfectly defined in base 10, there are numbers in base 2 that can't be perfectly defined



Decimal Data Type

- The decimal data type solves this issue by storing each individual digit in multiple bits
- Important in accounting applications
- Don’t use it unless you need it: slow

```
from decimal import Decimal
print Decimal(‘1.1’)+Decimal(‘2.2’)-
Decimal(‘3.3’)

# 0
```

Recursion

```
import os

def listFiles(mydir):

    print "Files in " + os.path.abspath(mydir) +  ": "

    subdirlist = []

    for item in os.listdir(mydir):

        if os.path.isfile(item):

            print item

        else:

            subdirlist.append(os.path.join(mydir, item))

    for subdir in subdirlist:

        listFiles(subdir)
```

Common Modules

- Some you have seen:
  - os, csv, numpy, scipy
- Popular modules that you have not seen:
  - sys, optparse, re, urllib2, json, xml, sqlite3, ...

Examples

- Economics Scripts:
  - https://github.com/tazzben/EconScripts
- Expense.txt (example of using Decimals, RE, etc)
  - https://github.com/tazzben/expense.txt

Topics I'm not Covering

- Threading
  - http://en.wikipedia.org/wiki/Thread_(computing)
  - http://docs.python.org/library/threading.html
- Cython
  - http://cython.org/
  - http://einstein.drexel.edu/courses/Comp_Phys/General/C_basics/
  - http://www.amazon.com/Practical-Programming-3rd-Steve-Oualline/dp/1565923065/
- Objective-C: http://goo.gl/XN7i9

Jadrian's (and Mine) Program

- We want to analyze the content of tweets to find out if people are followed more are more accurate about their predictions
  - We need to download the tweets
  - Parse the tweets
  - Find the predictions in the tweets

Download the Tweets

- Twitter has an API that spits out JSON data
- We can use urllib2
- Then use the JSON class to parse the data
- Then search the data using regular expression

Get the Data

```
import urllib2
import urllib
...
url = "http://search.twitter.com/search.json?q=" + urllib.quote(search)
req = urllib2.Request(url)
try:
        response = urllib2.urlopen(req)
except URLError, e:
        print 'Something is wrong with that WebHook'
        print e.read()
        sys.exit()
else:
        the_page = response.read().strip()
        return the_page
```

Parse the Data

```
import json
...
structeddata=json.loads(data)
```

```
{
  "completed_in":0.031,
  "max_id":12207846184098201 6,
  ...
  "results":[
    {
      "from_user_id":14093707,
      "id":12203244826669875 2,
      ...
      "profile_image_url":"http://...normal.jpg",
      "text":"Reminder: Blue Angels practice today“
      "to_user_id":null,
      "to_user_id_str":null
    }
  ]
}
```

Search the Data

```
import re
...
pattern = re.compile(u"(bears)([^\\.]+)(stink)",re.U)
find = pattern.search(line)
```