Automated Data Collection Exercises

In [33]:
```python
import urllib
from bs4 import BeautifulSoup
```

Exercise 1

In [ ]:
```python
url = "https://www.google.com/finance?q=AAPL"
sourcecode = urllib.urlopen(url)
soup = BeautifulSoup(sourcecode, 'html.parser')

content = soup.find("span",{"class", "pr"})
print content.text
```

Exercise 2

In [ ]:
```python
url = "https://news.google.com/"
sourcecode = urllib.urlopen(url)
soup = BeautifulSoup(sourcecode, 'html.parser')

contents = soup.findAll("span",{"class", "titletext"})
for content in contents:
    print content.text
```

Exercise 3

In [ ]:
```python
url = "https://www.whitehouse.gov/briefing-room/press-briefings"
sourcecode = urllib.urlopen(url)
soup = BeautifulSoup(sourcecode, 'html.parser')

contents = soup.findAll("a",{"href" : re.compile('/the.*')})
for link in contents:
    print link['href']
```

Exercise 4

```
In [ ]:  stocks = ["AAPL","GOOG","MSFT"]
         data=[]
         for stock in stocks:
             url = "https://www.google.com/finance?q="+stock
             sourcecode = urllib.urlopen(url)
             soup = BeautifulSoup(sourcecode, 'html.parser')
             content = soup.find("span",{"class", "pr"})
             data.append(content.text)
```

```
In [ ]:  for stock in data:
             print stock
```