

HTML Scraping

The url we will use is <http://econpy.pythonanywhere.com/ex/001.html> (<http://econpy.pythonanywhere.com/ex/001.html>).

We want to scrap the buyers names and prices in the url and store the data in a csv file.

```
In [1]: from lxml import html
import requests
import numpy as np
import csv
```

```
In [2]: pages=np.array(range(1,6,1))
buyers=[]
prices=[]
```

```
In [3]: for j in pages:
    i = str(j)
    text1 = "http://econpy.pythonanywhere.com/ex/00"
    text2 = ".html"
    url = text1 + i + text2
    page = requests.get(url)
    tree = html.fromstring(page.content)
    buyers_new = tree.xpath('//div[@title="buyer-name"]/text()')
    buyers = buyers + buyers_new
    prices_new = tree.xpath('//span[@class="item-price"]/text()')
    prices = prices + prices_new
```

```
In [4]: table = np.transpose([buyers,prices])
file = open("buyers_prices.csv", "w")
with file:
    writer = csv.writer(file)
    writer.writerow(["buyers", "prices"])
    writer.writerows(table)
```