Fetching Webpages with the Requests Library To Demonstrate

Fetching Webpages with the Requests Library To Demonstrate

Requests is good for times when you just need to fetch a webpage and do something with the raw HTML. It doesn't give you a whole lot more, but does do that incredibly well! The Requests homepage has lots of good examples and full documentation.

Fetching Webpages with the Requests Library To Demonstrate

Requests is good for times when you just need to fetch a webpage and do something with the raw HTML. It doesn't give you a whole lot more, but does do that incredibly well! The Requests homepage has lots of good examples and full documentation.

In [1]:

```python
import requests
r = requests.get('http://www.imdb.com/name/nm0000125/')
```

Fetching Webpages with the Requests Library To Demonstrate

Requests is good for times when you just need to fetch a webpage and do something with the raw HTML. It doesn't give you a whole lot more, but does do that incredibly well! The Requests homepage has lots of good examples and full documentation.

In [1]:

```
import requests
r = requests.get('http://www.imdb.com/name/nm0000125/')
```

You can make sure the request actually worked (ie. HTTP status code 200)

Fetching Webpages with the Requests Library To Demonstrate

Requests is good for times when you just need to fetch a webpage and do something with the raw HTML. It doesn't give you a whole lot more, but does do that incredibly well! The Requests homepage has lots of good examples and full documentation.

In [1]:

```
import requests
r = requests.get('http://www.imdb.com/name/nm0000125/')
```

You can make sure the request actually worked (ie. HTTP status code 200)

In [2]:

```
r.status_code
```

```
Out[2]:  200
```

You can check what type of content the webpage returned (ie. text, json, csv, etc)

You can check what type of content the webpage returned (ie. text, json, csv, etc)

```
In [3]: r.headers['content-type']
```

```
Out[3]: 'text/html;charset=UTF-8'
```

You can check what type of content the webpage returned (ie. text, json, csv, etc)

```
In [3]: r.headers['content-type']
```

```
Out[3]: 'text/html;charset=UTF-8'
```

You can check the character set (sure hope it is utf-8!)

You can check what type of content the webpage returned (ie. text, json, csv, etc)

```
In [3]: r.headers['content-type']
```

```
Out[3]: 'text/html;charset=UTF-8'
```

You can check the character set (sure hope it is utf-8!)

```
In [18]: r.encoding
```

```
Out[18]: 'UTF-8'
```

Of course, you can get the actual HTML text too!

Of course, you can get the actual HTML text too!

```
In [4]: r.text[0:200]
```

```
Out[4]: u'\n\n\n\n<!DOCTYPE html>\n<html\nxmlns:og="http://ogp.me/ns#"\nxmln
        s:fb="http://www.facebook.com/2008/fbml">\n    <head>\n        <meta
        charset="utf-8">\n        <meta http-equiv="X-UA-Compatible" content
        ="IE=ed'
```

Of course, you can get the actual HTML text too!

```
In [4]: r.text[0:200]
```

```
Out[4]: u'\n\n\n\n<!DOCTYPE html>\n<html\nxmlns:og="http://ogp.me/ns#"\nxmln
        s:fb="http://www.facebook.com/2008/fbml">\n    <head>\n        <meta
        charset="utf-8">\n        <meta http-equiv="X-UA-Compatible" content
        ="IE=ed'
```

But sometimes you just want the header content, for instance if you want to resolve redirects without actually downloading the full webpage content.

Of course, you can get the actual HTML text too!

In [4]:

```
r.text[0:200]
```

Out[4]:

```
u'\n\n\n\n<!DOCTYPE html>\n<html\nxmlns:og="http://ogp.me/ns#"\nxmln
s:fb="http://www.facebook.com/2008/fbml">\n    <head>\n        <meta
charset="utf-8">\n        <meta http-equiv="X-UA-Compatible" content
="IE=ed'
```

But sometimes you just want the header content, for instance if you want to resolve redirects without actually downloading the full webpage content.

In [5]:

```
r = requests.head("http://feeds.foxnews.com/~r/foxnews/national/~3/vZ_mHFtNHa
r.url
```

Out[5]:

```
u'http://www.foxnews.com/us/2014/11/21/nypd-rookie-kills-unarmed-man-
in-brooklyn.html?utm_source=feedburner&utm_medium=feed&utm_campaign=F
eed%253A+foxnews%252Fnational+%2528Internal+-+US+Latest+-+Text%2529'
```