

Simple Web Scraping with Python

First: Install BeautifulSoup

If you have pip, just run `pip install beautifulsoup4`

Otherwise, download [BeautifulSoup here](#), then extract the file and move the bs4 to wherever your project is located.

Python Basics

Create a string: `activity = 'scraping'`

Create an integer: `year = 2014`

Arrays (in python, called lists): `some_numbers = [1, 2, 3, 4, 5]`

Functions

```
def say_hello(name): print 'hello ' + name
```

Examples

Get the contents of a web page

```
import urllib

url = 'http://alibaba.com/catalogs/products/CID100003006'

data = urllib.urlopen(start_url).read()

print data
```

Print the text of all the links in a web page

```
import urllib from bs4 import BeautifulSoup

url = 'http://alibaba.com/catalogs/products/CID100003006'

data = urllib.urlopen(url).read()
```

```
soup = BeautifulSoup(data)
```

```
for link in soup.select('a'): print link.text.strip()
```

Resources

- [Learn Python The Hard Way](#)
- [Python Docs](#)
- [BeautifulSoup](#)
- [CSS Selectors](#)
- [Scrapy](#)
- [Casper](#)