# 使用Pyquery

## 初始化

像Beautiful Soup一样，初始化pyquery的时候，也需要传入HTML文本来初始化一个PyQuery对象。它的初始化有很多方式，我们一一介绍：

### 字符串初始化

实例：

参考：

##### 4.3.4.1.py

### URL初始化

初始化的参数不仅可以以字符串的形式传递，还可以传入网页的URL，此时只需要指定参数为url即可：  
参考：

##### 4.3.4.2.py

### 文件初始化

就是传入本地文件名。

参考：

##### 4.3.4.3.py

## 基本CSS选择器

用一个实例来感受下pyquery的css选择器的使用：  
参考：

##### 4.3.4.4.py

## 查找节点

我们介绍一些常用的查询函数，这里和jQuery中函数的用法完全相同

### 子节点

查找子节点需要用到find()方法，此时传入的参数是CSS选择器。

参考：

##### 4.3.4.5.py

### 父节点

我们可以用parent()方法来获取某个节点的父节点。

参考：

##### 4.3.4.6.py

### 兄弟节点

使用siblings()方法

参考：

##### 4.3.4.7.py

## 遍历

我们发现pyquery的选择结果可能是多个节点，可能是单个节点，类型都是PyQuery的类型，并不是像Beautiful Soup的那样的列表

对于单个节点来说：可以直接打印输出，也可以直接转化成字符串：

参考：

##### 4.3.4.8.py

但是对于多个节点来说，我们需要遍历来获取了。我们调用items()方法。

参考：

##### 4.3.4.9.py

## 获取信息

### 获取属性

在提取到某个PyQuery类型的节点后，就可以调用attr()方法来获取属性。

参考：

print(a.attr('href'))或者  
print(a.attr.href)

##### 4.3.4.10.py

### 获取文本

调用text()方法来实现。

参考：

##### 4.3.4.11.py

如果想要获取这个节点内部的HTML文本，就要用html()方法

## 节点操作

Pyquery提供了一系列方法来对节点进行动态修改，比如为某个节点添加一个class，移除某个节点等，这些操作有时候会为提取信息带来极大的便利。

### add\_class和remove\_class

参考:

##### 4.3.4.12.py

add\_class和remove\_class()方法可以动他改变节点的class属性

### attr,text和html

除了操作class这个属性外，我们也可以用attr()方法对属性进行操作。此外，还可以用text()和html()方法来改变节点内部的内容。

参考：

##### 4.3.4.13.py

attr()，text()，html的使用很灵活

### remove()