

pyquery

初始化

字符串初始化

```
In [ ]: html = '''
<div>
    <ul>
        <li class="item-0">first item</li>
        <li class="item-1"><a href="link2.html">second item</a></li>
        <li class="item-0 active"><a href="link3.html"><span class="bol
d">third item</span></a></li>
        <li class="item-1 active"><a href="link4.html">fourth item</a><
/li>
        <li class="item-0"><a href="link5.html">fifth item</a></li>
    </ul>
</div>
'''
from pyquery import PyQuery as pq
doc = pq(html)
print(doc('li'))
```

URL初始化

```
In [ ]: from pyquery import PyQuery as pq
doc = pq(url='http://www.baidu.com')
print(doc('head'))
```

文件初始化

```
In [ ]: from pyquery import PyQuery as pq
doc = pq(filename='demo.html')
print(doc('li'))
```

基本CSS选择器

```
In [ ]: html = '''
<div id="container">
    <ul class="list">
        <li class="item-0">first item</li>
        <li class="item-1"><a href="link2.html">second item</a></li>
        <li class="item-0 active"><a href="link3.html"><span class="bol
d">third item</span></a></li>
        <li class="item-1 active"><a href="link4.html">fourth item</a><
/li>
        <li class="item-0"><a href="link5.html">fifth item</a></li>
    </ul>
</div>
'''
```

```

        </ul>
    </div>
'''
from pyquery import PyQuery as pq
doc = pq(html)
print(doc('#container .list li'))

```

查找元素

子元素

```

In [ ]: html = '''
<div id="container">
    <ul class="list">
        <li class="item-0">first item</li>
        <li class="item-1"><a href="link2.html">second item</a></li>
        <li class="item-0 active"><a href="link3.html"><span class="bol
d">third item</span></a></li>
        <li class="item-1 active"><a href="link4.html">fourth item</a><
/li>
        <li class="item-0"><a href="link5.html">fifth item</a></li>
    </ul>
</div>
'''
from pyquery import PyQuery as pq
doc = pq(html)
items = doc('.list')
print(type(items))
print(items)
lis = items.find('li')
print(type(lis))
print(lis)

```

```

In [ ]: lis = items.children()
print(type(lis))
print(lis)

```

```

In [ ]: lis = items.children('.active')
print(lis)

```

父元素

```

In [ ]: html = '''
<div id="container">
    <ul class="list">
        <li class="item-0">first item</li>
        <li class="item-1"><a href="link2.html">second item</a></li>
        <li class="item-0 active"><a href="link3.html"><span class="bol
d">third item</span></a></li>
        <li class="item-1 active"><a href="link4.html">fourth item</a><
/li>
        <li class="item-0"><a href="link5.html">fifth item</a></li>
    </ul>
</div>
'''

```

```

from pyquery import PyQuery as pq
doc = pq(html)
items = doc('.list')
container = items.parent()
print(type(container))
print(container)

```

```

In [ ]: html = '''
<div class="wrap">
  <div id="container">
    <ul class="list">
      <li class="item-0">first item</li>
      <li class="item-1"><a href="link2.html">second item</a></li>
    >
      <li class="item-0 active"><a href="link3.html"><span class=
"bold">third item</span></a></li>
      <li class="item-1 active"><a href="link4.html">fourth item<
/a></li>
      <li class="item-0"><a href="link5.html">fifth item</a></li>
    </ul>
  </div>
</div>
'''

from pyquery import PyQuery as pq
doc = pq(html)
items = doc('.list')
parents = items.parents()
print(type(parents))
print(parents)

```

```

In [ ]: parent = items.parents('.wrap')
print(parent)

```

兄弟元素

```

In [ ]: html = '''
<div class="wrap">
  <div id="container">
    <ul class="list">
      <li class="item-0">first item</li>
      <li class="item-1"><a href="link2.html">second item</a></li>
    >
      <li class="item-0 active"><a href="link3.html"><span class=
"bold">third item</span></a></li>
      <li class="item-1 active"><a href="link4.html">fourth item<
/a></li>
      <li class="item-0"><a href="link5.html">fifth item</a></li>
    </ul>
  </div>
</div>
'''

from pyquery import PyQuery as pq
doc = pq(html)
li = doc('.list .item-0.active')
print(li.siblings())

```

```

In [ ]: html = '''
<div class="wrap">

```

```

    <div id="container">
        <ul class="list">
            <li class="item-0">first item</li>
            <li class="item-1"><a href="link2.html">second item</a></li>
        </ul>
        <li class="item-0 active"><a href="link3.html"><span class="bold">third item</span></a></li>
        <li class="item-1 active"><a href="link4.html">fourth item</a></li>
        <li class="item-0"><a href="link5.html">fifth item</a></li>
    </ul>
</div>
'''
from pyquery import PyQuery as pq
doc = pq(html)
li = doc('.list .item-0.active')
print(li.siblings('.active'))

```

遍历

单个元素

```

In [ ]: html = '''
    <div class="wrap">
        <div id="container">
            <ul class="list">
                <li class="item-0">first item</li>
                <li class="item-1"><a href="link2.html">second item</a></li>
            </ul>
            <li class="item-0 active"><a href="link3.html"><span class="bold">third item</span></a></li>
            <li class="item-1 active"><a href="link4.html">fourth item</a></li>
            <li class="item-0"><a href="link5.html">fifth item</a></li>
        </div>
    </div>
    '''
    from pyquery import PyQuery as pq
    doc = pq(html)
    li = doc('.item-0.active')
    print(li)

```

```

In [ ]: html = '''
    <div class="wrap">
        <div id="container">
            <ul class="list">
                <li class="item-0">first item</li>
                <li class="item-1"><a href="link2.html">second item</a></li>
            </ul>
            <li class="item-0 active"><a href="link3.html"><span class="bold">third item</span></a></li>
            <li class="item-1 active"><a href="link4.html">fourth item</a></li>
            <li class="item-0"><a href="link5.html">fifth item</a></li>
        </ul>
    </div>
    '''
    from pyquery import PyQuery as pq
    doc = pq(html)
    li = doc('.item-0.active')
    print(li)

```

```

        </div>
    </div>
'''
from pyquery import PyQuery as pq
doc = pq(html)
lis = doc('li').items()
print(type(lis))
for li in lis:
    print(li)

```

获取信息

获取属性

```

In [ ]: html = '''
<div class="wrap">
    <div id="container">
        <ul class="list">
            <li class="item-0">first item</li>
            <li class="item-1"><a href="link2.html">second item</a></li>
        >
            <li class="item-0 active"><a href="link3.html"><span class=
"bold">third item</span></a></li>
            <li class="item-1 active"><a href="link4.html">fourth item<
/a></li>
            <li class="item-0"><a href="link5.html">fifth item</a></li>
        </ul>
    </div>
</div>
'''
from pyquery import PyQuery as pq
doc = pq(html)
a = doc('.item-0.active a')
print(a)
print(a.attr('href'))
print(a.attr.href)

```

获取文本

```

In [ ]: html = '''
<div class="wrap">
    <div id="container">
        <ul class="list">
            <li class="item-0">first item</li>
            <li class="item-1"><a href="link2.html">second item</a></li>
        >
            <li class="item-0 active"><a href="link3.html"><span class=
"bold">third item</span></a></li>
            <li class="item-1 active"><a href="link4.html">fourth item<
/a></li>
            <li class="item-0"><a href="link5.html">fifth item</a></li>
        </ul>
    </div>
</div>
'''
from pyquery import PyQuery as pq

```

```
doc = pq(html)
a = doc('.item-0.active a')
print(a)
print(a.text())
```

获取HTML

```
In [ ]: html = '''
<div class="wrap">
  <div id="container">
    <ul class="list">
      <li class="item-0">first item</li>
      <li class="item-1"><a href="link2.html">second item</a></li>
    </ul>
  </div>
</div>
'''
from pyquery import PyQuery as pq
doc = pq(html)
li = doc('.item-0.active')
print(li)
print(li.html())
```

DOM操作

addClass、removeClass

```
In [ ]: html = '''
<div class="wrap">
  <div id="container">
    <ul class="list">
      <li class="item-0">first item</li>
      <li class="item-1"><a href="link2.html">second item</a></li>
    </ul>
  </div>
</div>
'''
from pyquery import PyQuery as pq
doc = pq(html)
li = doc('.item-0.active')
print(li)
li.removeClass('active')
print(li)
li.addClass('active')
```

```
print(li)
```

attr、css

```
In [ ]: html = '''
<div class="wrap">
  <div id="container">
    <ul class="list">
      <li class="item-0">first item</li>
      <li class="item-1"><a href="link2.html">second item</a></li>
    >
      <li class="item-0 active"><a href="link3.html"><span class="
"bold">third item</span></a></li>
      <li class="item-1 active"><a href="link4.html">fourth item<
/a></li>
      <li class="item-0"><a href="link5.html">fifth item</a></li>
    </ul>
  </div>
</div>
'''
from pyquery import PyQuery as pq
doc = pq(html)
li = doc('.item-0.active')
print(li)
li.attr('name', 'link')
print(li)
li.css('font-size', '14px')
print(li)
```

remove

```
In [ ]: html = '''
<div class="wrap">
  Hello, World
  <p>This is a paragraph.</p>
</div>
'''
from pyquery import PyQuery as pq
doc = pq(html)
wrap = doc('.wrap')
print(wrap.text())
wrap.find('p').remove()
print(wrap.text())
```

其他DOM方法

<http://pyquery.readthedocs.io/en/latest/api.html>

伪类选择器

```
In [ ]: html = '''
<div class="wrap">
  <div id="container">
```

```

        <ul class="list">
            <li class="item-0">first item</li>
            <li class="item-1"><a href="link2.html">second item</a></li>
        >
            <li class="item-0 active"><a href="link3.html"><span class=
"bold">third item</span></a></li>
            <li class="item-1 active"><a href="link4.html">fourth item<
/a></li>
            <li class="item-0"><a href="link5.html">fifth item</a></li>
        </ul>
    </div>
</div>
'''
from pyquery import PyQuery as pq
doc = pq(html)
li = doc('li:first-child')
print(li)
li = doc('li:last-child')
print(li)
li = doc('li:nth-child(2)')
print(li)
li = doc('li:gt(2)')
print(li)
li = doc('li:nth-child(2n)')
print(li)
li = doc('li:contains(second)')
print(li)

```

更多CSS选择器可以查看 http://www.bejson.com/apidoc/jquery/next_1.html

官方文档

<http://pyquery.readthedocs.io/>