



2.2 BeautifulSoup的安装

深圳信息职业技术学院

Shenzhen Institute Of Information Technology

教师：黄锐军

目录

COMPANY

2.2.1

BeautifulSoup的安装

2.2.2

BeautifulSoup装载HTML文档



PART ONE

BeautifulSoup的安装

BeautifulSoup的安装



BeautifulSoup是第三方的工具，它包含在一个名称为bs4的文件包中，需要另外安装。安装也很简单，在命令行窗体中进入Python的安装目录（例如Python在c:\Python36），再进入Scripts子目录，找到pip程序，执行：

```
pip install bs4
```

我们可看到安装过程进行，成功安装后就可以在Python的命令窗体中执行语句：

```
from bs4 import BeautifulSoup
```

如果这条语句执行没有报错，就说明安装成功了。

PART TWO

BeautifulSoup装载HTML文档



例2-2-2 : BeautifulSoup装载有缺失的HTML文档

```
from bs4 import BeautifulSoup

doc="""
<title>有缺失元素的HTML文档</title>
<div>
<A href='one.html'>one</a>
<p>
<a href='two.html'>two</a>
</DIV>
"""

soup=BeautifulSoup(doc,"lxml")
s=soup.prettify()
print(s)
```



```
<html>
<head>
  <title>
    有缺失元素的HTML文档
  </title>
</head>
<body>
  <div>
    <a href="one.html">
      one
    </a>
    <p>
      <a href="two.html">
        two
      </a>
    </p>
  </div>
</body>
</html>
```





由此可见在HTML缺失<html>根元素时BeautifulSoup会自动补上；它发现<title>元素没有<head>元素后自动增加用<head>元素包含<title>元素；它发现没有<body>元素也自动补上；它还把one改成one；还有它把：

```
<p>
```

```
<a href='two.html'>two</a>
```

改成：

```
<p>
```

```
<a href="two.html">
```

```
two
```

```
</a>
```

```
</p>
```




THANK YOU