法律声明

口本淉件包括：渍示文稿，示例，代码，题库，视频和声 音等，小象羊院拥有完全知识产权的权利；只限于善意 争习者在本淉程使用，不得在淉程范围外向任何笫三方 散播。任何其他人或机构不得盗版、复制、仿造其中的 创意，我们将保留一切通过法律手段追究违反者的权利。

回•.•

口淉程详情请咨询

* 微信公众号：大数据分析挖掘
* 新浪微博： ChinaHadoop



互联网新技术在线教育领航者

E山蛊学陲

ChinaHadoop.cn



零基础Python入门

－－梁斌

互联网新技术在线教育领航者

E山蛊学陲

ChinaHadoop.cn

第九讲





AIR QUALITY INDEX I

空气质量指数计算5.0

互联网新技术在线教育领航者

E山蛊学陲

ChinaHadoop.cn

案例描述

* 为了能有效地提取并利用网络信息并工作提高效率，出现了网络爬虫
* 利用网络爬虫实时获取城市的空气质量

互联网新技术在线教育领航者

E山蛊学陲

ChinaHadoop.cn

网络爬虫

* 自动抓取互联网信息的程序；利用互联网数据进行分析、开发产品
* 步骤：

1 . 通过网络链接获取网页内容

2. 对获得的网页内容进行处理

URL

URL

URL/

•

｀

/.

、、

~/\; 守

咽-~l

URL -.









互联网新技术在线教育领航者

E山蛊学陲

ChinaHadoop.cn

requests模块

* requests模块是—个简洁且简单的处理HTTP请求的工具
* 支持非常丰富的链接访问功能，包括URL获取，HTTP会话，Cookie记录等
* requests网页请求



post()

• requests对象属性

对应HTTP的GET方式

对应HTTP的POST方式，用于传逼用户数据



HTTP请求的返回状态，200表示连接成功， 400表示失败

HTTP相应内容的宇符串形式，即url对应的 页面内容

• 更多方法参考： <http://docs.python-requests.org/>

status code

text

互联网新技术在线教育领航者

E山蛊学陲

ChinaHadoop.cn

Next?

* 是否有更高效的处理和解析HTML的库？
* beautifulsoup4



） s



互联网新技术在线教育领航者

E山蛊学陲

ChinaHadoop.cn

疑问

口问题答疑： <http://www.xxwenda.com/>

·可邀清老师或者其他人回答问题

小象问答邀请@Robin\_TY回答问题



互联网新技术在线教育领航者

E山蛊学陲

ChinaHadoop.cn

联系我们

小象学院：互联网新技术在线教育领航者 －微信公众号：小象

－亲Ii讳良很女博： ChinaHadoop



互联网新技术在线教育领航者

E山蛊学陲

ChinaHadoop.cn