

# “网络爬虫”课程内容导学

WS00

---



嵩天

[www.python123.org](http://www.python123.org)

# The Website is the API ...



## Requests

自动爬取HTML页面  
自动网络请求提交

## robots.txt

网络爬虫排除标准



## Beautiful Soup

解析HTML页面



## Re

正则表达式详解  
提取页面关键信息



## Scrapy\*

\*网络爬虫原理介绍  
\*专业爬虫框架介绍



## Projects

实战项目A/B

掌握定向网络数据爬取和网页解析的基本能力

**python**  
弹指之间 · 享受创新

# Python网络爬虫与信息提取

04X -Tian

# 内容组织

- 全课程包括：8个内容单元和4个实例单元，共12个单元
- 全课程总长4周，每周3个单元，前3周是必修，第4周是选修
- 单元之间有课间休息
- 大部分单元内容都包括程序实践

# 本课程实例

- 京东商品页面的爬取
- 亚马逊商品页面的爬取
- 百度/360搜索关键字提交
- 网络图片的爬取和存储
- IP地址归属地的自动查询
- 中国大学排名定向爬虫
- 淘宝商品比价定向爬虫
- 股票数据定向爬虫
- 股票数据专业爬虫
- 表情包专业爬虫