

# Python数据分析与挖掘实战

## 1、爬虫防屏蔽手段之代理服务器 实战 韦玮

# 自我介绍

- 天善商业智能和大数据社区 Python 讲师 – 韦玮
- 天善社区 ID - 韦玮
- <https://www.hellobi.com> – 学习过程中有任何相关的问题都可以提到技术社区 Python 版块。

# 本课概要

- 作业讲解
- 什么是代理服务器
- 使用代理服务器进行爬取网页实战

# 作业讲解

上节课的作业是如何爬取CSDN首页的所有文章。

接下来我们为大家讲解一下如何实现。

# 什么是代理服务器

所谓代理服务器，是一个处于我们与互联网中间的服务器，如果使用代理服务器，我们浏览信息的时候，先向代理服务器发出请求，然后由代理服务器向互联网获取信息，再返回给我们。

# 使用代理服务器进行爬取网页实战

使用代理服务器进行信息爬取，可以很好的解决IP限制的问题。

# Python数据分析与挖掘实战

## 2、图片爬虫实战

韦玮

# 自我介绍

- 天善商业智能和大数据社区 Python 讲师 – 韦玮
- 天善社区 ID - 韦玮
- <https://www.hellobi.com> – 学习过程中有任何相关的问题都可以提到技术社区 Python 版块。



# 本课概要

- 什么是图片爬虫
- 淘宝图片爬虫实战

# 什么是图片爬虫

所谓图片爬虫，即是从互联网中自动把对方服务器上的图片爬下来的爬虫程序。

# 淘宝图片爬虫实战

接下来我们为大家通过实战讲解如何爬取淘宝的图片。

# 作业

试试把千图网（ <http://www.58pic.com/> ）某个频道的所有图片爬下来，要高清版的哦。