



网络爬虫介绍

讲师：黄老师

1. 认识爬虫。
2. 开发工具准备。
3. Http协议和Chrome抓包工具。

1. 爬虫的实际例子。
2. 什么是网络爬虫。
3. 为什么用Python写爬虫程序。

1. 搜索引擎（ 百度或谷歌等 ）
2. 伯乐在线。
3. 惠惠购物助手。
4. 数据分析。
5. 抢票软件等。

1. 通俗理解：爬虫是一个模拟人类请求网站行为的程序。可以自动请求网页、并把数据抓取下来，然后使用一定的规则提取有价值的信息。

2. 专业介绍：

<https://baike.baidu.com/item/%E7%BD%91%E7%BB%9C%E7%88%AC%E8%99%AB/5162711?fr=aladdin>

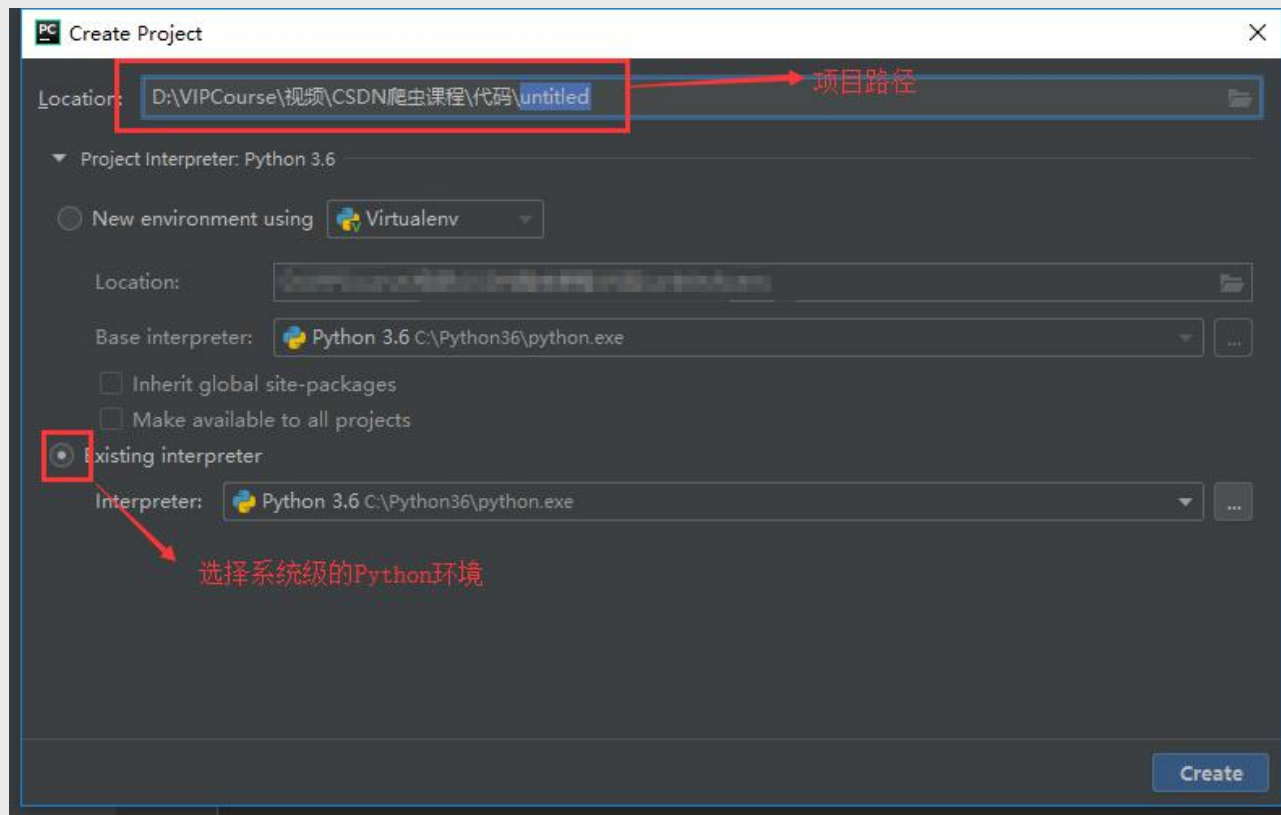
1. PHP：PHP是世界是最好的语言，但他天生不是做这个的，而且对多线程、异步支持不是很好，并发处理能力弱。爬虫是工具性程序，对速度和效率要求比较高。
2. Java：生态圈很完善，是Python爬虫最大的竞争对手。但是Java语言本身很笨重，代码量很大。重构成本比较高，任何修改会导致代码大量改动。爬虫经常要修改采集代码。
3. C/C++：运行效率是无敌的。但是学习和开发成本高。写个小爬虫程序可能要大半天时间。
4. Python：语法优美、代码简洁、开发效率高、支持的模块多。相关的HTTP请求模块和HTML解析模块非常丰富。还有Scrapy和Scrapy-redis框架让我们开发爬虫变得异常简单。

安装好爬虫开发所需软件

1. 安装好Python，需要3.6以上的版本。
2. 安装好Pycharm 2019 Community（社区版，免费的）。
3. 安装好Chrome软件，需要73以上的版本。

1. 下载地址：<https://www.python.org/>
2. 或者使用Anaconda：<https://www.anaconda.com/>
3. 哪个软件无所谓，只要是Python3.6以上的版本就可以了。

1. Pycharm Community (社区版) : 对于不需要做web开发, 不需要使用Professional (专业版), 使用Community版就足够了, 并且这个软件功能很强大, 而且免费!
2. 下载地址: <https://www.jetbrains.com/pycharm/download/#section=windows>
3. 创建项目:



1. 下载地址：<https://www.google.com/chrome/>
2. 如果是下载不了，说明被墙了，需要翻墙。

EDU

CSDN学院 IT实战派

