

## 实验 1: 金融数据获取实验指南

### 一、实验目的

1. 了解 python 开发环境构建方法, 学会环境变量配置和 python 依赖库安装。
2. 了解 Scrapy 框架各组件功能, 学会基于教程构建 Scrapy 爬虫程序。

### 二、实验步骤

#### 1. Python 环境安装

建议安装 python 最新版

MAC 或者 Linux 的同学特别注意, 系统本身自带 python2.7 版本, 不建议大家使用 python2 了, 目前 python2 已经不再维护。3 和 2 会有一些兼容性问题, 后续实验都是基于 python3 的。这一点大家安装 python 的时候特别注意。

##### 1. 安装原生 Python 环境

直接在官网下载对应平台 (Linux、Windows、Mac OS) 最新版本的 python, 参考文档和教程安装。

Python 官网: <https://www.python.org>

Python 文档地址: <https://docs.python.org/3/>

##### 2. 安装 Anaconda3 或者 Miniconda3 (推荐)

Anaconda 提供了很多基础依赖库, 并支持自己创建多个不同虚拟环境, 且通过 conda install 命令能自动安装相关依赖库, 十分方便。

Anaconda 下载地址 <https://www.anaconda.com/distribution/> 下载后直接根据提示安装即可。

Miniconda3 和 Anaconda3 的区别在于减少了一些不必要的包, 需要时可以通过 conda install xx 再安装即可。容量大概 600MB, anaconda 大概 2-3G。

建议大家尽量选择 **Anaconda3** 安装, 后续实验课会用到很多 python 包, 可以省掉一些安装的麻烦。

下面是 anaconda3 的链接, 可以在页面最下面找到下载链接, 对应于系统安装即可。

<https://www.anaconda.com/products/individual>



安装好 anaconda3 之后，conda 会默认创建 base 环境，我们直接使用这个环境即可。如需要创建多个 python 隔离的环境，可以查阅 conda 文档。

**检验安装成功：**在命令行输入 `python --version` 能够看到安装的对应版本 `python`。或者输入 `conda` 可以看到 `Anaconda` 相应的提示即可。（注意，强烈建议大家安装的时候选择写入 `PATH` 环境变量）。如果输入提示命令找不到，一般是 `PATH` 环境变量没有配置好，需要手动设置 `PATH` 环境变量到安装的 `anaconda` 目录。

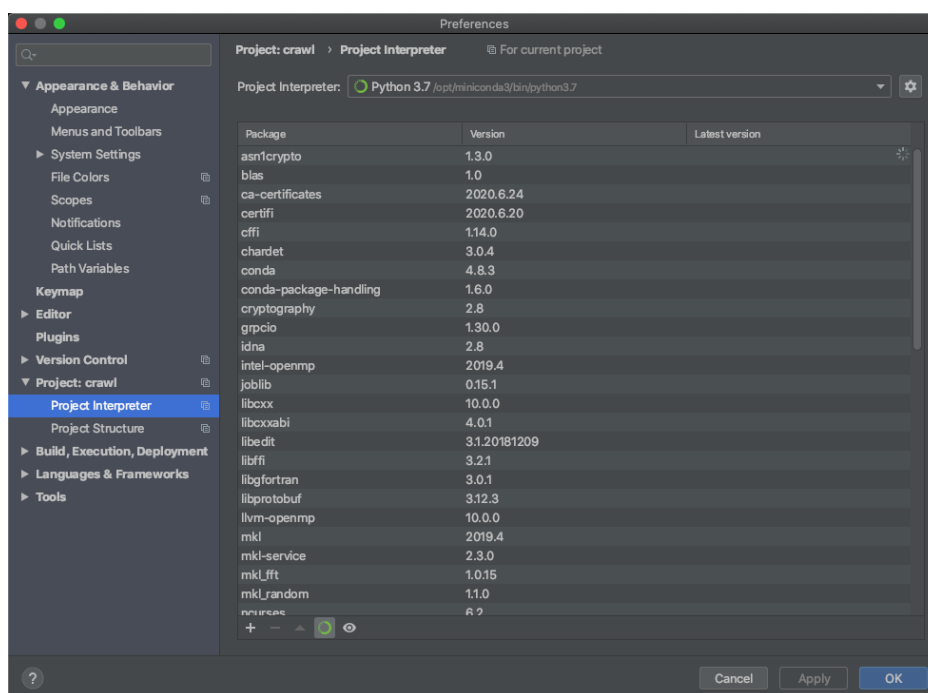
**PS：**如果输入提示命令找不到，同时下文使用 `pycharm`，可以通过配置解释器路径方式，不用配置 `PATH` 变量。

因为 conda 的默认源在国外，当我们使用 conda 命令安装包的时候，可能会失败，包括上面下载 anaconda 的安装包，如果网络有问题的同学，可以采取替换默认源的方法。具体参考 [清华源](#) 的替换教程[地址](#)。

## 2. IDE 环境配置

Python 本身是一门解释型语言，对于简单的程序随便一个编辑器都可以开发。目前流行的功能最强大的集成开发环境是 `pycharm`（[官网](#)），社区版本是免费的。当然，也可以使用 `vscode` 编辑器开发，还有 `python` 的 `jupyter` 包，可以在浏览器里完成开发执行。这里根据大家各自喜好，可以自由选择。

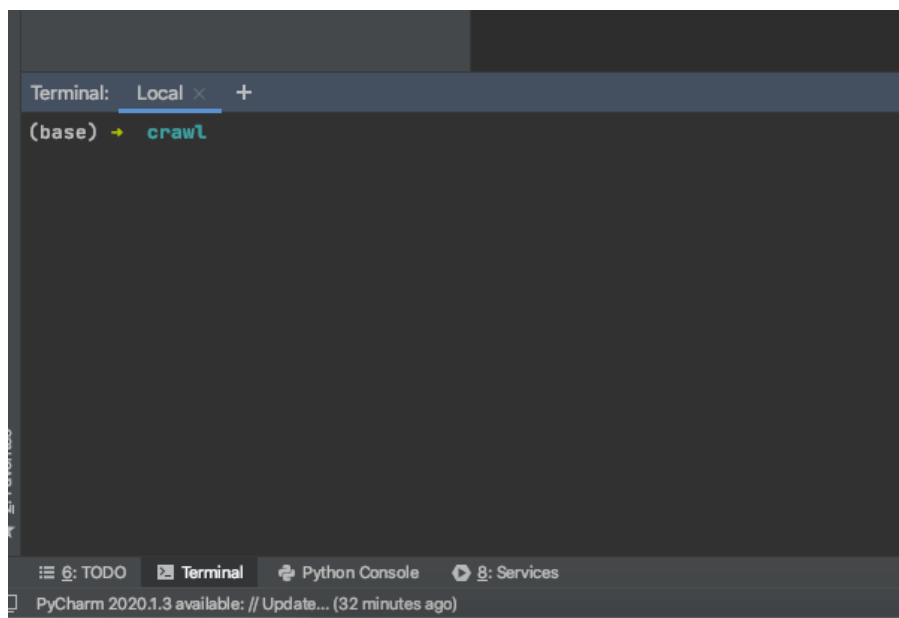
下面针对 `pycharm` 提一些可能会出现的问题：



安装好 pycharm 之后，在新建项目之后需要在设置中配置解释器路径，如上图。这里解释器路径和 conda 环境有关，默认使用 base 环境即可，对应路径在于 anconda 的根目录下 bin 文件夹。（上图配置的是 miniconda3，和 anaconda3 一样的）。

Pycharm 有可能会识别环境，但是不一定准确，大家记得校准一下。

配置好解释器之后，通过 terminal 打开命令行，可以看到当前环境名称（base）：



如果需要执行 python 的一些命令可以在这个 terminal 中执行。使用的 python 环境就是配置好的 python 环境。

### 3. Scrapy 框架安装

官方安装手册 <https://docs.scrapy.org/en/latest/intro/install.html>

1. 在原生 python 环境下安装：直接运行 `pip install Scrapy`，如提示因为某些依赖库缺失无法安装，则安装对应依赖库即可（可能需要安装的库：lxml、pyOpenSSL、Twisted、PyWin32）。有的库可能需要在官网下载 wheel 文件安装。
2. 基于 Anaconda 安装：直接运行 `conda install scrapy`，系统会解析需要安装的依赖库，确认安装即可。

确认安装完成：在命令行键入 `scrapy`，有相应提示即可，如果是在自己新建的虚拟环境下安装，则需要先执行 `conda activate` 虚拟环境，再运行 `scrapy` 命令。

### 4. 爬虫 Demo 编写

在确认安装完成 Scrapy 框架后，参考官方教程文档第三章内容，完成 `tutorial` 项目编写，该项目会从测试网站上抓取 2 页数据，将其中内容保存成 json 格式的文件。所有需要的代码和命令教程都有提供，仔细阅读即可。

官方教程文档：<http://docs.scrapy.org/en/latest/>

**加分项：**继续深入学习教程中 selector 用法，解析和提取金融相关网站平台的数据内容，并存储 MySQL 等关系型数据库中。

备选目标网站：网贷之家、雪球财经、新浪财经、同花顺/大智慧、众筹客、贷罗盘、互金协会、南京银行、光大银行

## 三、提交内容

1. 实验报告：包含上述四项实验内容。
2. 实验代码：爬虫程序的完整项目代码。

## 四、作业提交

点击以下链接，上传即可。

[链接](#)



或者扫描下方二维码：