## 常见的第三方库

1、网络爬虫

requests-对HTTP协议进行高度封装，支持非常丰富的链接访问功能。

PySpider-一个国人编写的强大的网络爬虫系统并带有强大的WebUI。

bs4-beautifulsoup4库，用于解析和处理HTML和XML。

Scrapy-很强大的爬虫框架，用于抓取网站并从其页面中提取结构化数据。可用于从数据挖掘到监控和自动化测试的各种用途

Crawley-高速爬取对应网站的内容，支持关系和非关系数据库，数据可以导出为JSON、XML等

Portia-可视化爬取网页内容

cola-分布式爬虫框架

newspaper-提取新闻、文章以及内容分析

lxml-lxml是python的一个解析库，这个库支持HTML和xml的解析，支持XPath的解析方式

2、自动化

XlsxWriter-操作Excel工作表的文字，数字，公式，图表等

win32com-有关Windows系统操作、Office（Word、Excel等）文件读写等的综合应用库

pymysql-操作MySQL数据库

pymongo-把数据写入MongoDB

smtplib-发送电子邮件模块

selenium-一个调用浏览器的driver，通过这个库可以直接调用浏览器完成某些操作，比如输入验证码，常用来进行浏览器的自动化工作。

pdfminer-一个可以从PDF文档中提取各类信息的第三方库。与其他PDF相关的工具不同，它能够完全获取并分析 P D F 的文本数据

PyPDF2-一个能够分割、合并和转换PDF页面的库。

openpyxl- 一个处理Microsoft Excel文档的Python第三方库，它支持读写Excel的xls、xlsx、xlsm、xltx、xltm。

python-docx-一个处理Microsoft Word文档的Python第三方库，它支持读取、查询以及修改doc、docx等格式文件，并能够对Word常见样式进行编程设置。

3、数据分析及可视化

matplotlib-Matplotlib 是一个 Python 2D 绘图库，可以生成各种可用于出版品质的硬拷贝格式和跨平台交互式环境数据。Matplotlib 可用于 Python 脚本，Python 和 IPython shell（例如 MATLAB 或 Mathematica），Web 应用程序服务器和各种图形用户界面工具包。”

numpy-NumPy 是使用 Python 进行科学计算所需的基础包。用来存储和处理大型矩阵，如矩阵运算、矢量处理、N维数据变换等。

pyecharts-用于生成 Echarts 图表的类库

pandas-一个强大的分析结构化数据的工具集，基于numpy扩展而来，提供了一批标准的数据模型和大量便捷处理数据的函数和方法。

Scipy: 基于Python的matlab实现，旨在实现matlab的所有功能，在numpy库的基础上增加了众多的数学、科学以及工程计算中常用的库函数。

Plotly-Plotly提供的图形库可以进行在线WEB交互，并提供具有出版品质的图形，支持线图、散点图、区域图、条形图、误差条、框图、直方图、热图、子图、多轴、极坐标图、气泡图、玫瑰图、热力图、漏斗图等众多图形

wordcloud-词云生成器

jieba-中文分词模块

4、WEB开发

Django-一个开放源代码的Web应用框架，由Python写成。 是Python生态中最流行的开源Web应用框架，Django采用模型、模板和视图的编写模式，称为MTV模式。

Pyramid是一个通用、开源的Python Web应用程序开发框架。它主要的目的是让Python开发者更简单的创建Web应用，相比Django，Pyramid是一个相对小巧、快速、灵活的开源Python Web框架。

Tornado-一种 Web 服务器软件的开源版本。Tornado和现在的主流Web服务器框架（包括大多数Python的框架）有着明显的区别：它是非阻塞式服务器，而且速度相当快

Flask是轻量级Web应用框架，相比Django和Pyramid，它也被称为微框架。使用Flask开发Web应用十分方便，甚至几行代码即可建立一个小型网站。Flask核心十分简单，并不直接包含诸如数据库访问等的抽象访问层，而是通过扩展模块形式来支持。

5、机器学习

NLTK-一个自然语言处理的第三方库，NLP领域中常用，可建立词袋模型（单词计数），支持词频分析（单词出现次数）、模式识别、关联分析、情感分析（词频分析+度量指标）、可视化（+matploylib做分析图）等

TensorFlow-谷歌的第二代机器学习系统，是一个使用数据流图进行数值计算的开源软件库。

Keras -是一个高级神经网络 API，用 Python 编写，能够在 TensorFlow，CNTK 或 Theano 之上运行。它旨在实现快速实验，能够以最小的延迟把想法变成结果，这是进行研究的关键。”

Caffe-一个深度学习框架，主要用于计算机视觉，它对图像识别的分类具有很好的应用效果

theano-深度学习库。它与Numpy紧密集成，支持GPU计算、单元测试和自我验证，为执行深度学习中大规模神经网络算法的运算而设计，擅长处理多维数组。

Scikit-learn-是一个简单且高效的数据挖掘和数据分析工具，它基于NumPy、SciPy和matplotlib构建。Scikit-learn的基本功能主要包括6个部分：分类，回归，聚类，数据降维，模型选择和数据预处理。Scikit-learn也被称为sklearn。

6、其他常用

IPython-一个基于Python 的交互式shell，比默认的Python shell 好用得多，支持变量自动补全、自动缩进、交互式帮助、魔法命令、系统命令等，内置了许多很有用的功能和函数

PTVS-Visual Studio 的 Python 工具

pydub-支持多种格式声音文件，可进行多种信号处理、信号生成、音效注册、静音处理等

TimeSide-能够进行音频分析、成像、转码、流媒体和标签处理的Python框架

dnspython-DNS工具包

pygame-专为电子游戏设计的一个模块

PyQt5-pyqt5是Qt5应用框架的Python第三方库，编写Python脚本的应用界面

PIL(Pillow)-PIL库是Python语言在图像处理方面的重要第三方库，支持图像存储、显示和处理，它能够处理几乎所有图片格式，可以完成对图像的缩放、剪裁、叠加以及向图像添加线条、图像和文字等操作

OpenCV-图像和视频工作库

Py2exe: 将python脚本转换为windows上可以独立运行的可执行程序。

WeRoBot 是一个微信公众号开发框架，也称为的微信机器人框架。WeRoBot可以解析微信服务器发来的消息，并将消息转换成成Message或者Event类型。