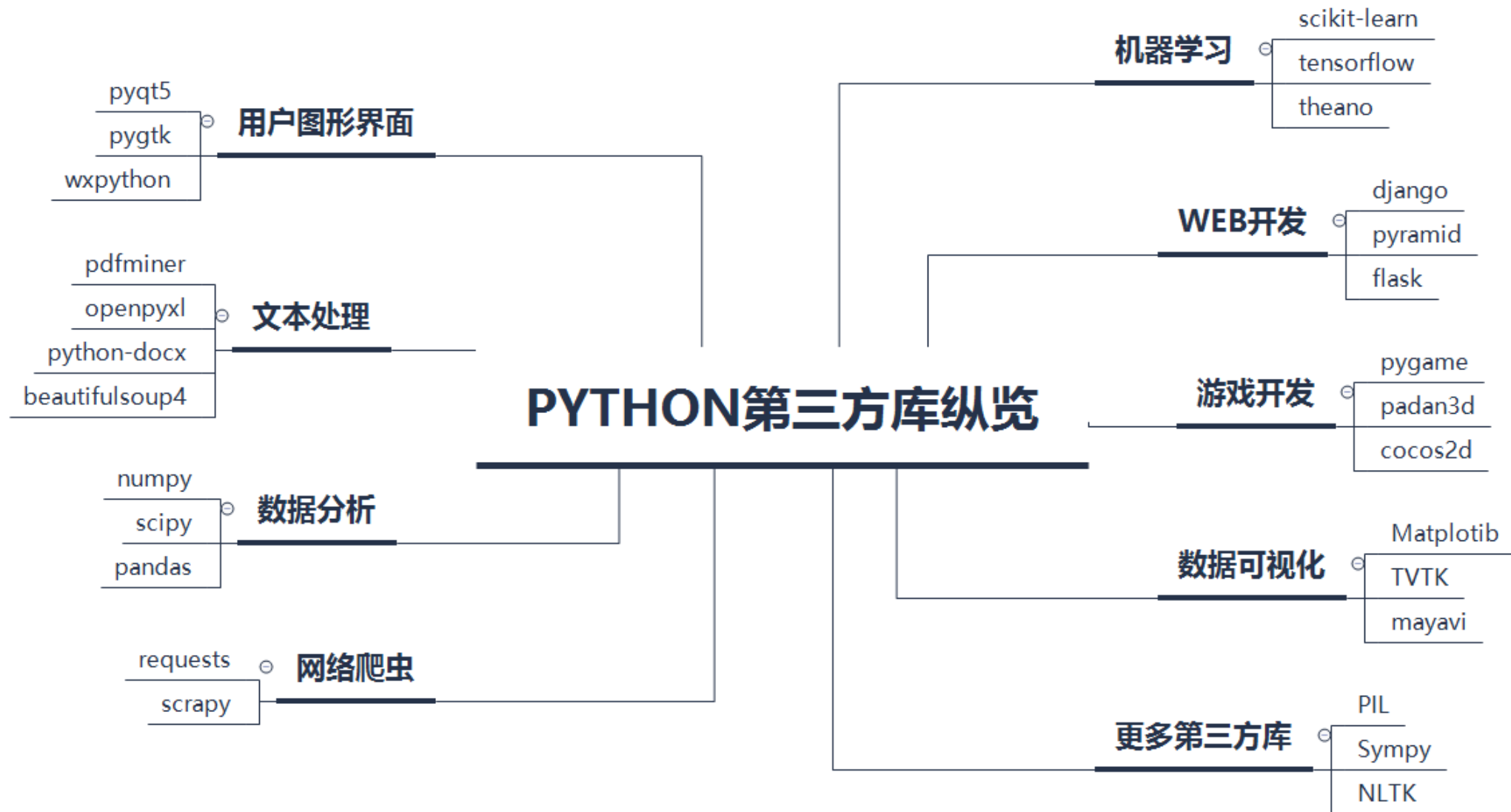


第十一章 python第三方库纵览



1.网络爬虫方向

requests:

requests库是一个简洁且简单的处理HTTP请求的第三方库，它的最大优点是程序编写过程更接近正常URL访问过程。这个库建立在Python语言的urllib3库基础上。request库支持非常丰富的链接访问功能。

scrapy:

scrapy是Python开发的一个快速的、高层次的Web获取框架。不同于简单的网络爬虫功能，scrapy框架本身包含了成熟网络爬虫系统所应该具有的部分共用功能，scrapy用途广泛，可以应用于专业爬虫系统的构建、数据挖掘、网络监控和自动化测试等领域。

2.数据分析方向

numpy:

numpy是Python的一种开源数值计算扩展第三方库，用于处理数据类型相同的多维数组（ndarray），简称“数组”。这个库可用来存储和处理大型矩阵，比Python语言提供的列表结构要高效的多。numpy提供了许多高级的数值编程工具，如：矩阵运算、矢量处理、N维数据变换等。

scipy:

scipy是一款方便、易于使用、专为科学和工程设计的Python工具包。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。它包括统计、优化、整合、线性代数、傅里叶变换、信号分析、图像处理、常微分方程求解等众多模块。

pandas:

pandas是基于numpy扩展的一个重要第三方库，它是为了解决数据分析任务而创建的。Pandas提供了一批标准的数据模型和大量快速便捷处理数据的函数和方法，提供了高效地操作大型数据集所需的工具。

pandas 提供两种最基本的数据类型：serise 和 dataframe，分别代表一维数组和二维数组类型。

3.文本处理方向

pdfminer:

pdfminer是一个可以从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PDF的文本数据。pdfminer能够获取PDF中文本的准确位置、字体、行数等信息，能够将PDF文件转换为HTML及文本格式。

openpyxl:

openpyxl是一个处理Microsoft Excel文档的Python第三方库，它支持读写Excel的xls、xlsx、xlsm、xltx、xltm等格式文件，并进一步能处理Excel文件中excel工作表、表单和数据单元。

python-docx:

python-docx是一个处理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，包括：字符样式、段落样式、表格样式等，进一步可以使用这个库实现添加和修改文本、图像、样式和文档等功能。

beautifulsoup4:

beautifulsoup4库，也称为Beautiful Soup库或bs4库，用于解析和处理HTML和XML。需要注意，它不是BeautifulSoup库。它的最大优点是能根据HTML和XML语法建立解析树，进而高效解析其中的内容。

4.数据可视化

matplotlib:

matplotlib是提供数据绘图功能的第三方库，主要进行二维图表数据展示，广泛用于科学计算的数据可视化。使用这个库可以利用Python程序绘制超过100种数据可视化效果。

TVTK:

TVTK库在标准的VTK库之上用Traits库进行封装的Python第三方库。视觉工具函数库（VTK）是一个开源、跨平台、支持平行处理的图形应用函数库，它是专业可编程的三维可视化工具。TVTK在Python生态系统中被等同于VTK。

5.用户图形界面

Python标准库内置了一个GUI库——tkinter，这个库基于Tcl/Tk开发，然而，这个库十分陈旧，提供的开发控件也很有限，编写出来的GUI风格与现代程序GUI风格相差甚远，从用户体验角度说，tkinter库并不成熟。

这里介绍3个高质量的用户图形界面Python生态库：pyqt5、wxpython、pygtk。

pyqt5:

pyqt5是Qt5应用框架的Python第三方库，它有超过620个类和近6000个函数和方法。它是Python中最为成熟的商业级GUI第三方库。这个库是Python语言当前最好的GUI第三方库，它可以在Windows、Linux和Mac OS X等操作系统上跨平台使用。

wxpython:

wxPython是Python语言的一套优秀的GUI图形库，它是跨平台GUI库wxWidgets的Python封装，可以使Python程序员能够轻松地创建健壮可靠、功能强大的图形用户界面的程序。

6.机器学习

Scikit-learn:

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。Scikitlearn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。

Tensorflow:

TensorFlow是谷歌公司基于DistBelief进行研发的第二代人工智能学习系统，也是用来支撑著名的AlphaGo系统的后台框架。Tensor（张量）指N维数组，Flow（流）指基于数据流图的计算，TensorFlow描述张量从流图的一端流动到另一端的计算过程。

Theano:

Theano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。Theano开发始于2007，可以理解它是一个运算数学表达式的编译器，并可以高效运行在GPU或CPU上。Theano是一个偏向底层开发的库，更像一个研究平台而非单纯的深度学习库。

7.Web开发

Flask: 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。

Django: Django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。Django中提供了开发网站经常用到的模块，Django的开发理念是DRY(Don't Repeat Yourself)，用于鼓励快速开发，进而减少程序员可以建立一个高性能Web应用所花费的时间和精力，形成一种一站式解决方案。

Pyramid:

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。Pyramid仍然面向较大规模的Web应用，但它更关注灵活性，开发者可以灵活选择所使用的数据库、模板风格、URL结构等内容。

8.游戏开发

Pygame:

Pygame是在SDL库基础上进行封装的、面向游戏开发入门的Python第三方库，除了制作游戏外，还用于制作多媒体应用程序。其中，SD（Simple DirectMedia Layer）是开源、跨平台的多媒体开发库，通过OpenGL和Direct3D底层函数提供对音频、键盘、鼠标和图形硬件的简洁访问。

Panda3D:

Panda3D是一个开源、跨平台的3D渲染和游戏开发库,简单说，它是一个3D游戏引擎，由迪士尼和卡耐基梅隆大学娱乐技术中心共同进行开发。Panda3D支持Python和C++两种语言，但对Python支持更全面。

cocos2d:

cocos2d是一个构建2D游戏和图形界面交互式应用的框架，它包括C++、JavaScript、Swift、Python等多个版本。cocos2d基于OpenGL进行图形渲染，能够利用GPU进行加速。cocos2d引擎采用树形结构来管理游戏对象，一个游戏划分为不同场景，一个场景又分为不同层，每个层处理并响应用户事件。

9.其他第三方库

Python语言有超过12万个第三方库，覆盖信息技术几乎所有领域。
即使在每个方向，也会有大量的专业人员开发多个第三方库来给出具体设计。

PIL:

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。

PIL库主要可以完成图像归档和图像处理两方面功能需求：

图像归档：对图像进行批处理、生成图像预览等；

图像处理：图像基本处理、像素处理、颜色处理等。

SymPy:

SymPy是一个支持符号计算的Python第三方库，它是一个全功能的计算机代数系统。SymPy代码简洁、易于理解，支持符号计算、高精度计算、模式匹配、绘图、解方程、微积分、组合数学、离散数学、几何学、概率与统计、物理学等领域计算和应用。

NLTK:

NLTK是一个非常重要的自然语言处理Python第三方库，它支持多种语言，尤其对中文支持良好。NLTK可以进行语料处理、文本统计、内容理解、情感分析等多种应用，具备非常可靠的应用价值。

WeRoBot:

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
Message或者Event类型。

MyQR:

MyQR是一个能够产生基本二维码、艺术二维码和动态效果二维码
的Python第三方库。