

全国计算机等级考试二级教程

Python语言程序设计

(2018年版)



【第11章】

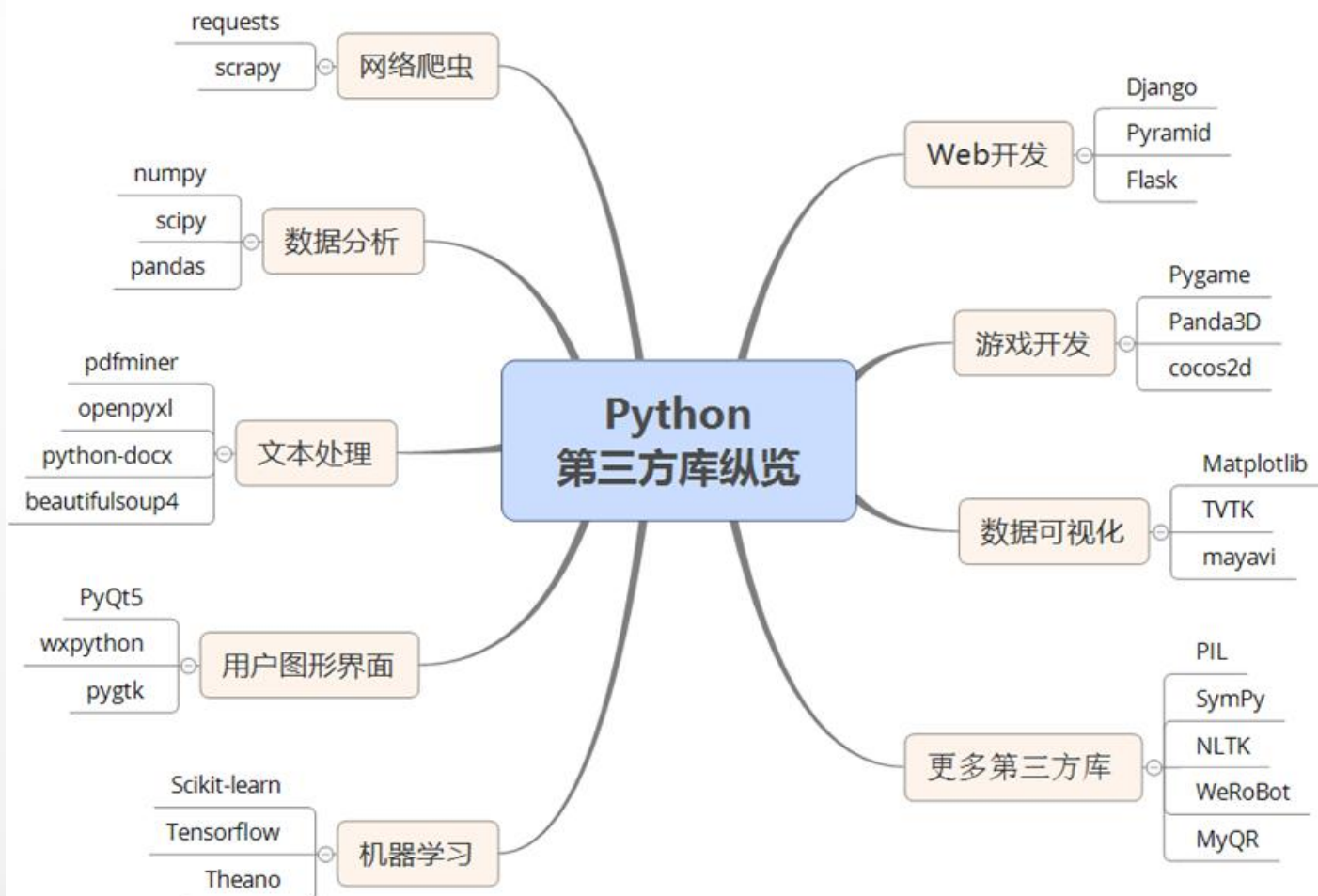
Python第三方库纵览



考纲考点

- 更广泛的Python计算生态，只要求了解第三方库的名称，不限于以下领域：网络爬虫、数据分析、文本处理、数据可视化、用户图形界面、机器学习、Web开发、游戏开发等

知识导图



A large, faint, circular logo of the Python programming language is centered in the background. It features two interlocking snakes, one blue and one yellow, forming a circle.

网络爬虫方向

网络爬虫方向

- 网络爬虫是自动进行HTTP访问并捕获HTML页面的程序。Python语言提供了多个具备网络爬虫功能的第三方库。这里，仅介绍2个常用的Python网络爬虫库：**requests**和**scrapy**。

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。
- 安装requests库在Windows的cmd命令行使用如下命令：

```
:\>pip install requests
```

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。
- 安装scrapy库在Windows的cmd命令行使用如下命令：

```
:\>pip install scrapy
```


The Python logo is centered in the background, consisting of two interlocking snakes, one blue and one yellow, forming a circular shape.

数据分析方向

数据分析方向

- 数据分析是Python的一个优势方向，具有大批高质量的第三方库。这里仅介绍3个最常用的生态库：**numpy**、**scipy**和**pandas**。

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。

numpy

- numpy内部是C语言编写，对外采用Python语言进行封装，因此，在进行数据运算时，基于numpy的Python程序可以达到接近C语言的处理速度。numpy也成为Python数据分析方向各其他库的基础依赖库，已经成为了科学计算事实上的“标准库”。
- 安装numpy库在Windows的cmd命令行使用如下命令：

```
: \> pip install numpy
```



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。
- 安装scipy库在Windows的cmd命令行使用如下命令：

```
: \> pip install scipy
```

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：Series和DataFrame，分别代表一维数组和二维数组类型。
- 安装pandas库在Windows的cmd命令行使用如下命令：

```
: \> pip install pandas
```

A decorative graphic is centered on the page. It consists of a grey circular background containing a white bird icon, which is a stylized representation of a bird in flight. The bird is positioned in the center of the circle, with its wings spread. The circle is surrounded by a white border.

文本处理方向

文本处理方向

- Python语言非常适合处理文本，因此，在这个方向也形成了大量有价值的第三方库。这里介绍4个最常用的生态库：**pdfminer**、**openpyxl**、**python-docx**、**beautifulsoup4**。

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。

pdfminer

- pdfminer 包含两个重要的工具：pdf2txt.py 和 dumppdf.py。
- pdf2txt.py 能够从PDF文件中提取所有文本内容。dumppdf.py 能够把PDF文件内容变成XML格式，并进一步提取其中的图片。
- 安装pdfminer库在Windows的cmd命令行使用如下命令：

```
: \> pip install pdfminer
```

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。
- 安装openpyxl库在Windows的cmd命令行使用如下命令：

```
: \> pip install openpyxl
```

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。
- 安装python-docx库在Windows的cmd命令行使用如下命令：

```
: \> pip install openpyxl
```

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。

beautifulsoup4

- beautifulsoup4库将数据从HTML和XML文件中解析出来，它能够提供一种符合习惯的方法去遍历搜索和修改解析树，将专业的Web页面格式解析部分封装成函数，提供了若干有用且便捷的处理函数。
- 安装beautifulsoup4库在Windows的cmd命令行使用如下命令：

```
: \> pip install beautifulsoup4
```

The Python logo, featuring two interlocking snakes, is centered behind the title text. The snakes are rendered in a light gray color.

数据可视化方向

数据可视化

- 数据可视化指根据数据特点将其展示为易于理解图形的过程。Python语言在数据可视化方面具有较强的优势。这里介绍3个最常用的生态库：**matplotlib**、**TVTK**、**mayavi**。

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。
- 安装matplotlib库在Windows的cmd命令行使用如下命令：

```
: \> pip install matplotlib
```

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。
- 安装matplotlib库在Windows的cmd命令行使用如下命令：

```
: \> pip install vtk
```

TVTK

- mayavi基于VTK开发，完全用Python编写，提供了一个**更为方便实用的可视化软件**，可以简洁地嵌入到用户编写的Python程序中，或者直接使用其面向脚本的API快速绘制三维可视化图形。值得注意的是，mayavi也被称为mayavi2。
- mayavi库在Windows平台上暂时不能方便的使用pip命令安装，建议使用文件安装，请下载mayavi对应版本文件，然后在本地安装。

A decorative frame surrounds the title text. It consists of two horizontal lines with small circles at the ends. In the center of the frame, there are two circular icons: the top one is a stylized bird (Twitter) and the bottom one is a stylized 'G' (GitHub).

用户图形界面方向

用户图形界面方向

- 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。

pyqt5

- pyqt5采用“信号-槽”机制将事件和对应的处理程序进行绑定。pyqt5窗体有很多内置信号，也可以自定义信号。
- 安装pyqt5库在Windows的cmd命令行使用如下命令：

```
: \> pip install pyqt5
```



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。
- 安装wxpython库在Windows的cmd命令行使用如下命令：

```
: \> pip install wxpython
```




wxpython

- pygtk是基于GTK+的Python语言封装，它提供了各式的可视元素和功能，能够**轻松创建具有图形用户界面的程序**。pygtk具有**跨平台性**，利用它编写的代码能够不加修改地稳定运行在各操作系统中，如Windows、MacOS、Linux等。
- 安装wxpython库在Windows的cmd命令行使用如下命令：

```
: \> pip install pygtk
```



机器学习方向

机器学习方向

- 机器学习是人工智能领域的一个重要分支，Python语言也是机器学习和人工智能的重要基础语言。这里介绍3个高质量的机器学习框架：**Scikit-learn、TensorFlow、Theano。**



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。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。
- 安装Scikit-learn库在Windows的cmd命令行使用如下命令：

```
: \> pip install scikit-learn
```

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。
- 安装TensorFlow库在Windows的cmd命令行使用如下命令：

```
:\>pip install tensorflow
```

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。
- 安装Theano库在Windows的cmd命令行使用如下命令：

```
: \> pip install theano
```

The Python logo is centered in the background, rendered in a light gray color. It features two interlocking snakes, one above and one below the text, forming a circular shape.

Web开发方向

Web开发方向

- Web开发是Python语言流行的一个重要方向，主要用于服务器后端开发。根据Python Web开发框架的复杂程度，这里依次介绍3个Python第三方生态库：**Django**、**Pyramid**、**Flask**。



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。
- 安装Django库在Windows的cmd命令行使用如下命令：

```
:\>pip install django
```



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。
- 安装Pyramid库在Windows的cmd命令行使用如下命令：

```
:\>pip install pyramid
```

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。
- 安装Flask库在Windows的cmd命令行使用如下命令：

```
: \> pip install flask
```



游戏开发方向

Web开发方向

- 游戏开发是一个有趣的方向，在游戏逻辑和功能实现层面，Python已经成为重要的支撑性语言。这里介绍3个Python第三方生态库：**Pygame**、**Panda3D**、**cocos2d**。

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，SDL（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。



Pygame

- Pygame是一个游戏开发框架，提供了大量与游戏相关的底层逻辑和功能支持，**非常适合作为入门库理解并实践游戏开发。**
- 安装Pygame库在Windows的cmd命令行使用如下命令：

```
: \> pip install pygame
```

Panda3D


- Panda3D是一个开源、跨平台的**3D渲染和游戏开发库**，简答说，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。
- Panda3D支持很多当代先进游戏引擎所支持的特性：法线贴图、光泽贴图、HDR、卡通渲染和线框渲染等。
- 安装Panda3D库在Windows的cmd命令行使用如下命令：

```
:\>pip install panda3d
```


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。
- 安装cocos2d库在Windows的cmd命令行使用如下命令：

```
:\>pip install cocos2d
```

A large, faint, gray GitHub logo (Octocat) is centered in the background of the slide.

更多第三方库

更多第三方库

- 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。除了本章所提到的方向外，这里再列出5个有趣且有用的Python第三方库，展示Python在工程实践方面强大的魅力。

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。
- 安装PIL库在Windows的cmd命令行使用如下命令：

```
: \> pip install pillow
```

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。
- 安装SymPy库在Windows的cmd命令行使用如下命令：

```
: \> pip install sympy
```

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。
- 安装NLTK库在Windows的cmd命令行使用如下命令：

```
: \> pip install nltk
```

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。
- 安装WeRoBot库在Windows的cmd命令行使用如下命令：

```
: \> pip install werobot
```

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。图11.1给出了一些MyQR生成二维码的实例。



- 安装MyQR库在Windows的cmd命令行使用如下命令：

```
:\>pip install myqr
```


本章小结

本章通过8个具体方向30个Python功能库的简要介绍纵览Python语言计算生态的丰富性，希望读者能够从Python基础语法出发，看到更广阔的程序设计生态，进一步“理解和运用计算生态”，掌握符合信息时代需要的程序设计能力。

国家计算机等级考试不仅仅是一场考试，更是检验能力提升的手段，加油！