Python类库

环境管理

* p：非常简单的交互式 python 版本管理工具。
* pyenv：简单的 Python 版本管理工具。
* Vex：可以在虚拟环境中执行命令。
* virtualenv：创建独立 Python 环境的工具。
* virtualenvwrapper：virtualenv 的一组扩展

包管理

* pip：Python 包和依赖关系管理工具。
* pip-tools：保证 Python 包依赖关系更新的一组工具。
* conda：跨平台，Python 二进制包管理工具。
* Curdling：管理 Python 包的命令行工具。
* wheel：Python 分发的新标准，意在取代 eggs。

**自然语言处理**

用来处理人类语言的库。

* [NLTK](http://hao.jobbole.com/nltk/)：一个先进的平台，用以构建处理人类语言数据的 Python 程序。[官网](http://www.nltk.org/)
* jieba：中文分词工具。[官网](https://github.com/fxsjy/jieba)
* langid.py：独立的语言识别系统。[官网](https://github.com/saffsd/langid.py)
* Pattern：Python 网络信息挖掘模块。[官网](http://www.clips.ua.ac.be/pattern)
* SnowNLP：一个用来处理中文文本的库。[官网](https://github.com/isnowfy/snownlp)
* TextBlob：为进行普通自然语言处理任务提供一致的 API。[官网](http://textblob.readthedocs.org/en/latest/)
* TextGrocery：一简单高效的短文本分类工具，基于 LibLinear 和 Jieba。[官网](https://github.com/2shou/TextGrocery)

### 图像处理

用来操作图像的库.

* [pillow](http://hao.jobbole.com/pillow/)：Pillow 是一个更加易用版的 [PIL](http://www.pythonware.com/products/pil/)。[官网](http://pillow.readthedocs.org/en/latest/)
* hmap：图像直方图映射。[官网](https://github.com/rossgoodwin/hmap)
* imgSeek：一个使用视觉相似性搜索一组图片集合的项目。[官网](https://sourceforge.net/projects/imgseek/)
* nude.py：裸体检测。[官网](https://github.com/hhatto/nude.py)
* pyBarcode：不借助 PIL 库在 Python 程序中生成条形码。[官网](https://pythonhosted.org/pyBarcode/)
* pygram：类似 Instagram 的图像滤镜。[官网](https://github.com/ajkumar25/pygram)
* python-qrcode：一个纯 Python 实现的二维码生成器。[官网](https://github.com/lincolnloop/python-qrcode)
* Quads：基于四叉树的计算机艺术。[官网](https://github.com/fogleman/Quads)
* scikit-image：一个用于（科学）图像处理的 Python 库。[官网](http://scikit-image.org/)
* thumbor：一个小型图像服务，具有剪裁，尺寸重设和翻转功能。[官网](https://github.com/thumbor/thumbor)
* wand：[MagickWand](http://www.imagemagick.org/script/magick-wand.php)的Python 绑定。MagickWand 是 ImageMagick的 C API 。[官网](https://github.com/dahlia/wand)

### 模板引擎

模板生成和词法解析的库和工具。

* [Jinja2](http://hao.jobbole.com/jinja2/)：一个现代的，对设计师友好的模板引擎。[官网](https://github.com/pallets/jinja)
* Chameleon：一个 HTML/XML 模板引擎。 模仿了 ZPT（Zope Page Templates）, 进行了速度上的优化。[官网](https://chameleon.readthedocs.org/en/latest/)
* Genshi：Python 模板工具，用以生成 web 感知的结果。[官网](https://genshi.edgewall.org/)
* Mako：Python 平台的超高速轻量级模板。[官网](http://www.makotemplates.org/)

### 搜索

对数据进行索引和执行搜索查询的库和软件。

* django-haystack：Django 模块化搜索。[官网](https://github.com/django-haystack/django-haystack)
* elasticsearch-py：Elasticsearch 的官方底层 Python 客户端。[官网](https://www.elastic.co/guide/en/elasticsearch/client/python-api/current/index.html)
* elasticsearch-dsl-py：Elasticsearch 的官方高级 Python 客户端。[官网](https://github.com/elastic/elasticsearch-dsl-py)
* solrpy：[solr](http://lucene.apache.org/solr/)的 Python 客户端。[官网](https://github.com/edsu/solrpy)
* Whoosh：一个快速的纯 Python 搜索引擎库。[官网](http://whoosh.readthedocs.org/en/latest/)

### 缓存

缓存数据的库。

* Beaker：一个缓存和会话库，可以用在 web 应用和独立 Python脚本和应用上。[官网](http://beaker.readthedocs.org/en/latest/)
* django-cache-machine：Django 模型的自动缓存和失效。[官网](https://github.com/django-cache-machine/django-cache-machine)
* django-cacheops：具有自动颗粒化事件驱动失效功能的 ORM。[官网](https://github.com/Suor/django-cacheops)
* django-viewlet：渲染模板，同时具有额外的缓存控制功能。[官网](https://github.com/5monkeys/django-viewlet)
* dogpile.cache：dogpile.cache 是 Beaker 的下一代替代品，由同一作者开发。[官网](http://dogpilecache.readthedocs.org/en/latest/)
* HermesCache：Python 缓存库，具有基于标签的失效和 dogpile effect 保护功能。[官网](https://pypi.python.org/pypi/HermesCache)
* johnny-cache：django应用缓存框架。[官网](https://github.com/jmoiron/johnny-cache)
* pylibmc：[libmemcached](http://libmemcached.org/libMemcached.html) 接口的 Python 封装。[官网](https://github.com/lericson/pylibmc)

### URL处理

解析URLs的库

* furl：一个让处理 URL 更简单小型 Python 库。[官网](https://github.com/gruns/furl)
* purl：一个简单的，不可变的URL类，具有简洁的 API 来进行询问和处理。[官网](https://github.com/codeinthehole/purl)
* pyshorteners：一个纯 Python URL 缩短库。[官网](https://github.com/ellisonleao/pyshorteners)
* shorturl：生成短小 URL 和类似 bit.ly 短链的Python 实现。[官网](https://github.com/Alir3z4/python-shorturl)
* webargs：一个解析 HTTP 请求参数的库，内置对流行 web 框架的支持，包括 Flask, Django, Bottle, Tornado和 Pyramid。[官网](https://github.com/sloria/webargs)

### 反垃圾技术

帮助你和电子垃圾进行战斗的库。

* django-simple-captcha：一个简单、高度可定制的Django 应用，可以为任何Django表单添加验证码。[官网](https://github.com/mbi/django-simple-captcha)
* django-simple-spam-blocker：一个用于Django的简单的电子垃圾屏蔽工具。[官网](https://github.com/moqada/django-simple-spam-blocker)

### 机器学习

机器学习库。 参见: [awesome-machine-learning](https://github.com/josephmisiti/awesome-machine-learning#python).

* Crab：灵活、快速的推荐引擎。[官网](https://github.com/muricoca/crab)
* gensim：人性化的话题建模库。[官网](https://github.com/piskvorky/gensim)
* hebel：GPU 加速的深度学习库。[官网](https://github.com/hannes-brt/hebel)
* NuPIC：智能计算 Numenta 平台。[官网](https://github.com/numenta/nupic)
* pattern：Python 网络挖掘模块。[官网](https://github.com/clips/pattern)
* [PyBrain](http://hao.jobbole.com/pybrain/)：另一个 Python 机器学习库。[官网](https://github.com/pybrain/pybrain)
* [Pylearn2](http://hao.jobbole.com/pylearn2/)：一个基于 [Theano](https://github.com/Theano/Theano) 的机器学习库。[官网](https://github.com/lisa-lab/pylearn2)
* [python-recsys](http://hao.jobbole.com/python-recsys/)：一个用来实现推荐系统的 Python 库。[官网](https://github.com/ocelma/python-recsys)
* scikit-learn：基于 SciPy 构建的机器学习 Python 模块。[官网](http://scikit-learn.org/)
* pydeep：Python 深度学习库。[官网](https://github.com/andersbll/deeppy)
* vowpalporpoise：轻量级 [Vowpal Wabbit](https://github.com/JohnLangford/vowpalwabbit/) 的 Python 封装。[官网](https://github.com/josephreisinger/vowpalporpoise)
* skflow：一个 [TensorFlow](https://github.com/tensorflow/tensorflow) 的简化接口(模仿 scikit-learn)。[官网](https://github.com/tensorflow/skflow)
* Caffe: 一个[Caffe](https://github.com/BVLC/caffe)的python接口。[官网](http://caffe.berkeleyvision.org/)

### MapReduce

MapReduce 框架和库。

* [dpark](http://hao.jobbole.com/dpark/)：Spark 的 Python 克隆版，一个类似 MapReduce 的框架。[官网](https://github.com/douban/dpark)
* dumbo：这个 Python 模块可以让人轻松的编写和运行 Hadoop 程序。[官网](https://github.com/klbostee/dumbo)
* luigi：这个模块帮你构建批处理作业的复杂流水线。[官网](https://github.com/spotify/luigi)
* mrjob：在 Hadoop 或 Amazon Web Services 上运行 MapReduce 任务。[官网](https://github.com/Yelp/mrjob)
* PySpark：Spark 的 Python API 。[官网](http://spark.apache.org/docs/latest/programming-guide.html)
* streamparse：运行针对事实数据流的 Python 代码。集成了[Apache Storm](http://storm.apache.org/)。[官网](https://github.com/Parsely/streamparse)

### 算法和设计模式

Python 实现的算法和设计模式。

* [algorithms](http://hao.jobbole.com/algorithms/)：一个 Python 算法模块。[官网](https://github.com/nryoung/algorithms)
* python-patterns：Python 设计模式的集合。[官网](https://github.com/faif/python-patterns)
* sortedcontainers：快速，纯 Python 实现的SortedList，SortedDict 和 SortedSet 类型。[官网](http://www.grantjenks.com/docs/sortedcontainers/)

参考资料：<https://github.com/jobbole/awesome-python-cn>