

词频-词频表 (研究一个词在一个语料库中重要程度)  
TF-IDF 算法: jieba.analyse.extract-tfids  
算法- [ 基于字典 (南大) 词库  
基于词频统计分词 (Viterbi 算法)

TextRank 算法: jieba.analyse.text\_rank  
语料库: 哈工大  
分词 speed  
(def) 词频统计: jieba.analyse.paddle模式  
全模式: 搜索引擎 搜索引擎 搜索引擎 [分词准确性差]  
提取关键词 jieba.analyse (stop-words)  
① 关键词抽取 (算法) - supervised  
② 关键词抽取 (算法) - unsupervised

wordcloud 预处理  
wordcloud 和 matplotlib 默认只支持英文: 修改 wordcloud.py 30 行代码字体路径 'PngFont.ttf'

smtplib MIME  
数据库 系统初始化时, 对数据库进行操作, 把连接放到内存中  
底层: 访问数据库, 用 SQLAlchemy 解决数据库问题  
建立连接, 从数据库获取数据  
取一个实例的数据库  
对象给你, 使用数据库  
数据库会自动返回数据库对象  
PyMySQL 是 Python 的 MySQL 数据库接口, 如何使用 ThreadingPoolExecutor 与 PyMySQL 实现多线程访问数据库

ORM  
From sqlalchemy import  
sessionmaker  
DBSession = sessionmaker(bind=engine)  
session = DBSession()  
new\_data = Mytable(name='John', age=10)  
session.add(new\_data)

数据入库  
ORM  
From sqlalchemy import  
sessionmaker  
DBSession = sessionmaker(bind=engine)  
session = DBSession()  
new\_data = Mytable(name='John', age=10)  
session.add(new\_data)

语意情感分析

未登录词: 训练语料中未出现的词 (HMM 模型)  
Viterbi 算法  
DeepLearning: PageRank  
pprint 打印出任何 Python 数据结构的值和方法  
前缀词典: 静态加载 - jieba.load\_userdict (user\_dict)  
动态添加、删除 - jieba.add\_word('')

(并发): jieba.enable\_parallel(4)  
优化 jieba 算法生产环境: ① 词频统计分析优化 降低误差、提高准确率  
② 降低低价值的词语 降低 ooad-usedict  
③ 句子的预处理

读取背景图片: background\_image = np.array(Image.open(path.join(dir, ...)))  
提取背景图片颜色: img\_colors = ImageColorGenerator(background\_image)

开始连接 (指定数据库) → 创建 Cursor (发数据) → 提交数据 (写入) < 数据库增删改查, 读取数据  
↓  
关闭 Cursor → 关闭连接

奥情监控

奥情监控

奥情监控

奥情监控

奥情监控

奥情监控