

SnowNLP: * (Continue)
 整体结构图: s.han
 提取关键词
 信息衡量
 训练

舆情监控

jieba

生僻字归类
 (Cutehub) 词频
 中文拼音: yin会生 yhs
 拆字: 耶一耳、耶
 单词在 GloVe 向量中的索引 (坐标)
 词处理和向量计算: 向量运算放在 GPU 上 深度学习多边形计算 矩阵计算

wordcloud

smtplib

pymysql

- ① 收集数据集 ② 设计文本表示模型 ③ 选择文本特征
- ④ 选择分类模型 (未训练, 贝叶斯决策) (提取关键词)

Computer vision
 语意情感分析

训练模型:
 From snownlp import sentiment
 sentiment.train('neg.txt', pos.txt')
 sentiment.score ('sentiment.harshal')

验证码识别

resnetact - Pyresnetact, image-to-string
 (其名字 数据识别能力) (th, lang='chi-sim+eng')

SNOWNLP
 词向量 CPU
 情感分析 (朴素贝叶斯)
 中文分词
 词性标注 (隐马尔可夫模型)
 拼音 (Trie树) *

背景与验证码颜色要有较大差别
 字母之间进行重叠
 验证码点字母且有一定顺序
 滑块验证码用 05
 提取数据和标注 - 深度学习训练模型 - 存入文件进行识别