Python 进阶训练营

尹会生



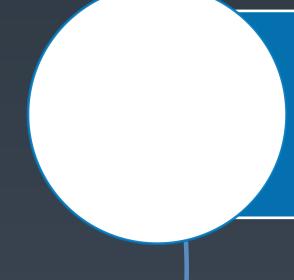
份分词与词云展示



- 01 jieba 的安装
- 02 分词原理
- 03 增加词向量
- 04 wordcloud 制作词云







1.掌握 jieba 的使用和分词原理

1.使用 wordcloud 制作并展示词云

自然语言处理

什么是自然语言处理? 用计算机算法、数学概念、统计学处理人类的自然语言(说、读、写) 语言、语音、文字

什么是自然语言处理(NLP)? NLP 是 AI 的分支学科,计算机通过计算技术和计算语言学理论构建, 实现像人一样处理自然语言

NLP 能做什么? 理解文字、语音并与人进行自动(半自动)交互处理



解决 NLP 问题的一般过程

什么是自然语言处理? 用计算机算法、数学概念、统计学处理人类的自然语言(说、读、写) 语言、语音、文字

什么是自然语言处理(NLP)? NLP 是 AI 的分支学科,计算机通过计算技术和计算语言学理论构建, 实现像人一样处理自然语言

NLP 能做什么? 理解文字、语音并与人进行自动(半自动)交互处理



jieba

jieba的3种分词模式精确模式全模式 搜索引擎模式

在 jieba 中,对于未登录到词库的词,使用了基于汉字成词能力的 HMM 模型和 Viterbi 算法



jieba

```
jieba调整词典
词语词频(可省略) 词性(可省略)
jieba.load_userdict(file_name) 载入词典
动态调整词典
add_word(word, freq=None, tag=None)
del_word(word)
调整词频
jieba.suggest_freq()
其他功能
  关键词提取
  词性标注
  并行分词
```



wordcloud

使用 wordcloud 制作并展示词云

下载地址: https://github.com/amueller/word_cloud



###