2018选题-备选

说明：

每位同学一个题目，不可重复。

以此契机学习python,用什么学什么。

实现可正常执行的程序。

**1、古诗词分析**

中华古诗词数据库<https://github.com/chinese-poetry/chinese-poetry>

——数据分析：一些简单的高频分析：唐诗、宋诗、宋词高频词、词牌名、作者作品榜等。

2、谣言粉碎机

<https://github.com/thunlp/Chinese_Rumor_Dataset>

结合“结巴”中文分词 https://github.com/CasterWx/python-girlfriend-mood/

——计算相似度，判断某信息是否谣言。

3、**情感波动分析**

通过对话获得其情感值波动图谱

<https://github.com/CasterWx/python-girlfriend-mood/>

——通过获取电视剧/电影/小说/书/某商品/食物等的评论文本，获取针对商品的不同维度的情感值。

4、**中文词库、停用词、敏感词** [dongxiexidian/Chinese](https://github.com/dongxiexidian/Chinese)

此package的敏感词库分类更细：

[反动词库](https://github.com/dongxiexidian/Chinese/blob/master/%E6%95%8F%E6%84%9F%E8%AF%8D%E5%BA%93/%E5%8F%8D%E5%8A%A8%E8%AF%8D%E5%BA%93.txt)， [敏感词库表统计](https://github.com/dongxiexidian/Chinese/blob/master/%E6%95%8F%E6%84%9F%E8%AF%8D%E5%BA%93/%E6%95%8F%E6%84%9F%E8%AF%8D%E5%BA%93%E8%A1%A8%E7%BB%9F%E8%AE%A1.txt)， [暴恐词库](https://github.com/dongxiexidian/Chinese/blob/master/%E6%95%8F%E6%84%9F%E8%AF%8D%E5%BA%93/%E6%9A%B4%E6%81%90%E8%AF%8D%E5%BA%93.txt)， [民生词库](https://github.com/dongxiexidian/Chinese/blob/master/%E6%95%8F%E6%84%9F%E8%AF%8D%E5%BA%93/%E6%B0%91%E7%94%9F%E8%AF%8D%E5%BA%93.txt)， [色情词库](https://github.com/dongxiexidian/Chinese/blob/master/%E6%95%8F%E6%84%9F%E8%AF%8D%E5%BA%93/%E8%89%B2%E6%83%85%E8%AF%8D%E5%BA%93.txt)

结合“结巴”中文分词 https://github.com/CasterWx/python-girlfriend-mood/

——百度贴吧：农大人（泰安人/山东人/北方人/985、211高校计算机专业毕业生/普通高校计算机专业毕业生……）现在关注什么？

5、根据中文姓名猜测性别

<https://github.com/observerss/ngender>

——输入一个名字，判断名字的男性化/女性化程度。

6、同义词表，反义词表，否定词表

<https://github.com/guotong1988/chinese_dictionary>

结合“结巴”中文分词 https://github.com/CasterWx/python-girlfriend-mood/

——通过获取电视剧/电影/小说/书/某商品/食物等的评论文本，按评论分类用户，实现商品推荐。

7、**中文聊天语料**

https://github.com/codemayq/chaotbot\_corpus\_Chinese

该库搜集了包含:豆瓣多轮, PTT八卦语料, 青云语料, 电视剧对白语料, 贴吧论坛回帖语料,微博语料,小黄鸡语料

——设计一个chatbot，聊天内容可由键盘输入输出。根据chatbot自身人格（少女、少年、青年人、中年人、老年人等）的设置选择不同的聊天内容，即对同一个问题人格不同，回答也可能不同。