XNR项目文本分析代码

**1、关注用户观点分析**

已嵌入一期代码库，放入位置：

/home/ubuntu8/yuanhuiru/xnr/xnr1/xnr/weibo\_xnr\_operate/search\_weibo\_byuid.py

调用方法：

from search\_weibo\_byuid import follower\_analysis

输入数据：

uid\_list：uid列表，xnr关注的用户列表

text：热点微博文本

输出数据：

summary\_list：子观点列表，每个元素表示一个子观点的摘要

**2、根据关键词生成对应的观点**

调用方法：

from opinion\_search import opinion\_relevance

输入数据：

keywords 关键词（列表）

file\_name 语料库的类型：violence 暴恐，weiquan 维权

输出数据：

summary 观点文本

语料说明：

目前观点生成的语料只有两个特定领域，分别是“暴恐”、“维权”，对应corpus文件夹下“text\_ violence.csv”和“weiquan.csv”

之后的语料可以存入数据库中，方便检索

**3、根据问题搜索对应的答案**

调用方法：

from question\_search import search\_answer

输入数据：

text:问题文本

输出数据：

text\_question:字符串，匹配上的问题

answer:列表，推荐的答案

语料说明：

目前自动问答的语料对应corpus文件夹下“corpus\_answer.csv”

之后的语料可以存入数据库中，方便检索

**4、问答语料的收集**

from answer\_corpus import get\_weibo\_text\_bymessagetype

获取敏感度大于1的转发和评论文本

输入数据：

key\_list：消息类型列表，此处的key\_list=[ '2','3']

name：字符串，用来存储结果的文件名

输出数据：

输出的文件，每一个es存成一个文件

**5、观点语料的收集**

from opinion\_corpus import get\_weibo\_text

输入数据：

key\_list：领域关键词

name：字符串，用来存储结果的文件名

输出数据：

输出的文件，每一个es存成一个文件

**6、调用图灵机器人（有访问次数限制，1000/天）**

from tuling\_test import get\_message\_from\_tuling

输入数据：

massage：utf-8类型的字符串

输出数据：

string：字符串