**说明文档**

1. **软硬件环境说明**
   1. Python2.7及其相关python依赖包,缺啥补啥
      1. tornado
      2. requests
      3. pymysql
      4. urllib2
      5. cookielib
      6. zlib
      7. Queue
      8. bs4
      9. demjson
      10. textrank4zh
      11. elasticsearch
      12. jieba
   2. mysql，启动server
   3. elasticsearch-6.x以上，默认端口配置，启动elasticsearch
   4. elasticsearch 要求机器4G以上内存
2. **项目部署说明**
   1. 进入项目文件夹的duck\_prophet/config/db\_operate/目录里的mysql\_conf.py文件，修改mysql数据库连接配置。
   2. 进入项目文件夹的script目录
   3. 执行mysql\_prepare.py脚本，完成数据库建库建表工作
   4. 执行history\_prepare.py脚本，抓取部分历史新闻数据，执行完成之后
   5. 执行mysql2elastic.py脚本，完成全量导入elasticsearch简历索引工作
   6. 执行daily\_crawler.py脚本，开启日常爬虫程序，前台或后台运行
   7. 执行mysql2elastic\_yesterday.py脚本，开启每日索引新闻程序
   8. 执行tornado\_server.py脚本，开启web接口服务,端口为8888
3. **web接口说明，三个主要接口，**请求地址应为**host:8888**,下图中18888端口不用管
   1. **news\_search**，请求方式如下，POST类型，返回Json格式新闻列表以及结果数目。

**URL参数：**target=news\_search&page\_num=分页号

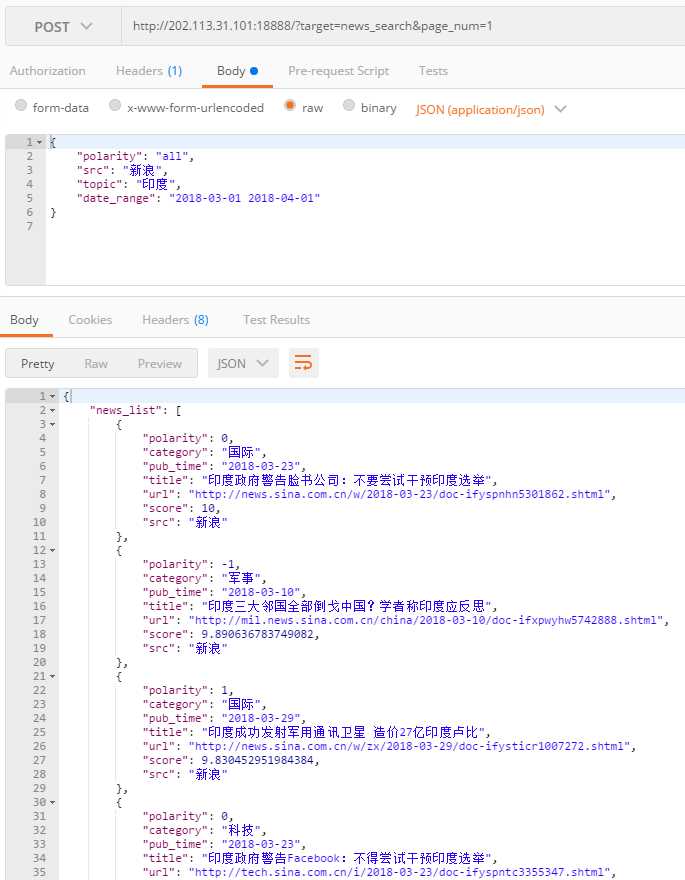
**POST请求参数：**

Polarity: 极性选项，可选“all”、“1”、“-1”、“0”，表示全部、正面、负面、中性

src: 新闻源选项，可选“全部”“新浪”“搜狐”“新华”“腾讯”“网易”

topic: 话题选项，短文本

date\_range: 日期可用两种类型，“2018-01-20 2018-01-30”这种表示某段时间范围，也可传“day”，“week”，“month”，“year”，“all”表示最近一天一周等等。



* 1. **news\_analyze**，请求方式如下，POST类型，返回某篇文章的摘要句以及带权关键词列表，Json格式。

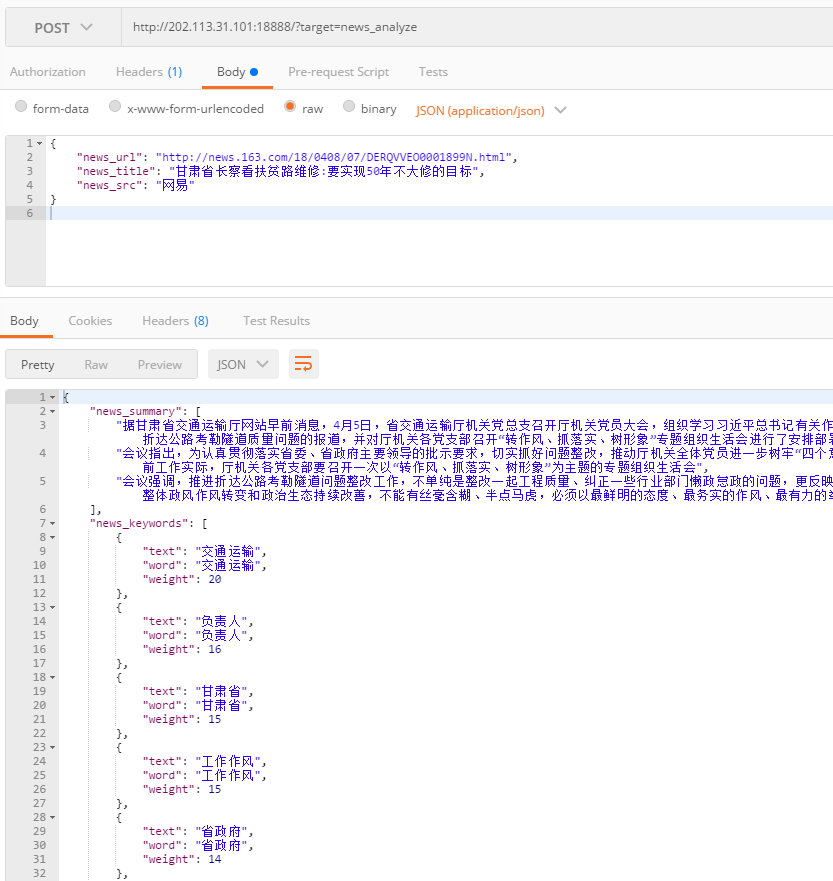
**Url参数：**target=news\_analyze

**POST请求参数：**

news\_url: 新闻文章的地址，也是它的唯一标识，可以来自与前一个接口返回的数据结果，也可以是未收录的新闻地址。

news\_title: 新闻标题

news\_src: 新闻源，写一个就行，不在五个预制来源内也没关系。



* 1. **topic\_trend，**请求方式如下，POST类型，返回某段时间某个话题的分时文章数量，表示话题热度趋势，Json格式。

**Url参数：**target=topic\_trend

**POST请求参数：**

topic: 关注的话题文本

time\_mode: 要关注话题的时间模式

day\_week: 最近一周每日

day\_month: 最近一个月每日

day\_year: 最近一年每日

week\_year: 最近一年每周

month\_year: 最近一年每月

