3、线索定期提取

1）投资类网站

* 投资界 <http://www.pedaily.cn/all/>
* 投资中国

<http://www.chinaventure.com.cn/cmsmodel/news/capital/list/11.shtml>

2、生物医药类资讯网站

* 生物谷 http://www.bioon.com/
* 丁香园 <http://www.dxy.cn/>

3、TMT类网站

* 36氪 http://36kr.com/
* IT桔子 https://www.itjuzi.com/

跟踪最近一周的新闻资讯，并对文章进行分词，计算词的TF-IDF值，提取出出现率最高的N个词输出，人工判断是否可作为研究线索。

3、涉及的算法：

* 分词
* TF-IDF计算（直接使用词频的话，会提取出大量的无用词）。