2020.2.7 会议内容

1. 查找到的推送状况：>1000篇 的推送素材

描述为两种风格：

1. 结构性强的（严肃）e.g.武大讲座
2. 活泼 e.g.小尼

Get上次推进： 可以做讲座类的

2.提议模板分类依据：

按风格分类

按模块分类

3.会议推进

（1）婧怡&清扬：

先做出严肃风格的模板 预计两周

（2）技术组：

主攻爬虫 （主要爬文本，也考虑一下图像）

为模拟根据用户所给关键词参数找素材的过程，三人都尝试实现以下三种功能：

<1> 查找：选择可靠信源，类似page rank

<2> 识别：在页面中搜索关键字，定位目标段落

<3> 提取：让爬的结果能讲通，从目标段落中提取出有效句子、核心词汇

#做出实验后把代码上传到JupyterLab，实现共享

\*一周后（2020.2.14）确认模板进度，视情况开会