

# 实训计划

内容：食品安全事件检测与舆情分析

## 1 / 第一周

- 余连玮：理清《网络评论文本的细粒度情感分析研究》和《文本话题识别算法的研究与实现》论文所用到的技术；
- 王子昂：理清《基于网络数据的食品安全事件检测与分析》和《社交网络推文情感分类系统的设计与实现》所用到的技术，将论文中涉及的技术整合到一个框架中，确定之后的研究和开发方向
- 两人共同完成：讨论并将上述论文的技术整合到一个框架，确定后续深入的方向。
- 输出：得到一个整合上述所有技术的大纲。

## 2 / 第二周

- 王子昂：配置环境，整理命名实体识别代码（标注、识别）、话题提取部分、事件检测部分的代码，理解代码结构和实现方式。基于源代码的基础上进行代码的重构，添加必要注释，修改其中存在的问题，使整体结构清晰，代码易于重用。
- 余连玮：寻找新的适合的视频或音频数据源，进行文本提取，形成可调用的 api，集成到邓娜代码中。帮助王子昂重构代码。
- 输出：得到更加丰富的数据集、等到重构后的代码。

### 3 / 第三周

- 数据源：社交媒体（微博、微信公众号）、新闻网站、视频网站（央视网、优酷）
- 共同完成：爬取数据，研读趋势分析相关论文，制定趋势分析策略
- 输出：来源更加丰富的数据

### 4 / 第四周

- 对爬取的数据进行预处理，包括清洗、去停用词、分词，得到语料库。结合得到的新的数据源，重新进行数据处理，命名实体识别、构建新的库语料、训练词向量、得到新的 LDA 模型，进行话题提取
- 从爬取的新闻中抽取一部分数据利用 YEDDA 进行命名实体的标记（时间、地点、食品专有名词）利用这些数据进行模型训练
- 输出：LDA、Bi-CRF 模型训练结果，得到重新训练的可用模型

## 5 / 第五、六周

- 共同完成：全部的事件按时间分类，对每一个月的新闻数据进行主题提取、事件检测，选出热点最高的 TOP-N 事件，绘制趋势图（根据事件发生的地点不同，可以针对地点、时间进行趋势分析）
- 输出：得到一个完整的食品安全事件检测与分析系统：采集文本、视频、音频数据，进行事件检测与提取，对不同时期相关的社交网络推文进行情感分析，得到食品安全热点事件的变化趋势。