**新闻分类项目讨论会议记录**

**时 间：**2018年4月10日

**地 点：**5-218

**参会人员：**赵义霞老师、刘利老师、王 程、曾庆圳

**本次会议主要对目前的项目所存在的问题进行讨论**

1、通过LDA算法可以产生两个矩阵，文本-主题和主题-词项，由于得到的结果显示的主题是隐含的，在不知道主题是什么的情况下如何得到主题。在讨论之后，决定先去查看一下是否有一种能根据词项得出隐含主题的库。

2、如何确定LDA算法最佳的主题个数，以及主题下对应词项的个数如何确定，进行了探讨。

**会议上对之后一星期的工作进行了安排，安排如下：**

1、通过lda算法可以得到文章-主题概率矩阵、主题-词项概率矩阵；

2、根据主题-词项概率矩阵，结合词料库，可以得到每个主题下的概率比较大的关键词，通过这些关键词总结出主题；

3、根据文章-主题概率矩阵，我们认为这篇文章中概率最大的主题就是这篇文章的主题；

4、根据用户选择的主题，把文章推送给用户。