**葛宇 毕业论文 数据处理需求**

1. 共需处理15本期刊（新闻与传媒类7本，出版编辑类8本）
2. 每一本期刊中包含两个文件夹,例如：附件一《现代传播》导出、附件二《现代传播》被引

附件一：现代传播被引里包含**篇名** **被引频次** **作者** 下载次数等

附件二：现代传播导出里包含摘要、关键词、作者等信息

首先处理“某X期刊(**如期刊：现代传播**)导出”Excel

1. 清除Source**-文献来源不是:**  **现代传播-北京广播学院学报** 或文献来源: **现代传播(中国传媒大学学报) 的文献**
2. 再次清除**没有Author-作者**的文献
3. 接着清除**没有Keyword-关键词**的文献
4. 接着清除：**来源库、文献来源、发表时间、年、卷、期、中图分类号** 信息
5. 接着将经过以上三轮筛检后剩余“现代传播导出” 与 “现代传播被引”用**“篇名+作者”**的形式匹配，得出**附件三**（附件三是将导出的信息横着排列在被引文件后面的）

附件三第一行： 题名、作者、来源、发表时间、数据库、被引、下载（被引文件原有的第一行内容）

**+题名、作者**、**单位、关键词、摘要、第一责任人、基金、页码**（这些内容）

其中（提名和作者我稍后抽查核对后自己会删除的，就是为了确保信息匹配暂时保留上）

1. 对附件3进行进一步的处理
2. 首先对“作者”纵列进行处理，首先在其右边插入一列表格

如果作者是两个人及以上的作者（作者之间是分号？——中英文的我有点不太清楚），在右边赋值1

如果是单独作者，在右边赋值0

1. 针对作者的“单位”，首先在其右边插入一列表格

如果是双一流高校赋值1 双一流高校名单（见附件）

研究所赋值2

如果是编辑部、杂志社赋值3

如果是公司，赋值4

学会、协会，赋值5

其他 赋值0

（3）针对“基金”，有基金赋值“1”，无基金赋值0

1. 针对“页码”

页码是个区间，要把他计算出来

如：“92-97”（从92页到97页），那么这篇文献的页码就是“97-92+1”

如“92-97+117”（从92页到97页，117页这篇文章还有1页），那么这篇文献的页码就是“97-92+1+1”