# 《大数据应用实战》课程考试题目

利用所提供的数据集，在2道大题中选取其中的一个题目进行解决，并按照下面要求提交作业（电子版和纸质版同时交）。

## 1 研究热点的抽取和演变

1.1 关键词的抽取和识别。

1.2 热点聚类（要求，一种是按照年度做聚类，一种是不分年度聚类；聚类结果存储在文件中）

1.3 研究按照年度的热点演变轨迹分析。

## 2 作者关系分析

2.1 作者画像（兴趣，单位，职称，邮箱，地址）。要求，作者画像的结果要存储在数据库中。

2.2 合作者关系挖掘与分析，包括：

要给出作者合作关系网络(合作强度，合作主题等)，并进行存储（文件或数据库都可）；

计算出合作网络社区；

并根据所发现的社区，分析社区结构关系和社区内部作者兴趣间的关系。

## 要求交的作业包括：

1 技术文档1篇，至少2000字。说明针对每个问题的建模，算法，实现代码及重要的注释，结果截图，结果分析。

2 项目文章1篇，至少5000字。按照附件2给出的论文格式来写。

### 3 源代码1套。

附件1：期刊近10余年的数据集；

附件2：可参照的论文模板。