**《信息论基础》第六次上机作业**

——从文本数据中发现知识

文本是一种数据形式，其中蕴含大量语义信息，借助统计学与文本分析方法，可以从中发现潜在的知识。

实验目标：基于 Excel 表格提供的100篇期刊论文的审稿人意见和作者的回应，解决如下问题：

1. 审稿人在多次审同一篇稿件时的情感倾向在发生什么变化？
2. 作者在回应中用了哪些语言表达方式，使得审稿人的情感倾向发生变化？

备注：一篇成功发表的论文在投到期刊编辑系统之后，编辑初审合格后，发送给若干位匿名审稿人审阅，审稿人审阅结束后需要提交评审报告，指出论文中的优点与存在的问题；作者收到评审报告后，会做出相应的修改与回应，之后再由编辑发送给审稿人进行第二轮审阅，作者再依此进行第二轮修改，这个过程会持续若干轮，直到审稿人不再提出任何问题，编辑就决定录用。

Excel表格 中 doi 是论文编号；round 是审稿轮次；reviewer 是审稿人编号；number 是审稿人审稿意见的编号（一个审稿人可以提多条意见）。

请于当天上机课结束前将**PDF文档**发放到邮箱： informatics\_course@163.com。请将文件命名为:学号\_姓名\_HW6。