

2015 年秋《信息系统建模理论与方法》

期末试题

一、 考试内容（小组团队完成，任选一题）

1、 请设计并编程实现一个小型的搜索引擎，具体任务包括：

- 1) 一个至少能支持 5 个以上网站的爬虫程序，且支持增量式数据采集；并至少采集 5000 个实际网页；
- 2) 针对采集回来的网页内容，能够实现网页文本的分类；
- 3) 可进行重复或冗余网页的去重过滤；
- 4) 对经去冗以后的内容建立倒排索引；
- 5) 采用 PageRank 算法、HITS 算法、SALSA 算法或其他你认为合适的算法，实现搜索结果的排序；
- 6) 支持自然语言的模糊检索；
- 7) 可实现搜索结果的可视化呈现；
- 8) 可以在线记录每次检索的日志，并可对日志数据进行统计分析和关联挖掘。

2、 请选择一个实际的社交网络，并开展必要的研究和开发，具体要求和研究内容如下：

- 1) 采集至少包含 1000 个以上用户 1 个星期以上的实际数据；
- 2) 实现该社交网络的可视化呈现；
- 3) 通过计算分析该社交网络是否符合小世界网络的 6 个基

本特性；

- 4) 如何表示社交网络中用户的地位和行为特征？
- 5) 如何表征（或用哪些特征参数）社交网络结构的演化？
- 6) 如何实现基于社交网络的个性化推荐？
- 7) 社交网络环境下舆情传播的动力学模型建模与分析。

二、 完成方式及要求：

- 1、 每道题由一个小组完成，每个小组有 5~10 人组成，每位同学加入一组参与完成一题即可，组织由组内同学产生，小组同学可以自由选择合作伙伴；
- 2、 请注明每个小组参与的全体同学的姓名、学号、班级等，以便登记成绩。
- 3、 作业完成后提交：1) 以电子文档（DOC 或 PPT）说明作业完成的内容，以及小组每一位同学所做的贡献；2) 软件代码；3) 采集的样本数据；4) 必要的参考文献或资料。
- 4、 验证方式：现场讲解 + 实战演示
- 5、 2016 年 3 月 5 日之前提交，过期无效。
- 6、 2016 年 3 月 6 日（星期天下午）2:30，每一小组半个小时，现场答辩和演示完成情况，并现场打分决定最后成绩。

三、 成绩计算：

平时成绩：50 分，本次大作业：50 分。