

毕业设计任务书

软件工程本科专业（2013 级）

题目： 基于 Hadoop 的个性化书籍推荐系统的设计与实现

学生姓名： 毛烨辉

班级： 13 级软件 2 班

学 号： 13270212

题目类型： 软件工程（R）

指导教师： 张秋余

一、设计目的和意义

传统的读者会选择在书店或图书馆寻找自己感兴趣的书籍，但面对各类书架上琳琅满目的书籍往往会无所适从，往往要翻阅几十本甚至上百本书籍才能找到自己感兴趣的书籍。这通常会花费读者大量的时间。随着互联网技术的不断发展，网络上书籍的数目和类别都将以指数型增加，所以对海量数据进行分析向读者进行个性化书籍推荐显得尤为重要。

个性化书籍推荐是根据用户的兴趣爱好及用户的个人行为计算相似度，并向用户推荐感兴趣的信息和书籍。这将改变读者传统寻找书籍的方式，通过有针对性的个性化推荐，系统从海量书籍中筛选出读者感兴趣的书籍，从而节省读者大量时间和精力去寻找感兴趣的书籍，还可以让读者更加方便、快捷地了解各种类的热门书籍和最新书籍。

通过该题目的分析和设计，使学生初步得到软件工程的训练，全面培养软件开发过程中的分析、设计、编码、测试及文档规范书写的能力，得到软件工程的综合训练，提高解决实际问题的能力。

二、设计任务及要求

1、完成基于Hadoop的个性化书籍推荐系统软件的软件可行性分析、项目开发计划、需求规格说明、概要设计和详细设计、设计开发、运行测试工作。

2、开发工具与环境： B/S架构、基于JSP技术、Hibernate、Spring、Struts框架等来实现前台后台交互，以MySQL和Hbase为数据库系统，将HDFS作为文件存储系统，使用MapReduce计算模型对数据进行离线计算分析。为了提高系统的可维护性可采用MVC框架技术，采用UML建模技术进行面向对象分析与设计。对比统计分析图插件建议采用开源HighCharts控件。

3、整个系统由账号管理，用户书籍，好友管理，精准推荐，书库管理模块等功能模块组成，各模块间耦合度较小；系统要实现用户对喜好书籍的录入、修改、查询；用户对朋友的关注、取消关注；用户对账号信息的增删改查；系统根据用户个性喜好的精准推送等功能。该任务书完成的具体功能为：

账号管理模块：注册模块：未注册用户需注册才能登陆，只有注册的用户才提供精准推送服务、登录模块：只有登录才能对各模块进行管理和查看系统推荐的书籍信息、信息显示模块：显示当前用户信息、信息修改模块：修改当前用户信息；

用户书籍模块：用户对喜爱书籍的信息进行录入、修改、查询和评价等操作，使其能分享给其他用户，并对精准推送提供数据支撑；

好友管理模块：用户对其他用户进行关注和取消关注，可以查看被关注者阅读的书籍列表及评价，对感兴趣的书籍进行收藏操作。作为精准推送的数据支持之一；

精准推荐模块：根据用户喜好书籍列表和关注用户所阅读书籍的列表进行分析，通过协同过滤算法计算相似度，产生用户潜在的兴趣书籍，并对其推荐；

书库管理模块：管理所有录入的书籍，并书籍进行分类，计算产生统计数据，如：各类别前10热门书籍、最新书籍

4、设计要求与技术指标

整个系统界面简洁、操作方便、统计图表直观。设计时要注意数据操作的安全措施的设计，每个读者只能看到自己的数据。

5、测试

系统要求采用实际数据作为测试用例进行功能性测试。

三、各阶段时间安排（共 13 周）：

设计内容	起止周
搜集参考资料，分析题目要求。	第 1 周
教师指导学生学习进行需求分析。	第 2 周
学生提出系统总体设计方案和详细设计方案。	第 3 周~第 4 周
功能实现（软件编程）。	第 5 周~第 8 周
软件调试。	第 9 周
撰写、修改毕业论文	第 10 周~第 12 周
准备答辩资料	第 13 周

四、主要参考文献

- [1] (美) Tom White 著. 华东师范大学数据科学与工程学院(译). Hadoop 权威指南（第 3 版）[M]. 北京：清华大学出版社，2015.
- [2] (美) Lars George 著. HBase 权威指南[M]. 北京：人民邮电出版社，2013.
- [3] (美) Edward Capriolo Dean Wampler Jason Rutherglen 著, Hive 编程指南[M], 北京：人民邮电出版社，2013.
- [4] (美) Sandy Ryza Uri Laserson Sean Owen 著. Spark 高级数据分析[M], 北京：人民邮电出版社，2015.
- [5] (葡) Luis Torgo 著. 数据挖掘与 R 语言[M], 北京：机械工业出版社，2008.
- [6] (挪) Magnus Lie Hetland 著. Python 基础教程（第 2 版 • 修订版）[M]. 北京：人民邮电出版社，2014.
- [7] (澳) Richard Lawson 著. 用 Python 写网络爬虫[M]. 北京：人民邮电出版社，2016.
- [8] (美) Matthew A. Russell 著. 社交网络的数据挖掘与分析[M]. 北京：机械工业出版社，2015.
- [9] 董启文 著. 基于语言处理技术的蛋白质结构和功能预测若干问题研究[M]. 博士论文,2007.
- [10] 赛奎春. JSP 工程应用与项目实践[M]. 北京：机械工业出版社，2008.
- [11] 李东艳,崔崇芳等. J2EE 应用开发实例精解[M]. 北京：清华大学出版社，2011.
- [12] HighCharts Demo[EB/OL], <http://www.highcharts.com/demo/>
- [13] 简朝阳 著. MySQL 性能调优与架构设计[M]. 北京：电子工业出版社，2009.
- [14] 谭火彬 著. UML2 面向对象分析与设计[M], 北京：清华大学出版社，2013.
- [15] Hashim N M Z, Mohamed S. Development of Student Information System[J]. International Journal of Science and Research (IJSR) Volume, 2013, 2: 256-260.
- [16] Andrea Steelman, Joel Murach. Murach's Java Servlets and JSP[M]. Mike Murach & Associates Inc. 2nd Revised edition,2008.
- [17] Li F. Design of University Instruction Administration System Based on JSP for Innovative Engineering[C]//Proceedings of the 9th International Symposium on Linear Drives for Industry Applications, Volume 2. Springer Berlin Heidelberg, 2014: 541-548.

附录：提交的成果

1. 毕业设计说明书一份，内容包括：
 - 1) 中、英文摘要 200 字；关键词 3-5 个；
 - 2) 序言；
 - 3) 系统分析（包含需要的系统流程图、数据流程图、数据字典）；
 - 4) 系统总体设计（包含总体软件结构图、总体数据结构）；
 - 5) 详细设计；
 - 6) 系统测试（包含测试方案、测试用例、测试结果及软件可靠性分析）；
 - 7) 软件使用说明书（核心界面说明）；
 - 8) 设计总结、参考文献、致谢等。
2. 完成 10000-30000 个印刷符号的外文文献翻译稿一篇。
3. 完成毕业设计简介一份，要求 4 页，科技论文格式。
4. 刻制光盘一张。