

数字档案馆数据交换系统的设计与实现

张金英, 李蜀瑜

(陕西师范大学 计算机科学学院, 陕西 西安 710062)

摘要:随着信息社会的迅猛发展和电子政务的普及,建设数字档案馆对实现档案资源的信息化管理有着重大意义。通过数据交换系统,可以为各部门提供数字档案和实体档案的归档和档案的接收;实现数字档案馆内部子系统间的数据交换。数据交换系统是数字档案馆相关子系统之间,以及数字档案馆和电子政务平台之间的档案数据交换平台,是数字档案馆的核心系统之一。文中分析了基于 XOOPS 的数字档案馆网络拓扑结构以及数据交换系统架构,从业务层、中间层和数据层对数字档案馆的关键子系统——数据交换系统进行了研究。

关键词:数字档案馆;XOOPS;软件体系架构;网络拓扑结构;数据交换

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2012)11-0100-05

Design and Implementation of Digital Archives Data Exchange System

ZHANG Jin-ying, LI Shu-yu

(School of Computer Science, Shaanxi Normal University, Xi'an 710062, China)

Abstract: With the rapid development of the information society and the popularity of the electronic government affairs, the construction of digital archives has a great significance on realizing file resource management. Through data exchange system digital archives can provide archiving and receiving of the digital files and physical files, at the same time it is the exchange platform between the digital archives internal subsystems which are related. The data exchange system is the file data exchange platform among the digital archives subsystems which are related, and is also the platform between the digital archives and the electronic government affairs. From the above, the data exchange system is the core system of the digital archives application. In this paper, analyse the network topology structure of digital archives application based on the XOOPS foundation and the data exchange system architecture. Research the key subsystem of digital archives—data exchange system on the business layer, middle layer and data layer.

Key words: digital archives; XOOPS; software architecture; network topology structure; data exchange

0 引言

数字档案馆^[1]建立在现代化信息技术普遍应用的基础上,利用数字化手段,以综合档案信息资源为处理核心,对数字档案馆中的信息资源进行收集、管理,通过高速宽带通信网络设施相连接和提供利用,实现资源共享的超大规模、分布式数字化信息系统。随着国民经济的发展和人民物质文化水平的不断提高,为了促进社会经济发展与和谐,全面提升档案管理的数字化水平,以及提高管理效率,档案管理的信息化越来越重要。

面向对象的可扩展门户系统(eXtensible Object

Oriented Portal System, XOOPS^[2])是一个强大的基于Web的内容管理平台。XOOPS集成了Apache、MySQL、FileZilla、Mercury软件,在安装到一个支持PHP语言、有像Apache一样的Web服务器、有mysql数据库的互联网主机上后,通过XOOPS就可以轻松创建动态的、复杂的以及内容丰富的Web站点,从简单如个人blog到复杂如公司门户。数字档案馆正是在XOOPS环境下设计并开发的,利用XOOPS使数字档案馆项目的开发工作轻松、功能强大。

文中在分析数字档案馆数据交换网络拓扑结构与系统架构的基础上,采用B/S的设计模式,基于XOOPS和使用动态网页设计技术,对数字档案馆系统中的关键系统——数据交换子系统进行了设计和开发。

1 系统网络拓扑结构

数字档案馆系统支持档案馆内部用户和外部社会

收稿日期:2012-03-19;修回日期:2012-06-24

基金项目:中央高校基本科研业务费项目(GK201002011)

作者简介:张金英(1986-),女,山东济南人,硕士研究生,研究方向为Web服务;李蜀瑜,博士,副教授,研究方向为嵌入式系统、Web服务、数据挖掘。

用户使用,并且提供与电子政务网络之间的网络接口,系统的网络^[3]拓扑结构如图 1 所示。

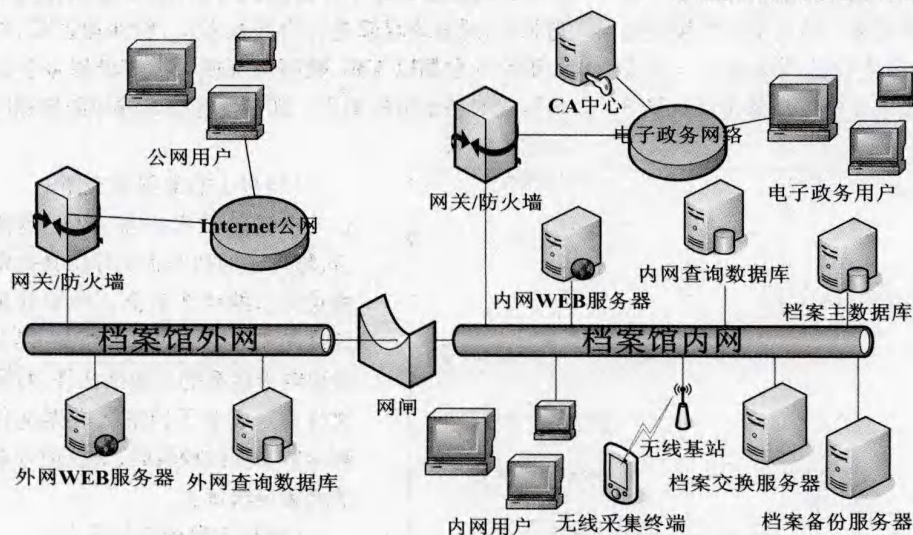


图 1 数字档案馆系统的网络拓扑结构图

整个档案馆网络分为內网和外网,其中内外网之间通过网闸设备,即网闸的外部主机连接外部网络,网闸的内部主机连接内部网络,网闸的外部主机和内部主机是完全物理隔离的,支持文件、数据或信息的交换。档案馆內网中,在 Web 服务器中部署系统软件,在內网查询数据库中存放电子档案数据,供档案查询利用。在档案主数据库中,则存放所有的档案对象数据和元数据。无线采集终端则通过无线基站接入到档案馆內网,档案馆內网用户则通过內网,直接访问系统服务。档案馆內网,通过网关设备与电子政务网络连接,能够直接接收电子政务网络內用户访问,以及访问 CA 认证中心服务器。同时,通过档案交换服务器接收来自电子政务平台的档案。档案馆外网上,也部署有 Web 服务器,部署外网用户访问相关应用程序。外网上同时部署有存储档案查询元数据的外网查询数据库。档案馆外网通过网闸设备,与 Internet 网络相连接,能够提供给 Internet 广域网用户访问 Web 服务。

数据交换系统的主要功能就是进行物理隔离^[4]环境下两个网络之间的数据同步,负责数据的导入、导出等,下文将研究数据交换系统的设计。

2 数据交换系统架构

数据交换系统要为各部门提供数字档案和实体档案

的归档和档案的接收;实现数字档案馆内部的子系统间的数据交换;数据交换系统中提供相应的数据接口,实现数字档案馆与同级和上级档案馆的档案数据交换以及与省内市(区)县间的档案文件信息交换;通过使用组织机构、服务器和个人数字证书对档案数据进行

签名,将馆藏实体档案生成成为数字档案。通过对所接收的数字档案进行数字签名验证,完成数字档案的接收、存贮和管理;同时支持联机交换与脱机交换数据。基于此,文中设计的数据交换系统架构如下。

数据交换系统架构分为三个层:业务层、中间层、数据层。数据层采用 PostgreSQL 数据库管理系统。中间层数据交换平台的组件包括: XOOOPS 内容管理框架、XML 解析引擎、安全认证组件、Mail 组件、PRADO PHP 组件。业务层在数据层与中间层基础上完成数据交换系统的业务。数据交换系统架构图如图 2 所示。

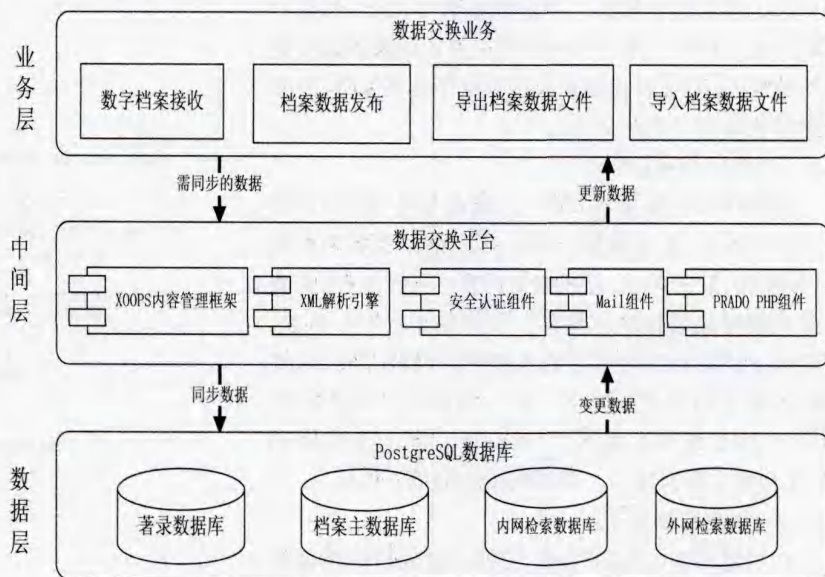


图 2 数据交换系统架构图

2.1 数据交换系统业务层

该系统数据交换业务流程为:数字档案接收、数字

档案发布、导出档案数据文件、导入数据档案文件。通过电子政务网络,或者采用U盘等载体,档案馆可以接收或发送给各级单位电子档案。档案发送与接收是数据交换的关键部分,文中重点设计档案接收与发送的业务流程。电子档案的接收发送流程泳道图如图3所示。

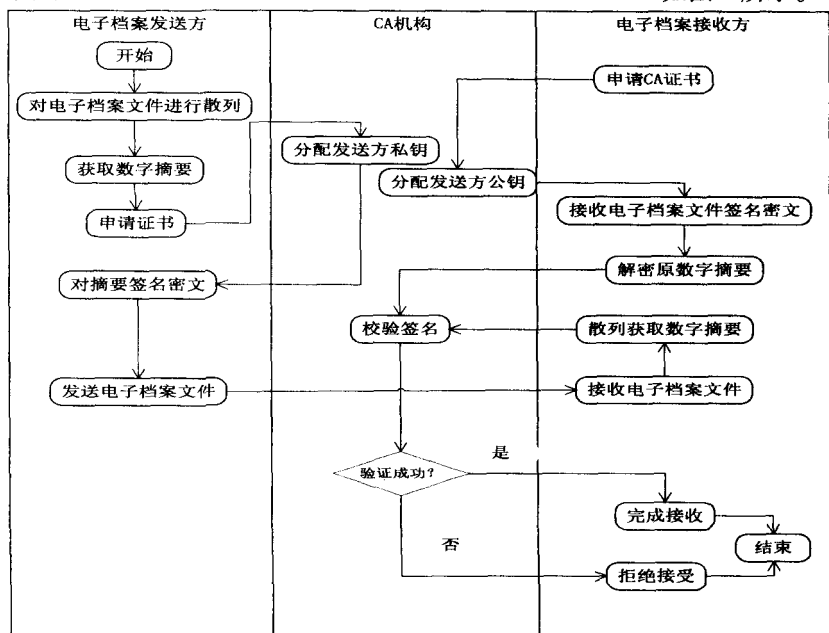


图3 电子档案接收泳道图

档案发送方对电子档案对象文件进行散列获取数字摘要并通过向认证中心CA申请证书,用证书^[5]中的发送方私钥对数字摘要进行签名;档案接收方通过向CA中心获取发送方的公开证书,获取发送方公钥对获取的签名密文解密^[6]得到原始数字摘要,然后对接收的电子档案对象文件进行散列获取的数字摘要进行对比校验,验证成功则完成接收电子档案文件,否则拒绝接收该电子档案文件。

(1) 数字档案接收。

档案接收主要通过邮件^[7,8]接收的方式,邮件接收大致流程为:首先解析Email,获取电子档案对象数据、校验码、发送单位、元数据等信息。用校验码,检查电子档案对象数据的完整性。然后向CA中心请求,校验电子档案元数据签名的正确性。检验成功后,对接收的电子档案进行签名。最后对电子档案进行归档^[9,10],补充其档案编号^[11]等信息。将电子档案归档,保存到主数据库中。其详细流程如图4所示。

(2) 档案数据发布。

在划定发布内容和时间范围后,将规定的档案数据导出为档案数据文件,档案管理员可以手工选择发布受控档案到内网和外网查询数据库中,为每个档案选择发布目标:内网查询数据库、外网查询数据库。对于档案主数据库中的公开档案,将实时发布到内网查

询数据库和外网查询数据库中,提高公开档案的数据发布实时性和利用率。在接收到第三方档案数据查询请求后,对查询请求进行合法性验证,验证通过后,将所请求的数据以XML数据服务或数据邮件服务传送给第三方档案数据用户。邮件发送档案的详细流程图如图5所示。

(3) 导出档案数据文件。

在划定导出内容、时间范围后,将规定的档案数据导出为指定格式要求的档案数据。档案管理员首先选择导出的电子档案,以及导出电子档案的元数据文件、对象文件。生成电子档案元数据文件和对象文件的校验码,系统记录电子档案导出日志。

(4) 档案数据文件导入。

首先要对需要导入的档案数据文件格式进行人工分析,并对数据规则进行手工定义,然后由本系统根据数据格式规则和数据转换定义选择需要导入的电子档案对象文件,将电子档案导入到档案主

数据库,系统记录电子档案导入日志。

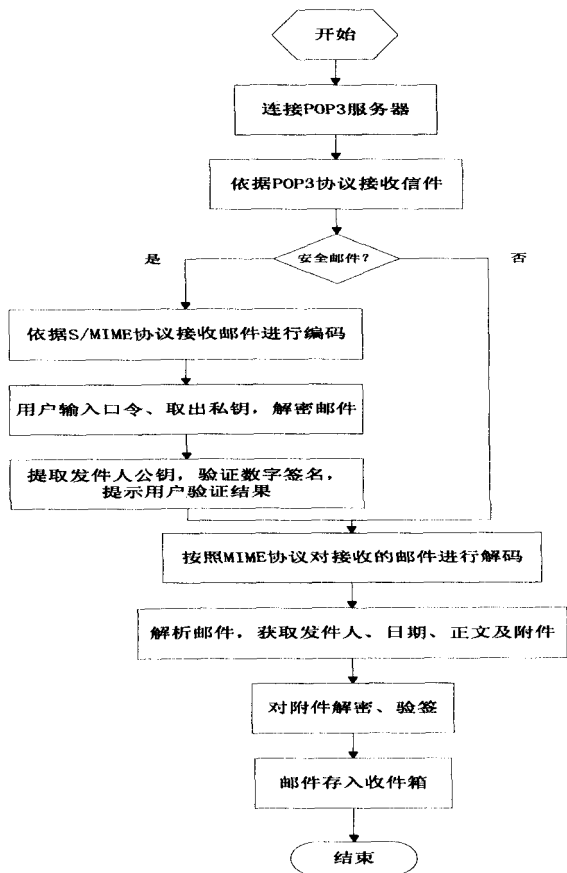


图4 邮件接收档案文件流程

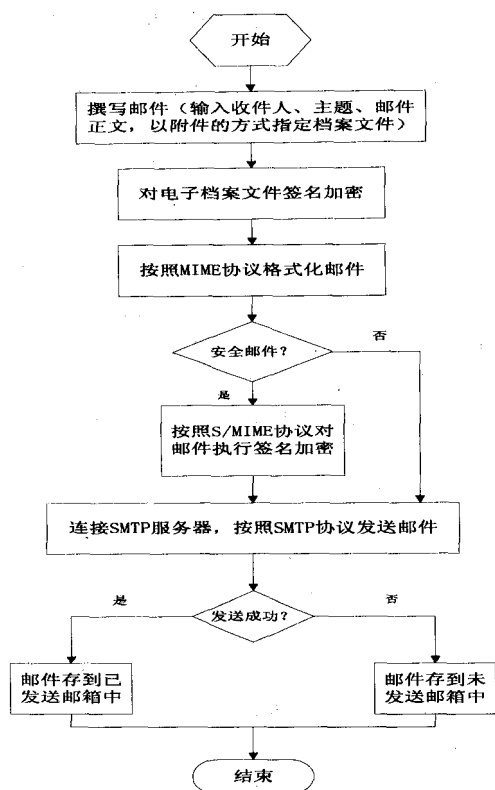


图 5 邮件发送档案文件流程图

2.2 数据交换系统中间层

中间层也称逻辑层,为业务层提供支撑和组件,联系业务层与数据库,使得业务层可以访问数据库。数据交换平台组件包括:XOOPS 内容管理框架、XML 解析引擎、安全认证组件、Mail 组件、PRADO PHP 组件。

XOOPS 是成熟的 Web 应用框架体系。已经在数字档案馆的著录和网站发布子系统中成功应用,XOOPS 基于 GPL 许可协议,不需要缴纳任何的许可和授权费用。采用 PHP 脚本语言编写,PHP 具有开发效率高、性能高、稳定性好、跨平台、维护成本低等突出优点。

XML 解析引擎,用于对 XML 格式的档案元数据提供数据解析和打包服务,本系统中文书档案目录数据文件交换使用 XML 格式^[12,13],解决数据异构问题^[14],方便数据的检索和利用。

安全认证组件,支持基于 CA 的数字签名和认证,能够通过与 CA 中心交互,进行签名认证,以及完成数字签名,本系统中数字档案接收模块向 CA 验证数字档案上的数字签名是否合法和有效。该数据交换系统通过使用组织机构、服务器和个人数字证书对档案数据进行签名,将馆藏实体档案生成为数字档案。通过对所接收的数字档案进行数字签名验证,完成数字档案的接收、存贮和管理。

Mail 组件,即利用 SendMail (基于 SMTP 协议的因特网邮件交换系统)实现邮件交换等功能,本系统

中利用 Mail 实现了邮件接收与发送档案的功能。

PRADO PHP 组件是基于组件的 Web 应用开发框架,使用面向对象方法,通过使用已有的组件,设置组件属性,为控件组件的各种事件编写对应的处理函数,模仿桌面消息处理模式,以事件驱动方法处理数据。

利用中间层设计的物理隔离下实现内外网之间数据交换的三种方式如下:

(1) 内网档案数据库与外网检索数据库的交换。

a. 浏览器上管理员筛选出需要交换的档案数据,采用 XML 格式封装选出的数据。

b. 将 XML 与相关的档案文件压缩后,在页面上提供链接供用户下载。

c. 下载后用户用 U 盘等存储设备将压缩包拷贝到另外能连接外网的计算机上,通过 sftp 上传至外网服务器,外网服务器检测到有新的文件上传,解压文件后按照增删该数据库修改外网查询数据库,同时将档案文件拷贝至文件目录下。

(2) 内外网查询业务数据库之间的交换。

交换的方式同(1)描述。内外网业务数据交换是双向的,外网的预约信息需要交换到内网、内网的审批信息需要交换到外网,不需要用户筛选数据,系统自动地筛选出哪些该交换,自动组织成 XML 文件,没有文件的交换。

(3) 内外网用户信息交换。

交换的方式同(1)描述。内外网数据交换是单向的,只把外网的用户信息交换到内网上,供管理员审批统计时使用。

2.3 数据交换系统数据层

数据层采用 PostgreSQL 数据库管理系统,对数字档案对象数据、元数据以及各种业务数据进行统一管理。该系统数据库分为著录数据库、主数据库、内网数据库与外网数据库。

著录数据库是通过著录系统著录的数据库。主数据库内容涵盖著录数据,以及可扩展的其他数据库。外网数据库是主数据库的子集,供公网上的用户检索预约。内网数据库也是主数据库的子集,供内部用户检索查询预约等。在联网状态下,各数据库分布在网络上,通过数据交换平台达到数据交换的目的,脱机状态下将数据库导出到文件,再用存储介质,比如 U 盘等拷到目的数据库。

数据交换依据 GB/T 7408-1994^[15] 标准,为了保证电子档案数据的安全性和可靠性,系统每周对整个数据库进行全局数据备份,生成全局备份文件。系统每天进行增量备份,生成增量备份文件,备份采取静态备份方式。假如系统崩溃,系统管理员可以通过备份文件,恢复整个数据库。

3 结束语

文中设计开发的基于 XOOPS 的数字档案馆数据交换子系统,使用了 FreeBSD 操作系统和 PostgreSQL 数据库管理系统,采用标准化的信息交换格式,利用 CA 证书和签名验证技术以及 SendMail 邮件交换系统,开发了数据导入、数据导出、数据注入、地方档案馆邮件接收等功能模块。本系统设计合理,功能实用,稳定性好,可靠性高,符合国家对数字化档案馆建设的要求。该系统已经投入使用,用户反馈良好。目前该数字档案馆系统功能仍在不断改进中,后期以提高数据交换效率为主要目标,改进的方面还包括增加管理监控的功能以及增强消息传递过程中的稳定性和健壮性等。

参考文献:

- [1] 国家档案局. 数字档案馆建设指南[EB/OL]. 2010-06-28. <http://wenku.baidu.com/view/246eaf87bceb19e8b8f6babf.html?from=rec&pos=0&weight=29&lastweight=15&count=5>.
- [2] 站长百科教程组. XOOPS 教程[EB/OL]. 2009-10-18. <http://www.zzbaike.com/wiki/Xoops>.
- [3] SJ/T11268-2002, 网络信息分类系统标准[S]. 2002.
- [4] 杨雪微. 物理隔离数据交换系统的设计[D]. 上海: 同济大学, 2009.
- [5] 张卫东. 网络安全理论与技术[EB/OL]. 2010-06-29. <http://wenku.baidu.com/view/97b6e01614791711cc7917a1.html>.

(上接第 99 页)

二是遵循有关标准和规范,如电气和电子工程师协会的软件项目管理计划标准以及美国卡耐基·梅隆大学 SEI (Software Engineering Institute) 的 CMM/PSP/TSP 软件过程改进体系;

三是利用支持工具,如美国微软公司的 Microsoft Project、北京维普时代软件有限公司的 Visual Project 以及获得《电脑杂志》“编辑选择奖”的 Scitor 公司的基于 Windows 的项目管理软件 Project Scheduler 等^[12]。

软件项目进度计划控制研究当前在国内外缺乏系统性的理论及方法。文中从软件项目进度计划管理的实际出发,解决软件项目管理过程中进度计划制订与控制的关键性科学问题,旨在提高软件项目的管理水平,提高软件项目开发的成功率,为软件应用企业提供高质量的软件产品,为软件开发企业创造更多的经济价值。

参考文献:

- [1] 韩承双,程再玲,程和侠. 软件项目过程管理矩阵模型研究与实践[J]. 计算机技术与发展, 2008, 18(4): 199-201.
- [2] 刘国靖. 21 世纪新项目管理[M]. 北京: 清华大学出版社,

<http://wenku.baidu.com/view/97b6e01614791711cc7917a1.html>.

- [6] 佚名. Java 加密解密实现[EB/OL]. 2010-01-17. <http://www.iteye.com/topic/572581>.
- [7] 禹水琴,史燕,洪飞龙. 基于 SMIME 的安全电子邮件系统的设计与实现[J]. 计算机与数字工程, 2005, 33(10): 137-139.
- [8] 陈建奇,张玉清,李学农,等. 安全电子邮件的研究与实现[J]. 计算机工程, 2002, 28(6): 121-122.
- [9] GB/T 7156-1987, 文献保密等级代码[S]. 1987.
- [10] GB/T18894-2002, 电子文件归档与管理规范[S]. 2002.
- [11] DA/T 1-2000, 档案工作基本术语[S]. 2000.
- [12] Ralf R, Carmen C, Uwe H, et al. Champagne: data change propagation for heterogeneous information systems[C]//Proceedings of the 28th VLDB Conference. Hong Kong, China: [s. n.], 2002: 1099-1102.
- [13] Carmen C, Uwe H, Ralf R, et al. A system for data change propagation in heterogeneous information systems[C]//Proceedings of the 4th International Information Systems. Ciudad Real, Spain: [s. n.], 2002: 73-80.
- [14] Hakimpour F, Geppert A. Resolving Semantic Heterogeneity in Schema Integration: An Ontology Based Approach[C]//Proceedings of the International Conference on Formal Ontology in Information Systems. Maine: [s. n.], 2001: 297-308.
- [15] GB/T 7408-1994, 数据元和交换格式、信息交换、日期和时间表示法[S]. 1994.
- [16] 解登峰. 中国海洋大学新校区图书馆建设项目进度计划与控制管理[D]. 青岛: 中国海洋大学, 2009.
- [17] 黄立威, 黄伟, 冯径. 支持软件质量控制的软件配置管理研究[J]. 计算机技术与发展, 2011, 21(7): 50-53.
- [18] Gray E, Sampaio A. An Incremental Approach to Software Process Assessment and Improvement[J]. Software Quality Journal, 2005, 13(1): 7-16.
- [19] 任永昌, 邢涛. 基于挣值管理的软件开发成本控制方法的研究[J]. 中国管理信息化, 2007, 10(10): 33-35.
- [20] 张家浩. 软件项目管理[M]. 北京: 机械工业出版社, 2005.
- [21] 任永昌, 邢涛, 鄂旭. 软件项目开发过程管理[M]. 北京: 北京交通大学出版社, 2010.
- [22] 范勇. 一种改进的软件项目计划管理方法[J]. 南京大学学报(自然科学版), 2005, 41(10): 699-703.
- [23] 范勇, 陈波. 基于 PSP 的软件项目计划管理[J]. 计算机应用与软件, 2007, 24(3): 81-83.
- [24] 况彬. 软件项目管理的计划和控制方法研究[D]. 贵阳: 贵州大学, 2006.
- [25] Li D Y. Expanded Study on Criticality in Schedule Plan of Project[J]. Advances in Intelligent and Soft Computing, 2012, 127(1): 899-911.