

科技项目档案数字化管理系统平台设计实践

王 杉, 刑 军, 孙 滔

(陕西省科学技术信息研究所, 陕西 西安 710054)

摘 要:随着信息技术的发展,大数据、云计算技术的成熟,海量科技项目档案的管理和利用,已经由纸质收藏与手工检索模式向数字化存储互联网共享服务模式转变。基于此,利用面向对象的软件设计方法,构建了基于互联网的科技项目档案信息系统平台,提出了平台设计目标、建设内容,对系统功能需求和系统安全性进行了分析,介绍了系统平台的整体框架结构和系统功能模块设计。旨在为科技项目档案的收藏、管理、挖掘、共享提供一套先进经验,让海量的基础科技项目档案资源能够为科技创新工作提供支持和服务。

关键词:大数据;云计算;科技项目档案;平台;功能

中图分类号:TP302

文献标识码:A

文章编号:1673-629X(2016)08-0143-03

doi:10.3969/j.issn.1673-629X.2016.08.030

Design and Practice of Digital Management Platform for Scientific and Technological Project Archives

WANG Shan, XING Jun, SUN Tao

(Institute of Scientific and Technical Information of Shaanxi, Xi'an 710054, China)

Abstract: With the development of information technology and mature of big data and cloud computing technology, the management and utilization of massive scientific and technological project files have been changed from the paper collection and manual retrieval mode to the digital storage based on Internet sharing service mode. The object oriented software design method is used to construct the information system of science and technology project based on Internet, and the design goal and content of the platform is put forward, and the system function requirement and security is analyzed, and the overall framework and the function modules design of the system are introduced. The aim is to provide a set of advanced experience for the collection, management, mining and sharing of scientific and technological project files, which makes the mass of the basic science and technology project file resources to provide support and services for scientific and technological innovation.

Key words: big data; cloud computing; archives of science and technology project; platform; function

0 引 言

科技项目科研档案是指在科技项目科研活动中形成的应当归档保存的网纸、网表、文字材料、计算材料、照片、影片、录像、录音带等科技文件材料,是科技工作人员进行项目科学研究活动的真实记录。它记载和反映了项目科学研究活动的全部过程和具体成果,是项目科研人员的劳动成果和智慧的结晶。科技项目科研档案归档材料应包括研究课题的可行性论证材料、申请立项材料,立项后各阶段的试验原始记录材料,最后的总结报告、成果报告、成果鉴定、推广示范和反馈信息材料等^[1]。长期以来,对项目受理和项目管理中形成的科技档案采用简单分类和纸质化集中存放的方式

进行管理和目录式手工检索利用。随着信息技术的发展,大数据和云计算技术脱颖而出,使得对海量科技数据资源的挖掘和分析变成可能,而科技档案作为基础的海量科研数据资源,如何对其进行加工分析、利用已显得尤为迫切。同时国内信息资源共享研究也在逐步深入,档案信息资源共享研究的领域和范围不断拓展,已从馆藏资源电子化、数字化等纯技术研究拓展到档案信息资源开发、管理、整合、规划、共享模式、共享平台建设、共享体系构建、共享效率、共享系统绩效评估、共享策略、共享隐私保护等各方面^[2]。

因此,针对陕西省的科技项目档案设计出一套数字化管理平台,对陕西省科技档案进行数字化加工并

收稿日期:2015-12-13

修回日期:2016-04-09

网络出版时间:2016-08-01

基金项目:陕西省科技统筹创新工程计划项目(2015PTXX-05)

作者简介:王 杉(1965-),女,馆员,研究方向为文献信息、科技查新。

网络出版地址:<http://www.cnki.net/kcms/detail/61.1450.TP.20160801.0907.060.html>

建立科技档案资源数据库系统,与陕西省科技业务综合系统无缝对接,实现数据交互,对省科技资源统筹大数据云平台本地科技资源数据库进行数据填装,通过互联网和移动互联面向政府决策、科研生产提供知识服务,实现海量科技项目档案资源的挖掘和共享。

1 设计目标

科技项目档案信息管理系统(见图 1),是依托海量的纸质科技项目档案资源,进行数字化加工,建立科技项目档案目录数据库和全文数据库,形成陕西省科技档案信息管理系统,实现与陕西省科技业务综合管理系统进行数据交汇的同时向陕西省科技资源统筹大数据云平台进行数据填装,实现科技档案资源的同步共享发布。从而加快科技档案的数字化进程,提升科技档案的管理与服务能力,促进陕西科学数据被科研生产部门公开使用^[3]。

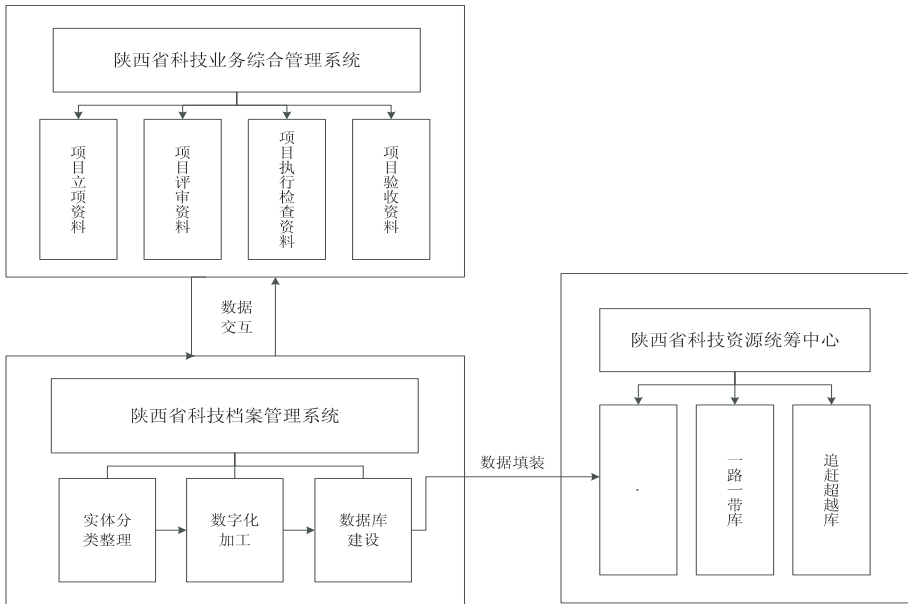


图 1 科技项目档案管理系统

2 建设内容

系统建设主要包括基础设施建设、数字化加工服务、应用系统建设,如图 2 所示。

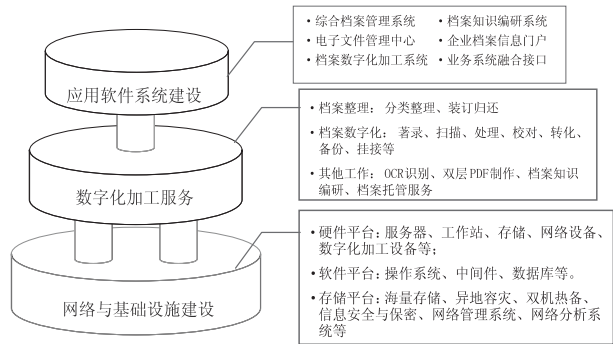


图 2 系统结构图

科技项目档案信息管理系统平台,是集海量科技项目档案数据资源优势、先进数字化加工处理工具、前沿服务模式优势于一体,具有理念新颖、功能完整、服务模式超前等优点。其主要目标如下:

(1)依据科学技术档案案卷构成的一般要求,建立科技项目档案的数字化模板和标准。

(2)利用先进的数据库技术,建立科技档案目录数据库和科技档案全文数据库,实现科技档案的快速全文定位检索。

(3)通过科技档案信息化管理平台,为科研人员、科技项目管理者提供科技项目档案数据信息的挖掘、分析与推送服务^[4]。

(4)通过与科技业务综合系统的无缝对接,实现科技项目数据的交互;通过对科技资源统筹大数据云平台进行数据的填充,实现科技项目档案数据同步共享发布^[5]。

2.1 基础设施建设

基础设施建设主要包括实体科技档案的存放空间建设和科技档案数字化存储以及科技档案管理系统运行所需要的软硬件网络环境的建设。主要包括档案密集柜的购置,服务器、存储设备以及操作系统、数据库软件的部署。同时要考虑科技档案信息的安全防护、容灾备份的建设。

2.2 数字化加工服务

数字化加工服务主要包括实体科技档案的分类整理、拆卸、著录、扫描、处理、校对、转化、挂接以及 OCR识别和双层 PDF 制作等^[6-7]。

2.3 应用系统建设

应用系统建设包括科技档案综合管理系统建设、科技档案系统与科技业务综合系统、科技统筹资源大

数据云平台的对接等模块。

3 系统功能分析

建立在海量科技计划项目档案数据上,面向科技管理和科研需求,科技计划项目档案管理系统必须有完善丰富的系统应用和管理功能解决方案。系统功能分为三个功能模块,包括数据访问层模块、业务逻辑层模块、界面表示层模块,涵盖了系统管理、实体案卷管理、库房管理、案卷信息管理、数据导入导出管理、案卷查询、数据统计与分析、报表管理等^[8-10]。

(1)系统管理:包括系日志管理、用户管理、权限管理、数据字典管理、统计分析、告警管理、身份卡管理等功能,通过系统进行设置^[11]。

(2)实体案卷管理:包括档案接收登记、借阅管理、档案盘点管理、编研利用管理、档案销毁管理等功能,实现对实体纸质案卷的数字化管理^[12]。

(3)库房管理:包括库房场地管理与排架管理,便于实体纸质档案的科学化存放和收藏。

(4)案卷信息管理:包括全宗管理、案卷管理、文件管理、档案影像管理功能,系统性地对案卷信息进行统计、管理。

(5)案卷查询:涉及档案检索服务、文件检索服务、案卷复制服务、案卷阅览服务,方便科技管理者和科研人员便捷地获取科技档案信息数据。

(6)数据导入导出管理涵盖数据导入管理、数据导出管理、防篡改验证服务、防篡改处理,系统具有开放性的数据接口,以便与OCR识别系统、全文检索系统等的集成,将现有档案数据向档案管理系统数据库的迁移或共享,同时保证系统与科技综合业务系统及科技资源大数据云平台进行安全的数据交互。

(7)数据统计与分析:包括对科技档案的数据与用户使用情况进行统计分析,供科技管理者进行决策。

4 系统安全需求分析

由于科技项目档案数据的保密性,因此对于科技项目档案系统的安全性有特别高的要求。系统的设计严格保障系统和项目档案数据的安全性能,包括对整个系统管理、设备、软件、网络等全方位的体现。

由于在互连网环境下,用户对任何资源(包括硬件和软件资源)都能够进行共享,所以必须通过制定相应的网络和系统安全策略来防止非法访问者访问数据资源,对数据资源的存储以及传输进行安全性保护。从网络级安全、传输级安全、系统级安全 and 应用级安全等几方面进行考虑,主要遵循下面的设计原则:

(1)标识与确认(Identification & Authentication)。

任何用户访问系统资源,必须得到系统的身份认

证以及身份标识,如用户的数字证书、用户号码、密码。当用户信息与确认信息一致时,才能获准访问系统。在科技项目档案管理系统中,对操作系统、数据库系统 and 应用系统都有相应的用户和权限的设置,也可与操作系统的安全策略集成。

(2)授权(Authorization)。

对系统资源,包括程序、数据文件、数据库等,根据其特性定义其保护等级;对不同的用户,规定不同的访问资源权限,系统将根据用户权限,授予其不同等级的系统资源的权限。系统提供功能授权、数据授权。

(3)日志(Logging)。

为了保护数据资源的安全,在系统中对所保护的资源进行任何存取操作都进行相应的记录,形成日志存档,完成基本的审计功能。系统的核心基础框架提供各种日志审计功能。

(4)加密(Encrypt)。

系统数据在传输过程中都采用加密方式。在系统的应用层面,系统对用户与权限提供完善的管理机制,每个用户登录后,只能在自己设定的权限范围内操作,并且操作权限有相互制约的机制,如负责采集、上载档案的用户,将不能具有审批的权限等^[13]。

5 结束语

科技项目档案数字化管理系统的设计以面向科技管理和科技创新的全过程出发,依托海量项目档案资料数据资源,利用系统平台先进的数据处理功能,深度挖掘分析档案数据资源,无缝对接科技业务综合系统与科技资源大数据云平台,有效整合互联网和移动互联网服务模式,既满足科技管理数字化、网络化需求,同时也满足了不同科技创新主体的个性化需求。实现了科技项目档案资料的数字化,档案管理流程的信息化、自动化、规范化,使得科研人员能够便捷地获取科技项目数据信息,又减轻了科技管理人员的工作量,提高了工作效率。

科技档案管理共享的实现,可进一步拓展档案资源共享空间,并且延伸科技档案共享实现科技咨询服务的扩充和链接,提高科研管理部门、科技咨询和科技档案管理部门的服务能力。

在建立健全科技档案管理内部组织结构的同时,需要加强对科研项目档案共享数据的宣传,通过科技档案数据共享有效地提高科技档案数据的使用率,为陕西省在科技创新中各学科的综合发展提供平台,为科研交流与互动创造良好的条件,使科技项目档案信息科技项目数据服务在推动陕西省科技创新中发挥重要作用。

(下转第151页)

械分类结果在查全率、查准率和 F_1 测度上表现的都比较差,原因可能是因为这两类与其他类具有一定的关联性,所以导致分类结果与其他相比较低。

实验结果表明,改进后的使用归一化方法计算中心向量的中心法的测试结果优于总和法和均值法,几乎各项指标都达到了 80%。由此可以推断出改进中心向量的计算方法对中心法的改良起到了一定的作用,得到的分类结果更好,准确率更高。

5 结束语

文中采用中心法对农业信息网页的分类进行了研究,实验结果表明归一化法计算中心向量对于农业信息网页的分类判别有着较高的准确率。下一步将继续调整中心向量的算法以及相似度算法,从而进一步提高模型在农业信息网页分类中的准确率。与此同时探索其他文本分类方法,并设计对比实验,以获得更佳的农业信息网页分类模型。

参考文献:

[1] 金一宁,王华兵,王德峰. 基于 KNN 及相关链接的中文网页分类研究[J]. 哈尔滨商业大学学报:自然科学版,2011,27(2):203-207.

[2] 江小平,李成华,向文,等. 云计算环境下朴素贝叶斯文本分类算法的实现[J]. 计算机应用,2011,31(9):2551-2554.

[3] 李琼,陈利. 一种改进的支持向量机文本分类方法[J]. 计算机技术与发展,2015,25(5):78-82.

[4] 胡金有,张健,游龙勇. 我国农业信息网站现状分析[J].

(上接第 145 页)

参考文献:

[1] 张炯森. 科技项目档案管理中存在的问题及对策[J]. 科技情报开发与经济,2009,19(1):98-99.

[2] 陈文英,陈开魁,徐迟默,等. 热带农业科技档案信息资源共享现状分析及可行性研究[J]. 安徽农业科学,2014,42(25):8846-8848.

[3] 肖琬蓉,张静. 科技档案全文数字化信息系统开发研究[J]. 计算机应用与软件,2013,30(3):145-147.

[4] 许雯倩,李娟. 论信息化背景下高校档案推送服务[J]. 兰台世界,2013(8):18-19.

[5] 王兰成,黄永勤. 大数据时代国防科技档案服务的信息技术研究[J]. 浙江档案,2014(11):6-9.

[6] 廖淑莉. 构建科技档案云平台 支撑科技创新驱动——以粤西高校科技档案云平台关键技术研究为例[J]. 档案时

农机化研究,2005(6):38-40.

[5] 王健,甘国辉. 多维农业信息分类体系[J]. 农业工程学报,2004,20(4):152-156.

[6] Feng Xia,Tang Xianchao. An improved dictionary-based chinese word segmentation approach in Lucene[C]//Proceedings of 2010 international conference on services science, management and engineering. [s.l.]:[s.n.],2010.

[7] 孙殿哲,魏海平,陈岩. Nutch 中庖丁解牛中文分词的实现与评测[J]. 计算机与现代化,2010(6):187-190.

[8] Zhang Qun,Cheng Yu. Research on Chinese word segmentation algorithm based on special identifiers[C]//Proceedings of the 2011 international conference on computing,information and control. [s.l.]:Intelligent Information Technology Application Association,2011.

[9] 王霜霜,张太红,冯向萍,等. 农业网站导航页面识别模型研究[J]. 新疆农业大学学报,2011,34(5):447-453.

[10] 苏金树,张博锋,徐昕. 基于机器学习的文本分类技术研究进展[J]. 软件学报,2006,17(9):1848-1859.

[11] Wang Bingkun,Huang Yongfeng,Yang Wanxia,et al. Short text classification based on strong feature thesaurus[J]. Journal of Zhejiang University-Science C (Computers & Electronics),2012(9):649-659.

[12] Zhang Yuntao,Gong Ling,Wang Yongcheng. An improved TF-IDF approach for text classification[J]. Journal of Zhejiang University Science A (Science in Engineering),2005,6A(1):49-55.

[13] 张保富,施化吉,马素琴. 基于 TFIDF 文本特征加权方法的改进研究[J]. 计算机应用与软件,2011,28(2):17-20.

[14] 宋枫溪,高林. 文本分类器性能评估指标[J]. 计算机工程,2004,30(13):107-109.

空,2016(2):16-18.

[7] 王兰成. 科技档案异构数据整合及其检索的研究[J]. 中国科技资源导刊,2009,41(5):36-41.

[8] 毕春华. “文档一体化”管理的建设及启示[J]. 江汉大学学报:社会科学版,2003,20(1):93-96.

[9] 张艳霞,齐兰英. 数字化档案管理与查询[J]. 黑龙江科技信息,2010(7):161-161.

[10] 张红,王敏. 利用 ACCESS 建立基建档案全文数据库管理系统[J]. 机电兵船档案,2007(3):24-26.

[11] 刘轩. 档案数字化的策略分析及系统构建[J]. 北京档案,2004(12):28-29.

[12] 杨春子,孙小康. 科技档案管理系统的研究与实现[J]. 武汉理工大学学报:信息与管理工程版,2006,28(3):62-65.

[13] 周丽. 电子档案加密探讨[J]. 国网技术学院学报,2015,18(1):85-88.