

海南省高校机构知识库联盟平台研究与实践^{*}

王小会, 张英

(海南大学图书馆, 海口 570228)

摘要: 本文采用案例研究法调研并剖析国内外典型机构知识库, 探讨海南省高校机构知识库联盟平台建设路径及技术路线。确定联盟平台采取集中采集式(集中建设、集中发布)建设模式, 并基于上述模式自行开发软件系统。该机构知识库平台的成功上线运行为我国机构知识库建设提供新范例。

关键词: 机构知识库; 联盟平台; 集中式模式; 软件系统; 自主开发

中图分类号: G250.74

DOI: 10.3772/j.issn.1673-2286.2017.09.010

20世纪90年代以来, 西方发达国家及我国港台地区的诸多大学和研究机构开展机构知识库联盟项目建设, 并已取得显著成果。机构知识库联盟指多所高校、研究机构通过合作的方式, 整合各成员机构的知识资源, 统一提供数字化服务的机构知识库群。机构知识库联盟本质是具有共享性质的机构知识库群, 其构建能够避免部分技术及设施的重复使用, 提高各成员机构知识资源的共享度和利用率, 在一定程度上促使联盟成员合作更有效地实现^[1]。从某特定区域或系统整体推进和规模效益的角度看, 联盟模式更有优势。进入21世纪, 中国科学院率先以联盟形式推进机构知识库建设^[2], 形成国内规模最大且使用效益最好的机构知识库联盟系统。同类项目还有CALIS系统机构知识库联盟平台^[3]、台湾大学机构知识库联盟TAIR^[4]以及中国农业科学院机构知识库^[5]。海南省高校有许多独特的知识成果需要借助机构知识库优势, 以提高高校学术影响力及社会影响力, 实现更大范围的知识成果共享。因此, 海南省教育科研数字图书馆及全省高校图书馆参考已有机构知识库建设经验, 建立海南省高校机构知识库联盟平台。本文拟就海南省高校机构知识库建设实践相关问题进行研究和探讨。

1 海南省高校机构知识库联盟平台建设思路

1.1 建设模式选择

迄今为止, 国内外机构知识库联盟建设有两种模式: 集中采集式和分布采集式。集中采集式指由中心机构统一建库和采集数据, 并统一发布知识成果及提供服务。成员单位二级库(子库)数据均集中在中心机构服务器。分布采集式指联盟成员先分别建立各自机构知识库, 由机构知识库联盟系统统一收割成员机构知识库元数据, 再开展统一检索及文献提供服务。上述两种模式各有利弊: 集中采集式集成度高, 因避免大量重复建设, 可将社会总成本控制在最低程度, 更易于体现规模效益及高性价比, 适合规模较小的机构协作共建; 分布采集式由各成员单位机构知识库独立运行, 需要各机构有较大的初期投入(开源软件的二次开发、硬件设施、前期数据采集)和长期持续性运维成本。在考察国内同行建设实践基础上, 基于对海南省高校体量小、学术成果总量少、项目投入经费少、独立建库运行维护困难的考量, 故选择以“集中采集式”开展机构知

^{*} 本研究得到海南省自然科学基金“海南省高校机构知识库联盟建设研究”(编号: 20167255)和海南省教育厅重点教改项目“海南高校文献信息资源体系建设模式研究”(编号: hngj2017ZD-6)资助。

识库联盟项目建设。

1.2 软件系统选择

国外很多机构知识库平台软件系统均为自主开发产品^[6-8]。OpenDOAR统计数据存储量前十位的机构知识库平台中,有一半以上均选择独自开发软件系统^[9]。对国内外机构知识库联盟及其成员使用软件调研表明开源软件所占比例较大,其中以DSpace居多^[10]。中国科学院机构知识库联盟平台是其所属各研究所基于二次开发的DSpace系统独自建库的前提下,开发建设了统一展示与服务等功能的集成网络服务系统。经过反复比较论证,海南省高校机构知识库选择自主开发软件系统,以满足建设模式及平台功能需求。

2 海南省高校机构知识库联盟平台架构

海南省高校机构知识库联盟平台由联盟总库与二级库(各高校机构知识库)构成机构知识库,采取一体化集中采集式模式建库来实现集中建设、统一发布和对外服务。该机构知识库联盟平台建立了一个统一、集中

的机构知识库中心系统,各高校机构知识库(二级库)数据均在知识库中心系统中。各成员机构将本机构各类知识成果数据上传到知识库中心系统中的本校机构知识库,并由知识库中心系统集中管理、维护,同时通过联盟总库平台提供统一检索与文献索取服务。建库初期,各高校的图书、论文、科研成果、专利等知识成果数据委托合作商集中回溯与更新;后续高校自建资源(教学资源、科研报告、实验报告等)由各机构成员上传到知识库中心系统中的本校机构知识库。该机构知识库联盟平台架构由基础层、整合层、用户层构成(见图1)。

2.1 联盟总库平台功能

联盟总库通过收割或导入各高校机构知识库成果元数据,经过加工整理形成集成数据,实现知识成果的集中展示和利用。联盟总库平台主要功能包括对二级库知识成果进行汇总,实现对知识成果的分类浏览及统计分析;统一收割联盟下各二级库数据,实现联盟总库平台数据定期更新;实现成果统一检索/查询并链接到全文下载;通过最新提交、浏览排行、下载排行等多项数据统计展示机构知识库联盟资源的利用情况。

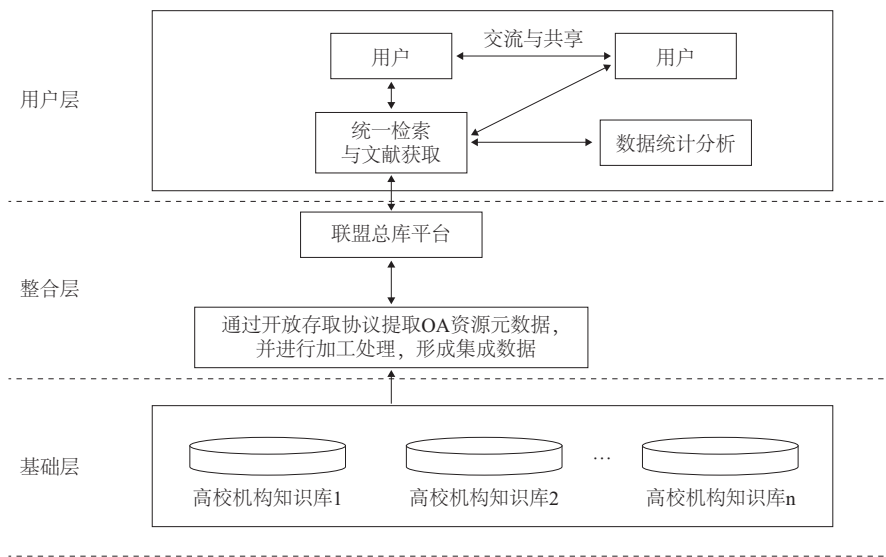


图1 海南省高校机构知识库联盟平台架构

2.2 二级库(子库)系统功能

海南省高校机构知识库联盟二级库功能设定与中国科学院机构知识库网络成员单位机构知识库基本功能类似。各二级库可独立运行,其核心功能是实现数据

收割与定期更新,包括支持多种资源导入模式;存储、展示多种类型知识成果(如图书、论文、专利、标准、科技成果、图片、视频、研究及实验报告等);实现知识成果一站式检索及获取;实现成果分类浏览、统计分析,成果排行,及对本校机构知识库利用趋势的可视化展

示;提供成果类型、重要期刊收录情况、各院系研究方向等信息;对科研趋势、学者发展趋势、研究方向等多种维度的可视化分析;为学者和管理人员提供管理后台,实现基于角色的各类管理和权限控制机制。

3 软件平台开发

自主开发软件系统是机构知识库联盟平台能否实

现预期建设目标的关键要素。该联盟平台软件开发的基本要义如下文所示。

3.1 系统技术设计

根据上述功能描述,海南省高校机构知识库联盟平台的软件系统总体技术框架如图2所示。

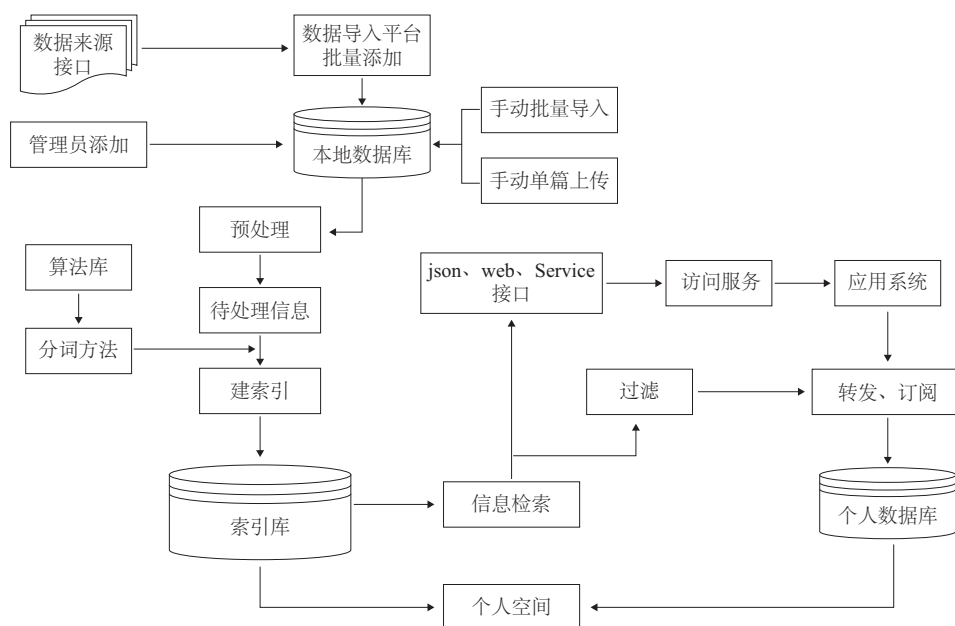


图2 软件系统框架

3.2 系统关键技术应用

本研究基于J2EE平台、Ajax、Lucene以及SpringMVC框架设计完成机构知识库系统开发工作。

(1) 采用多层架构的B/S结构和组件开发技术。基于J2EE分布式计算技术进行系统架构设计和开发,从而使多种应用程序间能交互,实现跨平台Web服务的互操作。

(2) 采用基于JavaScript和Xml技术的异步动态更新网页技术。通过Ajax核心JavaScript对象XmlHttpRequest,使用户可向服务器提出请求并处理响应,而不阻塞用户。

(3) 系统利用Lucene技术对平台自动更新和手动添加的成果进行检索。检索快捷、准确,且结果更清晰。

(4) 系统使用自主开发的中文分词,准确率超过90%。

3.3 二级库(子库)功能模块设计

各高校机构知识库作为二级库的功能设计是决定该机构知识库联盟平台运行及服务效能的最重要环节。海南省高校机构知识库联盟二级库主要功能模块如图3所示。

(1) 系统管理。用户管理,系统主要用户为学者和管理人员;参数设置,系统管理员可以对个人、院系、学术团体、机构的参数进行设置。

(2) 数据管理。数据回溯,系统根据各高校相关信息自动回溯历史数据;自动更新,系统定期根据设置的参数,自动更新数据;数据维护,管理员或学者可以在后台对已有数据进行删除、修改、上传附件等操作,并能通过检索添加、批量上传、单篇上传的方式新增数据。

(3) 检索模块。系统提供简单检索与高级检索,检索结果可通过语种、年代、院系、学者、数据类型、重

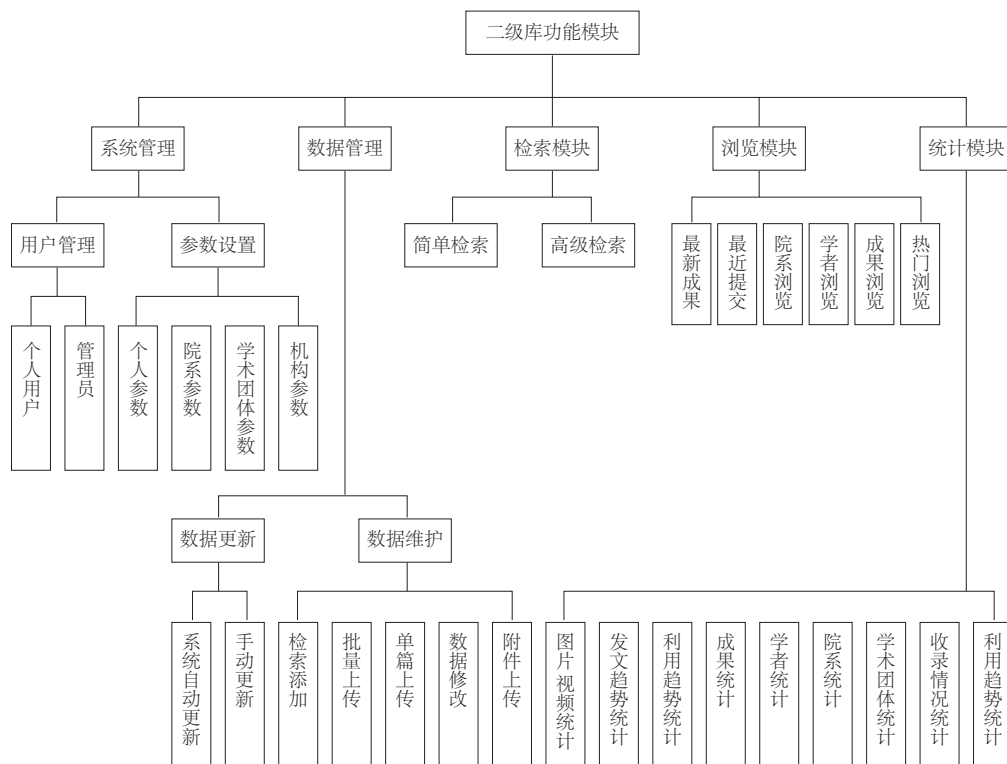


图3 二级库主要功能模块设计

要数据收录情况等进行聚类筛选;同时,也可对检索结果进行题录导出,且导出格式多样。

(4) 浏览模块。可对最新成果、最新提交、院系、学者、成果类型、统计分析、热门文献等进行浏览。

(5) 统计模块。对文献的成果、学者、院系、学术团体、重要数据库收录情况、利用趋势等进行可视化统计分析。

(6) 重点功能的相关说明。统计分析,主要有分类统计和利用趋势两个模块,通过对成果、年代、语种、点击量、访问量等不同维度的可视化分析可直观展示系统运行情况;个人后台,各高校机构知识库提供个人空间以实现个人成果维护,允许通过检索添加、批量上传、单篇上传来添加新数据,并提供个人收藏入口,收藏的数据不会在前台显示;权限控制,系统管理员作为各高校机构知识库的最高权限管理员可以对平台所有数据、学者、院系等信息进行维护。咨询馆员可根据系统管理员设置的不同权限,协助管理后台。用户可浏览、下载机构知识库知识成果,并对个人知识成果进行维护。

4 结论

海南高校机构知识库联盟于2016年年底完成集中

回溯建库工作。该机构知识库联盟包括12所高校,近8万条数据^[11]。经过一段时间试运行,与国内现有机构知识库平台相比,该机构知识库联盟虽有自身优势,但也存在待改进之处。

(1) 鉴于海南省高校自身状况,海南高校机构知识库联盟平台采取集中采集式(集中建设、集中发布)的运行模式。各成员机构将本机构知识成果数据上传到联盟平台知识库中心系统。由该系统负责管理、维护并提供服务,达到共享联盟应提供的技术服务要求,完成平台日常维护与管理。

(2) 与国内业界同类项目相比,海南省高校机构知识库自主开发的软件系统功能实现了预期目标。除具备集中式建库的特点与优势外,该平台还支持文献计量可视化分析和统计数据多维化分析,特别是基于研究方向、科研趋势和关系图谱揭示的院系、学科团队、学者等维度的深入分析,显示其在助力高校学科建设方面的功能优势。

(3) 海南省高校机构知识库联盟软件平台尚需完善。如各高校机构知识库模块未能根据学校特色进行区分性设计,为作者提供个人成果管理空间的功能尚需改善,深层次知识情报服务和战略情报分析服务尚有很大提升空间,科研管理系统深度集成、系统化学科

建设绩效分析与评估、学者与机构知识库的互动功能等板块均有待进一步开发。

(4) 从国内机构知识库发展态势看,已涉足本机构特色文献资源(本机构教学资源、科研项目、技术报告、实验报告、机构历史文档等)的机构知识库甚少。目前中国矿业大学、西安交通大学等高校机构知识库开始涉足机构特色文献资源系统建设;中国科学院系统机构知识库尚未发现同类案例。海南省高校机构知识库联盟现已启动该项工作,并取得初步成效。

基于集中采集式海南省高校机构知识库联盟软件平台在国内尚属首例。其成功上线运行为我国机构知识库联盟建设提供了新范例,实现了一站式区域机构知识库资源集成与共享的目标。

参考文献

- [1] 邵波,陈慧香,刘啸.基于联盟的高校机构知识库的构建研究[J].图书馆学,2016(24):33-38.
- [2] 祝忠明,马建霞,张智雄,等.中国科学院联合机构仓储系统的开发与建设[J].图书情报工作,2008(9):90-93,144.
- [3] 聂华,韦成府,崔海媛.CALIS机构知识库:建设与推广、反思与展望[J].中国图书馆学报,2013,39(2):46-52.
- [4] 闫永君,刘兰,陈艳.台湾TAIR与中科院机构知识库联盟CAS-IR的比较研究及启示[J].图书馆学刊,2013,35(12):133-136.
- [5] 赵瑞雪,杜若鹏.中国农业科学院机构知识库的实践探索[J].现代图书情报技术,2015(2):72-77.
- [6] EMERITUS P B.Open access developments in France:the HAL Open Archives System[J].Learned Publishing,2007(4):267-282.
- [7] DIJK E,BAARS C,HOGENAAR A,et al.Accessing grey literature in an integrated environment of scientific research information[C].[S.l.]:[s.n.],2007.
- [8] PETERS D,LOSSAU N.DRIVER:building a sustainable infrastructure for global repositories[J].Electronic Library,2011(2):249-260.
- [9] 陈美华,刘文云,毕煜,等.美国机构知识库联盟的建设研究及对我国的启示[J].图书馆,2015(11):59-64.
- [10] 付希金,田丽君,张艳婷,等.国内外区域IR联盟建设比较及我国区域IR联盟构建路径研究[J].情报科学,2015,33(3):35-40.
- [11] 海南省高校机构知识库[BD/OL].[2017-08-01].<http://hainanssp.dayainfo.com/>.

作者简介

王小会,女,1964年生,博士,研究馆员,研究方向:图书馆联盟及数字资源建设,E-mail:820997012@qq.com。
张英,女,1968年生,硕士,副研究馆员,研究方向:数字图书馆,E-mail:525421247@qq.com。

The Research and Practice of Institutional Repository Alliance Constructions in Hainan Universities

WANG XiaoHui, ZHANG Ying
(Hainan University Library, Haikou 570228, China)

Abstract: By using case study method, investigates and analyzes typical institutional repositories at home and abroad, the research addresses platform building and technology roadmap of Hainan Universities Institutional Repository Union. The union platform is based on the centralized mode (centralized construction, unified release). Based on the above model, the software system has been developed. Practice verifies that the launching operation of Hainan Universities Institutional Repository Union has provided a total new example for China institutional repository union building.

Keywords: Institutional Repository; Platform Union; Centralized Mode; Software System; Self-Development

(收稿日期: 2017-08-14)