

面向服务的高校机构知识库研究与实践^{*}

师俏梅, 李晶, 谭英
(西北工业大学图书馆, 西安 710072)

摘要: 机构知识库作为科研成果的存储共享平台已得到各高校普遍重视, 本文通过分析机构知识库在建设中的问题, 以问卷调查法和访谈法, 对西北工业大学科研人员和管理人员的需求进行广泛调研, 结合西北工业大学机构知识库服务的具体实践, 提出面向服务的机构知识库应具备数据典藏、统计分析、数据共享、查收查引、信息交流、论文开放级别自定义等服务性功能。该机构知识库的建立方便教职工、管理者及学生对科研成果的高效管理及统计, 是学校整体信息化的重要数据基础。

关键词: 面向服务的机构知识库; 用户需求; 数据共享; 平台功能

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2017.3.010

随着网络技术的迅猛发展, 数字文献资源受到广泛关注, 科研人员正是这些数字资源的生产者和创造者, 图书馆如何保证机构知识库 (Institutional Repository) 既具备资源典藏、成果展示等功能, 又能促进用户善用机构知识库的科研成果数据, 以及机构成员乐于维护科研成果数据的功能, 是机构知识库建设的重要问题。

1 机构知识库的建设现状及存在问题

1.1 机构知识库的建设现状

机构知识库又称机构仓储、机构典藏库等, 是学术机构为获取、保存机构内智力成果而建立的数字资源仓库, 是数字时代学术研究成果的基础平台, 也是高校为师生提供用于管理和发布机构数字资源的平台。

据OpenDOAR网站统计, 截至2017年3月1日, 全球共有机构知识库3 332个, 其中国39个、日本211个、美国494个。目前, 国内高校机构知识库 (如清华大学图书馆、厦门大学图书馆等) 陆续完成建设并取得良好成效, 并且中国高校机构知识库联盟已于2016年10

月成立, 同时受到中国高等教育文献保障系统的大力支持。

1.2 机构知识库建设存在的问题

国内对机构知识库的研究起步较晚, 研究内容主要侧重于机构知识库的概念、功能、意义、影响及国内外进展等。其中, 陈韬认为机构知识库有数字资源的呈缴、收集、描述、基于工作流的资源审核与发布等8项功能^[1]; 袁顺波等从机构知识库对传统出版模式、学术机构以及图书馆等方面的影响进行研究^[2]。近几年, 随着机构知识库建设的不断普及与深入, 对机构知识库的研究也由理论逐渐转至具体实施方案。如郭少友对机构知识库在建设过程中遇到的服务模式选择、法律、内容建设、资金等问题进行分析, 并提出建设性意见^[3]; 何琳等在分析我国发展机构知识库优势和主要障碍的基础上提出发展政策^[4]。

在系统平台建设方面, 已有研究多侧重于搭建机构知识库所使用的软件平台、互操作技术和协议等, 注重机构知识库“如何建”, 而不是“如何用”。在系统功能开

^{*} 本研究得到西北工业大学高等教育基金项目“高校机构知识库建设与服务实践研究” (编号: 2013GJZ04) 资助。

发与服务拓展方面的研究较少, 尚未形成可推广的模式。本文通过网络调研, 认为国内高校机构知识库普遍存在以下问题。

(1) 机构知识库建设者重建设轻应用。部分机构知识库以项目为导向, 项目结题, 数据库建成后常年不更新、不维护, 时效性差^[5]。

(2) 机构知识库数据开放程度较低。成果数据无全文, 或全文不对外开放; 高校教师对开放获取的认知度和认可度普遍不高^[5-6]。

(3) 机构知识库服务功能单一。对用户而言, 只能检索、浏览和成果维护, 无实用的附加功能^[7]。

(4) 机构知识库与校内其他系统间整合度有待提高^[7]。目前机构知识库与校内各信息平台(如科研管理系统、研究生管理系统、人事管理系统、统一认证系统等)的数据共享程度不足, 有待进一步提升。

(5) 用户访问量不高。机构知识库建设完成后, 由多方面原因导致使用率不高。

因此, 提出建设以服务为目标的机构知识库, 即机构知识库不仅具备收集机构科研成果的功能, 还能够保证用户充分利用机构知识库, 使其成为机构成员的必备工具。

2 机构知识库的用户需求分析

2.1 基于调查法的用户需求分析

机构知识库以“为用户服务”为宗旨, 根据上述问题, 为更好地了解用户需求, 设计更加符合用户需求的机构知识库, 研究以网络问卷形式向西北工业大学在校科研人员及管理人员开展机构知识库使用情况调查。调查问卷共设计13个问题。主要从参与人员情况、机构知识库普及性、用户建设意愿、成果开放范围、机构知识库功能、成果存缴措施等方面展开。

通过在线方式对在校的科研人员和管理人员进行调研, 内容主要涉及对机构知识库的认识、科研成果呈缴以及科研成果的开放共享等方面。本调研共得到有效问卷217份, 并将调研结果归纳为以下5个方面。

(1) 机构知识库普及性。从调研结果看, 69.0%的用户对机构知识库不了解, 78.8%的用户无法从机构知识库获取机构成果, 在一定程度上说明机构知识库的普及性有待进一步加强。

(2) 机构知识库建设意愿。91.2%的用户认为机

构知识库建设非常有必要, 55.5%的用户表示愿意支持并提交科研成果至机构知识库, 以期满足“长期保存个人学术成果”“管理个人学术成果”“个人学术成果统计平台”“增加科研成果显示度、提高成果使用率”等需求; 而44.5%的用户因“对机构知识库不了解、缺乏相关认知”“担心成果被剽窃或用于商业用途”“提交内容的手续过程复杂、操作不方便、增加工作量”等原因, 不愿意呈缴个人成果。

(3) 科研成果开放程度。14.5%的用户赞成“完全开放获取”, 59.6%的用户赞成在“有限范围内开放”。可以看出用户对于科研成果的完全开放获取还存在疑虑, 因此在机构知识库的建设中, 对于科研成果的开放共享, 可视情况设置开放范围和时滞期。

(4) 机构知识库功能。用户对“系统自动生成我的成果库”“自动生成收录引用证明”“设定成果访问权限”“SCI、EI收录邮件推送”“成果批量导入”“成果统计分析”等功能关注度较高。

(5) 成果存缴措施。以“制定政策规定成果必须存入机构知识库”和“利用网站、简报等多媒介宣传”等方式被多个用户推荐。

从调研结果看, 虽然用户对机构知识库的了解有限, 但大多数用户对机构知识库的建设还是非常认可的, 且愿意提交成果至机构知识库; 同时也可以看出, 尽管对机构知识库开放获取认可度不高, 却对机构知识库的服务性功能(自动生成收录引用证明, SCI、EI收录邮件推送等)比较认可和期待。

2.2 基于访谈法的用户需求分析

为更深入了解用户对机构知识库的需求, 多角度分析各类型用户行为, 从源头减少用户自存储负担, 深度挖掘机构知识库对教师和管理工作者的增值服务, 特补充访谈法以弥补网络调查法存在的样本数量不足、覆盖不全的问题。访谈对象包括西北工业大学各学院、科研管理部门、研究生院和图书馆, 具体结果如下。

(1) 教职工需求分析。①成果上报方面, 学院教师每年需填写与成果相关的表格, 表格核心内容基本一致, 仅存在格式差异, 但花费教职工大量时间。②数据共享方面, 科研人员需不定时进行数据分享, 但目前机构知识库在数据批量导入及格式转换方面存在欠缺, 不能完全满足需求。③个人权益方面, 开放共享虽能使科研成果得到更多关注, 但限于各种原因, 教职工希

望能自行设定开放等级(如根据个人需求将论文设置为完全开放、学院开放、团队开放或独有等)。

(2) 科研管理部门需求分析。科研管理部门的需求主要在于成果统计及奖励,目前针对学校和各学院的成果,尤其是高水平成果(如SCI、SSCI、ESI高被引论文等)的奖励,一般由校内外相关情报机构提供,通常存在统计不全或错误等问题。机构知识库应具有灵活的统计功能,不仅能快速、准确地实现科研成果的定位,也可以协助个人、学院、科研管理部门统计科研成果,为科研管理系统提供数据接口。

(3) 研究生院需求分析。研究生院的需求与科研管理部门相似,主要是统计方面的需求,如对学生发表论文的作者署名及期刊等级等情况的统计。机构知识库应为研究生管理系统提供数据接口,方便其调用数据,以便国家相关部门开展考核、研究生论文统计和奖励等工作。

(4) 图书馆需求分析。图书馆的需求主要集中在简化工作流程方面,目前很多图书馆均开展三大索引(SCI、CPCI、EI)收录工作,部分图书馆还提供将论文收录信息推送至用户电子邮箱的服务;同时随着国家项目申报、奖励政策等对论文要求的提升,机构知识库作为数据存储共享平台理应提供该功能,为用户提供多样化的服务,如学科服务、学科分析等。

综上,面向服务的机构知识库建设具有重要价值。对高校而言,既可以保存机构成员的研究成果,也可以展示学术成就,提高机构声誉;对研究者而言,既可以通过机构知识库管理学术成果,也可以扩大学术影响力、提高学术地位,便于交流共享学术成果;对科研管理人员而言,可以借以统计机构各类成果,提高科研管理效率。

3 面向服务的机构知识库功能设计

为建设“用户乐于使用,用户愿意维护”的机构知识库,将机构知识库的建设宗旨设定为“以用户为导向,以服务为核心”,即设计以服务为导向的机构知识库,站在用户角度完善机构知识库功能。

3.1 面向服务的机构知识库建设目标

为更好满足用户需求,需在了解用户需求的基础上,有针对性地建设机构知识库。对于教师而言,机构

知识库应该是“个人的成果存储空间、团队的信息共享平台”。教师在此建立个人成果档案,展现学术风采;开展学术交流,寻找合作伙伴;成立团队研究社区,共享研究信息。对于学生而言,机构知识库应该是“承前启后的学习空间,学术交流的信息平台”。学生可享受学者的创造与学术思想,保证研究的完整性和持续性。对于管理人员和馆员而言,机构知识库应该是“学术成果的管理平台,人才评价的支撑体系”。保证高校学术成果的长期保存;展示并传播本校智力资源;增加高校的可见度与声誉;培养高校对智力成果的管理意识;提供知识服务,进行学术评估。

3.2 面向服务的机构知识库功能模块设计

根据面向服务的机构知识库(Service Oriented Institutional Repository, SOIR)建设宗旨和目标,结合上述用户需求分析,提出机构知识库应具有8项功能模块。

(1) 基本数据功能。SOIR平台定期从主要文献数据库中收集基础数据,并自动匹配成果拥有者及其二级单位(即学院),对未匹配成功的数据,通过邮件推送至潜在成果拥有者账户,提醒其认领。最大程度地提高自动匹配成功率,尽可能减轻师生提交成果的负担,减少成果认领数量。

(2) 数据扩展功能。在基础数据的基础上,SOIR平台为机构成果提供更多附加信息。主要包括论文收录信息(如被SCI、EI、CSCD、CSSCI等收录,则给出收录数据库标记及论文被引次数)、期刊计量信息(如载文期刊影响因子)、ESI高影响力论文信息(定期查看文献是否成为ESI高被引论文或热点论文,并给予标记)。上述各种信息均由系统自动匹配,无需人工干预。这些信息在SOIR平台完全公开,用户可以通过浏览、筛选将所需数据单独列出。此外,还包括自有文献管理,用户不但可以管理自有成果,也可以在机构知识库的个人管理平台管理非个人成果,这些信息完全自有,不会在机构知识库平台上显示,却可以将个人平台的文献信息分享给他人或团队。成果全文的开放权限由用户自定义或由系统管理员统一设置。

(3) 检索功能。提供文献基本检索和高级检索,可检索字段包括题目、作者、来源、摘要。SOIR平台按学院、文献类型、作者、来源、年代等实现科研成果的浏览,此外,还提供检索结果的聚类分析功能。

(4) 信息服务功能。包括收录和引用通知、优秀成果展示、查收查引报告等。收录和引用通知指SOIR平台通过整理机构论文被EI、SCI等收录及ESI高影响力论文的情况,及时为用户推送相关信息;查收查引报告是用户利用SOIR平台的数据自动生成检索结果,由图书馆馆员确认盖章后生效。

(5) 分层管理功能。针对不同用户层设计管理功能。系统管理员可以管理机构知识库所有模块;学院、团队可以管理与其业务相关的模块,如数据批量导入和导出,数据增、删、改,学院、团队数据认领等;个人只能管理自身数据,如数据认领、修改、导出、建立团队等。数据导出功能,用户可以自定义字段,也可选择多个文件格式(如Excel、Word等)。

(6) 分层统计功能。统计与管理的角色划分基本相同,即系统管理员、学院管理员、团队管理员、个人。因角色不同,所统计的数据范围不同,如系统管理员可统计分析全学校各学院成果情况等,而学院管理员仅可以统计分析所在学院的成果情况等。该功能提高了机构成果的学科分布、部门分布、年度分布等科研统计信息效率。

(7) 科研团队协作功能。学校师生可以建立自身的科研团队,团队内成员拥有独立的共享空间和平台,相互分享文献、文稿、实验记录、设计文档、项目报告等知识资源,把原本分散于个人手中的科研文档汇总,形成团队的知识活动档案,促进知识和经验的传承与利用,提高科研团队的工作效率,同时解决了由于学生毕业导致的科研断档和成果无法传承等问题。

(8) 系统扩展接口功能。为提高SOIR平台中成果的高效利用,资源共享成为平台最重要的发展动力,在平台中设置扩展接口,方便系统平台间的知识共享;

同时,也为科研管理系统、研究生管理系统、人事管理系统等提供信息来源,此外,系统平台的身份认证模块将与校园一卡通系统对接,采用学校统一身份认证系统。

4 面向服务的机构知识库建设平台分析比较

机构知识库平台软件可以选择开源软件,也可以选择自定义开发,由于SOIR平台的建设目的是利用成果数据提供更多服务,开发难度较大,经项目组讨论和专家评审,决定采用自定义开发的模式建设机构知识库。

为保证机构知识库平台顺利完成,2014年上半年对已有机构知识库中较成功的公司进行广泛调研。经过意见征询,初步选定3家公司(甲公司、乙公司、丙公司),探讨其对项目建设方案的支持能力和公司的开发能力。其中,甲公司与乙公司在文献管理方面具有丰富的开发和实践经验,甲公司文献自动采集为特色,而乙公司在数据展示方面较有优势,丙公司在文档预处理方面具有专利技术。

本文分别从资源导入、检索、浏览、推送服务、文献操作、数据备份、机构管理、用户管理、认领审核、用户权限、开放级别、数据统计功能进行分析比较。从表1可见,3家公司的机构知识库基本功能具有良好的支持能力,而在服务功能方面均较弱。经与3家公司沟通后,其均表示可在现有基础上再次开发,提升现有机构知识库的服务功能,满足用户需求。在后期建设过程中,通过及时有效地沟通,系统开发顺利进行,建设方案顺利实施。

表 1 3家公司系统功能支持能力对比

功 能	甲公司	乙公司	丙公司
资源导入	个人提交、自动获取	个人提交、工具批量导入	在线上传、文档导入、文档预处理
检索	便捷检索、高级检索	简单检索、高级检索	初级搜索、高级搜索、二次搜索、智能搜索、相似性检索、指纹检索
浏览	根据用户需求制定发布模版,可更换电子文档管理平台的发布和检索样式	资源类型、资源年份、作者、机构,支持标题、时间排序	按学院、发布年度、收录情况、文献种类
推送服务	重要成果、最新收录、最新成果、通知邮件或RSS	-	用户可根据需求订阅电子文档、文档类型、检索条件

续表

功 能	甲公司	乙公司	丙公司
文献操作	查看摘要、下载全文、Like、评价、导出为文献管理软件格式、收藏为个人文献、编辑	下载全文/借阅、导出、收藏、查看元数据、认领、编辑,可以对多篇文献下载全文,导出,展示所有	下载、在线打印、收藏、分享、在线浏览
数据备份	-	√	-
机构管理	√	√	√
用户管理	√	√	√
认领审核	-	√	-
用户权限	√	√	√
开放级别	√	√	√
数据统计	√	√	-

注: “√”表示具备该项功能。

5 面向服务的机构知识库建设

SOIR平台经过一期、二期的建设及1年的内部测试和公测,于2015年10月21日正式上线提供服务,目前机构知识库三期建设正在紧张进行。在项目组、开发公司及共建学院的努力下,SOIR平台的功能已基本实现。

(1) 基本数据功能: 所有历史成果收集完成,并完成认领工作,部分教师也将尚未被机构知识库收录的成果自行导入。

(2) 数据扩展功能: 已有数据均完成数据库收录和期刊匹配工作,并定期更新成果的引用信息。随着成果数据的完善,图书馆逐步脱离人工收集、整理三大索引数据的工作,目前EI和CPCI-S已实现与机构知识库的融合。

(3) 检索功能: 基本检索功能已实现,但检索结果分析功能还有待完善,在机构知识库三期建设中将不断完善改进。

(4) 分层管理功能: 个人、学院、团队的管理已实现,个人可以增加、修改、批量导入、导出成果。为各学科科研秘书开通后台成果统计的权限。

(5) 分层统计功能: 实现个人、团队、学院、学校的四级数据统计模式。统计参数分别为年度、学院、收录情况、成果类型。

(6) 科研团队功能: 实现自由组建科研团队、分享文献,并组建70支科研团队。

(7) 扩展接口: 与校园一卡通系统顺利对接,用户

可以使用一卡通账号登录机构知识库,其他信息化的对接接口已预留,方便数据互通。

(8) 信息服务功能: 此项功能主要为让教师及时获取最新的论文收录信息和RSS信息,以及论文索引证明。目前,此功能将在三期建设中逐步完成和完善。

随着SOIR平台建设深入,一些新的功能(如学科贡献度分析、热点成果分析等)逐渐凸显,对图书馆、管理部门具有一定的参考价值。

6 结论与反思

随着SOIR平台的推出和图书馆的文献保障服务,机构知识库的重要性逐渐被越来越多的教职工和学生认可,而高校信息化进程加速了SOIR平台的发展。目前,SOIR平台已经向研究生管理系统、科研管理系统、教务管理系统等提供开放接口,可直接从SOIR平台中提取数据,在方便管理部门统计成果数据的同时,也减轻了教职工填表的压力。

SOIR平台是一个复杂的系统,在建设过程中,工作的方式、方法关系到系统建设的质量,在今后的建设中,以下4点需要重点关注。(1) 实施用户需求调查和行为分析,推动机构知识库建设。机构知识库是“以用户为导向,以服务为核心”为宗旨,只有充分了解用户需求,才能更好地提供具有良好发展前景的服务。(2) 良好地沟通不但可以保证系统开发与设计理念保持一致,也可以随时掌握开发进度,保证按时完成设计任务

(如设置固定的沟通周期、定期召开讨论会等)。(3) 设置系统平台扩展接口,方便高校各管理部门和中国高等教育文献保障系统等机构获取数据。(4) 机构知识库的宣传推广不可忽视,机构知识库建设是用户共建共享的工作,学科馆员可以通过发送邮件、发布宣传信息、召开专题宣讲会、培训学院科研秘书等方式展开。

参考文献

[1] 陈韬.北邮机构知识库原型系统设计研究[D].北京:北京邮电大学,2007.

[2] 袁顺波,董文鸾.机构库的起源、影响及图书馆的应对策略[J].情报资料工作,2006(2):44-46.

[3] 郭少友.机构库建设的若干问题研究[J].中国图书馆学报,2006(1):77-80.

[4] 何琳,李劲.国内外机构库研究综述[J].图书馆建设,2008(1):26-29.

[5] 李曦.我国高校图书馆特色数据库建设研究——以CALIS和CASHL中心馆为例[J].四川图书馆学报,2013(4):61-66.

[6] 庞红,阮小芳.机构库建设过程中教师抵制心理研究[J].上海高校图书馆情报工作研究,2013(3):21-24.

[7] 郎庆华.机构知识库服务:现状、问题及发展策略[J].图书馆论坛,2011(2):101-103.

作者简介

师俏梅,女,1970年生,副研究馆员,研究方向:数字图书馆、情报分析,E-mail:shiny@nwpu.edu.cn.

李晶,女,1980年生,馆员,研究方向:学科服务、网络技术、数字图书馆。

谭英,女,1976年生,馆员,研究方向:学科服务、情报分析,科技查新。

Research and Practice of Service-Oriented Institutional Repository for Colleges and Universities

SHI QiaoMei, LI Jing, TAN Ying

(Library of Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Colleges and universities attach importance to institutional repository which is a shared storage platform. With questionnaire method and interview method investigated and researched the needs of researchers and administrators at Northwestern Polytechnical University. This paper analyzed the problems of institutional repository construction. Combined with the practice of Northwestern Polytechnical University institutional repository services, the paper presented Service Oriented Institutional Repository. The Service Oriented Institutional Repository should have many service functions, such as data collection, statistical analysis, data sharing, database retrieval and cited reference retrieval, information exchange, user-defined open level ect. When the Service Oriented Institutional Repository established, it made teachers, students and administrators manage and statistic more efficient. It had been the important basement of informationization of colleges.

Keywords: Service-Oriented Institutional Repository; User Requirements; Data Sharing; Institutional Repository Function

(收稿日期: 2017-03-03)

■ 书 讯 ■

《汉语主题词表》(工程技术卷)

《汉语主题词表》自1980年问世以后,经1991年进行自然科学版修订,在我国图书情报界发挥了应有的作用,曾经获得了国家科学技术进步二等奖。为了适应网络环境下知识组织与数据处理的需要,2009年由中国科学技术信息研究所主持,并联合全国图书情报界相关机构,完成《汉语主题词表(工程技术卷)》的重新编制工作。

全书共收录优选词19.6万条,非优选词16.4万条,等同率0.84。在体系结构、词汇术语、词间关系等方面进行改进创新。为了方便工程技术领域不同专业用户使用,《汉语主题词表》(工程技术卷)按专业分13个分册出版,同时建立《汉语主题词表》服务系统,提供在线概念检索和辅助标引服务,通过可视化技术展示各类概念关系,是图书馆、档案馆、出版社、期刊杂志社、文献信息中心等专业工作者及科研、教育及工程技术领域人员必备的参考书。

《汉语主题词表(工程技术卷)》已于2014年由科学技术文献出版社出版,全书2300余万字,总定价3880元,可分册购买。