

2015; 1321–1331.

15 Qingqing Cai, Alexander Yates. Large-scale Semantic Par-

sing via Schema Matching and Lexicon Extension [C].

Proceedings of ACL, 2013; 423–433.

医院机构知识库构建

孙 蕾

谢志耘

(北京大学口腔医院图书馆 北京 100081)

(北京大学医学图书馆 北京 100191)

李晓霞

(北京大学口腔医院图书馆 北京 100081)

[摘要] 结合国内外机构知识库建设的现状,分析建设医院机构知识库的必要性和可行性,结合北京大学口腔医院的实际情况,提出建设北京大学口腔医院机构知识库的构想。系统分析医院机构知识库的框架与结构、内容维护与运营、特点以及具体建设实践及经验。

[关键词] 医院;机构知识库;构建

[中图分类号] R-056 [文献标识码] A [DOI] 10.3969/j.issn.1673-6036.2016.04.003

Construction of an Institutional Repository for a Hospital SUN Lei, Peking University School of Stomatology Library, Beijing 100081, China; XIE Zhi-yun, Peking University Health Science Library, Beijing 100191, China; LI Xiao-xia, Peking University School of Stomatology Library, Beijing 100081, China

[Abstract] Based on current situations related to the construction of domestic and foreign institutional repositories, the paper analyzes the necessity and feasibility of constructing an institutional repository for a hospital. According to the realities of the Peking University School of Stomatology, it proposes the idea of constructing an institutional repository for it. It makes a systematic analysis of the framework and structure of the institutional repository, content maintenance and operation, features, and specific construction practice and experience.

[Keywords] Hospital; Institutional repository; Construction

1 引言

近年来,机构知识库(Institutional Repositories, IR)和开放获取建设在国外蓬勃发展,国内对机构知识库的建设也做出了积极响应。但到目前为止,

国内机构知识库建设仅限于综合性大学和顶尖的科研院所,而且建设进度仍然处于初级阶段,成熟和完善仍需时日。有关医院机构知识库的设计和建设仍然未见文献报道。笔者结合北京大学口腔医院的实际情况,对医院机构知识库的建设提出设计思路。

2 机构知识库及其建设现状

[修回日期] 2015-12-07

[作者简介] 孙蕾,硕士,馆员,发表论文1篇。

2.1 概念与内涵

机构知识库是指利用网络及相关技术收集、整理和保存一个机构或多个机构所生产的知识产品,形成一个数字形式的学术数据库,允许机构内外的成员通过互联网免费或授权获取使用。机构知识库也称机构资料库或机构典藏。目前,对机构知识库的定义并没有统一的定论。最经典的定义是 Raym Crow 和 Clifford A. Lynch 分别从资源和服务角度的定义。从资源角度,机构知识库是学术机构为捕获并保存本机构的智力成果而建立的数字资源仓库(Raym Crow, 2002)。从服务角度,机构知识库是数字时代学术研究的基础设施,一个大学的机构知识库是学校为师生员工所提供的一系列服务,服务的目的是管理和发布由本机构及其成员所创造的数字资源(Clifford A. Lynch, 2003)。IR 的资源种类可以多样,如期刊论文、会议论文、学位论文、技术报告、实验报告、胶片、讲义等,资源形式丰富,可以是文档类、图像类以及超链接等。机构知识库是研究机构实施知识管理的工具,是机构有效管理其知识资产的工具,也是机构知识能力建设的重要机制。

2.2 国外

国外对机构知识库的研究和实践始于 21 世纪初,近年来一直处于线性增长趋势。截至 2011 年 8 月,世界范围内在 OpenDOAR 网站上注册的知识库数量已达 2 019 个,其中机构知识库 1 665 个(占 82.5%)。按国别计,美国排名第 1(占 19.8%),其次为英国、德国和日本,我国大陆及港澳台地区在该网站上注册的机构知识库数量有 42 个,其中台湾有 34 个^[1]。因此,国外机构知识库的建设已初具规模,积累了丰富的建设经验,产生了良好的社会效益。

2.3 国内

在国外大环境的影响下,我国的机构知识库建设也有了一定程度的发展,特别是港台地区的 IR 建设。但相对于国外来说,仍处于起步阶段,从 OpenDOAR 的数据看,中国开放获取的 IR 所占的比

例还非常小。在国内已有很多高校和科研单位建设了本校的学位论文库和教师文库,以收集保存本机构知识产出为目的,但由于元数据标准和数据开放性问题,未能在 OpenDOAR、OAster、ROAR 等开放仓储中注册^[2]。从知识库内容质量方面来说,我国的机构知识库建设尚存在着规模小、文献类型单一、资源共享能力差、功能不够完善和运行不稳定等问题。经查阅相关文献数据库,国内外还未见有医院机构知识库建设的案例与经验的文献报道。

3 医院机构知识库建设的必要性与可行性

3.1 必要性和意义

3.1.1 必要性 临床医学的迅猛发展和患者对于高水平医疗服务日益增长的要求使得医院必须对自己的知识产出进行积累和有效管理。在长期的医疗实践过程中,医院利用先进的医疗理论和技术积累了丰富的各种疑难杂症的诊疗方案和方法,创造了一系列的基础研究成果和疾病诊疗原则及标准。对这些知识产品进行有效管理、交流和互动,可以不断缩小各层次医生诊疗水平的差距,共同提高医院的诊疗水平,满足患者的医疗服务需求,降低医疗事故的发生和差错率,缓解医患矛盾。

3.1.2 意义 医院建设机构知识库的意义在于:(1) 长期保存和管理各种类型的科研成果、医疗和教学资源以及个人数据。如期刊论文、会议论文、专著教材、实验中间和最终结果、病例报告等;临床医生也可上传自己的个人数据并可长期保存;长期保存医生的学术数据,当医生评职称、申请项目时可以直接从机构知识库获取其发表论文和著作的列表,省去多次提交和审核的繁琐;按各种统计标准来统计医院内的科研成果,所有数据保证真实和可靠,有利于医院进行学科竞争力评价、职称评定中的论文审核、SCI 论文统计等科研评价工作。(2) 展示医院总的研究成果以及医教研人员的研究成果,提升医院及医教研人员的学术影响和地位。统计医院 SCI 论文数量、影响因子值、论文被引次数等,统计医院医教研人员数量、科研项目分类统计等;展示临床医生个人主页,包括个人的基本情

况、研究方向、研究项目、发表论文情况（包括论文被引情况）、教学科目等。（3）提高医院内科研成果的共享性，打破信息孤岛，作为医院内信息交流的数据中枢和数据保障；实现项目组内共享、各临床科室之间、各医生之间、各学生小组内的共享；对共享资源进行评价，如打分、点赞、推荐等，依据用户评价结果对共享学术资源进行排名等。（4）作为医院学科评价和建设的参考依据。机构知识库中长期保存了本院所有正式发表的论文和著作以及获取的研究项目等学术数据，为本院的学科竞争力评价提供有力的数据基础。机构知识库中的数据在提供给本院内部使用外，也向世界展示，有利于提高医院和研究人员的学术影响力和地位，找到国际上研究的合作者^[3]。

3.2 可行性

医院是一个知识密集型机构，尤其是大型高等级医院更是如此。例如笔者所在的北京大学口腔医院，实行口腔医学院、口腔医院、口腔医学研究所三位一体的管理体制。拥有众多国内外著名的口腔医学专家，是中国高层次口腔医学专业人才的培养基地和对外交流的重要窗口。在长期的学科建设和临床实践的基础上，已逐步形成了独具特色的口腔医学临床学科群。高等级医院丰富的知识产出为医院机构知识库平台内容的搭建提供了有力的保障。医院图书馆是医院机构知识库建设的主要力量，其信息资源的丰富程度和技术力量的高低决定了机构知识库建设的质量。笔者所在的北京大学口腔医院图书馆口腔医学信息资源丰富，无论是纸质还是电子资源在同类图书馆中均为前列，依托北京大学图书馆和医学部图书馆馆藏资源，大大扩展了文献信息资源的范畴和数量。另外，信息中心和图书馆的 IT 技术力量雄厚。这些都为机构知识库的建设提供了保障。

4 医院机构知识库构建

4.1 建设目标

机构知识库是对机构内部的知识产品进行有效

管理的平台。从现有的机构知识库来看，主要包括两大部分：知识产品的收集、加工、管理和保存；机构内部人员的交流互动模块。在知识库部分，首先对各种类型的文献进行系统全面的收集、整理和规范，这需要一种工作机制和流程，促使机构所有人员共同参与和维护。在交流互动部分，以知识内容为基础，展开不同层次人员的互动交流，如不同科室人员的科研合作和创新，师生之间的教学互动等。另外，科研管理人员还可以借助机构知识库充分了解本机构的科研产出和创新能力，从而制定规划和方案，促进创新^[4]。

4.2 建设原则

（1）标准化。选取成熟、通用的标准规范与协议进行系统建设，包括国际标准、国家标准、行业标准等。（2）成熟性。应用软件采用成熟的技术开发，充分利用已有的技术成果，通过基础组件与应用平台相结合完成产品的主体框架开发。（3）统一性。平台的各功能模块在统一的系统框架下运行，遵照统一的标准和规范，形成功能模块与扩展子系统之间的信息共享和交互机制。（4）组件化。功能组件化、模块化，便于系统的灵活配置和部署。（5）开放性。在不同层面实现对现有和新的资源与服务的开放性描述和有效集成，充分考虑与其他数据整合服务系统的集成与数据交换。（6）安全性。具有有效可信的安全保密机制，防止未经授权的信息访问或非法侵入，在信息访问和交换中确保数据的安全保密性。（7）可维护性。对系统管理员提供所有数据的查询、增删改的应用程序接口和网络服务接口，开放数据库表结构细节，后台数据维护、备份和管理长期性、便捷性等。（8）易用性。从用户角度来设计功能、界面和交互操作，响应速度快、用户体验好^[5-6]。

4.3 软硬件平台环境

医院机构知识库的软件平台是以机构知识库开源软件系统 DSpace 为基础，根据医院机构知识库的设计进行定制开发和优化。医院机构知识库的硬件平台由 3 台服务器（4 核以上，16G 内存，1T 硬

盘) 构成, 分别部署成果自动抓取、Web 服务、个人和团队服务。同时用作数据库的热备份。存储空间预留 10T, 以后视发展情况扩容。服务器运行 Window Server 2008 操作系统和 SQL Server 2008 数据库管理系统。

4.4 数据描述规范

医院机构知识库需要囊括各种类型的科研成果、科研资源和科学数据, 针对不同类型的数据需要设定各自的元数据规范, 允许医生提交新数据类型的元数据模板, 由馆员审核后即可使用。所有元

数据规范均需符合相关国际和国内标准, 如都柏林核心元素集、元数据国家标准、中国高等教育文献保障系统元数据规范等。元数据标准均通过图书馆审核, 数据类型包括期刊论文、会议论文、图书、研究报告、病例报告、图谱、音视频、动画、科学数据、专利、课题项目、课题申请书、信件等。

4.5 系统架构

针对北京大学口腔医院的规模和特点, 医院机构知识库系统架构设计, 见图 1。



图 1 医院机构知识库系统架构

4.6 建设内容

(1) 建立医院学术数据中心。采集整理并长期保存北京大学口腔医院所有产出的学术成果形成科研数据库。学术成果类型包括公开发表的期刊论文、会议论文、研究报告、病例报告、图书等, 也

包括未公开发表的学术资源, 如学术信件、课件、会议报告、实验结果、病理图谱、手术视频等所有类型的学术成果、科学数据和学术资源。(2) 建立学者库, 构建学术交流平台。展示北京大学口腔医院所有医生和学者的科研成果, 包括学者唯一识别

号(本地识别号、ORCID 等其他唯一识别号)、个人简历、研究专长、合作者图形展示、科研成果被收录引用下载情况(含替代计量学指标)、发表的论文、著作、研究项目/课题、教学课程(包括课时、学生情况、教学课件)、指导学生的论文(包括学位论文、学生在该专家指导期间发表的论文)、获奖情况、科学数据集等,临床医生还要包括出诊时间、挂号费、典型病例等。(3) 建立医院学术成果展示平台。医院科研成果数据的统计、挖掘与展示。对收集的科研成果进行各种角度的统计和挖掘(以图形方式展现),展示北京大学口腔医院整体科研成果、学科建设情况等,为医院的学科建设、科研成果管理等提供数据增值服务。系统提供多种统计分析功能,可以对专家、成果、学科分别进行评价与分析。针对专家可以进行综合性分析,包括其发布成果数量、影响力等。针对成果提供多维度计量指标来分析成果、专家的影响力。针对学科可以通过 ESI 指标、同类院校相关学科的各分类指标进行统计分析。(4) 数据的开放获取权限控制。平台可以灵活设置对不同类型资源的访问权限,可由用户来自行设定。默认情况下 pdf 格式的全文只能院内开放获取不对院外开放,只有经过作者允许并且是不涉及版权的全文或者数据才可以对院外开放。院内开放也要设不同级别,如全院范围内、科室范围内、课题组范围内等。(5) 与医院现有信息系统数据对接。与院内一卡通系统、医院信息系统等进行对接,进行数据交换。(6) 医院机构知识库可以通过多种终端浏览访问,如手机、平板电脑、PC 机^[7]。

4.7 功能实现

4.7.1 目标 对于医院整体而言,机构知识库不仅可以本院的学术成果进行长期集中管理和保存,更重要的是可以分析评价医院的研究方向、过程、趋势和结果,完善科研产出管理,提高绩效管理和评价工作的效率,提升医院学术影响力、成果展示度,提升声誉和品牌,体现医院的社会价值。对于临床医生来说,机构知识库可以长期保存和管理临床与科研资料,帮助临床决策和科研。对于临床研究人员来说,机构知识库能有效增加研究成果

的可见度,提高学术论文的引用率,助力科研工作。对于临床教学人员来说,机构知识库通过展示教师的教学课程、课件、讲课视频、专长、研究方向、学生情况等,提升教学影响力。

4.7.2 机构成果典藏 从 SCIE、SSCI、A&HCI、PubMed、CSCD、知网、万方等学术数据库中回溯机构成果,建成医院机构成果典藏库。机构名称支持曾用名和名称缩写及其变种,建立机构名称规范对照数据库。利用 SCI、JCR 免费应用程序接口,提供实时被引次数、实时引文列表、文章发表当年的影响因子。(1) 机构成果典藏支持多种成果的收集方法。成员个人提交,系统自动审核;馆员批量导入成果(支持 Excel、EndNote、BibTeX、RIS、NoteExpress 等格式的导入);成果自动获取;支持 OAI 接口收割。全面收录医院公开成果、有限公开与内部使用的学术成果。针对不同的科研成果类型,支持多种标准文件格式上传、保存、检索与浏览,如 BMP、TIFF、GIF、PSD、PNG、DOC、DOCX、PPT、EXCEL、PDF、JPG、TXT、EPUB、CAJ、FLV、WMV、MP3 等。(2) 科研过程、成果的自动积累。为成员提供文献、文档、笔记、实验数据等的管理服务,为团队提供科研协同服务,实现科研过程、成果的自动积累。提供途径允许用户将这些科研过程、成果与发表的论文或著作等相关联。在最终科研成果详细页面上显示相关联的科研过程、成果情况。(3) 重要成果自动检测。自动检测机构的代表性成果,形成新闻并进行彰显。(4) 重要成果界定范围。发表在特定期刊的文章;收录为 ESI 高被引文章;收录 ESI 热点文章;F5000;优秀博士论文;被引频次达到阈值;发布在高影响因子期刊的文章。(5) 检索和浏览。通过全文搜索引擎建立全文索引,支持一键检索与复杂检索、检索词智能联想、模糊检索、拼音或英文单词智能纠错;支持按科室、发表年代、学科、文献类型、语种、收录类型、有无全文等分面浏览。检索结果支持以相关度排序,包括题名、年份、被引次数、影响因子等排序方式,供用户选择。检索结果页面的每条成果显示题名、作者、来源、收录引用数量、替代计量指标图、推荐数量等。

4.7.3 医生学者库 系统全面地收集全院医生、学者以及研究人员的个人信息,建成医生学者库。可以按科室、学科、姓名字顺列出学者。学者栏目可列出所有机构学者,也可由馆员来推荐学者展示。学者名片可显示学者基本信息和成果列表,包括姓名、照片、科室、主要研究方向、教育背景、研究领域、联系方式、ORCID 等。学者成果列表中发布年度、成果类型、收录情况进行排序显示,可提供学者成果引用情况和列表。对学者学术历程、合作者、发表期刊进行可视化展现,支持用户关注该学者。被关注学者的最新成果会被关注者自动订阅。

4.7.4 个人学术空间 主要功能包括:认领个人成果,系统判断的疑似本人的文章,需要作者确认后,方能在学者库中展示;个人成果管理,包括没有提交到机构库的个人成果;论文收录/引用通知,当用户论文被数据库收录后,自动匹配后发送收录通知,检测机构成果被引用情况,当用户论文发现被人引用时,发送收录通知。此列提供我的关注、我的图书馆、我的科研助理、我的收藏等个性化服务。

4.7.5 个人科研助理工具 为所有成员提供文献管理软件并与机构知识库相关联。该软件的直接目的是为科研人员个人文献管理、论文写作、实验数据记录、数据分析挖掘、知识获取等提供科研助理工具,同时解决资源分享的“最后一公里”问题,让成员不用花费额外劳动,即可把资源分享到团队,提高成员对资源分享的支持。

4.7.6 机构学术成果展示平台 成果导出:可按年度(范围)、科室、收录情况、科室署名次序、作者署名情况(第一作者、通信作者等)导出成果。包括成果引证数据,如科室排名、论文收录情况、收录编号、期刊影响因子、JCR 分区等。英文文章列出其作者中文姓名和科室。成果统计:年度成果科室排名统计(可视化展现)、年度论文收录情况统计、年度论文收录情况排名、年度成果类型分布、年度成果类型排名比较、年度代表性成果统计、成果收录情况年度比较、成果引用情况年度比

较、ESI 排名统计。

4.7.7 其他统计功能 分享统计:对师生或研究小组在团队内部分享的资源进行统计。系统应用报告:对用户资源总数、用户登录次数、用户数、机构成员数、团队数、团队成员总数、浏览次数、评价、成果总数、成果全文数进行统计。对统计数字进行历史保存和回看。统计机构知识库平台被访问、浏览、下载次数,按访问的 IP 区分院内外。

5 结语

医院机构知识库的建设没有可供参考和模仿的成功案例和经验,因而是一个需要不断探索的课题。本文初步论述医院机构知识库建设的基本设计和思路,探讨实践的过程。机构知识库的建设还远没有达到预期理想的状态和目标,需要进一步完善和提高。也期望能够从其他同行处借鉴相关的经验,以不断完善机构知识库的建设,为医院的创新发展提供动力和支持。

参考文献

- 1 赵迎光,曾苏,马建霞.国内机构知识库发展现状与对策研究[J].情报杂志,2011,30(1):42-46.
- 2 万文娟.机构知识库在国内的实践及不足分析[J].国家图书馆学刊,2011,(2):69-74.
- 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,(2):46-51.
- 4 乔欢,姜颖,宋姝.机构知识库内容建设现状及思考[J].图书馆工作与研究,2011,(12):67-70.
- 5 聂华,韦成府,崔海媛,等.机构知识库相关政策研究——基于北京大学机构知识库的思考与探索[J].大学图书馆学报,2014,(2):24-27.
- 6 路莹,郝继英,陈锐,等.DRM 技术在机构知识库系统中的应用[J].医学信息学杂志,2014,35(6):27-29.
- 7 乔欢,姜颖.国内外机构知识库内容建设研究进展[J].图书馆理论与实践,2011,(8):18-22.