

开放获取在NSTL系统中的实现与服务模式研究*

□ 张晓丹 王莉 梁冰 / 中国科学技术信息研究所 北京 100038

摘要: 开放获取资源是一种非常重要的学术资源, 开放获取平台是获得开放资源的有效途径。文章通过对DOAJ、JST等国际权威开放获取平台资源进行分析, 得出了目前开放获取资源存在的信息增长与信息获取之间的矛盾结论。如何利用和整合国际上著名的开放获取资源为国内用户服务是目前研究的主要问题。文章针对国内科研用户对开放获取资源整合的需求进行了研究, 并将研究结果应用在国家科技图书文献中心(NSTL)平台上。

关键字: 开放获取, 资源整合, NSTL

DOI: 10.3772/j.issn.1673-2286.2013.10.012

1 引言

开放获取(Open Access), 是相对于传统出版模式而言的, 其目的是通过网络加速学术资源的共享, 使得任何人在获取资源时不受时间、费用等限制。它起源于20世纪90年代, 目前已经在世界各地得到了蓬勃的发展。

随着开放获取运动的发展, 开放获取资源得到了迅速增长, 使得信息增长与信息获取的矛盾日益突出。尽管国际上已经出现了DOAJ、OpenDOAR、Open J-Gate、Socolar等大型的开放获取资源统一访问平台, 但这些平台在资源收录方面各有偏重, 提供的检索方式参差不齐, 且多局限于题录级检索。如何利用国际上先进的开放获取资源, 为国内科研用户提供更好的服务是目前研究的热点问题, 也是本文要解决的主要问题^[1-4]。

国家科技图书文献中心(National Science and

Technology Library, NSTL) 是国际上具有一定影响力的现代科技文献信息服务机构^[5], 是我国权威的科技文献信息资源收藏和服务中心。本文就NSTL对国际上著名的开放获取资源进行整合与服务进行的相关研究进行阐述。

2 开放获取发展现状

自20世纪90年代《布达佩斯开

放获取计划》给予了“开放获取”的完整定义后, 开放获取运动就在全世界积极展开了^[5-8]。据开放存取知识库登记系统ROAR(Registry of Open Access Repositories)统计显示, 全球公开登记的机构知识库截至2013年5月已达6800个, 其中, 中国只有13个, 其余的都属于欧美知名大学和科研机构。ROAR对开放获取期刊的增长及资源分布情况统计结果见图1和图2所示。

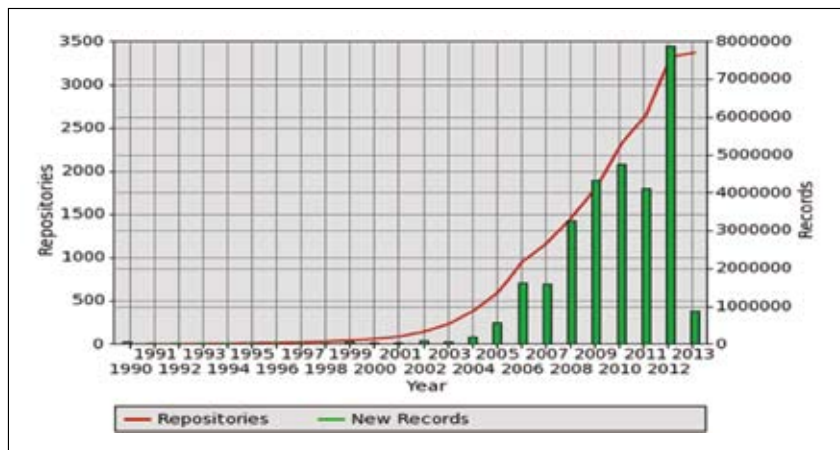


图1 ROAR不同年度开放获取期刊增长曲线图

* 本课题受“十二五”国家科技支撑计划项目(编号: 2011BAH10B05)和中国科学技术信息研究所预研项目(编号: YY-2013023)资助。

从图1中可以看出,开放获取资源是随时间的增长快速发展的,尤其是到了2012年度,开放获取资源增长迅猛,特别是2012年度,由2011年的400万增长到800万。由此可见人们已经普遍认识到开放获取对科技发展及科研活动的重要性。

从图2可以看出,开放获取资源目前在国外占有的比重比较大,而中国只占有2.61%的份额。因此,加快开放获取在国内的发展并整合世界上的开放获取资源,势在必行。

3 开放获取在NSTL体系中的实现

随着开放获取资源的爆炸性增长,信息增长与信息获取的矛盾日益突出。尽管已经出现了DOAJ、OpenDOAR、Open J-Gate、Socolar等比较大的开放获取资源统一访问平台,但是这些访问平台在资源收录方面各有偏重,提供的检索方式参差不齐,且多局限于题录级检索,整合开放获取资源并提供集成检索服务势在必行。

NSTL针对上述问题,提出了完善的解决方案,以期建立面向国内科研用户的完善的服务系统平台。

由于开放获取资源数量庞大,且全球范围内开放仓储数量超过2000个,开放获取期刊数量超过15000种,但质量不一而足。所以选择高质量、可靠的开放获取资源是开放获取建设的基础。我们选择资源的原则是:针对世界上较权威并可用的开放获取资源,其具有庞大的规模和稳定的运行状态,同时,所选资源要有代表性,能体现相关科研领域的研究现状和发展方向等。

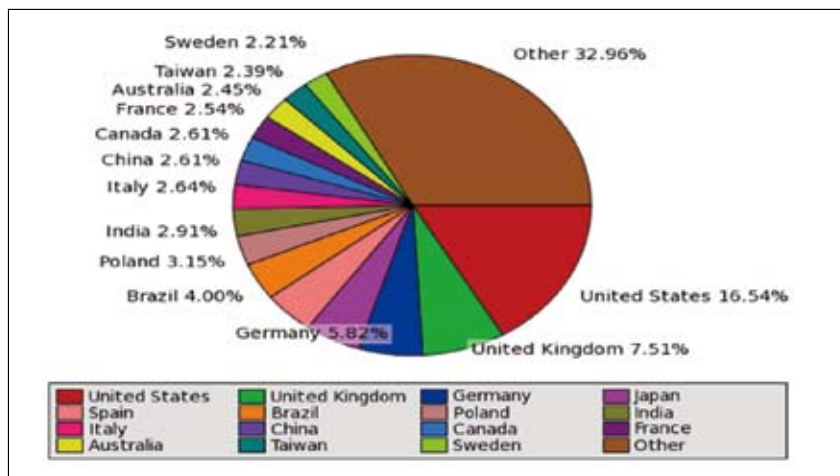


图2 ROAR 2012年度开放获取期刊世界分布图

(1) NSTL与权威的开放获取平台建立合作关系

遵循以上原则,目前NSTL已与世界上三个权威的开放获取平台建立了良好的合作关系,即DOAJ、JST及SciELO。

DOAJ (Directory of Open Access Journals, <http://www.doaj.org>) 是由瑞典隆德大学图书馆2003年5月推出的开放获取期刊的检索系统。目前该系统可提供9000多种开放获取期刊的访问(数据截至2013年1月)。该系统提供刊名检索、期刊浏览以及文章检索等功能。DOAJ按期刊的学科主题分为17类,包括农业及食品科学,艺术与建筑学,生物及生命科学,贸易与经济学,化学,地球与环境科学,总类,保健科学,历史与考古学,语言文字,法律与政治科学,数学与统计学,哲学及宗教,物理及天文学,科学总论,社会科学,技术与工程学。

JST: J-STAGE (Japan Science and Technology Information Aggregator, Electronic, 日本科学技术信息网络电子平台, <http://www.jstage.jst.go.jp>) 由日本科学

技术振兴机构(JST)开发,收录了日本各科技学会出版的文献(以英文为主),包括600余种电子期刊,多种会议录以及研究报告等。部分期刊、大部分会议录和研究报告可以免费浏览全文。

SciELO (Scientific Electronic Library Online, <http://www.scielo.org/>), 其主要目标是: 遵循开放存取模式, 也就是免费获取全文, 出版巴西在线科技期刊, 提高国家及国际间期刊显示度。目前, SciELO运动已经在拉丁美洲和加勒比国家(阿根廷、智利、西班牙、古巴、哥伦比亚、葡萄牙、委内瑞拉)广泛展开, 开放期刊数量达已达600多种, 有20多万篇文章, 内容覆盖农业、生物、化学、数学、地球科学、建筑、工程、健康科学、人文、法律、语言、艺术、海洋、社会科学等学科。

这些资源各有侧重, 因此对其进行有机的整合能为国内用户提供国际上较为全面的、先进的开放获取资源。

(2) NSTL开放获取系统

NSTL开放获取系统的主要流程为资源收割、资源加工、资源

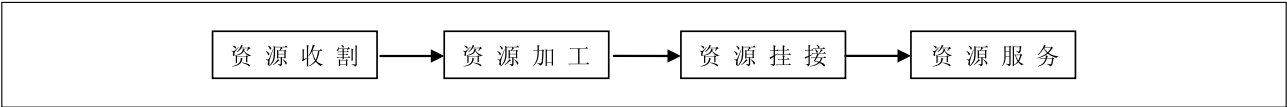


图3 NSTL开放获取系统流程图

挂接（与NSTL原有的类别进行挂接）、提供检索与浏览等服务，如图3所示。

从图3可以看出，首先对国外权威的开放获取平台进行资源的定期收割（根据更新的速度，收割的频率通常一周一次或三天一次）；对收割下来的资源进行数据加工，加工过程包括分类、去除噪声、提取特征元数据等；然后进行资源挂接，即将获取的资源与NSTL的原有类别进行挂接，实现NSTL统一的资源分类管理；最后是实现对外的检索和浏览等服务。

基于上述的开放获取流程，搭建为国内用户服务的开放获取服务平台的基本架构如图4所示。

从图4可以看出，为了满足NSTL开放获取期刊、全文集成服

务平台的总体要求，结构图中主要包括开放获取（OA）刊数据存储模块，实现对世界上开放获取资源的存储；OA刊数据管理模块，主要完成对资源的统计监测等功能；OA刊的服务模块，主要完成对OA刊及篇的浏览、检索等对外服务功能；对外接口模块，主要实现本服务系统与外界无缝的接口。

该平台的主要功能有：

实现会话管理功能。可以对访问系统的人进行身份识别，对会话操作进行记录，并根据用户的身份提供相关的服务（如基于会话的检索或浏览过程记录）；实现检索查询功能。可以按用户习惯的方式对OA期刊数据进行检索；实现期刊按字顺和按分类两种浏览功能；支持OA期刊数据中资源层次及相关

关系的有效组织和揭示；能够明确按期刊、卷、期、文章的次序浏览和组织期刊论文；实现期刊和论文内容揭示功能；提供Landing页，对于期刊和论文都需要有一个唯一的定位标识，可以及时定位到相应的Landing页上，Landing页提供期刊和论文的元数据揭示；实现全文链接功能。合法使用的用户可以在检索到篇数据之后链接到原始网站。

（3）NSTL开放获取系统元数据

对于收割的开放获取资源，要提取出对NSTL开放获取有效的元数据。根据系统需要，提供的开放获取期刊与篇章的元数据列表如表1所示。

表1 NSTL开放能够获取系统元数据列表

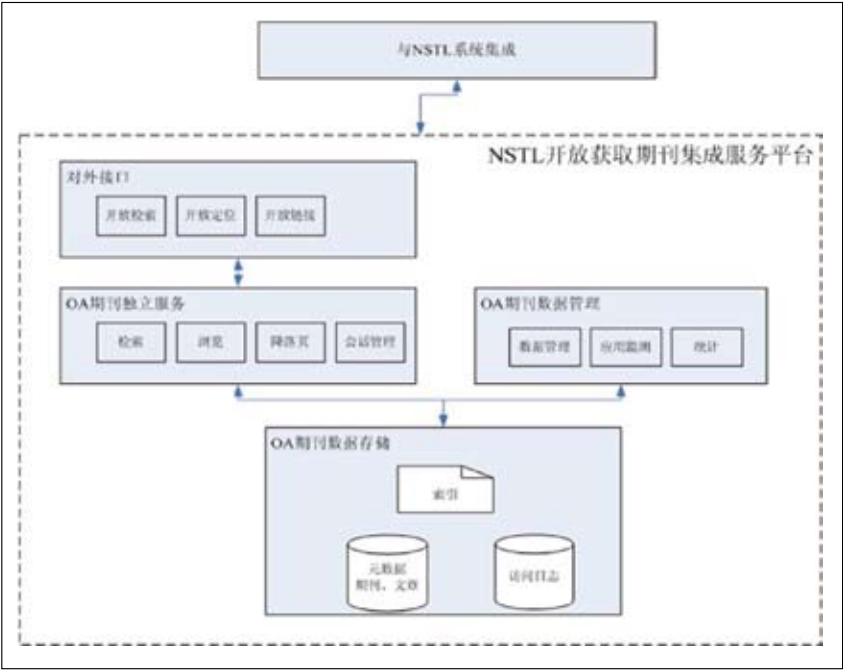


图4 NSTL开放获取期刊集成服务总体结构图

ArticleSet
Article
Journal
PublisherName
JournalTitle
Issn
Volume
Issue
PubDate
Year
Month
Season
Day
ArticleTitle
ArticleSubtitle
VernacularTitle
FirstPage

表2 期刊描述元数据

序 号	字段名称	描 述	必备性	可重复性
1.	JournalID	期刊唯一标识符	必备	不可重复
2.	JTitle	刊名	必备	不可重复
3.	ISSN	issn	可选	不可重复
4.	EISSN	Eissn	可选	不可重复
5.	Jlanguage	语种	必备	可重复
6.	Jclassification	分类号	可选	可重复
7.	JSubject	主题	可选	可重复
8.	Publisher	出版商	必备	不可重复
9.	Country	国家地区	必备	不可重复
10.	Start_year	起始年	必备	不可重复
11.	End_year	中止年	可选	不可重复
12.	JRights	开放权限	必备	不可重复
13.	Publication_fee	出版费用	必备	不可重复
14.	AJ_Identifier	期刊-篇标识符	可选	不可重复

表3 期刊-文章描述元数据

序 号	字段名称	描 述	必备性	可重复性
1.	AJ_Identifier	期刊-篇标识符	必备	不可重复
2.	AJTitle	刊名	必备	不可重复
3.	AISSN	issn	可选	不可重复
4.	AEISSN	Eissn	可选	不可重复
5.	AJlanguage	语种	必备	可重复
6.	AJclassification	分类号	可选	可重复
7.	AJSubject	主题	可选	可重复
8.	APublisher	出版商	必备	不可重复
9.	ACountry	国家地区	必备	不可重复
10.	AStart_year	起始年	必备	不可重复
11.	AEnd_year	中止年	可选	不可重复
12.	AJRights	开放权限	必备	不可重复
13.	APublication_fee	出版费用	必备	不可重复

4 结论

随着开放获取运动的发展,开放获取资源得到了迅速增长,使得信息增长与信息获取的矛盾日益突出。本文就如何利用国际上先进的开放获取资源,为国内科研用户提供更好的服务进行了研究并在NSTL中进行了实现。该开放获取平台可以进一步推动学术信息及资源的传播,扩大NSTL的服务范围系统。

为了将更完善的资源纳入我们的平台,我们将在下一步选取更多更全面的资源纳入到我们的平台中,并将个性化服务方式提供给用户以提供更好的服务。

参考文献

- [1] ZHANG B, CHEN Y, FAN W, et al. Intelligent GP fusion from multiple sources for text classification [C]// Proceedings of the 14th ACM international conference on Information and knowledge management. ACM Press, Bremen, Germany, 2006a: 477-484.
- [2] ZHANG G P. Avoiding Pitfalls in Neural Network Research. Systems, Man and Cybernetics, Part C: Applications and Reviews [J]. IEEE Transactions, 2007(37): 3-16.
- [3] ZHANG L, ZHU J, YAO T. An evaluation of information fusion techniques [J]. ACM Transactions on Asian Language Information Processing (TALIP), 2004(3): 243-269.
- [4] ZHANG Y, ZINCIR-HEYWOOD N, MILIOS E. Narrative fusion classification for automatic key phrase extraction [C]// Proceedings of the 7th annual ACM international workshop on Web information and data management. ACM Press, Bremen, Germany, 2005: 51-58.
- [5] XU L Y, DU Q D. Application of neural fusion to accident forecast in hydropower station [C]// Proceedings of the Second International Conference on Information Fusion. Vol2 Sunnyvale, 1999: 1166-1171.
- [6] SCHAPIRE R E, SINGER Y, SINGHAL A. Boosting and Rocchio applied to text filtering [C]// Proceedings of SIGIR-98 21st ACM International Conference on Research and Development in Information Retrieval. ACM Press, New York, US, 1998: 215-223.
- [7] SEBASTIANI F. Machine learning in automated image categorization [J]. ACM Computing Surveys, 2002(34): 1-47.
- [8] ZHANG XIAO-DAN. A system of file Automated classification: China, 201020200043 [P]. 2010-12.

作者简介

张晓丹 (1975-), 研究方向: 信息挖掘、信息融合等。E-mail: zhangxd@istic.ac.cn

Study on Open Access Implementation and Service Pattern of NSTL

Zhang Xiaodan, Wang Li, Liang Bing / Institute of Scientific and Technical Information of China, Beijing, 100038

Abstract: Open access resources are very important academic resources. In this paper, DOAJ, JST and other international authoritative platform for the open access resources are analyzed, and there is contradiction between the growth of information and the access to them of current open access resources. It is important to make use of and integrate internationally famous open access resources. Based on the background of NSTL, The present study establishes the integration of open access resources service platform for domestic users, improves the level of domestic users of research and the quality of the open access service.

Keyword: Open access, Resources integrating, NSTL

(收稿日期: 2013-06-06)