

科技期刊数字化建设策略:数据融合与信息共享*

郭飞 薛婧媛 胡志平 罗敏 吕建斌

重庆大学期刊社,400044,重庆

摘要 为了适应新形势下科技期刊的数字出版和体制改革的要求,科技期刊均尝试建造具有自身特色的数字化设施。虽然国内外已经有相当多的科技期刊在数字化方面取得了一些成功,积累了实践经验,但目前数字化建设仍然存在诸多问题。制约数字化发展的主要因素有数字化部件相对孤立、数字出版资源关注度差、新型数字媒体传播模式缺乏,等等。为此,需要建立适宜的互动模式来推进数字化建设,并为所有出版参与者服务。文章将数据融合、信息共享的理念应用于科技期刊的数字化建设中,采用高聚合、低耦合的方式来解决目前数字化建设中存在的一些问题。

关键词 科技期刊;数字化建设;数据融合

Digital construction strategies for scientific and technological journals: data fusion and information sharing//GUO Fei, XUE Jingyuan, HU Zhiping, LUO Min, LYU Jianbin

Abstract To adapt to the requirement of digital publishing and structural reform, scientific and technological journals tried to build different digital facilities based on their own characteristics. The attempts have achieved some success and practical experience, but there still are many problems in the digital construction process. The main factors that restrict the development of digital construction include isolated digital components, poor concern on digital publishing, and insufficient new digital media communication mode. Therefore, an appropriate interactive mode should be built to prompt the digital construction and serve for all publishing participants. We carried out an idea of achieving data fusion and information sharing in the digital construction of scientific and technological journals, and used a high cohesion and low coupling mode to solve existing problems in the digital construction.

Key words scientific and technological journals; digital construction; data fusion

Authors' address Journals Department, Chongqing University, 400030, Chongqing, China

目前,国内科技期刊的数字化建设主要围绕3个方面进行,分别是网络数字出版、门户网站及投审稿平台的搭建、新型数字媒体传播模式的创新。据新闻出版总署网站统计,截至2012年6月28日,我国共有期刊9906种^[1]。调查显示,我国科技期刊网络数字出版发展势头良好,加入CNKI的期刊有7920种^[2]、加入万方数据库的期刊有7397种^[3],普及率超过79.9%。另据姚远等的抽样调查,高校科技期刊加入网络数据库的

比例更是高达97.73%^[4]。然而,相比网络数字出版的蓬勃发展,科技期刊的门户网站和投审稿平台建设现状却不尽如人意,拥有自己门户网站的科技期刊仅占总数的40.45%。杜亮等对中国医学核心期刊的投审稿平台进行了抽样调查,结果显示,启用投审稿系统的期刊仅占26.1%^[5]。此外,随着传播模式的转变,自媒体传播、微传播等新型数字媒体传播模式悄然进入科技期刊的传播范畴。科技期刊纷纷开拓各种传播渠道,以求得到更广泛的关注;然而,此类传播模式的数字化建设并不完善,尚处在探索阶段^[6]。以往的研究表明,我国科技期刊数字化建设的某些方面还很薄弱,包括门户网站及投审稿平台的搭建、传播模式的创新。

在科技期刊致力于填补上述发展空白的同时,一些新兴网络技术又为科技期刊带来了更广阔的发展空间和新的服务途径;然而,绝大多数科技期刊并没有有效地利用数字化资源来完善自我建设,因此,本文运用现有的数字化资源和技术手段,来实现科技期刊的数字化建设,以构建适合我国科技期刊数字化发展的蓝图,更好地为作者、审者、编者以及学术团队服务,全面提升科技期刊的影响力。

1 科技期刊数字化发展进程

近10几年来,随着计算机技术与网络技术的迅猛发展,科技期刊也与时俱进地迈入了数字化时代。它们服务于科技期刊的各个方面,孕育而产生了许多编辑辅助应用工具,如科技期刊的投审稿系统、学术不端检测系统^[7]。此外,科技期刊的出版模式也演变为网络数字出版和开放存取模式^[8]。

门户网站从最初的静态页面转变为动态的Web 2.0交互模式,众多科技期刊纷纷宣布进入Web 2.0时代^[9]。门户网站是作者和读者了解该期刊的最重要的途径,已成为科技期刊对外宣传的重要窗口。除门户网站的建设外,科技期刊的投审稿方式也发生了重要的转变,大致经历了3个阶段:1)传统纸质邮件寄送;2)电子文档的E-mail投送;3)基于B/S模式的在线投审稿平台。这3种方式中,无论是论文发表的时效性、便捷性,还是论文管理的优越性,基于B/S模式的在线投审稿平台都是最适合的方式。它是以稿件为中心的业务处理流程,其优点在于实现了信息共享,

* 中国高校科技期刊研究会基金项目(GBJXC1251)

提高了科技期刊的通用性和灵活性。

从传统的纸质出版到数字出版,从订阅科技期刊到及时获取,科技期刊的出版模式全面进入了网络数字出版和 OA 时代。相对于签约特定期刊数字全文系统的网络出版,OA 期刊有其自身的优越性:出版快捷,出版费用低,便于传送或刊载大量数据信息,具有广泛的读者群和显示度。

网络数字出版和 OA 的建设使科研人员能更快地获取资源,但同时也出现了抄袭、剽窃等学术不端行为。为了有效地避免重复发表,防止学术不端行为,国内外机构相继开发出各种学术不端检测系统,供授权期刊或单位使用。其中最具有代表性的是中国知网的学术不端检测系统和 CrossCheck 论文检测系统^[7]。前者是基于快速的全文对比,后者是基于全文比对和语义识别技术。学术不端检测系统的引入,可以在第一时间发现抄袭、剽窃等现象,高效地防止了此类学术不端行为的发生,同时节省了出版单位的人力和物力。

2 科技期刊数字化建设的不足

虽然,目前科技期刊数字化建设研究取得了一些成果,但还存在很多不足之处。主要表现如下。

1) 数字化部件的独立存在和数据孤岛。科技期刊的数字化部件相对独立,具体表现为:首先,门户网站和投审稿平台不属于一个系统,需要手工定期上传全文数据到门户网站上;其次,有的科技期刊仍然采用纸质或者 E-mail 投审稿,这种情况造成了数据独立,形成了数据孤岛;第三,不同信息化部件的数据定义可能存在巨大的差异,容易滋生冗余数据,使得论文数据既繁琐又难于管理,不利于学术论文同行评议的追踪与分析——因此,需要在各信息部件中建立统一的通用协议,以保持数据的完整与关联。

2) 交互性及关注度差。第一,虽然门户网站提升了科技期刊的显示度,但作者或读者大多只是获取已有的资源,而不会主动向科技期刊提出需求;第二,大多数作者通过投稿系统所获得的仅仅是审稿进度和评审意见,缺乏与编辑部的交流;第三,编辑部对已发表论文的后续研究缺乏跟踪,重要研究成果的报道难有延续性,这样势必会影响作者投稿的积极性,以至于对该刊的关注度逐渐下降。

3) 数字出版资源或者 OA 资源的被动发现。搜索引擎技术和网络数字出版的建设让学术资源的获取变得方便快捷,然而它们在时间上都有短暂的滞后,被动发现这些资源往往不够及时,不能让学者在第一时间获取最新的学术成果。

4) 论文中数据的高度精炼。科技期刊因受版面

和载体的限制,不能把众多学术论文中完整的实验过程、实验数据,或者其他多媒体数据(例如声音、高分辨率的图像、影视片段等)呈现出来^[10],而是进行了高度的精炼或简单叙述;因此,应该提供额外的辅助途径让读者更详细地了解作者所进行的研究。

5) 引文获取途径单一。引文分析是用来定性评价期刊学术水平的重要手段,而引文的获取来自引文数据库或网络互引。众所周知,各个引文数据库都是独立存在的,它们提供的引文数据都是基于自身数据库产生的,这样就会出现错引或遗漏的情况,不能全面了解文章的被引全貌。

3 科技期刊数字化建设的融合

通过对科技期刊数字化建设现状和信息技术发展趋势的分析,笔者认为应当以数据融合、信息共享的理念来完善数字化建设,以弥补当前科技期刊数字化建设存在的缺陷。曾建勋等提出“实现全方位的数字化出版流程;与用户共同建设一个网络知识社区;提供基于全文的知识服务;积极参与知识创新的管理”,展望了我国未来泛在知识环境中科技期刊的发展方向^[11]。笔者认为可以从以下几个方面着手,让现有技术能充分服务于科技期刊。

1) 学术不端检测系统与 B/S 投审稿平台的融合。目前这 2 个系统是独立运行的,耦合度差。可以利用 Web Service^[12]的方法来封装学术不端检测服务,提供统一的接口供不同的投审稿平台调用。这样,既方便获取对应的检测结果,又不用改变投审稿系统的内部结构。Web Service 是目前较成熟的服务重用方法。

2) 给作者或潜在作者主动推送 OA 及其他数据。通过推送技术在第一时间将最新的研究论文推送至作者或潜在作者,把资源的被动发现变成主动推送,消除了搜索引擎和图情数字化的滞后性^[13];但是,我们也不仅仅依靠诸如 RSS 之类的推送技术,需要将之与搜索引擎和图情数字化融合,以实现资源主动推送和被动发现的结合。

3) 科技期刊的自传播模式与微传播模式的结合。作者与作者、作者与审稿专家、团队与团队的关系映射到网络就构成虚拟学术社区。在此社区中,大部分都是以个人或团队的博客为传播节点而形成的自传播网络。自传播模式促进了学术交流和讨论,成为研究成果产生的催化剂,优秀的学术成果能迅速地在自传播网络中扩散;因此,科技期刊应加入自传播网络环境中,融入到对应的学术关系网,提供相应的应用服务^[14]。此外,微传播模式的出现也对科技期刊产生了深远的影响。微传播模式加快了最新成果的报道,一方面科技期刊的关注

者能即时获取科技期刊的最新咨询,另一方面,科技期刊可以关注最新的研究成果和学科发展动向。

4) 自动关联参考文献及引文分析。多数科技期刊仅仅把网络数字出版看作一种出版模式,然而网络数字出版的外延却蕴含了众多可挖掘的信息。充分利用网络数字出版,可以将参考文献自动关联到相应的数据库或网络资源,还可以利用搜索引擎技术从不同的引文数据库和 Google Scholar 中自动加载论文的引文数据^[15]。

在当今信息技术的大环境中,科技期刊数字化建设不应该孤立地运用某1项或某几项技术,而应整合现有资源,加快数字化建设的步伐,推动数字化建设的进程。

图1完整地展示了如何融合现有技术进行科技期刊数字化建设,体现了数据融合、信息共享的理念和思想,使用户体验到各信息部件所带来的便捷,显著提高

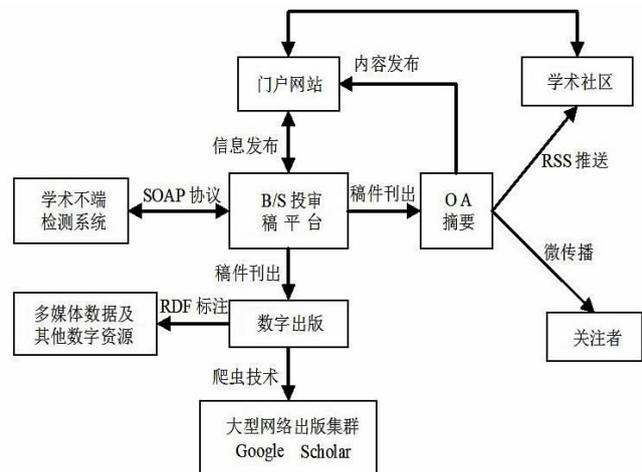


图1 科技期刊数字化建设的融合

了科技期刊的形象。各部件的数据交换和控制都是基于XML技术的,双向箭头表示信息部件之间有数据交互,单向箭头表示一信息部件向另一信息部件传递或者推送信息。如图所示,科技期刊的门户网站与投稿系统应具有良好的衔接,用户能够方便地获取期刊信息和投稿。当作者投稿时,投稿系统自动调用学术不端检测服务,返回结果作为编辑部审核稿件的重要依据,他们的交互可采用 Web Service 技术,学术不端检测服务提供统一的 WSDL 描述,并通过 SOAP 协议交互数据。经过同行评议合格的稿件进入数字出版阶段,以优先出版的方式自动将摘要及全文发布到门户网站,在数字出版的同时收集整理与稿件相关的多媒体及其他数字资源一并发布并进行 RDF 标注,同时利用主动推送技术 RSS 和调用微传播 API 接口推送至相关网络,提升最新研究成果的关注度。此外,期刊门户与作者群的博客建立良好的网络社会关系,有利于跟踪作者群的后续研究

成果。利用爬虫技术获取互联网数据,数字出版的论文将自动关联参考文献,定期从各个数据库和 Google Scholar 更新引文数据。以上过程是不断叠加循环的。

4 结束语

通过分析研究,笔者提出了如何融合现有技术来进行科技期刊数字化建设,展示了该结构的运作模式。该结构具有一定的可行性,将作为一种新的研究方法解除目前科技期刊数字化建设的桎梏,加快科技期刊的数字化变革。

5 参考文献

- [1] 新闻出版总署. 新闻出版机构查询[DB/OL]. [2012-06-28]. <http://www.gapp.gov.cn/cms/cms/website/zhrmgh-gxwebzsww/SearchService/xwjgex.html>
- [2] CNKI 中国知网. 中国数学期刊网络出版总库[DB/OL]. [2012-06-28]. http://acad.cnki.net/Kns55/oldnavi/n_Navi.aspx?NaviID=100
- [3] 北京万方数据股份有限公司. 万方数据知识服务平台[DB/OL]. [2012-06-28]. <http://c.wanfangdata.com.cn/Periodical.aspx>
- [4] 姚远,汤颀,赵军平,等. 中国高校科技期刊现状调查与分析研究报告[EB/OL]. [2012-06-28]. <http://www.chjjh.com/UploadFile/200726204318838.pdf>
- [5] 杜亮,陈耀龙,李晓,等. 中国医学核心期刊在线投稿与审稿系统调查[J]. 中国科技期刊研究,2008,19(4):608-611
- [6] 王亚新,史春薇,仲崇民,等. 在数字化条件下微博对科技期刊的影响[J]. 编辑之友,2011(10):71-72
- [7] 王新英,赵艳静,赵阳. 彻查科技论文学术不端的编辑策略[J]. 编辑学报,2011,23(3):231-232
- [8] 程维红,任胜利. 中国科技期刊开放存取出版现状[J]. 编辑学报,2007,19(3):196-198
- [9] 郭雨齐,窦红光,王桂颖. Web 2.0 时代下的科技期刊网站[J]. 中国科技期刊研究,2008,19(3):426-427
- [10] 吴根范,王青. 网络环境下科技期刊出版的特点、问题及发展建议[J]. 编辑学报,2009,21(3):252-253
- [11] 曾建勋,刘华. 泛在知识环境中科技期刊的发展方向[J]. 编辑学报,2010,22(6):471-474
- [12] 何清林,杨森,徐泽同. 基于元数据和 Web Service 中间件的分布式资源库集成[J]. 计算机工程与设计,2009,30(9):2201-2204
- [13] 张伟,陶小雪,林家乐. 利用网站实现学术期刊的主动宣传[J]. 中国科技期刊研究,2007,18(2):287-291
- [14] 陈鹏,张美琼,黄历,等. 学术博客与传统学术期刊的竞合发展[J]. 编辑学报,2012,24(1):19-22
- [15] 洪道广. Google Scholar 的数据整合研究[J]. 现代情报,2010,30(7):39-45

(2012-07-18 收稿;2012-09-01 修回)