

## 我国百强中文科技期刊 XML/HTML 出版现状调研与分析

■周小玲<sup>1)</sup> 侯春梅<sup>2)</sup> 黄爱华<sup>1)</sup> 迟秀丽<sup>2)</sup> 李 慧<sup>1)</sup>

收稿日期:2018-08-26

修回日期:2018-11-15

1) 中国科学院兰州文献情报中心《黄金科学技术》编辑部,甘肃省兰州市城关区天水中路8号 730000

2) 中国科学院兰州文献情报中心《地球科学进展》编辑部,甘肃省兰州市城关区天水中路8号 730000

**摘要** 【目的】全面调研国内外优秀科技期刊的 XML/HTML 出版现状和功能建设,为我国中文期刊应用 XML/HTML 出版提供参考。【方法】以第三届百强中文科技期刊为样本(有效样本数为70种),采用网络调研法分析我国优秀中文期刊的 XML/HTML 出版现状和功能建设情况,并与国际优秀期刊网络平台的 XML/HTML 出版情况进行比较。【结果】只有17种百强中文期刊网站推出了 HTML 全文阅读,仅占总样本数的24%,反映出我国中文期刊对 XML/HTML 出版不够重视。我国中文期刊 XML/HTML 出版的内容结构化程度低、知识关联方式单一、多维论文评价数据和补充材料欠缺,与国际优秀出版机构存在较大差距。【结论】XML/HTML 出版是科技期刊实现内容延伸、知识关联和知识重组的必要途径,能够满足读者的碎片化阅读需求,有助于显著提高文章的显示度和传播力,应受到国内期刊出版界的足够重视。

**关键词** XML/HTML 出版;内容结构化;知识关联;出版技术;网络出版平台;中文期刊

DOI:10.11946/cjstp.201808260756

随着网络技术的快速发展以及移动设备的普及,人们的阅读习惯和方式正在发生巨大变化<sup>[1-2]</sup>。在媒体融合、复合出版的环境下,读者对于多媒体、跨平台的数字内容资源的需求越来越大<sup>[3-4]</sup>,传统以 PDF 文档为主的单一数字出版形态已经无法满足读者需求,取而代之的便捷化、碎片化阅读需求对科技期刊数字化出版提出了更高要求<sup>[5]</sup>,因此适应碎片化阅读的可扩展标记语言(Extensible Markup Language, XML)和超文本标记语言(Hyper Text Mark-up Language, HTML)出版逐渐进入人们的视野。与传统单一的 PDF 全文阅读模式相比,XML/HTML 出版实现了文章内容的结构化标引、知识互联、跨平台资源自动链接等<sup>[6]</sup>,具备增强和延伸文章内容的优势,章节跳转、知识互联等功能有助于读者便捷获取感兴趣的内容,因此越来越受到期刊出版界的关注和重视。

近年来,我国期刊工作者开始关注和研究国际优秀期刊或知名出版集团的先进数字出版模式,包括英国皇家化学学会和 Elsevier 集团推出的语义出版<sup>[7-8]</sup>、国际四大医学期刊(《新英格兰医学杂志》《柳叶刀》《美国医学会杂志》《英国医学期刊》)网

络平台的富媒体出版<sup>[9]</sup>、Springer Nature、Wiley 等国外出版社采取的增强出版<sup>[10]</sup>,认为文章内容的结构化标引、知识挖掘和关联、音视频和数据集的嵌入等富媒体出版新形态是期刊发展的重要方向,值得国内期刊界借鉴学习。同时,国外视频实验期刊<sup>[11]</sup>和数据期刊<sup>[12]</sup>的相继推出,使我国学术出版界认识到数字环境下期刊出版形态已经发生了巨大变化,相比纯文字内容,可重复的实验数据、音视频带给读者的信息量更大、内容更丰富。基于此,中华医学会系列期刊集群率先尝试采用视频技术来增强文章内容<sup>[13]</sup>,以期实现内容的富媒体出版。

目前,有关国内外科技期刊的 XML/HTML 出版现状和模式研究相对较少,杨郁霞<sup>[14]</sup>对国外期刊应用 XML/HTML 的出版状况进行了初步研究,还有部分学者仅对国内单种期刊的 XML/HTML 全文出版模式进行了一些探索与实践<sup>[15-16]</sup>。而现阶段我国中文期刊 XML/HTML 出版的总体情况和功能开发尚鲜有报道。为此,本研究以全国百强科技期刊为研究样本,重点对其中具有代表性的优秀中文科技期刊的 XML/HTML 出版现状进行系统调研和分析,并通过对比国际优秀出版社在线出版平台的 XML/

**基金项目:**中国科学院兰州文献情报中心文献情报创新能力建设项目“期刊集群出版管理及规范体系建设”(Y7AJ012020)。

**作者简介:**周小玲(ORCID:0000-0001-8551-5203),硕士,编辑,E-mail:zhouxl@llas.ac.cn;侯春梅,硕士,编审,期刊编辑出版部主任;黄爱华,硕士,编辑;迟秀丽,学士,副编审;李 慧,学士,编辑。

HTML 功能,指出我国中文期刊在 XML/HTML 出版方面存在的差距,以期为更多中文期刊应用 XML/HTML 出版提供参考。

## 1 研究方法 with 概念界定

### 1.1 研究方法

选取入选第三届全国百强科技期刊的中文期刊作为研究对象,初步筛查剔除了 21 种英文期刊和 9 种科普期刊,最终确定的有效样本总数为 70 种。采用网络调研和统计分析方法,对 70 种百强中文期刊网站的论文阅读模式进行调研分析。具体调研步骤为:(1)通过互联网搜索 70 种百强中文期刊官方网站,逐个打开各刊网站来查看近一年发表论文的在线阅读模式,如元数据浏览、PDF 下载和 XML/HTML 全文阅读等;(2)随机选取提供 XML/HTML 全文阅读的期刊近一年发表在不同刊期的若干篇文章( $\geq 20$  篇),打开每篇文章的 HTML 阅读页面,记录文章要素(如图表、公式、参考文献等)的结构化程度和主要功能,从而较全面地了解我国优秀中文期刊 XML/HTML 出版现状。通过对比国际优秀出版社在线出版平台的 XML/HTML 出版特点,指出我国中文期刊 XML/HTML 出版存在的不足之处及未来的发展方向。

### 1.2 概念界定

HTML 是目前网络上应用最广泛的语言。HTML 文档是由 HTML 命令组成的描述性文本,通过静态页面展示文字、图形、表格和链接等,可被多种网页浏览器读取<sup>[6,17]</sup>。HTML 文档的超链接(各章节跳转、参考文献链接)、碎片化阅读及全文检索等功能,有助于提高用户交互性、知识传播效率及论文显示度和引用率<sup>[17]</sup>。但是,HTML 对信息语义及其内部结构、数据表现(矢量图形、科学符号)的描述较为欠缺。为解决这些问题,XML 诞生了,成为 HTML 的补充。

XML 的显著特征是结构化,即通过标识符使文档中各元素相互关联。基于 XML 的内容生产加工方式具有一次制作、多元多次发布,以及便于传输、存储等优势,可直接生成网刊数据、Word、PDF、XML、HTML 等格式文档,因此国际知名出版机构很早就将 XML/HTML 应用在数字出版领域。

## 2 国外优秀期刊 XML/HTML 出版总体情况

许多优秀国际出版社的网络出版平台很早就推

出了 XML/HTML 全文阅读模式,采用 XML 结构化内容加工方式,实现了文章的结构化标引、知识关联和跨数据库资源链接等。在先进数字出版技术的支撑下,ScienceDirect、Royal Society of Chemistry (RSC)、Public Library of Science (PLoS)、Science 等网络出版平台的 XML/HTML 出版日趋成熟,全文碎片化和结构化程度更加精细,功能建设更健全。本研究以这 4 个国际优秀网络出版平台为例,重点介绍其 XML/HTML 出版的特色功能(表 1),以期为我国中文期刊 XML/HTML 出版提供若干有益参考。

由表 1 可知,除论文各章节之间跳转、图表独立呈现、作者信息关联检索、参考文献链接等 HTML 全文基本特征之外,国际优秀期刊平台的 XML/HTML 出版还有其特色功能。

(1) 术语语义关联。ScienceDirect 网络出版平台发布的 HTML 全文建立了术语链接,即该平台构建了术语语义本体库,在底层 XML 数据文件中对专业术语进行标引后,点击 HTML 全文中相应的专业术语,便可自动跳转至平台术语语义本体库,便于读者查阅专业术语的权威名词解释及出现过该术语的相关文献资料。

(2) 作者身份标识。RSC 出版社要求每位作者注册用于识别作者身份的开放研究者和贡献者身份识别码(Open Researcher and Contributor identifier, ORCID),并在 HTML 全文中作者姓名处标注了“绿色 ID 标识”,若读者感兴趣可点击该标识链接至 ORCID 解析网站,获取该作者的学习、工作经历及全部文献链接等,有助于提升作者的影响力。

(3) 参考文献链接。ScienceDirect、RSC、PLoS、Science 通过网络出版平台发布的 HTML 全文中的参考文献均与 CrossRef 引文链接系统、Scopus、PubMed、Web of Science 数据库及 Google Scholar 等建立了链接,多数英文参考文献通过数字对象唯一标识符(Digital Object Unique Identifier, DOI)实现了跨平台或跨资源库的链接。

(4) 单篇论文评价数据。PLoS、Science 在线出版的 HTML 文章页面中提供论文浏览量、下载量、被引频次和基于社交网络的 Altmetric 分值等衡量论文质量和影响力的多维指标数据,根据这些定量数据,读者可快速了解所浏览论文的关注度和学术影响力,从而决定是否深入阅读全文。

(5) 增值服务模块。Science 在线出版的 HTML 文章末尾处链接有与论文相关的数据集、音

视频、图表等补充材料,文件格式包括 PDF、Excel、JPG、avi 和 mp4 等,读者点击链接可直接下载或先跳转至相应的数据仓储中再进行免费下载,这些补充材料对于读者理解文章内容和今后开展相关

研究有很大帮助。此外,ScienceDirect 和 Science 网络出版平台在 HTML 文章页面中设置有推荐服务模块,即自动推荐相关研究文献链接,供读者点击阅读。

表 1 国际优秀期刊网络出版平台 XML/HTML 出版的特色功能

出版平台	特色功能
ScienceDirect	①术语链接:点击 HTML 全文中出现的专业术语标签,可链接至术语语义本体库,查阅该术语及与之相关的其他术语的名词解释(来自权威图书),以及曾出现过该术语的相关文献资料; ②推荐服务:自动推荐相关文献链接及引用该文的文献链接; ③参考文献链接:全部参考文献与 CrossRef 引文链接系统和 Scopus 数据库建立了链接; ④作者关联:点击作者姓名会弹出一个窗口,显示作者单位,并链接了 Scopus 数据库中作者发表的文章
RSC	①作者 ORCID 标识链接:点击作者姓名上角标处的“绿色 ID 标识”,可链接至 ORCID 网站,在该网页中可获取该作者的 ResearchID、Scopus Author ID、学习和工作经历,以及已发表的全部文章链接; ②参考文献链接:全部参考文献与 CrossRef 引文链接系统、PubMed、RSC 数据库建立了链接; ③术语链接:部分期刊文章中出现的化学术语和化学结构式与本体库和 ChemSpider 数据库建立了链接
PLoS	①提供论文补充材料:与 Figshare 开放资源平台建立合作,将与文章相关的原始数据集、音视频等材料存储在 Figshare,并建立链接,供读者免费下载; ②论文评价数据:提供论文在线浏览量、下载量、分享量、被引频次等评价数据,支持在线评论; ③参考文献链接:全部参考文献与 Google Scholar、PubMed 数据库建立了链接; ④图片浏览方式:集中展示全部图片,支持用户下载 PPT、PNG、TIFF 格式图片,支持在线缩放图片
Science	①补充材料:提供与文章相关的原始数据集、音视频等材料,以 PDF、Excel、avi、mp4 等附件形式链接在页面上,供读者免费下载; ②推荐服务:自动推荐相关文献链接; ③参考文献链接:全部参考文献与 CrossRef 引文链接系统、PubMed、Google Scholar、Web of Science 数据库建立了链接; ④提供引用工具:可免费下载适用于 BibTex、EndNote、Mendeley 等引用管理器的文献信息; ⑤论文评价数据:提供论文在线浏览量、下载量、Altmetric 等评价数据,支持在线评论

### 3 我国百强中文期刊 XML/HTML 出版现状分析

#### 3.1 XML/HTML 出版概况

对 70 种百强中文期刊网站的 XML/HTML 出版情况调研可知,只有 17 种期刊在网站平台推出了 HTML 全文阅读,分别是《气象学报》《水产学报》《推进技术》《岩石学报》《中国科学院院刊》《中国石油大学学报(自然科学版)》《重庆大学学报(自然科学版)》《草业学报》《地球科学进展》《金属学报》《力学学报》《清华大学学报(自然科学版)》《中国水稻科学》《中国医学科学院学报》《药学报》《中国药理学通报》和《光学学报》,仅占总样本数的 24%,而绝大部分中文期刊只提供 PDF 阅读模式,局限于简单重复纸刊内容,未充分利用和发挥出互联网环境下数字内容资源结构化关联、跨平台链接、多媒体呈现及实时分享的巨大优势。

从上述调研数据可知,相比国际优秀期刊出版社,我国中文期刊在数字出版领域对新技术的关注和应用明显不够,绝大多数期刊偏重于前期的选题策划、组稿和编辑出版工作,而对于后期内容生产加工及传播仍然固守传统模式,满足不了互联网环境

下用户的实际需求。

#### 3.2 HTML 功能建设

HTML 全文即以网页形式呈现文章内容,通过页面内知识关联、跨平台资源链接、添加音视频等方式增强和丰富文章内容,为读者提供便捷的一键获取与文章内容相关的各类知识资源。本研究重点调研了上述 17 种百强中文期刊的 HTML 功能建设情况,从碎片化知识关联、文献资源链接、图表呈现方式、分享推广及论文评价等方面进行详细介绍。

(1) 碎片化知识关联。通过调研发现,17 种百强中文期刊的 HTML 全文碎片化要素及程度基本相同,仅在碎片化知识关联方式上有所不同。总体上,这些期刊文章的碎片化要素主要包括关键词、图表标引处和参考文献标引处,其中关键词、图表标引处与文章结构框架建立了关联,即从文章结构框架中点击关键词或图表标识,便会跳转至相应部位。同样地,图表标引处、参考文献标引处也与相应内容建立了页内关联,即点击正文中图表标引序号,会直接弹出相应图表或跳转至相应图表处;点击参考文献标引序号,会跳转至文末相应文献处,或在右侧“工作空间”弹出文献详情。此外,文中关键词与外部知识资源建立了索引链接,根据各期刊需求不同,

分别与万方数据库、百度学术资源或期刊官方网站建立了链接。

(2) 文献资源链接。绝大多数期刊文章的中文参考文献与中国知网、万方、维普等国内数据库建立了链接,只有个别期刊如《药学报》《地球科学进展》,可能是考虑到涉及版权问题,未建立中文参考文献的链接;英文参考文献通过 DOI 与 CrossRef 引文链接系统、国际知名数据库及期刊网站建立了链接,其中医学类期刊《中国医学科学院学报》所引用的大部分英文文献都有 PubMed 唯一标识码(PubMed Unique Identifier, PMID),与 PubMed 数据库建立了链接。另外,调研发现图书、报告、学位论文等参考文献并未建立链接,是待解决的问题。

(3) 图表独立呈现方式。本次调研的 17 种期刊 HTML 全文的图表独立呈现方式各异,其中 6 种期刊 HTML 全文的图表可点击浏览,但一次只能查看一张图片,读者无法下载原图,表格数据可直接复制,但不能下载;还有 11 种期刊 HTML 全文的图表不仅能点击浏览,还能下载 JPG、PPT 等格式的原图,并可通过左、右方向箭头一次性浏览全部图片,另外,还可以直接下载 CSV 格式的表格数据。XML/HTML 出版提供的多样化图表呈现和保存方式便于读者获取并引用图表数据,有助于促进知识资源的再利用和多途径传播。

(4) 分享推广及论文评价等功能。为提高文章的曝光度和促进文章内容快速广泛传播,部分期刊在 HTML 全文阅读页面增添了一键分享功能,即读者点击“分享”图标便可立即将文章内容分享到社交媒体上,供更多人浏览阅读。此外,还有少部分期刊在网站平台的 HTML 文章页面嵌入了单篇论文评价功能,能够为读者提供论文浏览量、下载量、被引频次和社交媒体分享量等评价数据,但尚处于开发探索阶段,相关功能有待进一步完善。

## 4 国内外 XML/HTML 出版差距分析

与国际优秀出版社网络出版平台相比,我国中文期刊 XML/HTML 出版尚处于起步阶段,许多功能有待进一步开发完善,存在的差距表现在以下几个方面。

(1) XML/HTML 出版仅实现了章节跳转、图表独立呈现、作者信息和参考文献外部链接等基础功能,在文章内容的挖掘深度和结构化程度方面还有待加强。ScienceDirect 网络平台发布的各学科文章

中的专业术语已经实现了语义关联,读者可点击获取专业术语的解释,及与其相关的其他术语和文献资料,显然我国中文期刊在这方面还存在较大差距。

(2) 与国内外数据库、大型搜索引擎未建立深入合作,资源共享受限。国际优秀网络出版平台的参考文献跨平台链接非常完善,这得益于他们与 CrossRef 引文链接系统、PubMed、Google Scholar、Web of Science 等数据库的深入合作。而我国多数中文期刊仅实现了部分文献链接,其原因主要有两点:①中文文献可能未注册 DOI,或 DOI 解析不成功,导致无法通过 DOI 实现文献跨平台链接;②我国文献数据库本身的开放接口限制等问题制约了文献跨库链接。

(3) 缺少基于单篇论文的多维评价指标数据。PLoS 和 Science 出版平台能提供论文浏览量、下载量、分享量、被引频次等评价数据,供读者了解文章的关注度和影响力;而国内多数期刊网站平台只有浏览量和下载量统计,并未嵌入社交网络和索引数据库的分享量、被引频次等评价数据。

(4) 与文章相关的数据集、音视频等辅助材料很少,未深入到科研全过程。PLoS 和 Science 几乎在每篇文章末尾处都链接有相关的原始数据、音视频等材料,直接链接附件或存储在 Figshare 开放共享平台,供读者免费下载,以便开展重复性验证实验或深入了解各研究环节;而国内中文期刊极少要求作者提供论文相关材料,缺少对论文补充材料的规范要求,更没有与面向科研人员的开放研究平台建立合作。

## 5 启示及建议

(1) 在传统媒体与新媒体融合环境下,我国科技期刊工作者应改变过去较为单一的传播理念和思维方式,建立能满足读者碎片化阅读需求的 XML/HTML 出版与传播模式,增强科技期刊的知识服务能力和传播效率。后期制作 XML/HTML 全文的成本较高,导致其在中文期刊的应用较少,鉴于此,我国中文期刊应积极向期刊主管、主办单位反映 XML/HTML 技术在数字内容生产加工和传播中的显著优势,主动争取相应的政策和经费支持。

(2) 在 XML/HTML 出版功能建设方面,我国中文期刊应借鉴国际优秀出版机构的先进做法,加强与国内外数据库、数据仓储、搜索引擎、社交媒体等互联网开放资源平台的合作,为实现跨平台知识资

源链接、数据关联重组和论文评价铺垫基础。

(3) 国际优秀出版平台在 HTML 全文页面添加作者身份标识 (ORCID、ResearchID) 及其链接网页, 便于读者全面了解作者及其合作者的研究成果, 而我国大部分中文期刊对作者身份标识不够重视, 即使有些作者已经注册了 ORCID 或 ResearchID, 但中文期刊 HTML 文章页面并未添加其 ORCID 个人主页链接, 从而在一定程度上影响了作者成果的传播和分享推广。因此, 建议我国中文期刊注重作者身份标识的注册与应用。

(4) 国际优秀出版平台在 HTML 全文页面单独展示与文章相关的数据集、音视频等补充材料, 以丰富和延伸文章内容, 为读者提供更多知识服务。而我国多数中文期刊的 HTML 文章页面并未设立补充材料模块, 也未发布详细的论文补充材料提交标准和规范, 更没有与第三方数据仓储、机构数据库等知识存储与共享平台建立合作, 建议今后加强这方面工作。

## 6 结语

通过对我国百强中文期刊 XML/HTML 出版现状的调研和分析, 并与国际优秀期刊出版平台 XML/HTML 出版现状的对比可知, 我国中文期刊 XML/HTML 出版还有很长的路要走, 必须在构建语义本体库、加强优质资源合作及主动服务科研全过程等方面投入更多的资金和精力, 集中力量推动中文期刊的 XML/HTML 出版, 从而加快提升中文期刊在国际上的影响力和传播力。

尽管我国中文期刊 XML/HTML 出版与国际出版机构存在较大差距, 但相信随着网络出版技术的迅速发展, 以及中文期刊与技术外包公司、国内外数据库、数据仓储等开放资源共享平台合作的不断深入, 中文期刊 XML/HTML 出版功能将得到不断完善, 在不久的将来能为用户提供更多优质的知识服务。

## 参考文献

[1] 武晓耕. 科技期刊发展中的碎片化趋势[J]. 科技与出版, 2015(9): 70-72.

- [2] 陈鹏, 叶宏玉, 梁凯, 等. 移动阅读环境下学术期刊的发展启示[J]. 中国科技期刊研究, 2015, 26(3): 300-304.
- [3] 武小菲. 泛阅读时代富媒体在数字出版中的应用[J]. 出版发行研究, 2014(8): 54-57.
- [4] 张虹. 互联网时代传统行业性学术期刊内容生产方式的创新[J]. 中国编辑, 2017(6): 62-67.
- [5] 祁丽娟. XML 在线排版在科技期刊出版中的应用实践[J]. 科技与出版, 2018(5): 65-69.
- [6] 苏磊, 张玉, 蔡斐. 科技期刊全文阅读模式 PDF 和 HTML 的对比与分析[J]. 编辑学报, 2015, 27(S1): S17-S19.
- [7] 翁彦琴, 李苑, 彭希珺. 英国皇家化学会 (RSC) ——科技期刊语义出版模式的研究[J]. 中国科技期刊研究, 2013, 24(5): 825-829.
- [8] 翁彦琴, 彭希珺. 爱思唯尔 (Elsevier) 语义出版模式研究[J]. 中国科技期刊研究, 2014, 25(10): 1256-1261.
- [9] 史强, 包雅琳, 姜永茂. 国外四大医学期刊富媒体数字内容开发现状及对国内医学期刊的启示[J]. 中国科技期刊研究, 2018, 29(2): 148-152.
- [10] 李小燕, 田欣, 郑军卫, 等. 科技期刊增强出版及实现流程[J]. 中国科技期刊研究, 2018, 29(3): 259-264.
- [11] 李苑. 视频实验期刊 (JoVE) ——视频出版的启示[J]. 中国科技期刊研究, 2014, 25(9): 1157-1161.
- [12] 雷秋雨, 马建玲. 数据期刊的出版模式与发展研究[J]. 图书与情报, 2015(1): 112-116.
- [13] 李鹏, 卜延明, 夏爽, 等. 视频技术在中华医学会系列期刊中的应用[J]. 中国科技期刊研究, 2016, 27(3): 292-295.
- [14] 杨郁霞. Rich HTML 在科技期刊出版中的应用与思考[J]. 编辑之友, 2018(1): 86-88.
- [15] 张正峰, 刘少华. 《地球物理学进展》网络平台建设和数字出版新进展[J]. 地球物理学进展, 2015, 30(6): 2963-2974.
- [16] 边钊, 韩向娣, 闫珺. “遥感学报”HTML 全文出版模式实践[J]. 编辑学报, 2016(S2): S53-S55.
- [17] 王玥, 刘谦, 范晨芳. 内嵌 PDF 文档在科技期刊网络发布平台全文展示中的应用[J]. 中国科技期刊研究, 2013, 24(1): 136-140.

## 作者贡献声明:

周小玲: 设计研究思路, 撰写和修订论文;

侯春梅: 指导选题和审改论文;

黄爱华, 迟秀丽, 李 慧: 调研数据和修改论文。

## Investigation and analysis on XML/HTML publishing status of top100 Chinese scientific journals

ZHOU Xiaoling<sup>1)</sup>, HOU Chunmei<sup>2)</sup>, HUANG Aihua<sup>1)</sup>, CHI Xiuli<sup>2)</sup>, LI Hui<sup>1)</sup>

1) Editorial Office of *Gold Science and Technology*, Lanzhou Information Center, Chinese Academy of Sciences, 8 Middle Tianshui Road, Chengguan District, Lanzhou 730000, China

2) Editorial Office of *Advances in Earth Science*, Lanzhou Information Center, Chinese Academy of Sciences, 8 Middle Tianshui Road, Chengguan District, Lanzhou 730000, China

**Abstract:** [Purposes] This paper aims to comprehensively study the status and function of XML/HTML publishing of excellent scientific journals at home and abroad, and provide reference for the application of XML/HTML publishing in Chinese scientific journals. [Methods] Taking the Third Top100 Chinese Scientific Journals as samples (70 valid samples), we analyzed the status and function construction of XML/HTML publishing for representative Chinese excellent journals by using network investigation method. [Findings] Only 17 Chinese journal websites have launched full-text XML/HTML reading, accounting for only 24% of the total samples, which reflects that Chinese journals do not give enough attention to XML/HTML publishing. The content of XML/HTML published in Chinese journals is low in structuralization, single in knowledge chain, and lack of multi-dimensional paper evaluation data and supplementary materials, and there is a big gap with international excellent publishing institutions. [Conclusions] XML/HTML publishing is a necessary way for scientific journals to achieve content extension, knowledge chain, and knowledge reorganization. It can satisfy the fragmented reading needs and help to significantly improve the display and dissemination of articles, which should be given enough attention by domestic journals.

**Keywords:** XML/HTML publishing; Content structuralization; Knowledge chain; Publishing technology; Webwork publishing platform; Chinese journal

(本文责编:刘晶晶)