

DOI: 10.13718/j.cnki.xdzk.2021.10.023

基于 XML 结构化数据的期刊信息化建设路径探析

——以西南大学期刊社为例

崔玉洁, 文娟, 包颖, 周雨蕾

西南大学期刊社, 重庆 400715

摘要: XML 作为一种描述数据和数据结构的语言, 已经逐渐应用到期刊的出版流程中, XML 结构化数据是一种全新的期刊展示形态, 是期刊实现全文检索的必经之路。文章对比分析了国内 XML 结构化排版的现状, 通过搭建集约化期刊管理系统和结构化数据网刊发布平台、实现 HTML 全文的 XML 结构化展示, 以及融合 XML 结构化数据的 AMiner 学术推广等方式, 实现期刊内容生产、流程管理、发布模式、传播模式的创新, 提升了期刊的数字出版水平, 为期刊信息化建设服务。

关键词: XML; 结构化排版; 网刊发布; 集约化; 学术推广

中图分类号: TP391 **文献标志码:** A

文章编号: 1673-9868(2021)10-0192-07

开放科学(资源服务)标识码(OSID):



Analysis of the Pathways for Periodical Informatization

Construction Based on XML Structured Data

——Taking Journal Press of Southwest University as an Example

CUI Yujie, WEN Juan, BAO Ying, ZHOU Yulei

Journal Press of Southwest University, Chongqing 400715, China

Abstract: As a language to describe data and data structure, XML (extensible markup language) has been gradually applied to the publishing process of periodicals. XML structured data is a new form of periodical display, which is the only way for periodicals to realize full-text retrieval. This paper compares and analyzes the current situation of XML structured typesetting in China. By building an intensive journal management system and a structured data online publishing platform, realizing XML structured display of HTML full text, and integrating AMiner academic promotion of XML structured data, innovations of journal internal production, process management, publishing mode and communication mode are realized, and the

收稿日期: 2020-12-01

基金项目: 教育部高等教育司 2020 年产学研合作协同育人项目(202002325025); 重庆市教育科学“十四五”规划 2021 年度课题(2021-GX-289); 中央高校基本业务费专项资金资助重点项目(SWU1809117)。

作者简介: 崔玉洁, 硕士, 副编审, 主要从事数字出版、计算机应用等方面的研究。

digital publishing level of journals is improved, thus contributing to the construction of periodical informatization.

Key words: XML(extensible markup language); structured typesetting; online publication; intensive; academic promotion

随着科技的进步, 按需出版、在线发布不断促使传统期刊向数字出版、数据出版转型, XML 作为一种描述数据和数据结构的语言, 也逐渐应用到期刊的出版流程中. 在国外, 基于 XML 的研究已有 20 多年, 近年来国内基于此方面的研究也在逐渐增加, 越来越多的学者开始关注 XML 结构化排版^[1-2]. 如陈肖玉等^[3]在基于 XML 的科技期刊排版技术探究中, 根据实践摸索出一套基于 XML 的相对高效的跨媒体出版流程. 刘冰等^[4]提出我国科技期刊应尽快实现基于结构化排版的生产流程再造. 谭雪静^[5]指出我们需要利用 XML 技术来完善编排校一体化出版模式, 文章重点介绍了一体化出版流程的优势, 既能提高工作效率, 又能实现全流程的数字化出版, 同时还可以实现多渠道宣传. 霍振响等^[6]撰写的基于用户角度的科技期刊数字化出版战略中, 提出了建设 XML 一体化出版流程和创办纯数字化期刊的策略. 本文以西南大学期刊社信息化建设为研究对象, 探索了适合高校期刊信息化建设的转型路径, 以期为期刊社信息化建设提出可行性的建议.

1 XML 与 HTML

XML(Extensible Markup Language)是一种可扩展的标记语言, 可扩展是相对于我们接触的 HTML(超文本标记语言)而言的, 网页编辑时可以使用 HTML 标签都是预先规定好的, 不能根据个人喜好随意更改或者添加规定之外的标签, 而 XML 是一种可以自定义标签的标记语言; 另外 HTML 是人为设计的, 即使不用浏览器查看 HTML 的内容, 人们也能大概知道其要表达的意思. 遗憾的是机器不能识别, 尽管这个文档标记告诉浏览器如何显示该信息, 但是标记并没有告诉浏览器信息是什么, 为了能让机器识别, 这就需要使用 XML^[7]. 并且如果需要在 HTML 中显示动态数据, 那么当数据变化时将需要花费大量的时间来修改此 HTML 文档, 而 XML 可以优化此不足, 因为 XML 中数据是独立存在的, 可以通过几行 JavaScript 代码, 读取一个外部的 XML 文件, 从而更新 HTML 中的数据内容. 由此可见, XML 的创建主要是用来克服 HTML 的局限性, 就像 HTML 的作用是显示数据, XML 的作用是便于不同应用、不同平台之间的数据共享和传输, 也可以说 XML 是为 Web 而设计的, 其本身不会做任何事情, 它仅仅是包装在 XML 标签中纯粹的信息, 人们需要编写软件或者程序, 才能传送、接收和显示这个文档^[8].

2 XML 结构化排版

常用的 XML 排版、转换工具有以下 4 种: XSL Formatter 可视化工具、SXparser-XML 效验器、XML Editor 编辑器、XSL Template Designer-XSL 模板设计软件. XSL Formatter 可视化工具是用来对 XML 和 XSL 进行排版的软件, 在报纸、期刊、电子图书排版中应用广泛, 另外该工具可以将 XML 内容通过多版面显示, 在无 PDF 打印软件的环境下可以直接将排版文件转换成 PDF 文件保存, 还可以在 SVG 中输出排版文件. SXparser-XML 效验器是一种检验 XML 格式的工具, 可以对 XML 文件中的 SGML 文件及 DTD 进行检验. XML Editor 是一款专门用来编辑 XML 文档的编辑器, XML 编辑器通过智能化的操作界面, 使其比以往的编辑器更加有效, 会强制使用者创建合法的 XML 结构, 还可以自动对 XML 语法的代码标注对应的颜色. 另外编辑器还提供预设和可配置方案, 只需单击就可以生成相同源的 PDF, Epub, HTML 和其他格式文件, 扩展性强, 支持并增强 XML 编辑器的内置 XML 发布框架, 甚至可以创建自己的框架. XSL Template Designer 是一款 XSL 模板设计软件, 该软件主要用来完成 XML 数据的输出布局, 可以轻松地实现复杂格式和公式的布局.

目前国内期刊界常用的 XML 排版来自 3 家公司: 中国知网、北京仁和汇智信息技术有限公司(以下简称仁和)以及北京北大方正电子有限公司(以下简称北大方正), 但是这 3 家公司的 XML 排版又各不相同。中国知网采用方正书版排版(因中国知网目前只有基于方正书版的排版, 故一体化排版及其延伸功能不做比较说明); 仁和主要采用的是自主研发的 XML 一体化排版技术, 技术优势有以下几点: ①编辑只需将原稿发给对方, 仁和在较短的时间内就会将该文章的 XML 文件返回^[9-11]; ②仁和自主研发的参考文献校对软件可对文章中的参考文献进行自动检测修改, 正确率高; ③XML 一体化排版中包含制图, 编辑在后期的修改过程中可以自行在平台上修改, 也可以将修改稿拍照传给仁和由其辅助修改, 该方式对那些无专门排版人员的单位适用性强; ④可实现作者微信实名关联, 实现文章状态及时提醒, 自动给编辑和作者发送文章排版的进度, 以及文章网刊发布后的点击量和下载量等。北大方正主要采用的是升级版的飞翔软件, 支持对 word, Excel 等文件的直接排版, 将原稿上传到系统后平台会自动将其转化成 XML, HTML 等格式的文件。北大方正排版的优势是排版速度快, 文章上传后即可获得转换后的 XML 文件, 2 种 XML 排版详细对比信息见表 1。

表 1 2 种 XML 排版对比表

	北大方正学术出版云服务平台	仁和 XML 一体化排版
缩短文章发表周期	有	有
促进扩展交流	平台可实现高质量 PDF, RichHTML, H5, 全文 XML 等多种格式文件的同步生成, 适用于传统印刷、网刊发布、移动阅读等多种场景, 并兼容第三方运营平台, 对接国内外数据库。	可以输出包含各类第三方数据平台要求的 XML 文件, 如 PubMed, PMC, DOAJ 等。
以 XML 为基础的网刊发布系统	无	有, 网刊发布系统较其他网站的优点: ①增加了被引文章的情况(施引文献); ②增加了本期刊发表过的相关文章的情况(相关文章)。
微信精准融合	排版有带微信版式的格式, 直接复制粘贴到微信公众号的编辑器中稍作调整可以发送全文, 但无法做到与作者直接微信交流文章情况。	有, 与微信平台自动对接, 可以自动给编者及作者发送文章排版的进度; 给作者发送个人论文网络版发布后 1 个月内的点击量及下载量; 可以自动给作者做精准推送, 给作者推送与之相关文章; 同时生产管理流程与微信平台自动对接, 可以时时在微信平台上展现论文最新录用状态或是当期状态, 并且可以自动给相关人群做推送。
XML 排版方式	升级版飞翔, 支持对 word, Excel 直接进行排版, 图由方正工作人员参与制作, 无需专门的排版人员参与, 适应于无专门排版员的单位。	XML 结构化排版, 图由仁和工作人员参与制作, 无需专门的排版人员参与, 适应于无专门排版员的单位。
智能校对	方正智能辅助审校系统包括方正智能辅助审校客户端和方正书畅智能辅助编排云服务平台, 可以做逻辑检查、上下文查重、公式识别、图像管理等辅助审校, 操作简单, 上手容易。	嵌入在一体化融合平台中, 可以实现全文格式、错别字与敏感词等的自动标识与校验, 但系统没法单独使用, 只有基于 XML 排版的期刊才可以使用。
收费情况	平台免费, 按页收制作费	平台收费, 另再按页收制作费
参考文献校对	无自主软件, 通过与善锋软件合作进行参考文献校对, 还需编辑辨别。	自主研发的参考文献校对软件, 有专利, 按条收费, 正确率高, 编辑只需通看 1 遍即可。

3 西南大学期刊社信息化建设整体实施方案

3.1 搭建集约化期刊管理系统

集约化期刊管理是现代期刊社提高效率与效益的方式之一。所谓集约化是指集合各个生产环节的人力、物力、财力、管理等要素,进行统一资源配置。同时以节俭、高效为价值取向,进而集中期刊社核心力量,保持其良好的竞争优势。任何一个单位要生存和发展,要高效地把内部活动有机地组织起来,就必须建立与自身工作相适应的管理系统^[12]。西南大学期刊社在信息化建设过程中,充分认识到延伸传播空间的必要性,敏锐地把握期刊信息化的发展方向,密切结合自身工作需要,以集约化的建设思路成功搭建了新一代的集约化期刊管理平台,将传统的期刊管理从静态转变为动态,从事后管理转变为全程管理,使得期刊信息化建设实现了全流程监督和各阶段逐层推进的顺利融合,为期刊社的信息化建设提供了良好保障。

西南大学期刊社的集约化期刊管理系统分为 5 个模块:定稿模块、编辑模块、付印管理模块、发行模块以及财务管理模块(图 1)。期刊集约化管理系统是对期刊社采用的原始稿件资源和数字化加工过程各类资源进行统一存储和管理的一种集约化平台,该系统不仅可以展示期刊出版的最终产品形式,还可以详细地记录稿件的处理状态,如排版文件、封面设计、校稿记录等。图 2 为定稿阶段定稿目录在该系统上的展示页面,定稿目录不仅展示了稿件的基本信息,还标注了稿件的状态,同时责任编辑还可以根据文章的情况动态调整栏目顺序。

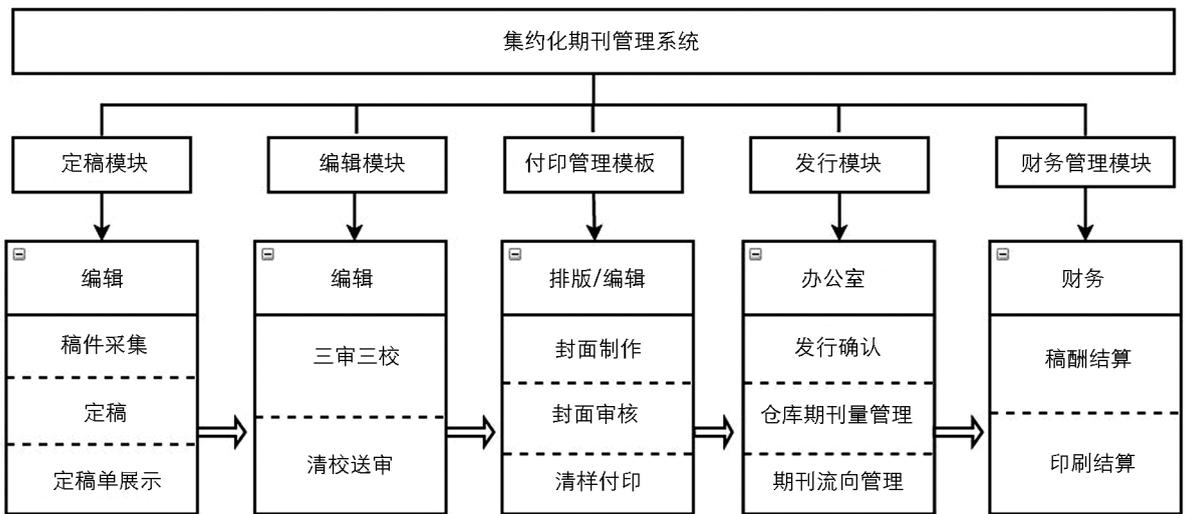


图 1 期刊集约化管理系统模块图

稿件编号	稿件题目	所属编辑	基金等级	作者	版面	状态	当前序号	稿酬数据	操作
应用研究(14篇)									
X20190469	收入分配与经济增长关系的SEM分析	夏靖	国家级	邵春玲	7	稿件定向发表	1		🔍 📄 🔄 🗑️
X20190377	光周期对艾角藤生长发育及繁殖的影响	王新清	其它	陈文龙	6	稿件定向发表	2		🔍 📄 🔄 🗑️
X20190578	佛山市城市森林生态系统服务价值核算研究	包颖	省部级	朱立安	7	稿件定向发表	3	yes	🔍 📄 🔄 🗑️
X20190400	雨量和坡度对新鲜岩石工程边坡侵蚀特征的影响	包颖	国家级	李伟	10	稿件定向发表	4	yes	🔍 📄 🔄 🗑️
X20180430	基于AHP的寿险影响因素评价分析	包颖	国家级		7	稿件定向发表	5	yes	🔍 📄 🔄 🗑️
X20190531	自适应网络防御的联合云计算框架	夏靖	省部级	彭劲松	5	稿件定向发表	6		🔍 📄 🔄 🗑️
X20180514	基于HOG-CSLBP与深度学习的人脸年龄识别算法	夏靖	省部级	胡若华	8	稿件定向发表	7		🔍 📄 🔄 🗑️
X20190605	基于小波去噪和岭回归的脱脂期鸡蛋蛋氨酸检测模型构建	夏靖	国家级	曲伟男	8	稿件定向发表	8		🔍 📄 🔄 🗑️
X20190289	通讯地址: 湖北省武汉市洪山区雄楚大道918号长江工业学院 李晋霞电话: 18186872121邮箱: 587194@qq.com	夏靖	国家级	李晋霞	5	稿件定向发表	9		🔍 📄 🔄 🗑️
20170295	小微企业信用等级分类模型研究	夏靖	国家级	李强	6	稿件定向发表	10		🔍 📄 🔄 🗑️
X20190465	商品价格的相对价值及配置策略研究——以马尔科夫链转移模型为例	夏靖	省部级	周亮	10	稿件定向发表	11		🔍 📄 🔄 🗑️
X20180919	网络能力对企业创新影响作用的实证研究	潘燕燕	其它	刘方润亚	7	稿件定向发表	12		🔍 📄 🔄 🗑️

图 2 期刊集约化管理系统界面图

依托期刊集约化管理系统, 期刊社实现了出版资源的结构化、数字化全生命周期的管理, 将传统的办公模式与新型管理模式进行动态组合, 各项工作进度都以无纸化办公形式存在, 同时加强了纸刊的管理, 详细地记录了纸刊的流向, 使得管理人员能够更好地进行期刊全流程的管理。

3.2 搭建结构化数据网刊发布平台

以 XML 为基础的结构化网刊发布系统是一种变革, 在真正意义上实现了全文检索。基于 XML 的结构化网刊发布系统, 使得平台上文章的网络传播效果得到了极大的提升。平台上所有的数据以结构化的形式存放, 真正意义上实现了元数据检索, 这种检索除了可以检索标题、作者、单位、关键词、摘要外, 还可以实现文章中任意章节、段落、图、表等的自动定位。网刊发布平台相比常规网站实现了搜索引擎的 SEO 优化, 极大地增加了网站搜索的显示度, 使得网站的浏览量实现指数级增长。网刊发布系统还会向期刊社提供流量监控报告, 此报告会显示出在一定周期内, 文章新增的被引量、科研机构访问量、文章访问及搜索热词的监控报告等信息。网刊发布平台还有另外的两大优点是: ① 过刊查询的目录中新增加了“施引文献”这一标识, 该字段可以自动统计文章在全网的被引量及详细信息, 施引文献信息在目录及文章详细信息页面都会有展示(图 3); ② 文章的详细信息页面还增加了同类文章推荐字段, 该字段会自动统计本刊所有发表过的同类型文章, 并逐条显示, 该功能极大地方便了作者, 缩短了其在搜索同类型文章中所耗费的时间(图 4)。

The screenshot shows the website interface for the Journal of Southwest University Social Science Edition. The top navigation bar includes links for '投稿主页', '首页', '期刊简介', '期刊在线', '编委会', '期刊荣誉', '作者指南', '留言板', and '在线订阅'. The main content area is divided into two sections: '投审稿系统' (Manuscript Submission System) on the left and '引用排行' (Citation Rankings) on the right. The '引用排行' section displays a list of articles with their citation counts for various years (2020, 2019, 2018, 2017, 2016). The first article is '精准扶贫视角下农村产业化扶贫政策执行逻辑的探讨--以Y村大棚蔬菜产业扶贫为例' by Li Bo, published in 2016, with 445 citations. The second article is '产业扶贫与扶贫“产业化”——基于广西产业扶贫的案例研究' by Xing Chengjun, published in 2017, with 57 citations. The third article is '资本下乡的政治逻辑与治理逻辑' by Wang Haijun, published in 2015, with 48 citations. Each article entry includes a '施引文献' (Cited Literature) icon and a count.

图 3 引用排行

3.3 HTML 全文的 XML 结构化展示

西南大学期刊社的《西南大学学报(自然科学版)》《西南大学学报(社会科学版)》《西南师范大学学报(自然科学版)》《教师教育学报》4 本刊物在期刊社官网和投稿系统首页开通了 HTML 全文在线阅读, 此全文展示与传统的 PDF 文件相比具有多种优势: ① 读者阅读体验得到提高, 并且不需要安装额外的阅读软件; ② 内容比 PDF 文件丰富, 在制作过程中可以通过添加二维码的方式实现内容的增强出版; ③ 网页

中可以通过添加超链接的方式实现与其他资源的互通互联; ④ 网页中的图、表可以单独呈现, 提高了读者的阅读感受; ⑤ HTML 文件中全文的参考文献都相互对应, 可以随时查找参考文献与原文的联系, 并可对其进行延伸阅读等。但是, 如果网站不是基于结构化的网刊发布系统, HTML 文件就只是网站上的一个链接, 无法实现全文的结构化检索, 另外, 在提高点击率和引用概率上面所起的作用不明显, 无法在其他网站上提高论文的显示度。基于此, 西南大学期刊社自 2020 年 11 月起开始搭建结构化网刊发布平台, 使得平台上的 HTML 以 XML 结构化的形式展示, 不仅可以实现碎片化阅读, 还可以实现全文的元数据检索, 同时网站的访问量也达到了指数级提升。

青少年社会支持与问题行为的关系: 心理素质和自尊的链式中介作用

张大均, 朱政光, 刘广增, 李阳

西南大学 心理学部、心理健康教育研究中心, 重庆市 400715

+ 详细信息

The Relationship Between Social Support and Problem Behaviors of Adolescents: The Chain Mediating Role of Psychological Suzhi and Self-Esteem

Dajun ZHANG, Zhengguang ZHU, Guangzeng LIU, Yang LI

+ More Information

HTML全文

- 摘要
- HTML全文
- 图(1)表(2)
- 参考文献(30)
- 相关文章
- 施引文献(15)
- 资源附件(0)
- 访问统计

一. 问题提出

在当前社会竞争不断加剧的大背景下, 青少年学生表现出对社会环境的适应不良, 心理冲突不断加剧, 由此产生了一系列的行为问题。因此, 掌握青少年的心理健康现状, 了解这一群体的心理需求和困难, 从而有针对性地做出应对, 对青少年的健康成长至关重要。

问题行为为个体身上出现的妨碍其社会适应的异常行为, 既涉及自身的情绪异常等, 也包括对他人和社会产生负面影响的行为。针对青少年群体, 心理学家林格伦认为青少年问题行为囊括了所有引起麻烦的行为, 也可以认为是该行为所产生的麻烦^[1]。国内关于问题行为研究主要有两种倾向: 一种是研究某一具体的问题行为(如言语侵犯), 另一种是将各类别问题行

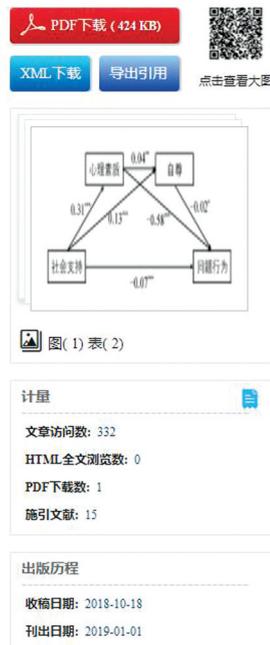
图 4 完全以 XML 为基础的结构化网站

3.4 融合 XML 结构化数据的 AMiner 学术推广

AMiner 平台是由清华大学计算机系研发的, 拥有我国完全的自主知识产权, 该平台的最主要优势是: ① 期刊推荐: 平台根据期刊的领域内容, 精准挖掘相关的作者, 并将期刊推送内容以邮件的形式发给作者; ② 编委推荐: 平台根据期刊指定的领域、H 指数范围、区域、语言, 在全球范围内推荐适合的、适量的编委名单; ③ 会议推荐: 根据会议内容的关键词, 精确找到与之相关或者交叉学科的参会者, 并通过邮件的形式向其发送会议通知; ④ 辅助期刊选题策划: 系统后台根据大数据的学科热点分析、交叉分析、前沿分析等, 帮助期刊完成下一年的选题策划。目前, 西南大学期刊社分管的 4 本学术刊物已经加入到 AMiner 平台中, 并开始每期主推文章的精准投送(图 5 为西南大学期刊社《教师教育学报》2019 年 6 期推荐文章的制作及推送展示), 读者点击文章题目将跳转到网刊发布平台的 HTML 页面, 这也将 AMiner 平台与 XML 网刊发布平台实现了融合, 期望此融合路径会在未来对提升期刊影响力起到积极的作用。

4 后期工作规划

为了更好地适应期刊移动端阅读的习惯以及数字出版和大数据发掘等的需求, 同时满足期刊的按需出版、优先出版、在线发布和加入全文数据库实现文献关联等要求, 期刊社需要不断地改变传统的出版方式, XML 结构化排版就是一种很好的方式。XML 结构化排版采用的是在线排版, 该排版方式满足排版人员、编辑、作者、校对人员等多角色的在线校对和修改功能, 文章从录用并上传到系统后采用全流程的结构化排版, 编辑直接在线修改稿件, 以所见即所得的方式提高了修改的效率, 编辑也可以采用传统的纸质稿件修改, 由排版人员或者仁和工作人员辅助在线修改稿件。该一体化排版还支持作者在线修改稿件, 流程中



编辑和作者的修改痕迹都会保留,方便回溯。XML 结构化排版是期刊满足现代信息化发展的必经之路。期刊社已开始打造集成化的网刊发布系统,系统布置完成后将首先在网站上展示全部文章的元数据信息,该发布平台也是 XML 结构化排版文件的展示窗口,后期加入 XML 一体化排版后,可以直接实现稿件信息在网站系统上的一键发布。



图 5 AMiner 推荐文章的制作过程及展示

参考文献:

- [1] 白杰, 杨爱臣. XML 结构化数字出版的特点与流程 [J]. 出版广角, 2015(5): 28-31.
- [2] 张光, 白雨虹, 刘文武. 利用 XML 技术提高期刊影响力的探索 [J]. 中国科技期刊研究, 2017, 28(6): 565-569.
- [3] 陈肖玉, 虞子治. 基于 XML 的科技期刊排版技术探究 [J]. 编辑学报, 2012, 24(Sup2): 57-60.
- [4] 刘冰, 游苏宁. 我国科技期刊应尽快实现基于结构化排版的生产流程再造 [J]. 编辑学报, 2010, 22(3): 262-266.
- [5] 谭雪静. 利用 XML 技术完善编排校一体化出版模式 [J]. 中国科技期刊研究, 2018, 29(7): 711-714.
- [6] 霍振响, 屈李纯, 常莹华, 等. 基于用户角度的科技期刊数字化出版策略 [J]. 农业图书情报学刊, 2018, 30(10): 128-131.
- [7] 陈焕东, 吴淑雷. 基于 Web 的数据挖掘技术浅析 [J]. 海南师范学院学报(自然科学版), 2004, 17(3): 243-245.
- [8] 周育洋. PDM 中面向信息集成的 BOM 多视图映射技术研究 [D]. 武汉: 武汉理工大学, 2010.
- [9] 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, 92.
- [10] 蒋晓, 谢暄, 叶芳, 等. 基于 XML 的科技期刊一体化数字出版流程 [J]. 传播与版权, 2018(1): 104-107.
- [11] 崔轶, 包雅琳, 姜永茂. 媒介融合背景下医学科技期刊的全媒体出版方案实践探究 [J]. 中国科技期刊研究, 2017, 28(6): 521-525.
- [12] 刘勇. 文印店管理信息系统的开发与应用 [J]. 江苏教育学院学报(自然科学版), 2013, 29(4): 19-21.