

国际研究者联系信息的批量采集和电子函件群发方法

张 明 明

海南医学院杂志社, 571199, 海口

摘 要 研究者是科技期刊的作者、审稿人、编委和主要读者。为了高效联系全球海量研究者,提出了3种利用期刊数字出版资源结合计算机工具软件快速搜集国际研究者姓名、邮箱和机构等联系信息的方法,即 PubMed 文献 XML 文件提取、商品化网站采集工具和定制 Java 采集程序。比较了常用的函件群发方法。实践证明,3种采集方法实用、高效、易操作,极大地节约了劳动力。函件群发平台是目前比较好的函件群发方法,可用于期刊电子函件营销;因此,通过计算机工具软件批量提取研究者联系信息并通过函件群发平台发送函件,极大地提高了与国际研究者沟通的效率。

关键词 研究者;邮箱;采集;函件群发;作者;科技期刊

Batch collection of contact information of international researchers and methods for Email group sending // ZHANG Mingming

Abstract This paper explores how to increase the efficiency to get in touch with massive researchers all over the world. These researchers are important to sci-tech journals since these researchers are potential authors, reviewers, editorial members and audiences. Three methods, combining digital periodical resources with computer tool software, are presented to collect researcher contact information including name, Email address and affiliation. These methods include extracting PubMed search result files in the format of XML, commercial web data collection tools, and customized collection programs in Java. Moreover, the paper compares the commonly used methods for Email group sending. As confirmed by practices, these three collection methods are useful, efficient, easy to operate and labor-saving. In addition, Email group sending system platform is a better way for journal Email marketing. Therefore, obtaining researcher contact information in large quantities with computer tool software and sending E-mail via Email group sending systems greatly help editors to communicate with international researchers more efficiently.

Keywords researcher; Email; collection; Email group sending; author; sci-tech periodical

Author's address Editorial Department, Hainan Medical University, 571199, Haikou, China

DOI:10.16811/j.cnki.1001-4314.2016.03.018

与国际研究者进行积极主动的沟通对科技期刊获得国际化稿源、建立国际化团队和实现全球范围的主动发行,尤其对缺少学术优势和专家资源却计划创办和发展英文科技期刊的出版单位有重要的意义。多家国内英文科技期刊已经实现了稿源和编委的国际化^[1-4]。

电子函件方便、高效、经济、可随时随地阅读,是多数编辑部对外联系最常用的通信方式,也被证实是一种联系已发表文章作者的有效方式^[5]。在韩国,电子函件也是专业技术人员最倾向使用的通信方式^[6]。随着数字出版的发展,国际出版商、数据库、索引、搜索引擎等提供了大量的研究者邮箱信息^[7];然而,如何从海量互联网资源中快速获取研究者的电子函件地址等联系信息并为之联系成为一项重要的技术。

信息采集终端技术和数据挖掘技术的应用,大大提高了信息采集的效率,推动了期刊编辑出版工作的变革^[8]。采集和处理后的期刊著录信息等元数据已经用于期刊的共享集成和数字化图书馆建设^[9-10],然而,多数传统编辑掌握这些信息技术有较大的难度。目前,尚未见到将采集技术用于高效收集研究者联系信息的报道。基于笔者的工作实践,本文结合相关背景知识介绍3种利用现有的数字出版资源,结合计算机工具软件,快速搜集研究者联系信息(姓名、Email和机构)的方法,并介绍函件群发平台的应用,以期在传统编辑提供实用、高效、易操作的与国际研究者进行沟通的工作方法。

1 研究者联系信息的快速搜集

1.1 通过 PubMed 批量获取研究者的联系信息

PubMed 提供了文献的著录项目,检索全面、准确、高效、免费。下面以《Human Reproduction (Oxford, England)》期刊为例介绍如何快速获取 PubMed 文献的作者信息。

1) 导出 XML 文件。XML 文档结构严谨,层次分明,语义更加明确^[11],是当前处理结构化文档信息的有力工具。检索《Human reproduction (Oxford, England)》[Journal],点击“send to”,选择“File”中 XML 文件格式输出 pubmed_result.XML 文件。除 XML 之外的其他格式文件或缺失了作者信息,或含有作者信息,但需更多后续处理。

2) XML 文件转换为 CSV 文件。使用格式转换工具,将 XML 文件转换为 CSV 文件。转换后,生成了多个著录项目的 CSV 文件(图1)。尽管可用 Excel 以“只读”方式打开 XML 文件,但由于该文件记录了多项信息,在文献数量较多时,表格通常比较庞大且难于处理。

接,最后内容页显示具体内容;因此,可以通过网站采集工具获取这些电子资源中的研究者联系信息。国内比较有名的采集工具有火车头、海纳、云采集、ET、三人行、狂人、八爪鱼等。其中,八爪鱼采集器使用了人性化的拖拽式采集流程,功能实用,操作简单,不需要写规则,适合计算机菜鸟。下面以使用该采集器在 ScienceDirect 平台采集《Immunity》期刊 2013 年至今的作者联系信息为例,介绍网站采集工具在研究者信息资源收集中的应用。

八爪鱼的采集流程见图 4,其流程设计界面和活

动界面分别见图 5 和图 6。其中,设计采集流程环节本质上是制定采集规则,也是采集中最重要的环节。值得注意的是,由于通信作者的位次不固定,需要首先通过 Xpath 定位人头像元素,再通过 Xpath 定位需要采集的作者姓名、邮箱和机构。XPath 即 XML 路径语言,是基于 XML 的树状结构用来确定 XML 文档中某部分位置的语言,提供了在数据结构树中找寻节点的能力。不熟悉 Xpath 的采集者可以用火狐浏览器提供的 FirePath 附加组件用于核查 Xpath 路径的唯一性,从而定位元素和数据的位置。

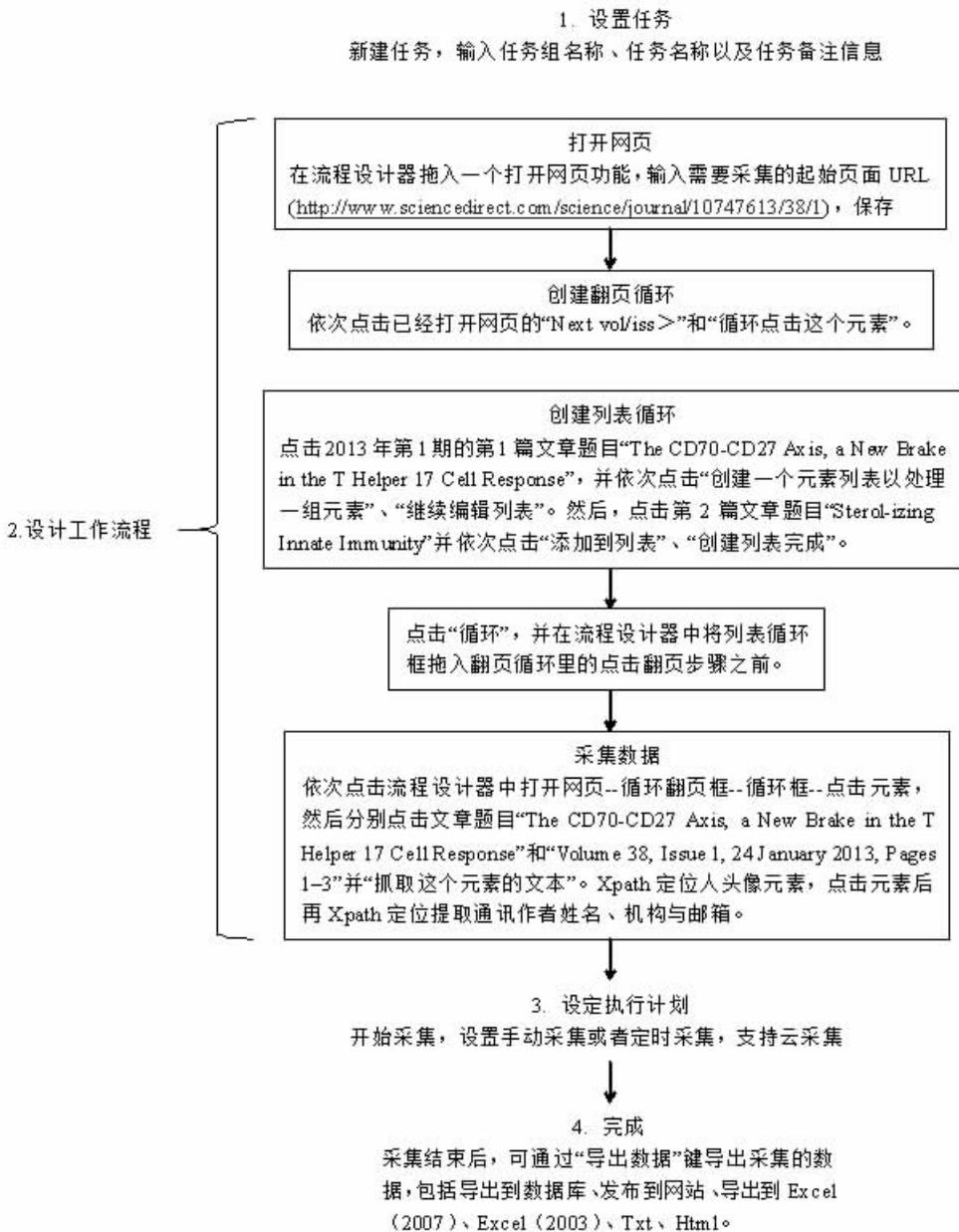


图 4 八爪鱼采集器的采集流程(以 Elsevier 期刊《Immunity》为例)

与从 PubMed 获取研究者联系信息相比,网站采集工具更加灵活和全面,可采集单篇文献的多项著录信息。实践中,人工复制粘贴速度约为 100 条/h,而采

集工具单任务采集速度通常可达 150 条/h,多任务同时运行可加快采集速度;然而,不排除某些网站可能采取防采集策略,以致无法使用采集工具。



图5 八爪鱼采集器的工作流程设计界面
(以 Elsevier 期刊《Immunity》为例)



图6 八爪鱼采集器的采集活动界面
(以 Elsevier 期刊《Immunity》为例)

1.3 定制采集程序 为追求更快的采集速度,开发相应的采集程序不失为另一种尝试。特请某软件开发工程师用 Java 语言编写了名为 Springer Collect 的程序。经过多次完善和实践检验,该程序实用性强,操作简单,只需要输入期刊名称和检索时间段,便可实现对 SpringerLink 平台研究者联系信息的快速采集,在 10 多 min 内可采集上千条数据,采集速度远快于八爪鱼采集器;然而,目前该程序不能采集非英语刊名的期刊,其成熟度、严谨性和稳定性都有待进一步考验和完善。此外,与商品化的网站采集工具相比,如遇网站调整、需求变更等采集过程中的变动因素,定制采集程序的应用可能需要计算机专业人员的后续技术支持。

2 函件群发平台的应用

科技期刊的网络营销方式多样,包括数字化的出版平台、功能强大的网站、博客、搜索引擎广告、社交网络、网上新闻发布、电子函件等。其中,电子函件营销是一种便捷、高效、经济和常用的网络营销方式^[13],免费的函件订阅对科技数字期刊的推广产生了非常大的影响^[14]。国内已经有期刊尝试电子函件营销,取得了不错的效果^[15]。

收集研究者联系信息后,通过邮箱推送信息给研

究者成为电子函件营销的下一个环节。传统发送函件的方式费时费力,短期大量发送函件容易导致邮箱被封。实践中,尝试利用 Word 的信件模板与 Outlook 相结合群发一对一函件,仍然不能解决这些问题。目前,也有很多函件群发软件,绝大多数是通过自己注册或者软件注册多个发件箱来实现群发的,所利用的是第三方邮箱服务商的服务器,软件本身不具备函件投递功能,本质上依然是客户端,发送数量过多或太频繁,会导致被禁号或 IP 被拉入黑名单。此外,发送日志以及是否发送成功等信息会被反馈到发件服务器上,无法统计函件送达率和收集用户行为数据。

函件群发平台,例如 WebPower(荷兰)、U-Mail 函件群发平台等,利用自身服务器群组、网络带宽资源、IP 资源等解决了函件群发软件的不足等问题。通过函件群发平台,可以实现高效、稳定、精准和安全的函件送达和自定义变量一对一群发,支持函件订阅、退订以及许可式函件营销,对群发函件的订阅和反馈进行全面监测和数据分析,也有助于发现函件营销中存在的问题和了解期刊的国际关注度;然而,电子函件营销也有不少难点和应注意的问题^[16-19],如何做好电子函件营销是另一门学问。例如,细化收件群体,尽可能向研究者的工作邮箱发信^[5],避免向同一机构的研究者同时发信,以期刊官方网址作为后缀的邮箱或通过平台申请的虚拟邮箱作为发件邮箱。

3 结束语

英文科技期刊工作者不应局限于“我认识的人才帮我”,而应在世界范围内去认识更多的研究者,寻找到“乐意且有能力帮助我的人”。本文为处于数字化转型阶段的传统编辑提供了高效、实用且易掌握的与国际研究者进行电子函件联系的方法,目的是充分利用数据资源和现代化工具,快速与专业技术水平高并热情支持期刊发展的国际学者取得联系。利用不同用途的收件人列表和函件群发平台,可以快速、有效地在全球范围内开展发布宣传信息、约稿与组稿、发起专题讨论、收集办刊建议等网络营销活动。其中,收件人列表至关重要。对于较高层次的约稿或专题组稿,推荐有条件的期刊使用爱思唯尔的 Find Reviewer 系统,该系统同时提供作者的 h 指数、邮箱和机构信息,可批量获取。由于知网、万方等数据库目前尚未将作者邮箱信息单独收录,批量获取中文文献研究者的联系信息有一定难度。此外,数据采集在期刊分析和出版中的其他应用尚需要更多研究和实践。

4 参考文献

[1] 余化刚,范晨芳.创办军事医学英文期刊的实践与思考[J].