

# 可发现性研究对学术期刊可见度提升的启示\*

张蓓 夏道家

《南京航空航天大学学报》编辑部,210016,南京

**摘要** 可发现性是针对学术界学者的搜索行为进行的调查研究,也是近2年国际学术出版界新兴的研究方向。本文总结现阶段可发现性研究的主要结论;并利用这一新兴的研究结果,提出了相应的技术策略。采用这些策略,学术期刊可以从传统的总结阅读量、下载量等数据间接了解读者行为的方式,变为从直接解读读者兴趣、习惯入手,提高学术期刊在目标读者群中的显示度。

**关键词** 可发现性;可见度;学术搜索引擎;图书馆发现系统

**Discoverability and strategies for enhancing international visibility of academic journals//ZHANG Bei, XIA Daojia**

**Abstract** Massive high-quality papers would not be even noticed, given the large magnitude of papers published in digital age. International academic publishing industry realized it and boosted "discoverability" research. We summarize typical results of those studies, and discuss reasons causing scholars' featured searching behaviors. Based on our analyses, we propose four strategies to help academic journals to enhance international visibility.

**Keywords** discoverability; visibility; academic search engine; library discovery system

**Authors' address** Nanjing University of Aeronautics and Astronautics, 210016, Nanjing, China

**DOI:**10.16811/j.cnki.1001-4314.2018.06.018

学术出版从一开始就有对论文和图书进行检索、分类的良好传统。但是,面对每年海量递增的论文数量、百花齐放的数字出版平台,以及越来越丰富的论文展现形式,传统的检索分类方式已不能有效地支持学术生态的运行<sup>[1-2]</sup>。了解信息时代的学者的搜索习惯,对学术期刊、出版商、图书馆等整个出版行业的各环节都有至关重要的作用。因此,很多学术出版从业者不约而同地展开了论文的“可发现性”(discoverability)研究,旨在了解读者搜索意图<sup>[3]</sup>,为学术论文的出版布局提供新的发展方向。

学术期刊长久以来都着重关注以影响因子为首的统计分析,通过统计数字论文的下下载量、阅读量等信息,制定相应的策略,提高期刊的可见度及影响力。而借助“可发现性”这一新的形势,可以变被动为主动,从学术期刊的根本目标——读者入手,深入解读读者行为,有效地助力学术期刊论文的精准投放,提高论文在数字海洋中的可见度。

“可发现性”是指如何定位到一篇搜索发起人需要的文章<sup>[3]</sup>;或者站在期刊的角度,“可发现性”是指一篇文章如何在众多文献中脱颖而出,出现在该出现的人面前<sup>[4]</sup>。

本文总结最近几年国际上的研究成果,然后针对这些研究结论,为学术期刊,特别是英文学术期刊在国际出版数字化时代的可见度的提升,提出相应的技术上的对策。

## 1 可发现性研究现状

“发现”(discovery)这一行为在学术出版界历史悠久,但是其定义却从来没有统一、确定过<sup>[5]</sup>。本文采用大多数业内人士默认的意义,定义如下:“发现”是指研究人员如何寻找并且使用电子资源的方法,以及这一过程带给出版商、图书馆及其他服务提供商的机会与挑战<sup>[5]</sup>。在“可发现”的行为研究中,需要着重区分2个概念,即“寻找”(looking for)和“获取”(finding<sup>[6]</sup>或者delivery<sup>[1]</sup>)的区别。“寻找”是指寻找文献的过程;“获取”是指当搜索人需要一篇文献时,如何定位到该篇文献,获取全文。往往“寻找”文献的平台和“获取”文献的平台不是同一个。因此,可发现性的研究基本都是按照这2个阶段分别展开的。

**1.1 寻找:Google和图书馆发现系统最受欢迎** 尽管科研人员的搜索习惯会受到工作资历、学科及所在地区等不同因素的影响,但是各种调查研究都指出,科研人员,尤其是年轻的学者,更倾向以Google和Google Scholar为首的搜索引擎开始寻找文献。文献[7]发现本科生、研究生、博士生和事业刚刚起步的青年学者对搜索引擎有更热烈的偏好;针对研究生(主要是博士研究生)的调查也表明对搜索引擎的喜好是近十年的发展趋势<sup>[8-9]</sup>。以往的这些调查研究的对象主要都集中在美国,而文献[10]开展的调查问卷涉及包括中国在内的8个国家,虽然样本体量较小(116位),但是也提供了一个跨国家的年轻学者的搜索倾向,他们都倾向于使用Google作为信息搜索的起点。

文献[11]展开了一项针对美国本科生长达4年的跟踪调查,发现随着研究的深入和成熟,学生会逐渐从Google过渡到图书馆发现系统,与针对其他资深的科研人员、青年学者的调查暗暗吻合。几乎所有的文

\*中央高校基本科研业务费青年科技创新基金(NR2017043)

献都强调,虽然搜索引擎(主要是 Google)是最受欢迎的起点,图书馆发现系统依然紧随其后,也是十分受欢迎的搜索起点。文献[1]在2015年对发达国家的各个学科的科研人员调查发现,资深研究员、讲师这些在学术上已经有所成就的科研人员,是除了图书馆员之外,最倾向于使用图书馆发现系统的学者。而文献[6]在针对各个群体的研究中发现,讲师及教授更倾向于从图书馆发现系统和特定的专业数据库中开始查找,这一倾向在教授中更加突出。

因此,在发起寻找的最开始,学者们最先打开的是 Google,随着研究年限的增长,他们逐渐倾向于图书馆发现系统。由此可见,学术期刊在进行数字出版时,应当着重注意其论文在搜索引擎和图书馆发现系统中的可见度。

### 1.2 题名中出现“搜索词”是确定目标的关键因素

“搜索词”是指搜索人发起搜索时输入的文字,它可能是一个词或者词组,或者是几个无语义关联的单词。当查询结果出现在眼前时,研究人员是通过哪些信息确定这就是自己要找的相关文献的呢?或者说,什么相关信息会引起读者的兴趣从而点开链接呢?文献[6]列出了题名中的关键字、文中关键字、是否能获得全文、引用率、作者姓名、出版日期、期刊名和关键字在全文中出现的次数这8个选项供受访人选择。结果表明,题目中出现相关的搜索词是研究人员确定相关文献时最重要的指标,作者姓名对于博士以上级别的研究人员是第二重要的指标,出版时间和期刊也是搜索者考虑的要点。令人意外的是,引用次数是所有指标中最不重要的。

由此可见,学者更加注重的是内容的相关性,而不是计量数据。学术期刊应当把重点放到如何精准地确定关键词汇,维护特定作者群上,才能增大论文的阅读量。

### 1.3 获取:图书馆授权的数据库平台是唯一途径

在发现一篇需要的文章后,几乎所有的人都会去图书馆平台,通过图书馆的授权,登陆各大出版商、数据库的网站,获得全文。因为几乎没有其他选择<sup>[12]</sup>。但是,随着近年来一些科研社交网站(如 researchgate)和一些免费的文献数据库的兴起,也吸引了部分学者,尤其是来自发展中国家的学者。文献[1]总结发现来自低收入国家的学者有接近75%的概率在免费资源中寻找全文;而来自高收入国家的学者的这一概率有60%。这说明免费资源的吸引力还是非常强大的,这应当促使期刊思考如何在OA环境中更好的生存。

### 1.4 可发现性研究的不足

1) 中国数据严重缺失。从2005年可发现性研究

萌芽开始,迄今为止,绝大多数研究几乎都是针对发达国家和地区,比如美国、英国、欧洲等的科研人员进行的。只有有限的几个调查中明确说明有来自中国的受访者,受访人数非常少。

2) 缺少针对特定期刊群体的调查。由于最先推动可发现性研究的机构大多是数据出版平台及大型出版商<sup>[13]</sup>,甚至最有力的研究来自财力雄厚的咨询公司,因此,这些研究的目的大多是站在出版商的角度;在组织调查时,虽然考虑到了期刊各特征对学者的影响,但是依然是站在学者的角度,根据学者的特征进行调查。所有这些研究都缺少站在某个期刊的视角,对某些特定目标群体来制定特定的可发现性策略。

## 2 学术期刊可见度提升策略

2.1 特别关注文章的主题词 如前所述,在可发现性的研究中,“搜索词”是指引读者寻找文章的关键因素。因此,整理文章中“主题词”与“搜索词”相对应,以适应学者的搜索行为,是提高期刊可发现性的最基本的因素。

由于读者在输入“搜索词”时没有固定规则及固定语法,因此,与之对应的主题词不能简单等同于文章的关键字。本文将主题词定义为:最能体现文章创新,最能定义文章所属方向的1个或几个关键词汇。主题词通常需要包括文章提出的新算法、新观点的名称,文章所属的专业小方向的名称,文章所可能应用的领域及工业方向等。

主题词编辑加工时要特别注意以下几点:

1) 主题词出现在文章题名、摘要、关键字和章节标题中时,应尽量使用全称。一般的搜索算法会将文章的关键部分——文章题名、摘要、关键字和章节标题——赋予更高的权值。也就是说,在以上几个部分中被检索到将会大大提高文章的可见度。而现在的语义搜索还不成熟,搜索算法更容易理解全称的名称,而难以区分各种缩写的意义。当读者了解到这一特征时,他们会输入全称以期得到更准确的结果。所以,文章的重点部分的主题词使用全称,文件被检索到和在检索结果中排名靠前的概率比较大。比如,CFD(计算流体力学)虽然在航空航天领域是一个不需解释的缩写、但是,搜索引擎不仅会理解成 computational fluid dynamics,还会理解成人名的缩写,机构名称等。如果题名中使用 CFD,在读者搜索 computational fluid dynamics 时,该篇文章将难以出现在结果中,或者出现在读者很难看到的第2页之后的搜索结果中。

2) 文章题名中一定要出现重要主题词。如前所述,如果搜索词出现在文章题名中,读者点开链接的可

能性最高;因此,文章题名与搜索词的匹配度是重中之重。所以,题名中的用词一定要能体现文章的核心创新。化学名称也尽量使用文字名称,尽量避免非文字形式的数学表达式出现在题名中。现在的搜索算法还很难达到智能语义搜索的水平,因此,需要编辑尽最大可能地将题名处理成搜索算法最容易理解和排序的模式。

3) 主题词在文章中一定要保持一致。搜索算法会考虑搜索词出现的频率;因此,文章主题词的英文表达一定要前后一致,确保在全文中出现时是相同的形式,这样会增加主题词的频率,进而增加文章在搜索结果中的排名,提高可见度。

**2.2 提高期刊文章在搜索引擎中的排名** 如前所述,鉴于 Google 及 Google Scholar 在科研人员,尤其是有巨大潜力的青年学者的心目中有不可撼动的地位,期刊文章在搜索引擎的搜索结果中排名靠前的话,也会大大提高文章的关注度<sup>[14]</sup>。

英文期刊应当重视 Google 学术搜索的索引与排名,首先应当申请加入 Google 学术的索引。同时,搜索引擎也有相应的技术特点。搜索引擎优化(search engine optimization, SEO) 技术<sup>[15]</sup> 特别推荐了提高搜索排名的方法,尤其针对如何提高 Google 算法中权重最高的 PageRank (PR) 值<sup>[16-17]</sup>。结合期刊的特点,除了 2.1 节针对主题词的特殊加工外,我们从技术上还可以做到以下几点:

1) 尽可能多地以 HTML 格式在各大论文出版平台投放文章。比如知名的开放获取平台及目录。这些大型出版平台本身的 PR 值很高,被检索的排名也会靠前。

2) 尽可能多地为期刊文章建立高质量的外部链接。比如:与其他同类期刊之间建立相关链接;与其他相关期刊、专业性的数据出版平台建立互相指引链接;鼓励作者、读者在社交网络上分享文章的全文链接。外部链接的质量与数量对 PR 值也十分重要。

3) 在 HTML 出版的文章网页的 HTML 源码中,注意在 title 域中放入文章的题名。在 Google 搜索结果的排名中,HTML 中的 title 域的权值排名也较高;所以,当单篇文章的 HTML 网页打开时,网页头部出现文章题名。比如,《Transactions of Nanjing University of Aeronautics and Astronautics》(TNUAA) 的 HTML 的出版网页中,就在头部域中加入了标题(即在网页的标题栏中显示的是文章题名)。这将会提高文章在搜索结果中的排名。

**2.3 提高期刊文章在图书馆发现系统的可发现性**

图书馆发现系统实际上是基于元数据联合预索引的网

络级发现系统。它可以在前端给读者提供一个类似搜索引擎的界面,在后端读取原本独立的各大数据出版平台的元数据,从而提供一个界面亲切、具有搜索功能的整合系统。

图书馆发现系统与搜索引擎最大的不同就在于,它是一个相对封闭、相对定制化的平台。每个图书馆都会针对本馆订购的数据库及馆藏内容建立属于自己的联合预索引数据库<sup>[18]</sup>。尽管它们都声称使用“相关性”作为结果排序的规则,但是具体的算法仍各有不同<sup>[18]</sup>。换句话说,在不同的图书馆的发现系统上,某篇文章在搜索结果中的排名有可能是不一样的。

例如,搜索 TNUAA 近年被引用最多的一篇文章是《Progress of grey system models》,不同的图书馆得到的结果是不同的。我们在这里不使用任何搜索技巧,仅仅按照文章题名在电子资源里搜索。在清华大学基于 Primo 的“水木搜索”图书馆发现系统中,该篇文献出现在第 11 位的位置,而使用南京航空航天大学图书馆的“智周搜索”搜索时,该篇文章排在首位。“智周搜索”是基于 EBSCO 图书馆发现系统的。

不仅不同的图书馆发现系统的排序结果不同,即便是基于相同的发现系统,每个图书馆也会选择属于自己的分面排序优先级。比如同为基于 Primo 的发现系统,清华大学图书馆水木搜索平台将按学科“定制检索结果”放在首位,波士顿大学图书馆则将“精炼检索结果”放在首位,牛津大学信息发现平台则将主题放在首位等<sup>[19]</sup>。

既然每个发现系统的排序不同,我们就需要做到:

1) 期刊要了解自己的目标人群主要集中在哪些大学和学术机构,以及这些机构的图书馆订购了哪些数据库,使用的哪一个图书馆发现系统。图书馆在安装使用发现系统时会有些定制化的操作,这些操作有可能会严重影响排序结果。因此,跟期刊读者所在的图书馆建立联系,了解他们购买的服务及相关服务的具体技术细则,从而有针对性地提高自己的文章的相应部分的曝光率,达到从图书馆入手,提高搜索结果的排序的目的。

2) 加入大型的全文检索数据出版平台,深耕专业性强的平台。3) 大图书馆发现系统 Summon、Primo、EDS,基本上都囊括了各大主流的全文数据出版平台;而购买发现系统的高校或者科研院所的图书馆,一般也会订购这些大型的数据库出版平台。因此,想要在发现系统的结果中出现,英文期刊的文章就势必要加入各大型数据库出版平台,如 Elsevier、Web of Science、Scope 等。

3) 编辑加工元数据,特别注意统一性及出现频

率。每一个发现系统的分面选项中,都有“主题词”这个重要选项,其中 Summon 最多,默认显示的主题词为 100 个左右<sup>[20]</sup>;因此,要想在图书馆发现系统中突显,也必须整理文章的“主题词”,使之最大可能地与相关学科接近。

另外,图书馆发现系统建立在各大数据库的元数据基础上,因此,在编辑加工时要特别注意元数据的处理。元数据一般包括文章的题名、摘要、关键字、作者、基金以及参考文献等。注意各名称在这些位置出现的频率,频率越高,相关性也越高;注意各名称的统一性,以便提高相关搜索结果的排名。

**2.4 深入了解具有中国特色、期刊特色的读者阅读行为** 如前所述,现在的可发现行为的研究都是由国外学者开展的,针对的也都是国外的读者群,这固然对中国的英文期刊拓展海外受众大有益处,但是我们对中国学者的搜索行为还一无所知,这也会导致国内期刊的发展立于无本之地。

并且,国内的学者跟国外的学者的生态环境有很多不同,期刊的推广环境也与国外期刊大为迥异。我们不能简单地将国外学者的行为类推至国内。因此,针对不同的期刊,若想提高期刊在国内读者群中的可见度,还要认真研究以下几点:

1) 了解国内学者的搜索行为。由于语言的不同,国内学者的搜索习惯可能会与国外读者倾向使用 Google 不同;并且国内的百度学术搜索在中文搜索中确实有可取之处。因此,国内作者是否会更多地使用中文搜索引擎,并且中文搜索引擎的排序算法是否跟 Google 不同,都是亟待解决的问题。只有掌握国内的本刊目标群体,由此为基点,制定相应文章投放和推广策略,才能在不丢失国内受众的基础上,走向国际。

2) 深入了解期刊目标群体人口分布及其搜索行为。国外研究显示,不同的学科,不同的研究阶段,学者的搜索行为都会有差异。因此,建立每个期刊的目标群体的侧写,分析专属本期刊的读者的行为,为期刊提高可发现性提供最可靠、最基本的数据证据。

3) 不同的新媒体环境给期刊的可发现性带来了新的机会。现在的研究中无一例外地都指出越来越多的读者在搜索文章时会转向新媒体社交平台。虽然数量无法跟搜索引擎及图书馆发现系统相比,但仍是一个有吸引力的方向。

国内新媒体生态跟国外完全不同。例如,国内的微信公众号平台是很多期刊推广文章的阵地。如何在了解读者的基础上更好地利用这个平台,以及其他微博、知乎等平台,也是期刊推广下一步需要深入研究的方向。

现阶段,南京航空航天大学学报编辑部下属 3 刊均注册了微信公众平台的服务号,定期推送当期目录、期刊专辑约稿等信息,作者可以在公众号内查询投稿文章的进度,读者可以在公众号内查阅过刊。这些举措也要跟读者搜索行为的调查结合,更深入地挖掘内部信息,以便期刊更好地利用国内新媒体发展的东风,提高期刊文章的可见度。

### 3 结束语

可发现性的研究是一个由市场催生的新的研究领域。虽然提供了一些十分珍贵的研究结果,比如搜索引擎、图书馆发现系统等的重要性出乎想象,并且对学术期刊提高可见度提供了相应的策略思路;但是仍需要更多国家比如中国,及更多期刊的融入和参与。对于国内的学术期刊来说,展开对本刊受众的可发现行为的研究,对期刊可见度的提升至关重要。

### 4 参考文献

- [1] GARDNER T, INGER S. How readers discover content in scholarly publications; trends in reader behavior from 2005 to 2015 [M]. USA: Simon Inger Consulting Ltd, 2016: 50-55
- [2] 付雅静, 钱俊龙. 数字出版时代提高科技期刊显示度途径的探讨[J]. 中国国际期刊研究, 2014, 25(10): 1262
- [3] SOMMER D. Creating impact resonance through a culture of holistic discoverability[J]. Learned Publishing, 2017, 30(1): 91
- [4] CONRAD L Y. Pathways to a new way of working: Discoverability[J]. Learned Publishing, 2017, 30(1): 3
- [5] CONRAD L Y. Headlines from the discovery files: Key publications on scholarly content discoverability[J]. Learned Publishing, 2017, 30(1): 31
- [6] DOSHI N. 'Finding' and 'discovering': how understanding researchers' search and discovery behaviour informed the building of Cambridge Core[J]. Learned Publishing, 2017, 30(1): 40
- [7] NIU X, HEMMINGER B M. A study of factors that affect the information-seeking behavior of academic scientists[J]. Journal of the Association for Information Science & Technology, 2012, 63(2): 336
- [8] CATALANO A. Patterns of graduate students' information seeking behavior: a meta-synthesis of the literature[J]. Journal of Documentation, 2013, 69(2): 243
- [9] SPEZI V. Is information-seeking behavior of doctoral students changing? a review of the literature[J]. New Review of Academic Librarianship, 2016, 22(1): 78
- [10] NICHOLAS D, BOUKACEM-ZEGHMOURI C, RODRIGUEZ-BRAVO B, et al. Where and how early career researchers find scholarly information[J]. Learned Publishing, 2017,