科技文献引用现状及趋势解读

作者：writer 来源：本站

本文原地址：https://www.iikx.com/news/article/368.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

导读：在这个获取的时代，科研工作者不能免费使用期刊论文上的参考文献进行学术交流，这着实是一桩不可思议的事情。科技文献的获取，必将是大势所趋。目前人们正在着力建设一个覆盖面广泛的公共数据平台，但是这条道还很长。

2012年5月，当Heather Piwowar开始着手调查研究数据公开是否能提高论文引用率时，她从来没有预料到会困难。作为总部设在的指标服务ImpactStory的联合创始人，那时的Piwowar还是美国北卡罗来纳州杜克大学的一名博士后。由于缺乏从机构层面进入Elsevier SCOPUS数据库的途径，Piwowar只有通过和国家科学图书馆的一份研究协议方才获得了使用权。但是由于她生活在美国，对方要求Piwowar提供自己的指纹作为证明。“我浪费好几天时间在获取研究所需的引文数据上。这太了。”Piwowar需要分析10,000多篇文章的引用次数，但是时下很多主要的引文来源，诸如汤森透科学网，并不支持使用PubMed文献服务系特数字标识符的查询。Piwowar说：“要是有公开的引文数据，我早就可以写自己的文章了！”

Steven Greenberg是美国州市哈佛大学医学院的神经病学家，他在通过对重复引用的统计，研究是如何被转变为“事实”的时候，也遇到了类似的障碍。Greenberg构建和分析了一个与某个特定的相关引文网络，包括242篇论文、675份引文、553个截然不同的引文径。可是研究的开展非常困难，如果这些引文数据能够轻易在网上获取，那将能使Greenberg少费许多周折。

在这个访问的时代，学者不能免费使用期刊论文上的参考文献进行学术交流，这着实是一桩不可思议的丑事。

为了改善这种情况，数据应该被视为公共资源的一部分，放置在一个的信息库里。为此，自2010年以来，英国信息技术研究与开发资金组织资助了价值13.2万英镑的项目，旨在建立并发展引文语料库(OCC)。在的学术引用数据方面，OCC还是一个“初出茅庐”的资源库，目前正在寻求可持续的资金，力求在将来成为数字研究基础设施，支持学术事业的基石。

虽然更为适合的评价指标正在摸索中，直接引用仍然是衡量产出重要性的一个重要指标。学术交流包括引用网络信息和观点的流动，以及随着时间的推移分析网络的变换，从而学者间沟通模式的变换和学科的发展与。这种信息对学术研究至关重要，对于制定正确的研究投资和战略、促进创新、增长和繁荣，特别是在日益国际化的研究合作中也显得越来越重要。

目前最权威的学术引用数据来源要属汤森透科学网，它的前身是美国科学家Eugene Garfield于1964年创建的科学文献索引，由科学信息研究所(ISI)出版。而它的主要对手Elsevier诞生于2004年。这两大巨头覆盖了主要的学术文献，但由于都不完整，它们也互为补充。

为了使用这两大数据资源，英国的每所研究型大学每年都需要支付数万英镑，这相当于其它发达国家研究机构所能获得的款项。此外，订阅这些数据库还需要严格遵守保密协议。如此种种，严重损害了那些不在上述机构工作的群体的利益，包括大多数企业和。其他引文信息的重要来源，例如谷歌学术搜索和微软学术搜索，也由商业公司运营，但是无需订阅即可获取。谷歌学术搜索的资源库比同类型的资源库大，因为它不仅包括书籍、论文、预印本、技术报告等等，也包括其他非同行评议的“ 灰色 ”文献。

所有这些资源都有许可，对其引文数据的重新发布。基于这个原因，文献计量学论文所能公布的数据非常之少。更糟的是，可用的引用数据还是不准确的。在汤森透科学网、斯高帕斯数据库、谷歌学术搜索和微软学术搜索上，Shotton的引用记录竟截然不同。例如，Shotton于2009年发表的一篇语义学论文，在这4家平台的引用次数分别是22、37、88和16。我们该相信哪一个?更令人担忧的是，一篇更早的关于蛋白质结晶学的论文在汤森透科学网上有3个的条目，每个条目的引用次数分别为59、19和0。据此，汤森透期刊影响因子的可信度也无法。

OCC，作为一个的学术引文数据库也许可以改善上述情况。它的目的是提供准确的引文数据，人们可以地查看、引用或用作其它目的，不受到版权或数据库法规的。

OCC数据库始建于2010年年中，并在2011年年中发布了第一个版本。在这个版本中，有204,637篇文章的参考文献能够通过包含3,373,961篇论文所引用的 6,325,178篇参考文献的PubMed OA-PMC式访问。虽然目前数据量较小，这个语料库依然包含了大约20%的1950年至2010年所有生物医学文献索引，其中包括了已在PubMed发表的每个生物医学领域的高引用频次论文。OCC的工作人员目前正在修改数据模型，提高其托管的基础设施，扩大其覆盖范围，从最初的672,442篇，更新OA-PMC文章量超过一倍。

理想的情况下，引用文献会直接从出版商在文章发布时获得。但是大多数出版商都将参考文献列表划归在付费订阅的类别里，因为这些文章要受到版权。目前OCC已经与几家主要的期刊出版商制定了合作协议，包括自然出版集团(NPG)，大学出版社，美国科学促进协会(科学)，英国皇家学会出版，波特兰新闻，麻省理工学院出版社等。

OCC的长期目标是汇集世界范围的学术文献，以及艺术、人文科学等学科的引文信息。当然这还需要很长的一段要走。

理想情况下，出版商会把自己的出版书目和引用数据拿出来分享，比如自然出版集团把自己的数据发布在。它是第一家也是唯一一家将数据共享的公司。

但是，如果能够将这类数据整合在一起，势必会获得更大的收益。OCC将为用户浏览引用数据提供完整的权限，这些数据的来源非常广阔，既有传统学术出版物的数据，也有其他数据，并且所有数据都会标明来源。OCC将把文章与文章、文章与数据库、数据库与文章引用的关系清晰展现出来。此外，还将提供附加信息，例如作者、机构、共同的资助人、文章间语义关系，以及数据来源。

一旦引用数据被公开，其它有益的分析服务也将开展，包括面搜索与浏览工具、与趋势识别服务，以及时间表可视化服务。对于这些服务，OCC目前已经开发出了其中几种的原型。OCC对统计引用指标的巨大价值也将随着其覆盖面的延伸而增加。

除此之外，还有另外一个服务，它尤其对作者和编辑有益，那就是错误引用更正服务。出版论文

的引用中有约1%存在着不同程度的错误，例如在引用标题中将“ β淀粉酶 ”误写成“ β-淀粉酶 ”，或者对作者名字中符号的疏忽;严重一些的包括年份、期刊号、页码或者DOI的错误。OCC已经在内部使用引用更正方法，以处理多次引用问题，或者通过外部资源获取权威书目记录来减少。

接下来，OCC会有怎样的发展呢?10年前的1999年到2002年，由英国南安普顿大学、美国纽约州伊萨卡市康奈尔大学以及arXiv网站共同构建的项目也有一个具有相似目标，旨在引文数据的项目存在，名为引文计划。该项目开发了一款名为引用库的软件，能将引用信息记录下来，该软件的设计者将它描述为“ 引文计划皇冠上的珠宝 ”。但是，如今再点击该项目的网址，得到的提示却是：该网址没有网站链接。

要想从一个雄心勃勃计划项目过渡到现实的全球性服务并将它长期维持下去常困难的。OCC为了避免与引用库相同的命运，使它得到全面的发展并最终向整个学术界各个学科提供令人信任的引用数据服务，需要者、管理者、开发者以及监督者共同的努力。它也需要怀有相同目标的同道中人精诚合作、来自资助者充足持续的资金、支持者和投资者提升社会福利的决心，而不是以经济回报为导向，还需要来自出版界的鼎力支持、重要机构或国际组织的认可。

地址：陕西省西安市灞桥区新寺569号唐都医院临床教学楼312室 邮箱： 联系电话 传线号
技术支持：陕西奈特星网络发展有限公司

更多 论文写作 请访问 https://www.iikx.com/news/article/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发