新兴科学引文索引数据库的比较研究*

□李洁　孟烨　金佳丽　姚翔宇　黄晨

摘要　随着网络爬虫、基于人工智能的文本语义分析技术的更新和研究范式的发展，越来越多的新兴科学引文索引数据库问世，不仅为科研工作者搜索和筛选学术文献资源提供了新的体验，也为科研评价提供了更多的选择。本文选取了7个科学引文索引数据库（Microsoft Academic、Dimensions、Semantic Scholar、Lens、Scinapse、Scilit、Scite）进行调研比较和案例分析，以期挖掘各数据库的收录范围、指标、人工智能技术、可视化分析、插件及导出等特色，供图书馆员做科研评价分析时参考和选择。通过调研和案例分析发现，对于机构和学者的评价，综合考虑文献收录范围和查准率，Microsoft Academic 和 Dimensions 是最优的选择。

关键词　科学引文索引　引文分析　语义引文　学术影响力评价
分类号　G255.75
DOI　10.16603/j.issn1002－1027.2021.06.011

1　引言

科学引文索引（Science Citation Index，SCI）通过论文之间的引用关系以及对论文的被引次数等的统计，对个人、科研机构乃至一个国家或地区的科研产出绩效进行评价。Web of Science（以下简称WoS）和 Scopus 是两大传统的科学引文数据库，但在经历多年发展后，科研人员越来越发现其不尽如人意的地方。首先，两者都是付费数据库且价格昂贵，订阅费逐年上涨，这也引起了“打破科研成果付费墙”事件[1]，研究者对出版商发表成果和获取文献双重收费感到不满，图书馆也对购买大量数据库的高额费用感到捉襟见肘。出版商对学术期刊垄断，并以此获取高额利益，严重影响了学术传播、分享学术资源的意义。其次，开放获取已成为趋势，学界对于自由获取科研文献需求迫切，而这些开放获取的文献是这两个传统引文数据库无法收录齐全的。最后，在传播途径相对单一的学术交流环境，传统的引文数据库尚可保证覆盖大多数的引用情况，遗漏一小部分也不会对分析结果产生明显的影响。但如今大数据科研时代的到来，学术交流渠道多样化，引文数据库所遗漏的非收录文献之间的相互引用情况已经无法再被忽略了[2]。开放获取的鼻祖 Google Scholar（以下简称 GS）于 2004 年诞生，覆盖的文献量巨大，完全免费开放，且已作为一个非常成熟的索引数据库而被全球广泛熟知，虽未纳入本文的调研对象，但在案例分析时，把它和 WoS、Scopus 一起作为对标比较的数据库进行讨论。

2015 年后，新一代的科学引文索引数据库如雨后春笋般涌现。新加坡管理大学的资深馆员谢亚伦（Aaron Tay）在他的博客中综述了“下一代引文发现索引数据库”[3]，提到了以下 9 个近年来崛起的引文索引数据库：Semantic Scholar（以下简称 SS，2015 年）使用机器学习将引文类型分为背景引文、方法引文和结果引文[4]；微软公司的 Microsoft Academic（以下简称 MA，2016 年）应用了最新的自然语言处理技术[5]；Lens（2017 年）将专利源与引文相融合，创建了一个免费的发现引文索引[6]；Digital Science 公司推出的 Dimensions（2018 年）使用更新的技术和范式，对传统引文评估提出了挑战[7]；Scite（2018 年）利用自然语言处理技术对引文立场进行分类，区分“支持性”“争议性”“中立性”引文[8]。另外还有 Scinapse[9]、NAVER academic[10]、Scilit[11] 和 1findr[12]。

初步考察发现 NAVER academic 的文献量不如传统引文数据库，而 1findr 在本文修订时已终止

* 浙江省教育厅一般科研项目（人文社科类）“新兴科学引文索引数据库调研”（编号：Y202146118）的研究成果之一。
通讯作者：李洁，ORCID：0000-0003-0739-8579，邮箱：jieli88@zju.edu.cn。

对外提供服务。综上，本次最终调研选取了 MA、Dimensions、SS、Lens、Scinapse、Scilit 和 Scite 这 7 个近年发展崛起的科学引文索引数据库作为对象，考察了以上数据库在收录范围、指标、人工智能技术、可视化分析、插件及导出等方面，并进行比较和案例分析，以期挖掘各数据库的特色功能，供不同需求的使用者选择科学引文索引数据库时参考。

2 研究现状综述

文献收录范围是潜在用户了解不同科学引文索引数据库的学科全面性的第一步。众多学者从文献量角度对 MA、Dimensions 与传统引文数据库做了比较，发现其文献量上的优势正逐年增加，比如和晋飞等和赵蓉英等都在文献中提及 GS 在文献量和文献类型上都是最全的，MA 次之[2,13]。另外，胡克（Hug S E）和布兰德勒（Brändle M P）以苏黎世大学为案例[14]，塞沃尔（Thelwall M）在分析了多学科 29 种期刊的 172752 篇文献后[15]，都得出 MA 在数量超过了 WoS 和 Scopus。2018 年，库沙（Kousha K）等人为了调研预印本在 MA 中收录情况，取样了 2016－2017 年的 65000 篇文献，发现其在 MA 的引文量是 Scopus 的 2－5 倍，这主要是因为 MA 像 GS 一样涵盖了预印本网站 arXiv 的大量文献[16]。到 2020 年，维瑟（Visser M）等人在大范围地比较了 Scopus、WoS、Dimensions、CrossRef 和 MA 的所有文献后，得出了 MA 文献量最大的结论，且与 Scopus 有最高的重合度（81%）[17]；得益于 CrossRef 的数据，Dimensions 与 Scopus 重合达到 78%。Dimensions 的文献量也从 2018 年的比 Scopus 略低[18]，到 97% 有 DOI 号的 Scopus 文献都同时被 Dimensions 收录[19]，再到 2019 年哈辛（Harzing A W）提出 Dimensions 与 WoS、Scopus 相较，有相似乃至更好的文献收录范围，且引文量也相似[20]。最新 2021 年的研究分学科对文献收录范围进行比较，得出 MA 在多个学科门类较之 WoS 和 Scopus 有优势，尤其在人文、社科和商业与经济学学科门类中；Dimensions 在各学科门类的引文量仅略低于 Scopus；在除了物理数学和化学材料的其他六大学科分类中，引文量都超过了 WoS[21]。由此可见，MA、Dimensions 在文献收录范围和引文量上较传统引文数据库有优势，是多方学者调研且认同的。

另外有学者对个别引文数据库的检索界面、指标等做了比较。谢智敏等对 SS 与 GS、MA、Bing 学术和百度学术在文献源、检索方式、检索结果呈现和指标做了比较研究后，指出 SS 基于引文内容的分析而得到的高影响力引文指标，给科研评价注入了一股新动力，但其在数据来源、检索功能和个性化服务功能方面还有待进一步完善[22]。同样地，研究 SS 高影响力引文指标的学者，验证了通过使用监督分类算法，该指标在召回率 90% 的情况下，准确率能达到 65% 以上[23]。另外，谢亚伦（Aaron Tay）在他的博客中多次提及 Scite 的引文立场指标，认为其直接用于评价还为时过早，但将其用于文献研究是可以接受的；虽然可能会产生不明确的偏差，但并不比使用其他基于引文的方法或依靠搜索引擎的黑匣子算法更糟糕[24]。对于 Lens，有学者认为其优势在于涵盖了大量专利文献，提供专利与学术文献间的引用信息，但其不足是数据之间的可靠性和 API 配置还需要完善[25]。

综上所述，国内外的学者大多从文献量的角度分析各引文索引数据库的情况，都一致赞同新兴引文索引数据库文献量上的逐年崛起，这其中一大部分归功于全球最大的预印本系统 arXiv，它目前总量约 186 万余篇，且每月以 1 万多篇的速度增长，可见 Wos 和 Scopus 在预印本和开放获取文献上的缺失，在如今这个要求信息快速及时的时代，是难以容忍的短板。对 Scinapse 和 Scilit 这两个科学引文索引数据库，国内外学者的系统研究都较少。本研究旨在从收录范围、指标、人工智能技术、可视化分析、插件及导出等方面，对 7 个调研的新兴引文索引数据库做更多的揭示。

3 新兴科学引文索引数据库的特色

新兴科学引文数据库较之传统的 WoS 和 Scopus 数据库的特色，不仅体现于不受期刊来源限定的更全的文献收录范围、更拓展性的文献类型、更个性化的可视化分析，还在于其利用人工智能技术提供文献的可发现性，创新性地提出了语义引文在内的多项评价指标。

3.1 收录范围特色

各科学引文索引数据库的文献来源都不外乎以下几个：网络爬取、与出版商或数据供应商合作、开放获取网站、机构知识库和个人主页等。总体而言，在文献量上 MA 占绝对的优势（见表 1），其他有些

数据库还直接与其进行数据合作，例如 Lens 和 Scinapse 都通过与微软图数据库（Microsoft Academic Graph）的合作扩充了自身的文献量。

表 1　新兴科学引文数据库的数据源和文献量比较

科学引文索引数据库	数据源	文献量*
MA	Bing（必应）网络爬虫，EBSCOhost 专业数据库、Springer 等出版社，政府、教育、组织等学术资源，ResearchGate	2.48 亿条
Dimensions	Crossref、PubMed、Altmetric（推文、博客文章、新闻文章和政策文件）、IFI Claims（专利）、ÜberResearch（基金）	1.15 亿条
SS	ArXiv 等开放获取平台和合作的期刊、数据合作方（AMiner）等	1.89 亿篇文献
Lens	PubMed、Microsoft Academic Graph、Crossref、USPTO、ORCID、EPO、OpenCitations、DOAJ、WorldCat、Unpaywall、GRID、WIPO、CORE、IP Australia	2.25 亿篇文献，1.27 亿条专利
Scinapse	Microsoft Academic Graph、SS、Springer Nature 和 PubMed	文献量未公开，含 4.8 万种期刊
Scilit	CrossRef、PubMed	1.28 亿篇文献
Scite	出版商、PubMed Central 数据库、大学存储库和作者	文献量未公开，含 8.13 亿条引用

注：统计数据截止到 2020 年 12 月 24 日。

大多科学引文数据库的基本文献类型包括：期刊论文、会议论文、图书、预印本等。Dimensions 较早觉察到整个研究生命周期应囊括各阶段的研究成果，即基金是将要发生的研究的早期信号，出版物代表了其研究过程，专利、临床试验和政策文件则是研究转化为实际应用的证明，所以，Dimensions 开创性地将其文献类型从期刊论文等出版物扩展到了基金、专利、临床试验、数据集和政策文件。相较之下，Lens 的特色是其 1.2 亿多条的专利数据，其来源于各国专利局和世界知识产权等组织，并在未来两年内，目标收集全球 95% 以上的专利信息[10]，并链接关联到大多数学术文献，从而建立专利与学术文献间的引用信息。

3.2　指标特色

和传统的引文索引数据库一样，新兴的科学引文索引数据库也提供发文量、引文量等基本指标，但也有不少数据库提供了学者 h 指数、期刊影响因子或自身数据库研发的新指标（见表 2）。

Dimensions 的特色指标是其在单篇文献评价中引入了领域引文率（Field Citation Ratio，FCR）和相对引文率（Relative Citation Ratio，RCR）。FCR 是指一篇文章在其学科领域内与同年的文章相比时的相对引文绩效；RCR 是指一篇文章在其研究领域（共引文献组）内与其他文章相比时的相对引文绩效。两者的区别是 RCR 的领域限定更加严格，仅限于同属共引文献组内。两者的标准化值都为 1.0，当值大于 1.0，则表明该文章的影响力高于其领域的平均值，反之亦然。不到 2 年的文章没有 FCR 和 RCR 值，引用次数为零的文章的 FCR 和 RCR 值为 0。

表 2　新兴科学引文索引数据库的指标比较

科学引文索引数据库	总被引次数	学者 h 指数	其他特色指标
MA	有	无	无
Dimensions	有	无	领域引文率 FCR、相对引用率 RCR，有些会显示 Altmetric 数据，Mendeley 文献管理软件中的读者数
SS	有	有	高影响力被引次数、引用速度、引用加速度、学者影响力分数、学者影响图
Lens	有	有	无
Scinapse	有	有	无
Scilit	有	无	Altmetric 评分
Scite	有	无	支持、争议、提及引用次数，SI 指数

目前通用的论文学术影响力评价指标是基于论文的被引次数，但并非所有的被引次数都具有同等价值，有的引用对研究具有重要参考价值，与仅作为背景或进行对比的引用截然不同。因此，只简单地统计被引次数并不能客观准确地衡量一篇论文的学术影响力。自引文分析理论建立以来，基于引用内

容的引文分析便已被学者们提出，但最初由于全文数据库还未出现，相关研究一直较少，随着全文数据库的不断完善和信息处理技术的不断发展，为进行全文层面的语义引文分析研究提供了良好的数据基础[26－27]。SS 基于引用内容分析，利用深度学习技术判断作者对所引用文献的态度，辨别引用的价值，建立了一个新的引文指标，即高影响力被引次数。它的判定是通过建立深度学习模型，理解和分析一系列的因素，包括：某篇论文在施引文献中的被引次数，每次被引时的位置及上下文情境等。由高影响力被引次数衍生，SS 还提出了学者之间的影响力分数和影响力图。

SS 还创新地提出了引用速度和引用加速度的单篇论文评价指标。引用速度是指某篇论文近 3 年被引次数的加权平均数（个别论文为近 1－2 年），可以表明该篇论文的被关注度和关注持续的时间。引用加速度是指某篇论文引用速度随时间的变化率，判断该篇论文的被引次数随时间的变化是增加还是减少了，其中计算引用加速度的时间跨度为近 2 年。但目前，SS 只提供利用引用速度和引用加速度进行排序，而未给出具体指标得分。

Scite 试图通过分析引用的上下文对引文立场进行人工智能分类，从而提供更细致的引用信息，即将引用分成支持引用、争议引用和提及引用，并分别统计一篇文献所获得不同类别的引用次数。依据引述文字的语义分析，辨别支持或争议性的立场，如果没有明显立场则归入提及引用。如果人工智能引文分类有误，用户可以标记认为正确的分类并转给专家审查，审查后该引文会被标记为人工标引。

由引文立场衍生，Scite 提供了期刊和机构 SI 指数。SI 指数是 Scite 提出的基于支持和争议引用对期刊和机构进行评价的指标，其计算公式为：

$$SI=\frac{SupportingCites}{SupportingCites+DisputingCites}$$

该指数包含 5 年、2 年和全生命周期的 SI 指数，并且可以展示评价对象 SI 指数的趋势图以及其在所有期刊或机构的排名和百分位。另外需要注意的是，被评价机构在评价周期内需要至少获得 100 个支持或争议引用才能进行 SI 指数计算。

替代计量学指标也是目前文献评价中不容忽视的指标之一。Dimensions 和 Scilit 都在各自数据库中嵌入式显示了补充计量学指标 Altmetric 评分。另外，Lens 融合了学术论文和专利，使得它能够获取两者之间的引用信息，提供每篇论文的论文引文数和专利引文数，同样也提供每篇专利的专利引文数和论文引文数，可以作为研究学术文献与专利之间相互影响力的一项评价指标。Scinapse 数据库提供学者 h 指数和期刊影响因子指标。

3.3 人工智能技术特色

在诸如 WoS 和 Scopus 之类的传统引文索引数据库中，通常以期刊为依据对文献进行学科分类，是从期刊级别来衡量学科领域（只有 SciVal 的研究主题是按单篇文献的引文网络来分类）。近年来，自然语言处理、机器学习和人工智能领域都取得了长足的进步。Dimensions 选择的路径是使用现有分类系统和基于机器学习的方法来自动为所有单篇文献分配其学科或研究领域，而不管其来源如何。

而 MA 则采用人工智能技术中的自然语言处理来理解和记忆数据库中每个文档传达的信息，然后应用一种称为语义推断的技术，来识别用户的意图并主动传递与用户的意图相关的结果。例如，当在搜索框中键入“Microsoft”时，它将推断用户的意思是检索机构，并显示由 Microsoft 所属研究人员撰写的出版物。同样地，输入“Artificial Intelligence”也能识别其为期刊标题、会议名称或研究主题。

相似地，SS 利用“机器阅读”、BERT 自然语言处理技术等从文本中挑选出最重要的关键词和短语，而不需要依赖作者或出版商键入这些关键词。SS 的人工智能还体现于利用深度学习实现机器对论文内容的理解，有利于帮助用户筛选有用信息，推荐影响力高、媒体（Twitter 数据等）关注度高以及相似作者的文献，另外它提供图表的直接截取预览，提高了学术信息搜索和过滤的效率。

3.4 可视化分析特色

除了展示相关文献列表及文献信息，有些引文数据库尝试在检索结果的基础上提供可视化的分析。MA 提供从会议、期刊、合作机构、发文机构多角度分析的学者发文影响力可视化图表。Lens 的可视化除了和 MA 一样通过发文期刊、会议、机构等字段对学者发文影响力进行分析之外，还提供关键词云图和学术文献及专利引文情况的分析图表，并允许用户对可视化图表的样式进行设置。

SS 在学者个人主页中嵌入了学者的学术影响图（见图 1），影响力分数是衡量一位学者 A 发表的

论文对另一位学者 B 工作的影响程度，其计算方式是基于 B 对 A 发表论文的被引次数和高影响力被引次数的加权平均。

图 1　SS 的学者学术影响图

Scite 的可视化特色是其提供基于文献的引用关系网络图（见图 2），页面左侧是筛选条件和文献列表，通过输入 DOI、题名、作者等检索词可以对引用文献列表进行检索和高亮显示，同时右侧可视化图形也会精炼变化为命中检索词的节点。以原始文献为中心，引用列表包括引用文献、引用的上下文片段、被引文献三个列表，当点击任意文献时右侧图形中会高亮显示相应文献与原始文献的图形节点，如果点击列表中被引文献右侧加号可以在图形中增加新的文献节点。右侧图形界面主要显示文献之间的关系，节点代表文献并标注作者和时间，通过线条数量显示该篇文章的引用次数，通过线条颜色显示引用立场（支持、争议和提及），通过箭头方向显示引用和被引用关系。图中默认不显示提及引用，可以通过选项选择隐藏或显示。另外图形还可以切换 8 种不同的布局方式。

3.5　插件及导出功能特色

Scite 提供驱动程序，在其他网站通过 DOI 或其他信息匹配是同一篇文献，就会在该文献中嵌入式显示该文献的引用立场，即支持引用次数、争议引用次数和提及引用次数。目前经测试在 WoS、Scopus、Dimensions 等数据库乃至维基百科中都能成功显示。

在数据导出方面，Dimensions 支持以 API 接口导出数据到 Google Sheets 中；Lens 支持一次性导出不超过 5 万条的数据；Scilit 支持一次最多导出 200 个结果。

4　案例调研分析

为了进一步调研各科学引文索引数据库的文献收录范围和测试学者页面的学者姓名消歧准确性，本研究选取了两位学者和临床医学学科中关于新型冠状病毒被引量排名前 10 位的文献进行检索测试，并同时选取了 WoS、Scopus 和 GS 的数据源进行对比。在初步调研中，发现 Lens 的文献收录量与其他相近，但无学者主页，无法考察学者的姓名消歧准确性；Scite 的文献量有 20%左右的缺失；Scilit 只有部分学者有个人主页，且文献量也远小于其他数据库。所以在表 3 中，只列出 MA、Dimensions、SS 和 Scinapse 与 WoS、Scopus 和 GS 的比较结果。在学者选取方面考虑学者的学术年龄，选取了一位发文较多且年限较长的资深学者和一位近几年开始发文的潜力学者来调研。

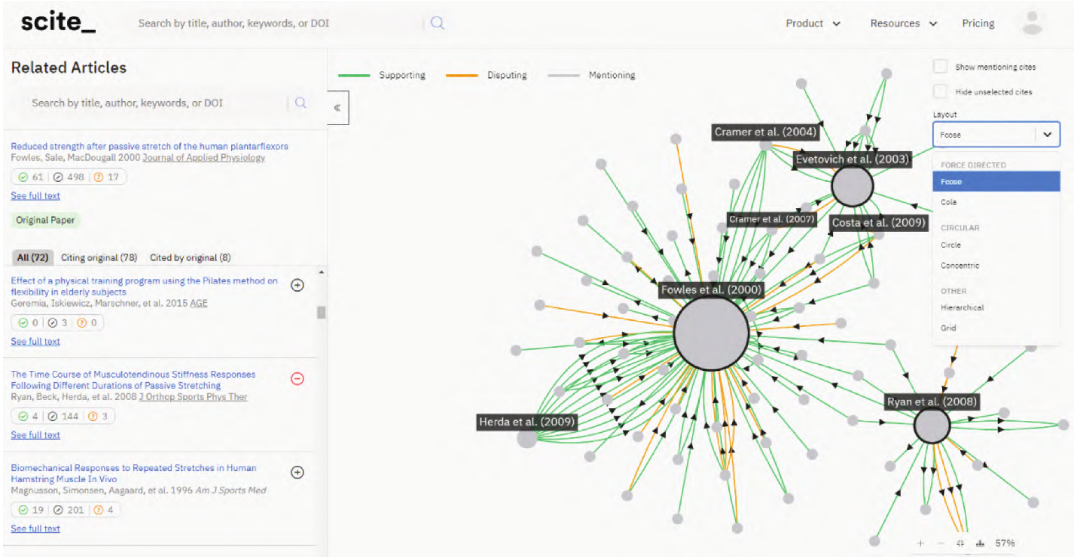

图 2　Scite 的文献引用关系网络图

4.1 学者案例

学者X是生物医学方面的专家，其论文在WoS中收录220篇，独立的学者页面中，收录数为212篇，但其中8篇经人工逐篇核查发现并非为X本人发文，其作者为电气与电子工程学科的同名X学者和同学科的相近名字学者（姓名拼音简称相同）。Scopus的X学者页面收录文献为216篇，且学者姓名消歧算法无误。在GS中，X学者页面的收录文献更多（含专利），但存在多个数据来源重复情况，去重后文献226篇，其中有3篇不属于X。

由表3可见，MA的发文量最多，为227篇，且经过人工确认均为学者X发文。Dimensions检索结果中混入了另一相似姓名学者的文献成果，需要二次筛选。SS包含5个该学者页面，其中发文量最多的是199篇，但也有4篇作者姓名消歧错误。Scinapse中检索结果有88个X相近名字的学者页面，文献数最多的包含219篇，且都为该学者发文。

表3 学者X在各数据库的文献收录量和被引次数

	WoS	Scopus	GS	MA	Dimensions	SS	Scinapse
学者页面的文献收录数（篇）	212	216	226（238篇有12篇重复）	227	出版物427种，基金24项，数据集52个，专利50项	199	219
学者页面收录文献的引用次数（次）	31266	33226	48983	48462	36928	36188	34200
学者页面中姓名消歧错误的文献数（篇）	8	0	3	0	出版物221种，基金6项，数据集50个，专利0项	4	0
数据库实际收录该学者的论文数（篇）	220	218	223	261	出版物206种，基金18项，数据集2个，专利50项	272	235

学者Y是一位光电科学领域并于2015年才开始陆续发文的学者。从表4中可见，在包括GS在内的新兴引文索引数据库中，文献量都比WoS和Scopus多，这与它们涵盖的数据集文件不无关系。引文方面，GS最多，MA次之。除了WoS有1篇文献学者姓名消歧错误之外，其他各数据库的作者判断无误。

表4 学者Y在各数据库的文献收录量和被引次数

	WoS	Scopus	GS	MA	Dimensions	SS	Scinapse
学者页面的文献收录数（篇）	20	22	31	28	出版物27种，数据集5个	30	28
学者页面收录文献的引用次数（次）	1047	1137	1389	1261	886	969	874
学者页面中姓名消歧错误的文献数（篇）	1	0	0	0	0	0	0
数据库实际收录该学者的论文数（篇）	23	23	31	29	出版物27种，数据集5个	34	29

从学者案例来看，有学者页面的MA、Dimensions、SS和Scinapse在查全率和查准率都表现不错，MA引文数据最大，且作者姓名消歧准确度最高，达到了100%；Dimensions包含了基金、专利、数据集等多种文献类型，但其有将相似名字学者文献混淆的情况，需要进行二次条件筛选才能分辨；SS的平台收录文献是最多的；Scinapse在引文数上偏低。

4.2 文献案例

为了调研各数据库的引文量，本研究选取了2021年3月期的ESI数据库，临床医学学科中关于新型冠状病毒被引量排名前10位的文献作为案例（检索时间为2021年5月6日），各引文数据库的引文数据见表5。

从调研结果看，各数据库的响应和文献更新程度都是比较及时的，传统的引文数据库中，WoS和

Scopus 的引文量相近，Scopus 略多，约为 WoS 的 1.1—1.2 倍，而 GS 的引文量是 WoS 的 2.2 倍左右。在调研的 7 个新兴引文索引数据库中，MA 的引文量遥遥领先，与 GS 相近甚至更多；Scite 的引文量是 WoS 的 1.8 倍左右；Dimensions、SS、Lens 和 Scinapse 的引文量是 WoS 的 1.4—1.5 倍左右；Scilit 的引文量约是 WoS 的 1.2 倍。

表 5　10 篇文献案例在各数据库的被引次数(次)

WoS 唯一 ID 号	发表时间	WoS	Scopus	GS	MA	Scite	Dimensions	SS	Lens	Scinapse	Scilit
WOS:000514576900032	20200215	12794	14376	27954	29526	24526	19408	18326	17736	16363	13585
WOS:000531351300014	20200430	8213	9560	17851	17124	15242	12741	12233	11597	10540	9650
WOS:000522650100033	20200328	7270	8069	15470	18194	14686	11129	10592	10520	9654	8151
WOS:000521968800018	20200317	7270	8036	15344	15316	13168	10903	10539	9715	8900	8198
WOS:000517119800008	20200220	6899	7884	14696	14296	9865	11244	10687	10178	9387	8853
WOS:000514576900033	20200215	6241	6974	13639	13676	10928	9173	9145	8430	7723	7015
WOS:000522357300007	20200326	4312	5211	11416	11191	7955	7443	7508	6939	6459	5501
WOS:000514849400027	20200222	3400	3804	7485	7317	5393	5146	5257	4823	4481	3879
WOS:000514576900034	20200215	2863	3316	6665	6549	4292	4374	4573	4159	3876	3549
WOS:000531069600035	202005	2756	3268	6151	7271	5213	4572	4122	4201	3863	3412

5　结论与建议

调研和案例分析显示 7 个科学引文索引数据库各有特色，不少引文索引数据库文献量和引文量领先于 WoS 和 Scopus，这与其全网收录范围和更多文献类型的数据汇聚有关，若其能提高检索功能上的不足，且完善文献计量学分析所需的 API 接口或导出功能，能像 WoS 或 Scopus 那样轻松地批量获取数据，其取代传统引文索引数据库将指日可待。

新兴科学引文索引数据库的全网收录范围数据、全文献类型和多维度指标都是传统的引文索引数据库所欠缺和无法比拟的。

目前，利用传统引文索引数据库 WoS 和 Scopus 及其衍生的科研管理分析平台 InCites 和 SciVal 所做的科研评价，从某种程度来说都属于“以刊评文”。其通过论文发表期刊的级别和数量来衡量评价对象的科研水平，即只关注论文“出身”而忽略了单篇论文的内容精细评价，不利于破除“唯论文”和“唯 SCI”的不良科技评价导向。Scopus 限定 2.2 万多种期刊范围，WoS 则以 SCI、SSCI 和 A&HCI 等 1.3 万余种期刊来限定文献成果，这都与近年来国家教育部、科技部提出的“破五唯”和改变“SCI 至上”的意见相悖。新兴科学引文索引数据库则以全数据采集和保存为理念，不分国界不分出版商地收集所有出版文献。有些发表在非正式期刊，如预印本上的论文也同样有卓越的科研价值，比如破解千禧年数学七大难题之一的庞加莱猜想就刊登在 ArXiv 预印本数据库上。

其次，新兴科学引文数据库涵盖和整合了研究生命周期的各种文献类型，如 Lens 建立专利与学术文献间的引用关系，Dimensions 包含了研究早期信号，如基金和数据集文件以及研究的成果转化，如专利、临床试验和政策文件。新兴科学引文索引数据库的更全文献类型，为单纯的论文评价扩展到整个研究成果影响力生命周期的评价服务提供了可能性。

最后，SS 的高影响力引文指标和 Scite 的引文立场指标是基于人工智能的深度学习和语义分析技术，通过分析引文的内容来判断引用者对所引用文献的态度，以此来辨别引用的价值；Dimensions 和 Scilit 融合了 Altmetric 评分。这些新提出的指标突破了仅追求引文量并结合语义引文的评价，以及替代计量学指标，扩充了成果评价的范畴，为传统的评价提供了更全方位的评价维度。

综上所述，对于仅以文献资源发现为需求的普通读者来说，对高级检索无过多要求，MA 是最佳选择，它具有最全的文献收录范围和一目了然的检索页面。对于科研人员，如需了解某个研究领域的最新文献或寻找同行合作学者，可以考虑 Dimensions，

斯坦福大学、加州理工学院和牛津大学都在其数据库导航中推荐了该数据库。其学者页面有较细分的学科领域和代表关键词，以及个人任职和教育经历，除此之外，Dimensions 整合了数据集、基金、专利、临床试验和政策文件等文献类型，能较全方位地了解学者科研的详细情况。

就图书馆而言，面对数字资源订购费的日益增长，完全可以尝试利用这些免费的新兴科学引文索引数据库来发现和获取文献资源，甚至可以借助这些数据库完成更细致和个性化的科研评价工作。比如科研评价时从宏观层面分析国家、机构的学术影响力，常常需要导出清单做进一步的筛选分析，Lens 的界面和 WoS 较接近，符合图书馆员的操作习惯，且允许按国家、机构进一步过滤，一次性导出 5 万条数据到 csv 格式，是最符合用户需求的选择。而 Dimensions 需要查询语句才能导出数据到 Google Sheets 中，MA 和 SS 的 API 接口也有较高的技术门槛，对于普通图书馆员来说可能不够友好。在学者个人评价时，SS 在传统的被引次数之外引入了高影响力引文指标，利用人工智能技术对施引文献中的上下文情境和被引时的位置进行分析得出的高影响力被引次数，为人才评价提供了一个更多维度的衡量指标。尚在探索中的代表性学术成果评价涉及到微观层面的单篇文献评价，Scite 提供单篇文献的支持、争议和提及引用次数，且在网页中有详细出处和论据支持，不失为代表作评价的一种参考维度。

参考文献

1 新浪网.加州大学系统停止订阅爱思唯尔：一场特殊的较量[EB/OL].[2021－04－16].https://t.cj.sina.com.cn/articles/view/3949237910/eb64969602000f16i.

2 和晋飞，吕锋昌，房俊民. 利用 Google Scholar 自动统计引文年度被引频次的方法[J].图书情报工作，2014，58(7)：88－94，118.

3 Tay A. The next generation discovery citation indexes—a review of the landscape in 2020 (I) [EB/OL].[2021－02－26].https://medium.com/a-academic-librarians-thoughts-on-open-access/the-next-generation-discovery-citation-indexes-a-review-of-the-landscape-a-2020-i-afc7b23ceb32.

4 Semantic Scholar[EB/OL].[2021－02－26].https://www.semanticscholar.org/.

5 Microsoft Academic [EB/OL]. [2021－02－26]. https://academic.microsoft.com/.

6 Lens[EB/OL].[2021－02－26].https://www.lens.org/.

7 Dimensions[EB/OL].[2021－02－26].https://www.dimensions.ai/.

8 Scite[EB/OL].[2021－02－26].https://scite.ai/.

9 Scinapse[EB/OL].[2021－02－26].https://www.scinapse.io/.

10 NAVER academic [EB/OL]. [2021－02－26]. https://academic.naver.com/.

11 Scilit[EB/OL].[2021－02－26].https://www.scilit.net/.

12 findr[EB/OL].[2021－02－26].https://1findr.1science.com/.

13 赵蓉英，陈烨.学术搜索引擎 Google Scholar 和 Microsoft Academic Search 的比较研究[J].情报科学，2014，32(2)：3－6，15.

14 Hug S E，Brändle M P. The coverage of Microsoft Academic：analyzing the publication output of a university[J]. Scientometrics，2017，113(3)：1551－1571.

15 Thelwall M. Microsoft Academic：a multidisciplinary comparison of citation counts with Scopus and Mendeley for 29 journals[J].Journal of Informetrics，2017，11(4)：1201－1212.

16 Kousha K，Thelwall，Abdoli M. Can Microsoft Academic assess the early citation impact of in-press articles？a multi-discipline exploratory analysis[J]. Journal of Informetrics，2018，12(1)：287－298.

17 Visser M，Van Eck N J，Waltman L. Large-scale comparison of bibliographic data sources：Scopus，Web of Science，Dimensions，Crossref，and Microsoft Academic [EB/OL]. [2021－02－26]. https://arxiv.org/abs/2005.10732.

18 Orduña-Malea E，Delgado-López-Cózar E. Dimensions：re-discovering the ecosystem of scientific information [J]. El Profesional de la Informacion，2018，27(2)：420－431.

19 Thelwall M. Dimensions：a competitor to Scopus and the Web of Science？[J]. Journal of Informetrics，2018，12(2)：430－435.

20 Harzing A W. Two new kids on the block：how do Crossref and Dimensions compare with Google Scholar，Microsoft Academic，Scopus and the Web of Science？[J]. Scientometrics，2019，120(1)：341－349.

21 Martín-Martín A，Thelwall M，Orduna-Malea E，et al. Google Scholar，Microsoft Academic，Scopus，Dimensions，Web of Science，and OpenCitations' COCI：a multidisciplinary comparison of coverage via citations[J]. Scientometrics，2021，126(1)：871－906.

22 谢智敏，郭倩玲.基于深度学习的学术搜索引擎——Semantic Scholar[J].情报杂志，2017，36(8)：175－182.

23 Valenzuela M，Ha V，Etzioni O. Identifying meaningful citations：WS-15-13 [R]. AAAI Workshop-Technical Report，2015：21－26.

24 Tay A. More cutting edge—Research tools for researchers—Oct 2020 by Aaron Tay [EB/OL]. [2021－02－26]. https://aarontay.medium.com/more-cutting-edge-research-tools-for-researchers-oct-2020-by-aaron-tay-b90b56f3dfbd.

25 Tay A. 7 reasons why you should try Lens.org (updated to version Release 5.16.0-March 2019)[EB/OL].[2021－02－26].https://aarontay.medium.com/6-reasons-why-you-should-try-lens-org-c40abb09ec6f.

26 刘盛博，丁堃，唐德龙. 引用内容分析的理论与方法[J]. 情报理论与实践，2015，38(10)：27－32.

27 刘盛博，丁堃，张春博. 引文分析的新阶段：从引文著录分析到引用内容分析[J]. 图书情报知识，2015(3)：25－34.

作者单位：浙江大学图书馆，浙江杭州，310027
收稿日期：2021 年 7 月 27 日
修回日期：2021 年 10 月 23 日

（责任编辑：关志英）
（转第 77 页）

—25.
22 周卓.回归与复兴:从现代纸书看出版融合的新方向[J].中国编辑,2020(Z1):97—101.
23 毕若旭.范璞文:短视频让名著阅读更有趣[EB/OL].[2021—07—15]. http://news.cyol.com/app/2020—12/24/content_18896502.htm.
24 官凤婷."使用与满足"理论视角下高校图书馆移动短视频服务机制构建[J].新世纪图书馆,2020(7):58—64.

作者单位:南京农业大学信息管理学院,江苏南京,210095
收稿日期:2021 年 8 月 17 日
修回日期:2021 年 9 月 25 日
(责任编辑:关志英)

Promoting the Integration of Online Video and Reading Services

Mao Yihong

Abstract: Media form is an important external factor affecting reading behavior. Online video has increasingly become a dominant media on the Internet, which is reshaping our reading behavior and reading service. Libraries and other reading service institutions should face up to the coming of the era of network video and grasp the trend of the integration of network video, reading and reading services. Network video brings not only new channels or new tools, but also a comprehensive reshaping of reading materials, reading behaviors and reading services. Its positive performance is to give full play to the advantages of network video to provide independent video reading services, or to carry out multi-media reading services in combination with audio and graphic reading services. At present, we need to actively promote the integration of online video and reading service, clarify the development direction, actively explore the application of online video in reading service, strive to establish or optimize the video reading service system, and constantly forge the ability of video reading service.

Keywords: Network Video; Short Video; Live; Reading Promotion; Reading Service; Smart Library

(接第 55 页)

A Comparative Study of Emerging Science Citation Index Databases

Li Jie Meng Ye Jin Jiali Yao Xiangyu Huang Chen

Abstract: With the development of web crawler technology, artificial intelligence-based text semantic analysis technology, and newer research paradigms, more emerging scientific citation index databases have come out. These databases provide a new experience for researchers to search and filter academic literature resources and create more scientific research evaluation options. Seven science citation index databases-Microsoft Academic, Dimensions, Semantic Scholar, Lens, Scinapse, Scilit, and Scite—were chosen to conduct comparative research and case analysis. This paper evaluated and explored collection coverage, indicators, artificial intelligence technology, visualization analysis, plug-ins, and export features of each database. These functions allow librarians to have more references and choices when doing scientific research evaluation. Based on investigations and case studies, it can be concluded that Microsoft Academic and Dimensions are the best choices for evaluating institutions and scholars when taking an enormous amount of papers coverage and optimal accuracy into account.

Keywords: Science Citation Index; Citation Analysis; Semantic Citation; Academic Influence Evaluation