



基于文献情报学的学术热点分析方法研究

□ 杨新涯 王莹* 尹伟宏

摘要 数据的海量增长,伴随而来的是信息贫乏时代向信息爆炸、信息富余时代的转变。领域前沿主题的识别工作不仅是学科领域关注的重点,也是图书情报领域研究的热点问题。文献调研发现,当前高校图书馆对学术热点的研究重视不足。聚焦数据获取、数据整理、数据分析和数据呈现等四个环节,形成文献情报学的学术热点研究方法体系,旨在综合运用图书馆的方法、众多的数据库资源和不同功能的分析平台,实现有效的学术热点数据管理,为科研人员提供产生高影响力研究成果的指导与参考。

关键词 学术热点 数据获取 数据分析 数据管理

分类号 G252

DOI 10.16603/j.issn1002-1027.2020.03.005

1 引言

学术热点,包括研究前沿、发展趋势等概念,往往是某一学科或领域在某一时间段被高度关注并产生较大影响力的研究方向,表现为该学科或领域细分后的研究主题。由于其创新性和未来发展空间,学术热点是科研人员把握科研动向和需求的重要参考。图书馆作为文献集散中心,其研究核心就是节省在文献、信息和知识的发现、获取、阅读、理解和消化等方面的时间,节省查准文献的时间,增加有效阅读时间^[1]。科研人员作为一大用户群体,其科研需求应该得到满足。

过去,高校图书馆对科研人员的支持主要体现在基本的文献资源保障。大数据及相关的科学数据、数据密集型科学发现范式的出现,加之众多数据库分析平台的兴起,为使用已有数据进行新知识的生产提供了可能^[2],图书馆事业重心也发生了新的变化,从“资源为王”转向“服务为王”,而“服务为王”的重要基础则来源于“内容为王”。当下,图书馆的多元化数据正成为“内容为王”的重要支撑,图书馆界亟待提升对所拥有的多样化数据的理解、分析和挖掘等能力,实现与目标用户的有效对接,提供精准服务。

本文是对学术热点研究的流程及各环节使用的研究方法进行探索的新尝试,着眼于学术热点研究中方法的综合运用和流程操作,使用不同的方法来

解决各环节遇到的问题,兼具针对性和综合性。

2 学术热点研究现状

准确把握学科研究热点和发展趋势,对科研人员的研究和管理者的决策具有重大意义。科研数据的海量升级,凸显了学术热点研究的现实意义。上海科技情报研究所在梳理科学知识图谱的发展中发现,科学知识呈爆炸式增长,吸引了一批数学、计算机、图形学等领域专家进入图书情报分析领域,与科研情报工作者协同合作,开发并研究出大量的新工具和新方法,形成了众多数据库平台和分析工具。Scopus 数据库显示,全球关于学术热点的研究正逐年增长。中国知网数据也表明,国内关于学术热点的研究成果在2006年以后出现明显增长,主要分布在图书馆学、经济学、医学、计算机科学及相关交叉学科。

在分析工具方面,针对学术热点研究的数据来源,Web of Science、Scopus、中国知网等均成为科研人员检索文献的重要平台。其他如 MEDLINE 等专业数据库也成为重要的专业数据获取平台。针对学术热点研究的评价与分析工具,Incites 的学科分析、ESI 的前沿领域、Scival 的热点关键词云图、CNKI 的文献计量分析等均提供不同程度、不同角度的量化分析,促使学术热点的研究更加精细化、精准化。另据不完全统计,目前国内外已开发的免费

* 通讯作者:王莹,ORCID:0000-0002-5560-4456,邮箱:201702021129@cqu.edu.cn。



供科研人员使用的文本分析工具有数十种之多,如CiteSpace、VOSviewer、Histcite等。

在研究方法与成果方面,蒋文钊等以Web of Science数据库为数据源,借助OmniViz可视化数据分析工具和Excel软件,对文献进行聚类分析和词频分析,研究国际运筹学领域的学术热点^[3]。喻国明等结合文献计量法中的内容分析法和共词分析法,对EBSCO数据库中收录的2016年的人工智能专题中与传播学相关的英文学术论文做定量分析,描绘出该领域的研究热点^[4]。李峰等综合使用文献调研法、文献计量法、德尔菲法、指标对比分析法和可视化分析方法等多种方法研究基础和交叉学科领域的学术热点,为学校职能部门的决策提供依据^[5]。

在实践应用方面,科睿唯安等数据库商依据自身资源,加入学术热点研究的行列,自发布《2013研究前沿——自然科学和社会科学的前100个探索领域》白皮书后^[6],2014年起又与中国科学院战略情报研究团队合作,连续5年发布当年研究前沿报告。

综上所述,多样化的数据库平台和分析工具为学术热点的研究提供了保障。随着文献计量学方法和众多新分析工具与平台的应用,学术热点的研究具备了充足的理论支持和具体的实践案例,但是仍存在对其价值认识不足、重视不够、研究方法不够系统等问题,有待进一步研究。

3 学术热点研究当前存在的问题

3.1 对学术热点在学科服务中的价值认识不足

对国内高校图书馆而言,在“双一流”的建设背景下,依托全球范围内的科研数据,提供与国际接轨的科研服务、决策支持,是其研究性、学术性定位的重要体现,在大学生态圈中具有不可替代的作用。当前,高校图书馆的学科服务仍集中于学科的影响力评价、院校对标分析,主要还是在对已有科研产出的评价研究,缺乏对学术热点研究的重视,因而也缺乏能力与竞争力,利用自身资源优势,率先淡化学科边界,最大限度融入科研过程并且正面影响科研过程。

3.2 对学术热点研究工具的培训支持不够

科研数据的指数级增加,导致科研人员难以在专注自身研究的同时,花费大量时间和经费兼顾权威性学术热点的发现及追踪。因此,对科研人员进行针对性的信息素养专业培训显得尤为重要和迫切。而现有的信息素养培训主要集中于文献检索的

基本知识与技能,难以应用于多样的数据库及分析平台的检索和分析,难以满足日益精准的科研需求,加之学科交叉愈加明显,数据库商产品竞争激烈,导致检索数据不准确甚至有重大遗漏。

3.3 学术热点的研究方法系统性不足

通过文献梳理发现,现有学术热点研究的实践多依赖于某一综合性数据库和某一分析软件的结合,科研产出也主要集中于结果阐述,对于如何发现等研究方法相关论述较少。从这些较少的关于研究方法的论述中可以发现,在学术热点研究的过程中,数据来源单一、数据分析维度单一、分析方法相对独立使用等均直接影响数据分析效果。如从时间维度关注学术产出(Papers)、论文被引频次(Cites)、篇均被引频次(Cites of Per Paper)、h指数等常见指标,可以发现一定时间内各项指标的起伏变化。而对于学术热点研究,科研人员更关心在此基础上的其他维度,如全球关注度、时效性、新颖性等有效指标。

此外,学术热点往往暗含预测未来的功能,发现数据背后的关联进而预测发展趋势对科研人员更有意义。独立指标的观察或某两者的简单交叉,难以捕捉其中的内在联系。伴随众多热门领域向交叉方向、综合方向的延伸,单一的数据库数据、简单的分析指标和相互独立的分析方法难以满足捕捉学术热点的需求,需要打破各环节研究方法系统性不足的现状,组合一套专业化流程与简单操作,为不同专业的科研人员所用。

4 学术热点的研究方法论构建

方法论的基本意义在于提高人们对周围世界的认识,其中包括采用怎样的理论去认识和解释世界,以及关于方法本身的理论^[7]。学术热点研究方法论的构建,是图情领域人员在当前科研环境直接或间接为科研人员提供精准服务的重要内容之一。其核心在于,系统整合已有研究方法和工具,为科研人员提供操作简洁、结果科学的针对性研究方法体系;重点在于提炼出解释和发现学术热点的方法和流程。

4.1 量化分析与质性分析相结合

研究方法直接影响一项研究的信度与效度。对于学术热点研究而言,核心离不开多样化的科研数据。本文立足于文献情报学已有方法和文献发展规律,充分利用图书馆的多种数据库资源和分析平台,解释并分析当前热门领域全球学术热点研究现状及



趋势,为该领域科研人员提供真实、可靠、直观、具体的热点研究方向,助力科研。对于科研数据的研究,综合运用量化分析和质性研究等社会科学研究方法,促使分析结果更加趋近事实本身。同时针对不同的科研需求,量化研究和质性研究应针对性地进行不同程度的结合。

对于高影响力论文选题等科研需求,从量化角度来讲,合理选择、组合运用中英文数据库收集原始数据,获得与学术热点研究相关的原始数据支撑。在收集数据过程中,通常用到的数据检索方法有布尔逻辑检索、截词检索等,关键在于检索词的选择和检索策略的运用。获取数据后,通过建立统一的标准,确定分析的维度,综合运用数据分析平台如 ESI、Scival 和文本分析工具,多维度描述原始数据,快速、高效、科学地从众多数据中梳理热点研究范围。对得到的数据分析结果,需借助文献发展规律、共被引分析方法、文献耦合规律等进行客观、科学的解释。从质性角度来讲,除获取原始权威数据本身外,还需收集研究领域在某一研究时间范围内的高水平论文。通过对高被引文献或核心文献等内容的整合分析和综合解读,形成有价值的综述、述评、文摘、提要等二次文献,作为数据分析结果的重要补充。二者相辅相成,缺一不可。

对于科研项目申请、学术热点趋势权威发布等,还需推荐给该领域专家,进行学者研讨、专家评议甚至投票,采用德尔菲法、专家访谈等方法,吸收专家意见,进一步确定具体研究问题、内容等。

4.2 以数据为核心的标准化流程控制

学术热点研究是一个流程化的过程,涉及不同的数据管理过程,从数据获取的多元性、数据处理的标准化、数据分析的深入化到数据呈现的精准化。围绕其核心关键词“数据”,可将该过程分为数据获取、数据整理、数据分析和数据呈现等四个标准化流程,如图 1 所示。

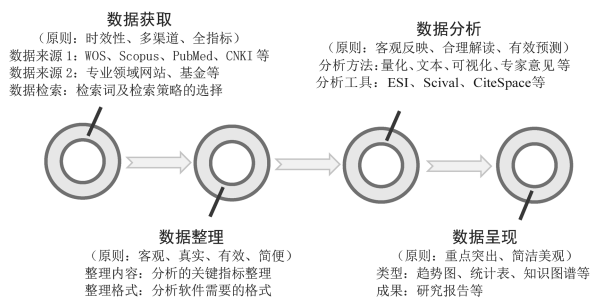


图 1 学术热点的研究方法与流程

数据收集流程中,围绕时效性、多渠道、全指标^[8]等原则,全面采集包括学术产出(Papers)、被引频次(Citation)、引文影响力(Citation Impact)、H 指数等常见指标在内的多维指标,如媒体关注度、研究前沿热度指数等不同数据库个性化的评价指标,明确各指标含义。同时对于项目申请等科研需求,除常见数据库数据外,还需收集各领域专业网站数据、相关基金网站数据等。

数据整理流程中,围绕客观反映、真实有效、方便操作等原则进行数据的内容整理、格式处理,具体包括数据预处理、数据内容中关键指标整理汇总和根据分析软件要求进行文件格式转换等。

数据分析流程则需遵从客观反映、合理解读、有效预测原则,注重量化分析软件、可视化软件和文本分析工具的综合运用,注重专业领域知识框架的学习,避免解读不足、过分解读或解读偏差等不当分析现象的发生。

最后是数据呈现流程,注重内容的简洁性,必要时辅之以专家意见,共同为科研人员提供学术热点内容。

5 计算机辅助诊断领域学术热点的实证研究

通过 Web of Science 核心合集数据,对计算机辅助诊断领域进行整体了解。数据显示,该领域覆盖学科 76 个,研究方向 46 个(均为 WOS 分类标准下),一方面研究方向较为宏观、涉及范围较广,如数据显示计算机辅助诊断领域的研究方向有 Engineering(工程学)、Computer Science(计算机科学)等,范围较为笼统;另一方面,WOS 学科、研究方向的分类标准与国内高校实际学科、研究方向不一致,难以精准映射。由此可见,当下对于计算机辅助诊断领域科研人员而言,最困惑的不再是没有科研方向的选择,而是面对众多科研方向难以把握,面对宏观大方向,如何细化管理、获得精准的学术热点分析结果。

5.1 数据获取

数据是学术热点分析的起源和支撑,科研数据的来源直接影响分析结果。因此,甄别和选取数据来源尤为重要。以“计算机辅助诊断(Computer Aided Diagnosis)”为关键词进行主题检索,对比不同数据库平台 2014—2018 年、文献类型为 Article 和 Review 的检索结果,相关统计见表 1。其中,数据库平台未提供限定主题检索的,则选择限定关键词、摘要等字段组合,以最大程度靠近限定主题结果。



表1 “计算机辅助诊断”领域不同数据库检索结果统计

名称	结果/条	简介
WOS 核心合集	2612	综合性文摘索引数据库。收录了18000多种世界权威的、高影响力的学术期刊,内容涵盖自然科学、工程技术、生物医学、社会科学、艺术与人文等领域,最早回溯至1900年。核心合集收录了论文中所引用的参考文献,并按照被引作者、出处和出版年代编制成独特的引文索引
Scopus	4169	文摘和引文数据库。收录了来自全球5000多家出版商的18000多种同行评审期刊、500多种会议录、600多种商业期刊以及125种丛书。内容涉及人文、科学、技术及医学方面的文献
Springer Link	9278	综合性外文期刊全文数据库。收录了Springer出版的1800余种电子期刊
Wiley Online	175	高质量学术期刊数据库。科技、医学为主

对比发现,由于定位、性质不同,不同数据库数据数量相差悬殊,其次由于元数据切割标准不同,检索功能设计存在差别。如Wiley Online期刊数据库中限定字段仅有全字段、标题、作者和关键词字段,且不同字段间不能使用布尔逻辑的“或”检索方式,以检索“Computer Aided Diagnosis”为例,检索全字段可获得14918条数据,而限定为关键词时,数据量锐减至175条,极大限制了数据选择。因此,科研人员进行学术热点的研究数据准备时,需要根据需求,综合对比不同数据库的数据情况,确定原始数据来源。

此外,在数据收集环节需注意抽取检索词并构建恰当的检索式。对于需精确检索的内容,科研人员需确定合适的限定字段,如限定摘要、主题、篇名/题名、作者或机构等。一般数据库均提供布尔逻辑检索选择,操作简便。对于非精确检索内容,通常在

检索框附近位置会有系统提供的相关检索词,如Ei Compendex检索结果左侧会根据检索痕迹,提供Controlled Vocabulary(控制性词汇)供参考。对于陌生领域,需借助熟悉的数据库检索该领域内的代表性文献,如高水平论文或该领域专家的学术成果等,提取3—5个关键词进行检索。

5.2 数据整理

不同分析工具在分析数据过程中对数据文件的格式有不同要求,为方便数据分析,需将原始数据转换为适用分析的数据类型,并保存为通用格式,通常包括CSV、BibTeX等。表2给出了常见的英文数据库数据获取规则, Springer、Wiley数据库需要登录进入才可以获取数量,因此未能获取。

5.3 关键环节:数据分析

数据分析包括对数据进行总体描述和具体分析两部分。在总体描述中,各数据库会提供如学科分布、时间分布、区域分布、机构分布、作者、来源期刊等维度的整体统计。以“Computer Aided Diagnosis”检索结果为例,Scopus数据库给出年份、作者、学科、国家/地区、来源出版物、出版阶段、归属机构、资金赞助商和语言等维度的描述统计包括图表界面,以便科研人员快速、清晰地了解该领域整体情况。其他数据库平台如Web of Science会提供领域中高被引论文、热点论文等精炼选项,方便科研人员聚焦高水平论文。

在具体分析时,不同软件提供了不同维度的分析,如ESI的前沿领域分析,Scival的热点关键词云图,CNKI的学术热点搜索等。以“Computer Aided Diagnosis”为例,在ESI中限定检索为“Research Fronts”,筛选领域包括计算机科学、临床医学等领域,保存文件格式为CSV,可通过Excel数据筛选工具找出相关度较高的该领域前沿研究方向。

表2 常见英文数据库的数据获取规则

数据库	下载选项	下载内容	数量要求
WOS	EndNote、Incites 和其他文件类型如BibTeX、HTML、纯文本、制表符分隔(Win)等	作者、标题、来源出版物、摘要、全记录、引用的参考文献等	每次500条
Scopus	MENDELEY、Refworks、EndNote、CSV/Excel、BibTeX等	引文信息(作者、标题、年份等)、题录信息(归属机构、ISSN等)、摘要和关键词、资金资助详情等	CSV仅支持前2000篇
Springer Link	Bibtex、EndNote、acmref、csv	数据库收录的全记录	
Wiley Online	文本、XML或Email、保存到EndNote等	题名、作者、机构、实验室、关键词、来源、基金、文摘、被引频次等供选择	



分析结果显示,ESI平台前沿分析的局限在于不能直接进行特定领域分析,需通过ESI学科分类体系进行与主题领域相关的学科限定,获得前沿研究方向列表后再进行筛选。在筛选过程中,领域交叉性过强,而学科涵盖不全时,容易造成遗漏。因此,ESI常用于学科前沿检索,进一步的研究主题或领域的分析还需借助其他工具。

Scival平台可进行学术热点关键词检索。进入Overview界面下的Topics and Research Areas检索“Computer Aided Diagnosis”,可聚焦该平台的主题词划分结果,如T.55899:Computer Aided diagnosis; Cytology; Microscopic Images,选择涵盖主题词的选项进行热点关键词分析。需注意在检索中,当检索词缺乏匹配结果时,可通过Web of Science、Scopus数据库等获取目标文献集合的DOI、ID编号等,在Scival平台的Overview界面,根据Publication Sets中创建新数据集的步骤提示,导入目标文献数据集合后再进行分析。除此之外,Scival数据分析平台还提供了Viewed即关注度的分析维度,也是学术热点分析的重要参考指标。

CNKI学术热点搜索等分析工具可提供不同分析视角。由于自然科学领域的研究成果多以外文形式发表,以“计算机辅助诊断”结果为例,中文数据量偏小,分析结果仅供参考。

通过对比不同的分析平台发现,单一的某一项工具难以满足多样化的学术热点研究需求,包括发表高影响力论文、申请项目、人才引进、学校职能部门决策等,因此面对不同需求甚至同一需求的不同阶段,需要系统性地综合运用不同的学术热点研究方法和工具,择优实现需求目标。

5.4 数据呈现

对于某一领域的学术热点分析结果,需根据科研需求选择呈现方式,如某领域学术热点的研究报告、热门研究主题表、冷门研究主题表等。以研究报告为例,通常需介绍目标领域的研究概况和学术热点研究情况。其中,研究概况包括该领域全球范围内在某一时间段的论文数量、被引频次、引文影响力等基本信息,也可根据需要介绍学科分布、国家地区分布、机构分布、作者分布、来源期刊分布等。学术热点研究情况主要包括量化分析结果和内容分析结果两个部分。量化分析结果中需提炼出科研人员应密切追踪的热

门研究主题或应避免的研究趋冷的主题。内容分析结果包括对领域内高水平论文的研究问题、方法及创新点等主要内容的整理总结,也可以纳入专家意见,共同为科研人员提供具体思考和科研指导。

6 总结与展望

实践表明,甄别数据库、获取目标数据是学术热点研究的基础,分解研究需求、综合运用不同分析工具则是关键。本文构建的学术热点研究方法体系弥补了单一平台分析的不足,同时通过实践,形成了领域学术热点的研究报告,为科研人员提供参考,也为学校职能部门的决策提供依据。当前,不少国内高校图书馆正在参与制定新的科研数据管理政策和服务,并将其视为未来服务的重要组成部分^[9]。基于图书馆海量数据的高端服务将是新时代图书馆服务的重要板块,学术热点研究则为支持科研工作、开展高端服务打开新的大门。讲好图书馆故事,从深化学科服务、重视学术热点的研究开始。

参考文献

- 1 陈有志,郑章飞,刘平,等.“创新与发展:新时代的图书馆与图书馆学”高端论坛会议综述[J].大学图书馆学报,2018,36(3):5-13.
- 2 司莉,曾粤亮.需求驱动的大学图书馆发展趋势研究[J].大学图书馆学报,2018,36(3):30-40.
- 3 蒋文钊,区晶莹,俞守华.国际运筹学学术热点研究[J].现代计算机,2015(7):33-38.
- 4 喻国明,梁爽,程思琪.当前国际传媒领域人工智能研究的学术热点与框架——基于EBSCO数据库2016年英文文献的文献计量学方法[J].东南学术,2018(2):55-65.
- 5 李峰,马芳珍,刘雅琼,等.面向基础和交叉学科的科研战略发展热点研究[J].大学图书馆报,2018,36(2):37-44.
- 6 中科院发布《2018研究前沿》[EB/OL].[2018-12-24].http://cx.xinhuanet.com/2018-10/25/c_137556756.htm.
- 7 仇立平.社会研究方法[M].重庆大学出版社,2008.
- 8 范少萍,安新颖,晏归来,等.医学领域前沿主题识别方法研究[J].情报学报,2018,37(7):686-694.
- 9 同2.

作者单位:杨新涯,重庆大学图书馆,重庆,400044

王莹、尹伟宏,重庆大学经济与工商管理学院,重庆,400044

收稿日期:2018年12月3日

修回日期:2018年12月26日

(责任编辑:支娟)



Research on Academic Hotspot Analysis Method Based on Literature and Information Science

Yang Xinya Wang Ying Yin Weihong

Abstract: The growth of massive data is accompanied by a shift from information scarcity era to information explosion and information surplus. The identification of the frontier topics in the research fields becomes not only the focus of the subject area, but also a hot issue in the field of library and information science. Through literature research, it is found that the current research on academic hotspots in college libraries is insufficient. Focusing on data acquisition, data processing, data analysis and data presentation, an academic hot research method system has been developed for literature and information science. The aim is to comprehensively use the library method, numerous database resources and different functional analysis platforms to achieve effective academic hotspot data management and provide guidance and reference for researchers to produce high-impact research results.

Keywords: Academic Hotspots; Data Acquisition; Data Analysis; Data Management

封面照片简介:国防科技大学图书馆

国防科技大学图书馆创建于1953年,其前身为哈尔滨军事工程学院图书馆,1970年随学校主体南迁长沙,1999年原长沙炮兵学院、长沙工程兵学院、长沙政治学院并入国防科技大学,三个学院的图书馆也先后并入国防科技大学图书馆。目前,图书馆有三座馆舍,分别坐落在一、三、四号院区,建筑总面积4.7万平方米,阅览座位4300余个,每周开放90个小时。

国防科技大学图书馆建立了以理工类重点学科为核心,以军事文献为特色,理、工、军、文、管、经、哲协调发展的文献资源体系。现有印刷型文献370余万册(件),年订购印刷型中文报刊3300余种/份,印刷型外文期刊170余种/份。建成了101个大型综合数据库,实现了SCI、EI、CPCI-S三大系统及其他各类外文文献数据库的国际同步检索,与国内外图书情报机构建立了广泛的资源共享与合作关系。坚持创新读者服务模式,通过实体馆、数字馆及微信公众号为读者提供文献信息服务。

国防科技大学图书馆坚持“以人为本,为战育人”的服务宗旨,紧密围绕学校“双一流”建设总体要求,拓宽服务渠道,创新服务模式,推进军校特色阅读文化建设,为全校提供用户培训、参考咨询、科技查新、文献传递等深层次服务。国防科技大学图书馆依托CALIS军队院校文献信息服务中心,为军队院校全面提高科技创新能力和国际科技竞争能力提供优质高效的文献信息资源和服务保障。

国防科技大学图书馆积极响应“全民阅读”号召,努力打造并全面推进“蕙风讲坛”等四大阅读推广品牌建设,并于2019年荣获中国图书馆学会“全民阅读先进单位”、湖南省教育厅普通高等学校阅读推广活动优秀组织奖及创新案例奖等各类荣誉。