

基于主题分析的交叉学科科技文献 资源遴选方法研究*

——以蜜蜂学为例

卢垚 王鸞飞 马鑫

(中国农业科学院农业信息研究所, 北京 100081)

摘要: 随着科研用户对信息资源需求的多元化和个性化,尤其是对交叉学科和新兴学科资源的需求日益旺盛,科研院所专业图书馆资源建设仅依靠传统的文献资源需求分析与遴选方法已无法满足用户的科技创新需求。本文以中国农业科学院(以下简称“中国农科院”)蜜蜂学2017—2019年发文及其引文为研究对象,开展文献资源需求分析和资源遴选方法研究。通过对文献进行主题组织、分析遴选相关主题,把握该学科细微的文献资源需求;再利用主题关键词提取、资源内容匹配和质量评估等系列方法,遴选出与该机构研究内容高度契合的高质量期刊和核心论文,为其科研人员提供精准的文献资源服务。本研究基于主题分析创新了科技文献需求分析和资源遴选方法,打破学科限制,能够精准定位相关内容的资源,为科研院所专业图书馆交叉学科的资源建设和服务提供新的思路。

关键词: 交叉学科;需求分析;主题组织;资源评估;内容匹配

中图分类号: G258.5; G254.2 **DOI:** 10.3772/j.issn.1673-2286.2020.11.005

引文格式: 卢垚,王鸞飞,马鑫.基于主题分析的交叉学科科技文献资源遴选方法研究——以蜜蜂学为例[J].数字图书馆论坛,2020(11):33-41.

随着科学研究的不断深入,不同学科间的沟通、交互、融合日趋深化,科研人员跨学科合作也越来越频繁,学科之间的关联度越来越高,学科发展的边界也日益模糊。当某一领域跨学科研究积累足够数量的知识点,并相互建立起成熟的关联和逻辑时,就形成了交叉学科。因此,交叉学科就是指两门及以上学科互相结合、彼此渗透交叉形成的新学科^[1]。近年来,国内外科学界一直倡导跨学科研究和促进新兴交叉学科发展^[2-4],而文献资源保障是交叉学科建设发展的基础条件,交叉学科的研究需要多学科领域的知识支撑,其文献需求也有别于传统学科,交叉学科信息资源建设对馆藏资源建设的意义也越来越重要^[5-7]。

1 研究背景

1.1 交叉学科信息资源建设的研究情况

很多图书馆和文献服务机构围绕交叉学科信息资源建设开展了一系列的研究和实践。如国家科技图书文献中心(National Science and Technology Library, NSTL)资源组通过定义交叉学科分类规则来组织相应学科的科技文献,其构建的《NSTL文献信息资源学科分类主题列表》根据交叉学科的研究对象及研究内容的需要,在《中图法》交替类目的基础上调整了交叉学科的应用规则,能够对交叉学科的科技文献进行很好组织^[8]。如对农业经济学这类自然科学和社会科学融合产生的学科、农业生物工程这种自然科学与技术科学交叉学科的相

*本研究得到中央级公益性科研院所基本科研业务费专项“中国农业科学院交叉与新兴学科文献资源保障体系建设研究”(编号:Y2020YJ04)资助。

应文献都能按此分类表归入适当的类目下。

该领域多数研究主要集中在交叉学科文献资源需求分析方面,利用不同的方法,分析交叉学科不同层面的文献资源需求。如厦门大学图书馆对交叉学科研究者进行调研,分析其图书借阅流通、馆际互借数据,尤其是中图分类号的分布情况,了解相关研究者的文献需求及文献使用特点,从而优化交叉学科的文献资源建设^[9]。沈丹等^[10]利用引文分析从期刊引文数据入手,将期刊间的引用转化为学科间的引用,分析学科间的交叉引用情况,从整体学科层面发现用户对交叉学科的资源需求。同样是基于引文分析,王婷等^[11]提出了期刊关联法,是对某一学科最具代表性期刊的参考文献进行共引分析,将共引期刊间引证强度进行量化,判断其相互联系紧密度,从而找到相关的期刊。东华大学图书馆则从学科竞争力分析的角度借助科研评价工具,发现该校具有竞争优势的交叉学科领域范畴,为馆藏资源建设提供宏观导向和针对性的意见^[7]。

这些研究从信息组织和需求分析方面,为交叉学科信息资源建设提供了实践应用指导。而在实际工作中,即便是掌握了学科需求,但由于期刊资源一般按照学科进行分类和组织,并且许多新出现的新兴学科和交叉学科研究领域在现有的期刊分类学科体系中未能及时体现,使得交叉学科期刊资源遴选仍面临困难。

1.2 农业交叉学科信息资源建设的困境

对于农业科学而言,交叉学科的文献资源建设一直都是保障的薄弱环节。农业科学因其研究对象多样,涉及的学科范围广泛、门类繁多,其中既有侧重基础理论的,也有侧重应用技术的,具有很强的学科交叉属性。农业学科科技文献资源远远超出了作物学、园艺学、兽医学等经典农业学科的范畴,涉及生物学、化学等基础学科,工程学、计算机科学等应用学科,经济学等人文社会学科。如果对这些所涉学科的资源按照学科为单元进行建设,需要巨大的建设成本,并且针对性不强,资源利用也将不充分。此外,科研人员所需的信息资源不仅在学科层面越来越多元化,在面对海量的资源时,用户需求也越来越个性化、精准化,即使是同一学科的资源,不同用户的需求也存在很大的差异。因此,仅以学科为单元的文献资源需求分析与遴选方法一定程度上已无法满足交叉学科发展要求,打破学科的界限,进行更细粒度的文献信息组织,探索新的资源遴选方法,

为科研人员提供更精准的学科资源,成为交叉学科信息资源建设亟待解决的问题。

1.3 按主题遴选期刊资源迎来机遇

主题组织是将具有相同主题特征的文献以表示其主题的受控词汇作为标识标引、描述、存贮和检索文献的方法^[12-14]。主题组织虽是图书馆信息资源管理常用的组织方式,但如何按照主题遴选期刊资源是新的议题。一方面,Scopus数据库利用文献间的直接引用关系识别相同研究主题的文献进行聚类,并将每一类用主题词表进行标引,形成了9.6万个主题组成的稳定的组织结构,并且Scopus中95%的文献都有对应准确的唯一主题^[15]。这就为利用发文引文进行文献需求分析提供了一种比学科粒度更细的文献组织方式和良好的大数据基础。另一方面,爱思唯尔的指纹引擎技术,采用先进的自然语言处理(NLP)算法从非结构化文本中提取信息,配合领域相关的主题词表的应用,创建加权术语索引定义文本,即“指纹”,再通过聚合和比较指纹,能够实现语义层面的内容匹配^[16],这也为在主题层面进行需求和资源匹配提供了技术上的可能。

2 研究对象和方法

2.1 研究对象和数据来源

本文拟选择中国农科院蜜蜂学为对象,针对其开展科技文献资源遴选方法研究。从中国农科院蜜蜂学主要研究群体蜜蜂研究所(以下简称“蜜蜂所”)2017—2019年发文及引文着手,对文献进行主题组织和分析,遴选出与该所科研人员研究内容高度契合的高质量期刊和核心论文。以典型的农业交叉学科作为对象,研究科技文献需求分析和资源遴选新方法,拟为科研人员提供精准的科技文献资源,也为科研院所专业图书馆的信息资源建设和服务提供新的思路。

中国农科院蜜蜂学学科致力于蜂学的理论与应用研究,其设置有7个主要研究方向,包括传粉蜂生物学与授粉应用、蜂产品加工与功能评价、蜂种质资源与育种、蜂产品质量与风险评估、蜜蜂蛋白质组学、蜜粉源植物产地环境识别与控制、蜜蜂病虫害生物学,具有专业性强且相对小众的特点,涉及生物学、昆虫学、生态学、蜜蜂病虫害、蜂产品加工等多个学科,是典型的交

叉学科专业。通过对蜜蜂所2019年发文进行分析发现所涉学科分布广泛且极其分散(见图1), 按照Scopus数据库学科分类体系, 主要涉及农业和生物科学, 生化、遗传和分子生物学, 化学等一级学科; 各一级学科内又涉及多个二级学科, 如农业和生物科学涉及食品科学, 昆虫学, 生态、进化、行为和系统学, 植物学, 农学

和作物科学等。而无论是一级学科还是二级学科都很难与蜜蜂学科主要研究方向进行映射, 若以学科为单元对蜜蜂所及其主要研究团队进行文献需求分析, 很难精确定位其所需的资源, 因此需要将学科资源按照更细粒度进行组织遴选。

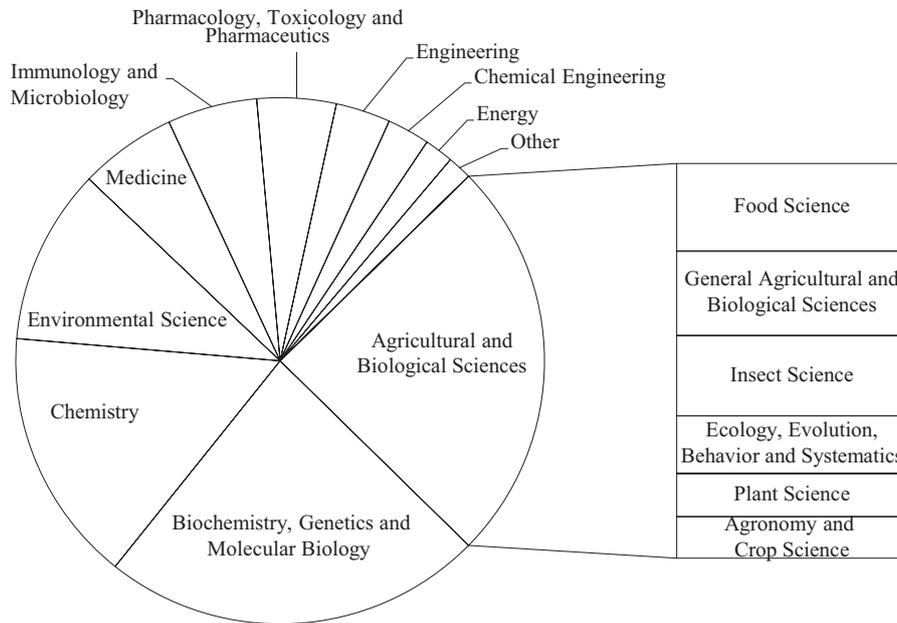


图1 中国农科院蜜蜂学发文学科分布

本文研究数据来源于Scopus引文索引数据库和Scival数据库, 通过机构检索从Scopus数据库中采集中国农科院蜜蜂所2017—2019年发文272篇, 引文11 836篇, 同时采集文献所属主题信息。各主题所涉全球论文数量、主题显示度等数据则采集自Scival数据库。

2.2 研究方法

将上述蜜蜂学科发文引文按主题进行组织并遴选, 从期刊和论文两个层面开展资源内容和质量评估(见图2), 其中期刊的遴选又分为来源期刊分析和期刊内容匹配两条路径, 各关键环节具体方法如下。

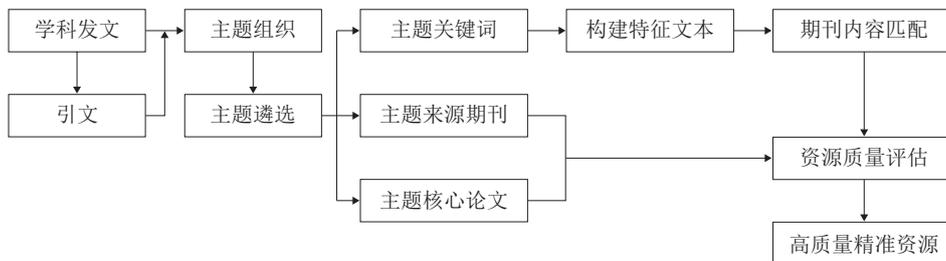


图2 基于主题分析的文献资源遴选方法技术路线

(1) 主题组织和遴选。本文利用Scopus数据库对文献按照引用关系进行聚类并用学科主题词表标引的主题组织方法^[15, 17], 继承数据库对蜜蜂学科发文、引文的主题组织结果, 并与该学科主要研究方向进行映射。随后, 分析蜜蜂学科发文及其引文在各主题的分布

情况, 综合比较发文/引文在主题内全球论文中的占比以及主题显示度指标, 遴选出蜜蜂学科发文/引文集中度、贡献度相对较高并且在全球范围内被关注程度较高的、发展势头较好的主题。

(2) 主题关键词识别和分析。针对遴选出的主题,

借助爱思唯尔指纹引擎技术识别提取主题内文献的关键词^[16]。指纹引擎技术采用先进的自然语言处理(NLP)技术从非结构化文本中提取信息,并配合领域相关的主题词表的应用,对主题内每篇出版物创建加权术语索引“指纹”,再基于逆向文件频率(Inverse Document Frequency, IDF)^[18]统计方法,判断术语的重要性,提取与主题相关性高的关键词。本研究借助Scival平台爱思唯尔指纹引擎后端处理组件完成主题文献关键词的识别、相关性分析及其在主题中的走势分析^[19]。

(3) 期刊遴选评估。本研究利用爱思唯尔期刊查找工具(Elsevier Journal Finder),查找与主题相匹配的期刊资源。期刊查找工具原本用于为作者找到适合发表其论文的期刊,需要作者提供论文标题、关键词和摘要^[20],经由爱思唯尔指纹引擎生成规范化、结构化的加权术语索引,再将“指纹”与Scopus数据库中论文进行匹配,再经期刊排序算法计算,得到推荐期刊列表^[21]。本研究利用该工具能够语义搜索相关期刊的功能,以所遴选主题的名称作为标题和关键词,将主题关键词依据其与主题的相关性赋予相应的权重,模拟生成摘要文本,输入该工具,通过主题特征文本的匹配定位内容相关度高的资源,并且为了兼顾交叉学科的期刊,并不预设期刊所属学科。最后再结合期刊影响力指标,从中遴选出高水平的期刊。

(4) 核心论文遴选评估。在单篇论文层面,本研究利用Scival平台主题特征分析功能,分析获得各主题的代表性论文。代表性论文的遴选依据核心论文分值^[22],核心论文分值反映论文在主题中的核心程度,由文章在主题内的链接数(包含被引和引用)、主题内链接数占其所有链接数之比、被引次数出版年规范化值三者相乘而得。核心论文分值高的论文,通常有许多主题内的引文关系,而且大部分引文关系都在主题内,被引用的次数也相对较高,能反映主题的中心研究问题,并且通过设置论文出版年份,还能够反映主题内较新的发展方向。

2.3 指标说明

(1) 主题显示度。其反映主题的被关注度及发展势头,是对每一主题中文献的被引次数、浏览次数和平均的期刊影响力加权求和的结果^[17]。

(2) 关键词相关性。其反映关键词与主题的相关程度,是关键词IDF^[18]统计结果的归一化值,用0~1表示,主题内相关度最高的关键词相关性为1,其余关键词

的相关性则根据统计结果取相对值。

(3) 期刊影响力指标。CiteScore是以4年区间为基准计算的每本期刊平均被引用次数;SNIP(Source Normalized Impact per Paper),即规范化影响力,是期刊中每篇论文平均被引次数与其所属学科领域潜在引用值之比,可以直接比较不同学科领域期刊的引用影响^[23-24];SJR(SCImago Journal Rank)是既考虑期刊被引数量,又考虑期刊被引质量的指标,它采用Google的PageRank算法,赋予来源于高声望期刊的引用以较高的权重,并以此规则迭代计算直到收敛^[25]。

(4) FWCI值。学科规范化引文影响力(Field-Weighted Citation Impact, FWCI)^[26]反映主题文献合集或主题内单篇论文的学术影响力,是论文被引频次与同类型论文平均被引次数的比值,排除了出版年、学科领域与文献类型的影响。

3 结果与分析

3.1 主题分析与遴选

引用参考文献、发表学术论文是科研用户最主要的科技文献利用方式,其发文和引文直接反映其文献资源需求,尤其是引文分析法被广泛应用于文献需求保障分析^[27-29]。通过对中国农科院蜜蜂学2017—2019年的272篇发文及其11 836篇引文进行梳理分析,发现其发文分布于149个主题、引文分布于2 253个主题。综合比较该机构发文、引文在各主题中的分布数量和占比,判断其对各个主题文献的需求,再结合主题全球表现情况和主题与该机构蜜蜂学科研究方向的映射结果,遴选出16个主题(见表1)。

可以看出2017—2019年中国农科院蜜蜂学科发文较集中的主题有“NosemaCeranae, Deformed Wing Virus, Sacbrood Virus”(21篇)、“Bombus, Pollinators, Neonicotinoids”(18篇)和“Honey, Stingless Bees, Hydroxymethylfurfural”(16篇)。从发文对主题同期论文贡献程度而言,该机构贡献较大的主题有“Royal Jelly, 10-Hydroxy-2-Decenoic Acid, Hypopharyngeal Glands”(6.71%)、“NosemaCeranae, Deformed Wing Virus, Sacbrood Virus”(4.42%)和“Honeybee, Lactobacillus Kunkeei, Bombus”(3.70%)。但由于主题的组织方式是基于引用关系,所以主题粒度大小不一,因此有的研究方向集中在一个主题内,如蜜蜂蛋白质组学;有的研究

表1 中国农科院蜜蜂学2017—2019年发文、引文主要主题分布

主题名称	发文量/篇	发文占主题同期论文比	引文/篇次	引文量/篇	引文占主题全部论文比	主题显示度	FWCI值	映射研究方向
NosemaCeranae, Deformed Wing Virus, Sacbrood Virus	21	4.42%	533	312	17.21%	96.623	1.15	蜜蜂病虫害生物学
Bombus, Pollinators, Neonicotinoids	18	0.96%	547	306	6.61%	99.804	2.09	传粉蜂生物学与授粉应用
Honey, Stingless Bees, Hydroxymethylfurfural	16	2.01%	316	248	8.63%	98.943	1.38	蜜粉源植物产地环境识别与控制; 蜂产品质量与风险评估
Propolis, Pinobanksin, Stingless Bees	12	1.57%	166	117	3.95%	98.200	1.74	蜂产品质量与风险评估; 蜂产品加工与功能评价
Royal Jelly, 10-Hydroxy-2-Decenoic Acid, Hypopharyngeal Glands	11	6.71%	394	187	22.21%	88.192	1.73	蜜蜂蛋白质组学
Thin-Layer Drying, High Temperature Air, Solar Dryers	8	0.53%	237	140	2.68%	99.372	5.16	蜂产品加工与功能评价
Honeybee, Lactobacillus Kunkeei, Bombus	6	3.70%	197	92	19.96%	94.492	1.19	蜂产品质量与风险评估; 蜂产品加工与功能评价
Bee Pollen, Honeybee, Propolis	6	2.76%	154	104	14.73%	91.584	1.35	蜂产品加工与功能评价
Varroidae, Honey Bee Colonies, ApisMellifera	5	1.78%	126	80	4.75%	91.838	0.64	蜜蜂病虫害生物学
Worker Honey Bees, Queen Pheromones, ApisMellifera	4	2.11%	176	115	9.41%	88.532	1.16	蜂种质资源与育种
Vomitoxin, T 2 Toxin, Trichothecenes	4	0.88%	119	87	4.53%	97.187	1.44	蜜蜂病虫害生物学
Vomitoxin, Zearalenone, 15-Acetyldeoxynivalenol	4	0.50%	118	96	3.03%	99.441	0.75	蜂种质资源与育种
Peptidoglycan Recognition Protein, Hemolin, Polypeptide Antibiotic Agent	4	1.01%	106	70	3.49%	96.367	0.85	蜂产品质量与风险评估
ApisMelliferaMellifera, Honeybee, Apiaries	4	2.15%	86	66	5.95%	81.849	0.43	蜂产品质量与风险评估
HelicoverpaArmigera, Insecticide Resistance, Cytochrome P-450	4	1.78%	72	57	5.38%	93.420	1.19	蜂种质资源与育种
Bombus, CrithidiaBombi, Apidae Enzyme System	3	2.80%	81	47	6.58%	82.654	0.80	蜜蜂病虫害生物学

方向则涉及多个主题,如蜜蜂病虫害生物学就涉及孢子虫、瓦螨、杀虫剂抗性、短膜虫等主题。因此,在考虑发文占比的同时,也应结合机构研究方向的设置情况考虑主题粒度,选择恰当的主题进行文献需求分析。

对引文的主题分布进行分析发现,无论是引文篇次还是引文量的分布规律,都与发文主题分布有相同的特征。引文较集中的主题同样是“NosemaCeranae, Deformed Wing Virus, Sacbrood Virus”(312篇)、“Bombus, Pollinators, Neonicotinoids”(306篇)和“Honey, Stingless Bees, Hydroxymethylfurfural”(248篇)。而引文占主题全部论文比排名前三的主题分别是“Royal Jelly, 10-Hydroxy-2-Decenoic Acid, Hypopharyngeal Glands”(22.21%)、“Honeybee, Lactobacillus Kunkeei, Bombus”(19.96%)和“NosemaCeranae, Deformed Wing Virus, Sacbrood Virus”(17.21%),充

分说明蜜蜂所利用的文献集中在这些主题内。

除发文、引文的主题分布情况,还通过主题显示度指标来判断主题的关注程度和发展势头。蜜蜂学涉及的主题“Bombus, Pollinators, Neonicotinoids”(99.804)、“Vomitoxin, Zearalenone, 15-Acetyldeoxynivalenol”(99.441)和“Thin-Layer Drying, High Temperature Air, Solar Dryers”(99.372)显示度较高,在全球范围内受关注程度高、发展势头较好,分析这类主题的文献内容构成将有助于识别相关研究领域的科技前沿热点信息资源和技术。此外,本文还将FWCI值作为遴选目标主题的参考,但机构用户的文献需求分析主要考量其与主题文献的相关程度,并不评价主题内论文的引文影响力高低,因此FWCI值在此仅作参考不作为遴选指标。综上所述,综合比较主题的发文量、对主题的发文贡献、引文量、引文占比以及主题显示度等因素,最

终将16个主题遴选为分析中国农科院蜜蜂学文献需求的主要研究对象。

3.2 主题关键词识别

以主题“NosemaCeranae, Deformed Wing Virus, Sacbrood Virus”为例,说明本研究对主题关键词的识别过程和分析结果。通过分析蜜蜂所引用该主题内312篇文献发现,引文的出版年主要集中在2010年以后,因此对该主题2010年至今的1 384篇文献进行关键词识别,提取出相关性排名前50的关键词,构建特征文本,代表该主题的主要内容。在提取出的关键词中(见表2,此处仅列出相关性排名前20的关键词)

表2 主题“NosemaCeranae, Deformed Wing Virus, Sacbrood Virus”主要关键词

序号	关键词	相关性	走势/%	序号	关键词	相关性	走势/%
1	Honeybee	1.00	97.8	11	NosemaApis	0.19	-43.5
2	NosemaCeranae	0.76	30.8	12	Apiculture	0.18	57.6
3	ApisMellifera	0.53	116.2	13	Colony	0.17	50.7
4	Varroida	0.50	112.2	14	Bee Virus	0.13	257.1
5	Nosema	0.47	-20.0	15	Israeli Acute Paralyse Virus	0.13	-5.6
6	Bee	0.35	43.3	16	Insect Virus	0.12	-75.0
7	Deformed Wing Virus	0.34	263.6	17	ApisCerana	0.12	188.9
8	Honey Bee Colony	0.27	26.7	18	Apiary	0.12	29.2
9	Sacbrood Virus	0.24	188.9	19	Microsporidia	0.11	-25.8
10	Colony Collapse	0.21	-54.3	20	Black Queen Cell Virus	0.10	228.6

3.3 相关期刊评估

本研究从资源内容和资源质量两个方面对主题相关的期刊进行了评估。仍然以主题“NosemaCeranae, Deformed Wing Virus, Sacbrood Virus”为例,通过分析主题文献来源期刊分布得到刊载该主题文章数量较多的主要期刊(见表3),可以看出这些期刊不仅涉及昆虫学,还涉及传染病学、病毒学,以及生态学、进化学、行为学和系统分类学等学科,由此说明利用主题来组织期刊资源能够突破学科的界限。除此之外,还用文本匹配的方法尝试寻找内容相关期刊,利用上文中识别提取出的主题关键词,将其按与主题的相关性赋予权重,构建特征文本,再利用爱思唯尔期刊查找工具,从语义上匹配内容相关的期刊。对主题“NosemaCeranae, Deformed Wing Virus, Sacbrood Virus”而言,期刊查找工具匹配出49种期刊(见图

除Honeybee、Bee、ApisMellifera(意大利蜜蜂)、ApisCerana(东方蜜蜂)等蜜蜂的不同常用名、种名之外,绝大多数是蜜蜂病虫害生物学的专业名词,包括NosemaCeranae(微孢子虫)、Deformed Wing Virus(畸翅病毒)、Sacbrood Virus(囊幼病病毒)等描述病害和虫害的关键词。关键词的相关性通过IDF统计方法算得,相关性越高越能代表该主题,将由此设置在特征文本中的权重。此外,关键词走势反映主题内关键词所涉文献过去10年的数量变化趋势,可作为判断主题主要内容变化的参考,如关键词Deformed Wing Virus,其相关文章的数量过去10年在此主题内增长率超过200%,说明相关的研究越来越受到重视。

3),文本匹配分数(5分为满分)较高的期刊有*Journal of Apicultural Research*(2.32分)、*Saudi Journal of Biological Science*(1.81分)和*Current Biology*(1.08分)。这3种刊如果按照传统的蜂学期刊资源遴选方法,从昆虫学期刊中寻找都不会被发现,因此也证明本研究采用的方法能够在其他学科精准地定位到与蜜蜂研究密切相关的期刊。此外,在期刊质量评估方面,本文则选取了CiteScore、SNIP、SJR期刊影响力指标,作为评估期刊资源质量的重要参考依据。

3.4 核心论文推荐

对于主题核心论文的遴选,本研究利用Scival主题特征分析,获得主题“NosemaCeranae, Deformed Wing Virus, Sacbrood Virus”中2017—2020年核心程度高的论文(见表4),这些论文反映主题的中心研究

表3 主题 “NosemaCeranae, Deformed Wing Virus, Sacbrood Virus” 主要来源期刊

期刊名称	文章量/篇	CiteScore	SNIP	SJR	所属学科
Journal of Apicultural Research	144	2.7	0.838	0.548	昆虫学
Journal of Invertebrate Pathology	102	3.8	1.077	0.877	生态学、进化学、行为学和系统分类学
American Bee Journal	100	0.1	0.081	0.104	昆虫学
PLoS ONE	96	5.2	1.123	1.023	综合学科
Apidologie	82	4.0	1.196	0.948	昆虫学
Scientific Reports	45	7.2	1.240	1.341	综合学科
Viruses	26	4.3	1.193	1.633	传染病学; 病毒学
Journal of Apicultural Science	23	1.4	0.645	0.270	昆虫学
Insects	22	2.2	1.166	0.838	昆虫学
Journal of Economic Entomology	19	3.0	1.030	0.784	昆虫学

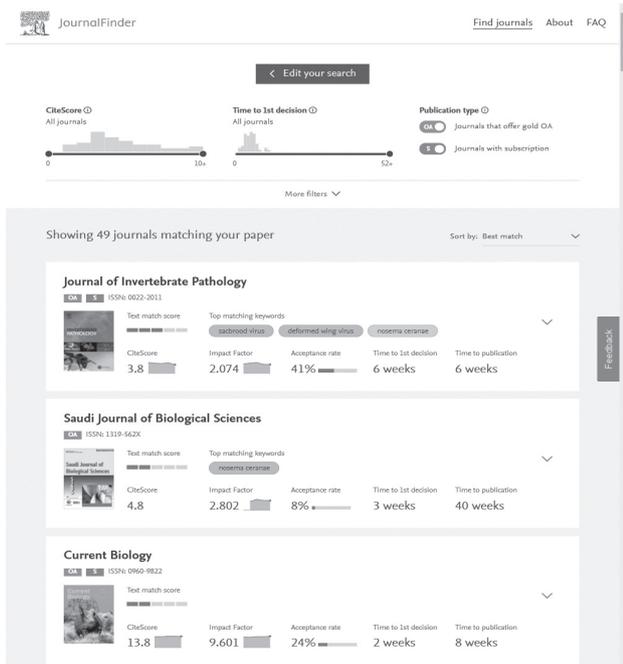


图3 爱思唯尔期刊查找工具查找相关期刊结果

问题, 与主题内的其他文章有较多的引文关系, 并且对其自身而言大部分引文关系都发生在主题内, 相对于其出版年份被引的次数也较高。在论文质量评估方面, 本文采用各论文的被引次数和FWCI值作为判断论文学术影响力的依据, 但并不作为核心论文遴选指标。因为被引次数受出版年份的影响, 但本研究并不是遴选主题内积累引用最多的论文, 而是在内容上是中心的, 且反映较新发展方向的论文, 作为较新的核心资源推荐给用户。而FWCI值虽然消除了出版年份的偏差, 但由于其是以同学科论文为基准进行计算, 粒度比主题粗, 也不适用。因此, 该两项指标仅用于评估论文质量。

3.5 遴选结果验证

为了进一步验证该方法和遴选结果, 笔者针对遴选出的期刊开展了用户需求分析, 通过分析电子期刊使用统计数据和引文数据反映用户对资源的需求程度,

表4 主题 “NosemaCeranae, Deformed Wing Virus, Sacbrood Virus” 核心论文

标题	出版年份	被引/次	FWCI值
Covert deformed wing virus infections have long-term deleterious effects on honeybee foraging and survival	2017	42	4.95
The virulent, emerging genotype B of Deformed wing virus is closely linked to overwinter honeybee worker loss	2017	39	2.71
Bee viruses: Ecology, pathogenicity, and impacts	2019	31	9.56
Viruses of commercialized insect pollinators	2017	30	2.12
Deformed Wing Virus in Honeybees and Other Insects	2019	25	10.89
Recently identified bee viruses and their impact on bee pollinators	2018	24	2.89
Nosemaceranae in Apismellifera: a 12 years postdetection perspective	2018	17	1.65
Nosemaceranae disease of the honey bee (Apismellifera)	2018	15	1.72
The two prevalent genotypes of an emerging infectious disease, deformed wing virus, cause equally low pupal mortality and equally high wing deformities in host honey bees	2019	12	5.74
Diversity and global distribution of viruses of the western honey bee, apismellifera	2020	3	3.37

进而佐证遴选结果。以中国农科院蜜蜂学的主要研究群体蜜蜂研究所为目标用户,分别统计其2019年各刊全文下载量,及其2017—2019年发文对各刊的引文量(见表5),可以看出,蜜蜂所对这些期刊都有一定程度的需求,尤其是对其中非OA刊的需求强烈。但下载量和引文量一定程度上仅反映用户对能获取资源的需求,因此仅是佐证用户的现实需求,而遴选结果中还出现了该用户未订阅的资源,则是挖掘出了潜在的需求。因此,笔者还通过走访蜜蜂所蜜蜂病虫害生物学创新团队,针对遴选出的期刊进行调研,征询相关研究人员的意见,也获得了正向的反馈。

表5 期刊遴选结果需求分析

期刊名称	订阅情况	开放获取情况	全文下载量/篇次	引用量/篇次
Journal of Apicultural Research	否	非OA		149
Journal of Invertebrate Pathology	是	非OA	622	135
American Bee Journal	否	非OA		9
PLoS ONE	否	OA		396
Apidologie	是	非OA	425	238
Scientific Reports	否	OA	987	130
Viruses	否	OA		17
Journal of Apicultural Science	否	OA		16
Insects	否	OA		12
Journal of Economic Entomology	是	非OA	355	54
Saudi Journal of Biological Sciences	否	OA	71	3
Current Biology	是	非OA	385	31

注:未订阅非OA资源和部分OA资源无法获取使用统计数据

4 讨论

本文以蜜蜂学为例,研究了基于主题组织和分析的科技文献资源遴选方法,并且证明该方法能够突破学科的限制精准地定位内容相关期刊资源。在本文采用的研究方法中,主题组织、关键词识别、资源匹配等环节都运用到了爱思唯尔指纹引擎技术,该技术采用先进的自然语言处理技术并结合多领域的主题词表应用,使分析超越元数据,能够从语义层面挖掘出版物内容间有价值的联系,本文将其应用于资源遴选,相对于传统按学科组织资源的方法,按主题组织资源颗粒度更细,将在跨学科、交叉学科资源遴选的过程中发挥优势。

本文对于期刊查找工具的利用实现了通过语义匹配遴选期刊资源的目的。虽然可以将主题的特征内容生成“指纹”与Scopus数据库中大量的文摘进行匹配,

但期刊推荐工具反馈的结果仅局限于爱思唯尔出版的期刊。本研究过程中也尝试使用了其他出版社类似工具,如SpringerNature期刊推荐工具(Springer Nature Journal Suggester)和Wiley期刊推荐工具(Wiley Journal Finder)等,但这类期刊查找工具原是为作者寻找发表论文的期刊而设计,各工具算法不一、不同的工具匹配出不同出版社的期刊,无法对匹配结果进行统一比较。因此,可以考虑利用指纹引擎技术和文摘大数据开发独立的期刊遴选应用组件,对更大范围更多来源的期刊资源进行匹配遴选。此外,通过这种方法遴选得到的期刊,也有待后续从资源利用效率和用户反馈等方面进行进一步验证,确保为该交叉学科需求的资源,针对资源遴选方法及其结果的评估开展更深入的研究,完善此方法的科学性。

参考文献

- [1] 肖敏,江宁,叶小梅,等. 农业科研院所视角下新兴交叉学科发展的思考[J]. 农业科技管理, 2020, 39(2): 17-20.
- [2] 湛群芳,陈积明. 交叉学科发展是“双一流”建设的新兴增长点[J]. 中国高等教育, 2018(10): 42-43.
- [3] 刘涛,陈敏. 面向问题与未来统筹分化与综合——浅议现代农业交叉学科发展[J]. 农业科技管理, 2017, 36(2): 22-24.
- [4] 章成志,吴小兰. 跨学科研究综述[J]. 情报学报, 2017, 36(5): 523-535.
- [5] 刘伟,郝继英,王松俊. 交叉学科信息资源优化流程构建[J]. 中华医学图书情报杂志, 2016, 25(6): 13-18.
- [6] 范慧婷,谢鑫. 跨学科文献需求与保障策略研究——以中国边界与海洋研究为例[J]. 图书情报工作, 2018, 62(5): 23-29.
- [7] 陈磊,陈惠兰,詹司佳,等. 学科竞争力视角下的馆藏资源建设探析——基于Scival跨学科科研评价工具[J]. 图书馆, 2015, 25(18): 83-86.
- [8] 阮学平,李军莲,唐小利,等. 《中图法》在NSTL文献资源建设中的改造应用[J]. 图书馆理论与实践, 2013(3): 69-70, 91.
- [9] 向佳丽,韩冬丽,应巧兰. 交叉学科研究者文献使用习惯和特点分析——以厦门大学法学与海洋交叉学科文献资源建设为例[J]. 农业网络信息, 2016(3): 101-104.
- [10] 沈丹,周锐,王晓丽,等. 基于引文分析的交叉学科资源建设研究[J]. 数字图书馆论坛, 2013(5): 20-26.
- [11] 王婷,颜蕴,续玉红,等. 交叉学科国外期刊需求分析方法探析[J]. 数字图书馆论坛, 2016(8): 50-53.
- [12] 阮守江,冯红岩. 网络信息资源组织方法的比较分析及其应用

- 探讨 [J]. 科技创新导报, 2011 (13): 18-20.
- [13] 卜庆书. 知识组织系统构建与知识服务研究 [M]. 北京: 国家图书馆出版社, 2014.
- [14] 孙海霞, 李军莲, 华薇娜, 等. 科技知识组织体系语义互操作网络协同工作平台设计与实现 [J]. 农业图书情报, 2019, 31 (1): 23-34.
- [15] KLAUVANS R, BOYACK KW. Research portfolio analysis and topic prominence [J]. Journal of Informetrics, 2017, 11 (4): 1158-1174.
- [16] VESTDAM T, RASMUSSEN H, et al. Black magic meta data-get a glimpse behind the scene [J]. Procedia Computer Science, 2014 (33): 239-244.
- [17] KLAUVANS R, BOYACK K W. Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? [J]. Journal of the Association for Information Science and Technology, 2017, 68 (4): 984-998.
- [18] SALTON G, BUCKLEY C. Term-weighting approaches in automatic text retrieval [J]. Information Processing & Management, 1988, 24 (5): 513-523.
- [19] How are keyphrases calculated? [EB/OL]. [2020-10-10]. https://service.elsevier.com/app/answers/detail/a_id/27763/supporthub/scival/.
- [20] Find journals [EB/OL]. [2020-10-10]. <https://journalfinder.elsevier.com/>.
- [21] KANGN, DOORNENBALM A, SCHIJVENAARS R J A. Elsevier Journal Finder: Recommending Journals for your Paper [C] //Association for Computing Machinery. In Proceedings of the 9th ACM Conference on Recommender Systems (RecSys'15). New York: 2015: 261-264.
- [22] Topic Prominence in Science FAQs [EB/OL]. [2020-10-10]. https://service.elsevier.com/app/answers/detail/a_id/28428/supporthub/scival/#panel12b.
- [23] MOED H F. Measuring contextual citation impact of scientific journals [J]. Journal of Informetrics, 2010, 4 (3): 265-277.
- [24] WALTMAN L, VAN ECK N J, VAN LEEUWEN T N, et al. Some modifications to the SNIPjournal impact indicator [J]. Journal of Informetrics, 2013, 7 (2): 272-285.
- [25] GONZÁLEZ-PEREIRA B, GUERRERO-BOTE VP, MOYA-ANEGÓN F. A new approach to the metric of journals' scientific prestige: The SJR indicato [J]. Journal of Informetrics, 2010, 4 (3): 379-391.
- [26] PURKAYASTHA A, PALMARO E, FALK-KRZESINSKI H J, et al. Comparison of two article-level, field-independent citation metrics: Field-Weighted Citation Impact (FWCI) and Relative Citation Ratio (RCR) [J]. Journal of Informetrics, 2019, 13 (2): 635-642.
- [27] 文金书, 黄梅. 基于读者需求分析的资源建设与优化 [J]. 数字图书馆论坛, 2015 (4): 64-68.
- [28] 刘一鸣, 蒋欣羽, 段驿智. 区块链技术助推大数据时代高校图书馆数字资源建设 [J]. 农业图书情报学报, 2020, 32 (6): 15-22.
- [29] 高建群, 张荣安, 陈吉平, 等. 期刊引文分析及其情报价值 [J]. 图书情报工作, 2004, 48 (1): 58-61.

作者简介

卢焱, 女, 1983年生, 博士, 副研究员, 研究方向: 农业信息资源建设与组织。

王鸞飞, 男, 1973年生, 硕士, 副研究馆员, 研究方向: 农业信息资源建设。

马鑫, 女, 1982年生, 硕士, 馆员, 通信作者, 研究方向: 农业信息资源建设, E-mail: maxin01@caas.cn。

Methodological Research of Interdisciplinary Scientific Literature Resources Selection Based on Subject Analysis: A Case Study of Apicultural Science

LU Yao WANG YueFei MA Xin
(Agriculture Information Institution, CAAS, Beijing 100081, China)

Abstract: Nowadays the demands of scientific research users on information resources have become more diversified and personalized, especially for the cross-disciplinary and emerging disciplines resources. The resources construction of special libraries in research institutes will no longer meet the users' needs of scientific and technological innovation if only relies on the traditional methods of demand analysis and selection for literature resources. In this paper, methodological researches of literature resource demand analysis and resource selection were carried out based on a case study of apicultural science, the publications of which from 2017 to 2019 and the corresponding citing documents were studied as objects. We organized the documents by subjects, analyzing and selecting related subjects to grasp the subtle demands on literature resource of the users of this discipline. Then we made use of subject keyword extraction, resource content matching and impact assessment to select high-quality journals and core papers that were highly consistent with the research contents of this discipline, and provided the precision services of literature resource to its researchers. Through this study, we innovate the methodologies of demand analysis and resource selection of scientific and technological literature based on the subject analysis. It breaks the disciplinary limits and can accurately find the resources of relevant contents which would inspire the interdisciplinary resource construction and service of special libraries in research institutes.

Keywords: Interdiscipline; Demand Analysis; Subject Organization; Resource Assessment; Content Matching

(收稿日期: 2020-09-29)