

基于水族文献的计量分析与知识图谱研究

杨秀璋

(贵州财经大学信息学院,贵阳 550025)

摘要:

随着科学技术迅速发展,学术成果呈爆炸式增长,挖掘论文的核心作者、识别高引论文、分析文献主题变得日益重要。通过 Python 抓取中国知网 1953-2018 年间 990 篇水族文献,基于文献的计量分析方法挖掘出我国水族文献的高影响力期刊、报纸及机构,分析水族领域高引论文、发文热点时间及核心作者;基于知识图谱的分析方法构建水族文献核心科研群体及作者间的合作关系,并揭示出水族文献的主题演化趋势。实验结果展示我国水族文献的发展历程、核心科研群体和高引论文,识别出水族文化、体育、音乐、医学和水书五大主题的关键词及历年的主题演化趋势,为水族文化的研究和进一步发展提供相关启示。

关键词:

水族文献;计量分析;知识图谱;高引论文;中国知网

基金项目:

贵州省教育厅青年科技人才成长项目(黔教合 KY 字[2016]172)

0 引言

随着科学技术迅速发展,学术成果呈爆炸式增长,基于文献的计量分析可以挖掘出一个研究领域的高影响力期刊和机构,识别出高被引论文的主题、出版社和核心作者;基于知识图谱的分析可以构建文献作者的合作关系,挖掘出核心科研群体,并揭示文献的主题演化趋势。

水族是一个具有悠久历史和古朴文化的民族,繁衍生息于贵州、云南、广西等省份,拥有独特的民族文化、风土人情和语言文字^[1],具有重要社会历史文化价值。近年来,学术界和社会媒体掀起了研究水族文化和水书的热潮,1953 年至今,中国知网共收录水族相关文献 990 余篇,涉及水族文化、水族体育、水族医学、水族音乐、水书文字等主题。水族文献作为水族文化和知识交流的重要载体,是水族相关的学术研究团队或科研工作者贡献的重大成果。这些学术成果将有效地推动水族学科和领域的发展,为进一步追踪水族源流、研究水族文化变迁、保护和传承水族文化提供相关

启示。

传统的学术成果研究方法是基于文献计量学和科学计量学的方法,分析内容包括文献的主题、地域分布、机构分布、核心作者、引文特点等。姜春林通过文献计量历时法对《科学学研究》做出全面的计量分析^[2]。梁永霞等基于 CSSCI 的中国引文分析进行了可视化研究,展现引文分析领域的知识结构关系^[3]。黄晓斌等统计分析我国情报学高被引论文,展示情报学的发展历程和学科主题^[4]。蔡文伯等通过计量分析方法研究我国民族教育文献态势^[5]。王宗水等基于 1998-2014 年中国社会科学引文数据分析社会网络范式的演化、发展与应用^[6]。徐庶睿等利用引文内容进行主题级学科交叉类型分析^[7]。近年来,国内外学者致力于通过数据分析与机器学习技术来研究学术文献,基于社交网络和知识图谱的引文分析技术也被用来辅助科研工作者和管理者把握学科发展脉络,挖掘主题关键词和核心作者。罗双玲等提出了基于半积累引文网络社区发现的学科领域主题演化分析方法,并应用于“合作演化”领域^[8]。马文博等通过文献计量方法和知识图谱分

析《经济研究》近十年载文^[9]。任晓松等归纳研究中国碳排放热点演化与知识图谱分析^[10]。

当前水族领域的研究更多的是采用传统的查阅资料、现场考察及问卷调查等方法分析水族文化、历史、习俗、水书内容,没有采用计量分析方法和知识图谱来研究水族文献,缺乏对水族文化深层次的主题演化和科研群体发现的研究。针对这些不足,本文提出了一种基于计量分析和知识图谱的水族文献分析方法,一方面统计分析水族文献的高影响力期刊、报纸及机构的基本情况,挖掘水族领域高被引论文、发文热点时间及核心作者;另一方面通过知识图谱构建水族领域核心科研群体及作者间的合作关系,并揭示出水族文献的主题演化趋势。本文的研究成果具有重要的理论意义和实用价值,对传承与弘扬水族传统文化具有重要意义,为水族文化的研究和进一步发展提供相关启示,为后续的水族引文分析、水族迁徙研究和水族文化传承提供有效支撑。

1 研究方法和数据采集

1.1 算法总体流程

当前科研成果的分析研究已经引起了社会和学术界的广泛关注和重视,所涉及的领域包括引文分析、论文自动分类、主题挖掘、学术成果测度、论文知识图谱构建等。本文主要是关于水族文献的计量分析和知识图谱研究,算法的总体流程如图1所示。

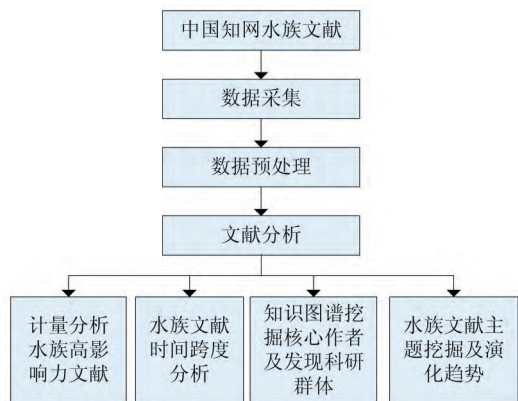


图1 水族文献计量分析和知识图谱研究流程图

(1)调用 Python、Selenium 和 XPath 技术自动抓取中国知网水族领域的文献共 990 篇。

(2)对所抓取的论文信息进行数据预处理,包括数据清洗、数值提取、中文分词、停用词过滤等处理,这是数据分析的重要处理环节。

(3)将预处理之后的文本数据转换为特征词矩阵,涉及特征提取和权重计算。

(4)通过计量分析和知识图谱方法研究水族文献,包括水族高影响力论文、知识图谱挖掘核心作者及科研群体、水族文献主题挖掘及演化趋势等。

1.2 数据说明及预处理

本文数据集采用 Python 自定义爬虫抓取中国知网 CNKI 水族领域相关的学术成果信息,包括 1953-2018 年共收录的 990 篇水族领域的学术文献,其中学术期刊论文 662 篇,会议论文 106 篇,博士论文 7 篇,硕士论文 91 篇,中国专利 5 篇,科技成果 3 个,报纸 116 篇,详细信息如表 1 所示。所抓取的水族文献共包括九个字段,如图 2 所示,包括文章标题、出版社、文献类型、出版年份、下载次数、引用次数、作者、摘要和关键词。

表 1 中国知网水族领域的学术成果汇总表

学术成果类型	刊载成果数量	总下载次数	总引用数量	篇均下载次数	篇均引用数量	单篇最高下载次数	单篇最高引用数量
学术期刊	662	91478	1869	138.18	2.82	951	82
会议论文	106	4628	8	43.66	0.08	176	1
博士论文	7	7478	50	1068.29	7.14	3664	18
硕士论文	91	28324	192	311.25	2.11	1430	23
中国专利	5	0	0	0.00	0.00	0	0
科技成果	3	1	0	0.33	0.00	1	0
报纸	116	1589	5	13.70	0.04	65	2
总计	990	133498	2124	134.85	2.15	3664	82

A	B	C	D	E	F	G	H	I
文章标题	出版社	文献类型	出版年份	下载次数	引用次数	作者	摘要	关键词
1 水族文化中的“水书”研究	《中国水书》	《中国水书》	2012年 第01期	100	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
2 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
3 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
4 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
5 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
6 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
7 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
8 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
9 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
10 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
11 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
12 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
13 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
14 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
15 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
16 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
17 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
18 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
19 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
20 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
21 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
22 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
23 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
24 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化
25 水族文化多样性与生态文明建设	《中国水书》	《中国水书》	2006/11/1	60	1	李德胜 黄其强	《中国水书》田野调查	水族 水书 水书文化

图2 水族文献数据集部分数据

在进行数据分析之前,通常需要对所爬取的信息进行预处理操作,包括缺失值填充、异常值处理、数据清洗、中文分词、停用词过滤等步骤,其目标是为了得到高标准、高质量的数据,从而提升分析的结果。接着

从六个方面分析水族文献,包括水族文献的高影响力期刊报纸及机构计量分析、高被引水族论文计量分析、水族文献的时间跨度分析、高引论文的核心作者分析、核心作者知识图谱分析及科研群体发现、水族文献主题挖掘及演化趋势。

2 水族文献分析及研究

2.1 高影响力期刊、报纸及机构计量分析

本文将分析水族文献的高影响力期刊、报纸及机构,所采用的指标是刊载成果数量、总引用数量、总下载数量、篇均引用数量和篇均下载数量。

(1)引文期刊分析

对水族学术期刊论文的计量分析,可以从宏观层面揭示出刊载期刊及研究群体的力量分布情况,也在一定程度上挖掘出高影响力的期刊,它们对水族文化研究做出了重要的贡献。表 2 列出了刊载水族论文六篇及以上的所有期刊,其中《黔南民族师范学院学报》刊载论文数量最多,共刊载 45 次;《贵州民族研究》总引用次数最多,42 篇论文共被引用 203 次;《人类学学报》篇均引用数量最高,刊载 10 篇论文,平均每篇论文被引用 16.40 次。

表 2 刊载六篇及以上水族论文的期刊

学术期刊名称	刊载成果数量	总引用数量	篇均引用数量
黔南民族师范学院学报	45	142	3.16
贵州民族研究	42	203	4.83
当代贵州	22	5	0.23
贵州民族学院学报(哲学社会科学版)	21	87	4.14
黔南民族医学学报	12	53	4.42
人类学学报	10	164	16.40
中国民族民间医药杂志	10	15	1.50
广西民族研究	9	51	5.67
歌海	9	12	1.33
贵州文史丛刊	9	6	0.67
凯里学院学报	8	20	2.50
中国妇幼保健	8	20	2.50
大众文艺	8	7	0.88
中国民族	8	2	0.25
贵阳医学院学报	7	13	1.86
民族论坛	7	4	0.57
贵州大学学报(艺术版)	6	45	7.50
贵州社会科学	6	12	2.00

(2)引文机构分析

引文机构分析主要研究来自全国各大高校、各类研究所和图书馆等机构的水族文献。本文根据毕业论文的第一单位进行统计分析,得出了如表 3 所示的计量分析结果。发表水族领域相关论文两篇及以上的机

构共 16 个,其中刊载水族论文数量排名前三位的分别是:贵州民族大学(共刊载 16 篇)、贵州师范大学(共刊载 8 篇)、贵州大学(共刊载 8 篇);论文被引用最多的四个机构是:贵州民族大学(共被引用 32 次)、贵州师范大学(共被引用 24 次)、云南大学(共被引用 19 次)、昆明理工大学(共被引用 19 次);论文被下载最多的三个机构是:云南大学(共被下载 4193 次)、贵州民族大学(共被下载 3715 次)、中央民族大学(共被下载 2793 次)。

该结果表明大学是水族领域研究机构的中坚力量,这些大学主要分布于贵州、云南、广西等省份,这些省份也是水族的聚居区;同时民族类和师范类大学也对水族文化有重要的贡献,如贵州民族大学、中央民族大学、华东师范大学、广西民族大学等。

表 3 刊载水族论文两篇及以上的机构

研究机构名称	刊载成果数量	总引用数量	总下载数量
贵州民族大学(贵州民族学院)	16	32	3715
贵州师范大学	8	24	1907
贵州大学	8	11	1756
中央民族大学	7	17	2793
贵州财经大学(贵州财经学院)	4	6	2048
中山大学	4	5	1736
西南大学	4	3	845
云南大学	3	19	4193
昆明理工大学	3	19	2561
华东师范大学	3	14	603
广西民族大学	3	11	1493
云南艺术学院	3	1	366
南京师范大学	2	0	83
上海师范大学	2	14	1206
西南民族大学	2	5	652
西南政法大学	2	0	251

(3)引文报纸分析

报纸作为学术成果的组成部分,同样对水族文化和水族传承具有一定的贡献,它能让社会大众接触到水族文化内涵、水书艺术和迁徙漂泊的历史。

图 3 是报道水族领域相关的报纸计量分析结果图,横坐标为报纸的名称,纵坐标从上至下依次为成果总数、下载总数和被引用总数。其中《贵州民族报》报道的水族领域文章数量、下载次数和被引用次数都是最多,分别为发表 45 篇、被下载 633 次和被引用 2 次;接下来是《黔南日报》报道了 20 篇,《贵州日报》报道了 19 篇,《中国民族报》报道了 9 篇。

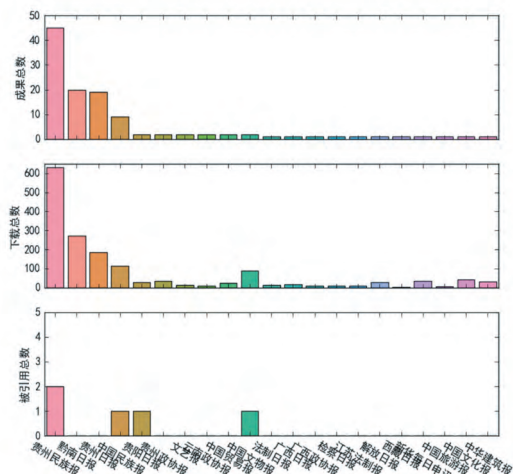


图3 各类报纸关于水族领域文献的计量分析图

2.2 高被引水族论文计量分析

论文被引用频次越高,说明论文的学术价值越高。本文研究的 990 篇水族学术成果共被引用 2124 次,平均每篇学术成果被引用 2.15 次。被引用次数最多的文献类型是学术期刊和硕士/博士论文,其中学术期刊 662 篇共被引用了 1869 次,硕士/博士论文 98 篇共被引用了 242 次,二者共占总被引用次数的 99.4%。

表 4 是被引频次前 10 位的学术论文,全部 10 篇高引论文的被引频次都是 20 次以上,是所有水族文献篇均被引用次数(2.15 次)的 10 倍。其中被引频次最高的论文是李培春等在 1994 年发表在《人类学学报》上的“水族的体质特征研究”一文,被引频次为 82 次;排名第二的是苏和平所作,2004 年发表在《贵州民族研究》上的“水族审美意识探源”,被引频次为 48 次;随后是孙志国等于 2011 年发表在《贵州民族学院学报(哲学社会科学版)》上的“水族非物质文化遗产保护的探讨”,顾晓艳等于 2006 年发表在《中国体育科技》上的“传统体育文化在水族山寨中的生存状态——水族‘端节’赛马活动的变迁”,两者被引频次均为 38 次。

表 5 是被引频次前 10 位的硕士/博士毕业论文,这 10 篇高引论文总计被引 120 次,平均每篇论文的被引频次为 12 次。其中被引频次前三位的分别是:陕西师范大学陈永娥的“黔南本土文化语文校本课程资源开发的研究”,被引用 23 次;云南大学尤伟琼的“云南民族识别研究”,被引用 18 次;上海师范大学韦学纯的“水语描写研究”,被引用 14 次。这 10 篇论文主要分布于民族类、师范类、人文社科专业较强以及西南水族

聚居区所在的高校,一定程度上反映了这类大学对我国民族文化研究的影响程度。

表 4 水族领域被引用频次最高的 10 篇期刊论文

论文题目	作者	来源期刊	发表 时间	被引 次数	下载 次数	论文 第一位
水族的体质特征研究	李培春;梁明康等	人类学学报	1994年 第01期	82	240	广西石江民族医学院
水族审美意识探源	苏和平	贵州民族研究	2004年 第03期	48	321	中央民族大学
水族非物质文化遗产保护的探讨	孙志国;黄莉敏等	贵州民族学院学报(哲学社会科学版)	2011年 第06期	38	490	咸宁学院
传统体育文化在水族山寨中的生存状态——水族“端节”赛马活动的变迁	顾晓艳;石国义等	中国体育科技	2006年 第05期	38	743	黔南民族师范学院
西南六个少数民族儿童重民族文化认同的比较研究	陈世联;刘云艳	学前教育研究	2006年 第11期	37	951	重庆师范大学
中国水族地区传统体育文化变迁的调查研究——以贵州省三都水族自治县为个案	顾晓艳;张东秀等	武汉体育学院学报	2006年 第10期	30	665	黔南民族师范学院
贵州南部6个民族5对遗传性状的基因频率	张勋;余跃生等	遗传	2006年 第04期	29	187	黔南民族医学高等专科学校
水族传统体育文化概述	顾晓艳	体育科学	2004年 第11期	25	658	黔南民族师范学院
水族文化传承的学校教育策略研究	黄胜	民族教育研究	2009年 第01期	23	634	西南大学
贵州汉族、苗族、布依族和水族人群线粒体DNA多态性研究	晋华鹏;吴立甫等	遗传学报	1995年 第01期	23	154	贵阳医学院

表5 水族领域被引用频次最高的10篇期硕士/博士论文

论文题目	作者	来源期刊	论文类型	发表时间	被引次数	下载次数
黔南本土文化 语文技术课程 资源开发的研究	陈永娥	陕西师范大学	硕士论文	2011年	23	1095
云南民族识别 研究	尤伟琼	云南大学	博士论文	2012年	18	3664
水语描写研究	韦学纯	上海师范大学	博士论文	2011年	14	1027
民族主题公园 景观设计研究	罗艳艳	昆明理工大学	硕士论文	2005年	12	1430
三都水族马尾 绣艺术初探	张婧	贵州民族学院	硕士论文	2011年	10	361
茂兰自然保护区 的保护与当地 社区和谐发 展研究	蔡泽东	贵州师范大学	硕士论文	2007年	10	338
村落社区视野 下的少数民族 政治认同研究	王茂美	复旦大学	博士论文	2010年	9	2040
水文造字机制 研究	翟宜疆	华东师范大学	博士论文	2007年	9	357
水族马尾绣工 艺特征及传承 研究	章婧	中央民族大学	硕士论文	2013年	8	1025
广西10个少数 民族Y染色体 17个STR基因座 的遗传多态性	梁祚仁	广州医学院	硕士论文	2009年	7	436

2.3 水族文献的时间跨度分析

时间跨度分析展示了学术成果的扩散过程,有效地反映出水族领域的研究进程及主题演化趋势。图4是所有水族相关学术成果的年度分布图,最早一篇论文发表在1953年,是凌慰民老师发表在《中国金融》的“少数民族地区的农贷工作不能违反民族政策”,作者深入彝、水、苗、侗等族聚居的农村,研究贷放工作与民族政策的关联;第二篇论文是12年之后,发表在《中国

民族》的“水族人民破旧立新过‘端节’”，接着是1979年王连和老师文章“西南地区少数民族天象历法调查报告”。1991年是水族论文量产的一年，共发表了32篇；2004年至2017年水族领域的学术成果越来越多，其中2016年发表的81篇学术成果为历史最高。

表6是水族文献2001年至2017年被引用次数的分布情况。从表中可以看出，2001年至2017年被引用过的水族文献共397篇，占总水族文献数量的40.1%，被引用文献的总频次为1748次。其中被引用10次及以上的文献共45篇，被引用5至9次的文献共64篇，被引用2至4次的文献共151篇，仅被引用1次的文献共137篇。同时，2012年被引用过的水族文献最高，共45篇；2006年被引用的总频次最高，为277次，其中10篇文章被引用频次10次以上，为高被引论文；2004年单篇均被引用频次最高，平均每篇论文被引用10.69次。

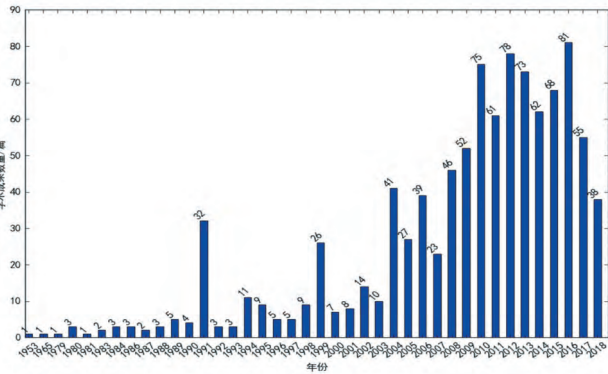


图4 水族文献1953-2018年的学术成果数量分布图

表6 水族文献2001-2017年被引用次数分布表

年份	被引用 论文总数	引用 总数	单篇均被 引频次	被引用10 次及以上 论文篇数	被引用 9-5次论 文篇数	被引用 4-2次论 文篇数	被引用 1次论 文篇数
2001	7	48	6.13	2	1	3	1
2002	8	57	7.13	2	2	3	1
2003	6	42	7.00	2	1	2	1
2004	16	171	10.69	8	1	2	5
2005	12	74	6.17	3	3	5	1
2006	29	277	9.55	10	5	10	4
2007	13	82	6.31	3	3	3	4
2008	31	164	5.29	6	5	14	6
2009	38	141	3.71	2	8	17	11
2010	38	164	4.32	2	14	13	9
2011	37	177	4.78	4	4	18	11
2012	45	123	2.73	1	6	20	18
2013	40	102	2.55		9	11	20
2014	25	47	1.88		1	12	12
2015	25	41	1.64		1	8	16
2016	23	34	1.48			10	13
2017	4	4	1.00				4

2.4 高引论文的核心作者分析

水族领域的核心作者是对水族研究发展具有较大贡献的科研工作者，同时也是期刊学术影响力、竞争力的重要贡献者，他们影响着水族文化的繁衍与发展。本文主要采用发文量、被引频次、篇均被引频次三个指标评估核心作者。该方法从论文的量 and 质两方面来确定核心作者，表7水族文献核心作者统计表，共包括了40位核心作者。其中余跃生发表了水族相关的文章23篇，被引频次为174次，篇均被引频次为7.57次；顾晓艳发表了水族文献16篇，被引频次为176次，篇均被引频次11次；赵凌发表了水族领域的论文14篇，被引频次为25次，篇均被引频次1.79次。

表7 水族文献高被引用频次的核心作者

排名	作者	发文量	被引 频次	篇均被 引频次	排名	作者	发文量	被引 频次	篇均被 引频次
1	余跃生	23	174	7.57	21	朱晓君	7	25	3.57
2	顾晓艳	16	176	11.00	22	杨胜文	7	16	2.29
3	赵凌	14	25	1.79	23	林先燕	7	10	1.43
4	潘朝霖	13	39	3.00	24	冉景丞	7	1	0.14
5	何燕	12	59	4.92	25	石国义	6	80	13.33
6	吴昌学	11	57	5.18	26	陆玉炯	6	65	10.83
7	刘世彬	11	41	3.73	27	戎聚全	6	58	9.67
8	张振江	11	22	2.00	28	齐晓凤	6	38	6.33
9	王亚琼	10	132	13.20	29	陈国安	6	32	5.33
10	黄胜	10	106	10.60	30	文永辉	6	32	5.33
11	单可人	10	61	6.10	31	张兴雄	6	31	5.17
12	任锡麟	10	61	6.10	32	赵艳	6	29	4.83
13	谢渊	10	57	5.70	33	陆兴斌	6	9	1.50
14	蒙耀远	10	5	0.50	34	文毅	6	9	1.50
15	罗春寒	9	17	1.89	35	张婷	6	9	1.50
16	张东秀	8	116	14.50	36	赵俊华	6	8	1.33
17	蒙爱军	8	36	4.50	37	苟利波	6	2	0.33
18	张超	8	25	3.13	38	曹显明	5	72	14.40
19	韦宗林	8	20	2.50	39	王克松	5	48	9.60
20	谢传红	8	8	1.00	40	吴晓黎	5	31	6.20

2.5 核心作者知识图谱分析及科研群体发现

知识图谱分析主要是对1953-2018年刊发在中国知网的所有水族文献作者之间的关系进行分析，挖掘出水族科研核心作者以及对水族文化作出重要贡献的科研群体。其步骤如下：

(1)首先计算水族学术成果作者的共现矩阵，当两名作者合作完成一篇学术成果时，则认为共现并构建一条相关联的边，否则没有。

(2)接着采用Gephi构建核心作者的知识图谱如图5所示，图中圆圈代表发文作者，圆圈越大发文量越多，反之越少；连线代表作者间的共现关系。

图 5 所对应的关系图谱共构建了 497 个核心作者和 1095 条关系,它将经常合作的科研群体聚集在一起,形成了以余跃生、顾晓艳、赵凌、潘朝霖、何燕、吴昌学、刘世彬、单可人等学者为核心的学术研究团体。同时,图 5 过滤了关系较为简单的群体,例如硕士论文或博士论文中学生与导师的关系,这些两两关系对水族文献群体发现贡献较少。

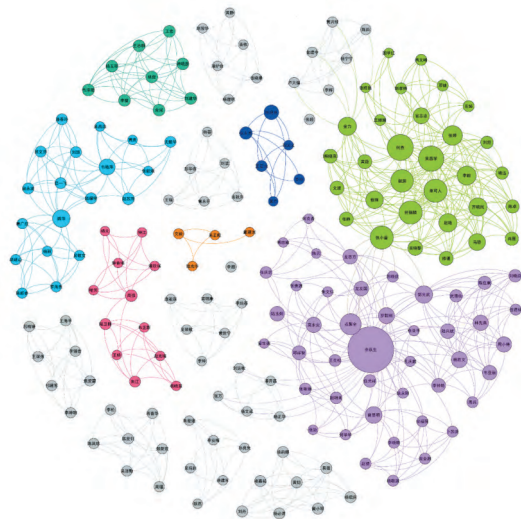


图 5 水族文献发文作者共现知识图谱

图 6 是通过知识图谱分析发现的水族文献核心科研群体,这些科研群体对应核心成员、科研主体机构及单位、主要研究方向和代表学术成果如表 8 所示。其中以余跃生为首的团队以水族基因、水族医学、遗传学方向为主;以顾晓艳和王亚琼为首的团队主要研究水族体育、水族文化传承;以何燕为首的团队研究方向是水族医学、水族基因、心血管疾病;以赵凌为首的团队主要研究水族音乐和水族乐器;以潘朝霖老师为首的团队主要研究水书、水族图腾、水族祝词;以张振江为首的团队以水族村落、水族祖灵和民族语言为主。这些团队主要来自于黔南民族医学高等专科学校、黔南民族师范学院、贵阳医学院、黔南民族师范学院、贵州民族大学、中山大学等机构。

2.6 水族文献主题挖掘及演化趋势

本文采用 LDA 模型对水族文献主题进行深层次挖掘,得到了水族文献五大主题对应的关键词,如表 9 所示。其中,水族文化主题是关于水族民俗、文化变迁、农耕文化、水族图腾、水书文字、特色旅游的文

献。水族体育主要以传统民族活动为主,典型的包括水族舞蹈、赛马、吞口舞、水族武术、棋类活动等。水族音乐是用于祭祀、节庆、丧葬仪式的少数民族音乐,在水族传统的卯节、端节、敬霞节通过乐器铜鼓、大皮鼓演奏古歌、酒歌等。水族医学是针对水族聚居群体的研究,具有代表性的包括水族本土人群的遗传性状、基因频率、血型、民族族源、地中海贫血等研究。水书主要用来记载水族民俗、天文、地理、文化、婚丧嫁娶等信息,于 2006 年被列为国家级非物质文化遗产。但是很多水族文献显示了这些特有的传统水族文化有逐步消退的趋势,保护水族文化刻不容缓。

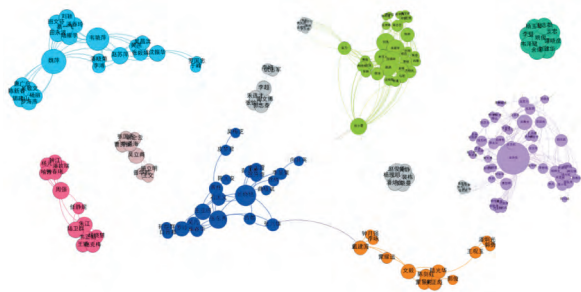


图 6 水族文献核心科研群体

表 8 水族文献核心科研群体表

科研群体 核心作者	科研主体 机构	主要研究 方向	代表学术成果
余跃生;陆玉娟;戎聚全; 张庆忠;孔庆鹏;王克松; 杨胜文;罗裁刚;张勋;	黔南民族 医学高等 专科学校	遗传学 水族医学 水族基因	1) 贵州水族人群线粒体 DNA 序列多态分析 2) 贵州苗族、水族 5 项群体遗传学特征的研究 3) 水族老年人高血压流行病学研究
顾晓艳;王亚琼;张东秀; 黄 胜;张兴雄;张兴奇; 石国义;罗 玲;单春华;	黔南民族 师范学院	水族体育 水族文化 体育文化	1) 水族传统体育舞蹈的保护与传承 2) 对水族山寨原生态传统体育文化的调查研究 3) 水族传统体育文化的价值及其传承策略初探
何燕;单可人;任锡麟;谢 渊;张小蕾;张 婷;吴昌 学;熊晓燕;齐晓岚;	贵阳医学 院	水族医学 水族基因 心血管疾病	1) 贵州三都水族 Y 染色体单倍型频率分析 2) 贵州三都水族 B-地中海贫血筛查及基因分析 3) 少数民族葡萄糖-6-磷酸脱氢酶基因突变检测
赵凌;谢传红;石维武;	黔南民族 师范学院	水族音乐 音乐舞蹈 水族乐器	1) 贵州三都水族端节铜鼓音乐文化考察与分析 2) 马联村水族端节铜鼓音乐文化初探 3) 水族音乐研究综述
潘朝霖;韦宗林;王炳江;	贵州民族 大学	水书 水族文化 水族图腾	1) 水族鱼图腾析 2) 水族汉族二十八宿比较研究 3) 纸上化石——水书
张振江;陈颖照;何自成; 林嘉莉;杨国章;李云霞;	中山大 学	水族村落 水族祖灵 水族卯节	1) 贵州荔波现代水族体质研究 2) 水族祖灵信仰研究 3) 水族村落的民族传统空间结构——以贵州三洞乡为主要对象的调查与研究

图 7 展示了 1953 年以来水族文献的主题演化趋势,包括 1953-1995 年、1996-2000 年、2001-2005 年、2006-2010 年、2011-2015 年、2016-2018 年中不同年份的水族文献所研究的热点主题关键词。从图 7 可以发现,1953-1995 年主要研究水族基因、遗传医学、水书等主题;1996-2000 年主要研究水族文化、服饰、传统节日等主题;2001-2005 年主要研究水族文化、水族医药、

民族旅游等主题;2006–2010 年主要研究水族文化、水书、水族体育等主题;2011–2015 年主要研究水族文化、水族传承、贵州发展等主题;2016–2018 年主要研究水族文化、民俗保护、水族体育、水书和医学等主题。

表 9 LDA 模型识别水族文献主题关键词

主题	主题关键词
水族文化	水族文化; 民族文化; 民俗; 传承保护; 非物质文化遗产; 信仰; 文化产业; 原生态; 民族认同; 演变; 语言文字; 农耕文化; 水书; 马尾绣; 族群; 艺术; 特色旅游; 文化变迁; 图腾; 火耕水耨; 铜鼓;
水族体育	水族; 传统体育; 体育文化; 民族传统体育; 舞蹈; 文化变迁; 赛马; 非物质文化遗产; 进化机理; 棋类活动; 少数民族; 吞口舞; 文化传承; 农耕民族; 传统体育舞蹈; 村落集体活动; 水族武术;
水族音乐	水族; 音乐; 少数民族音乐; 民歌; 铜鼓音乐; 酒歌; 音乐教育; 音调特征; 丧葬仪式; 祭祀; 音乐文化; 流变; 蹻跳圆; 古歌; 乐器; 三都水族; 音乐语境; 铜鼓; 传统音乐; 四音音组; 端节; 敬霞节;
水族医学	水族医学; 冠心病; 民间习俗; 中医体质; 水族医药; 民族族源; 遗传; ABO 血型; 基因频率; 群体遗传学; 少数民族; Rh 血型; 遗传性状; 医药; 聚合酶; 染色体; 人类学; 地中海贫血; 水族人群;
水书文字	水族; 水书; 水语; 民族文化; 马尾绣; 水书教育; 水族源流; 水族银饰; 水书学; 水书文化; 文化遗产; 非物质文化遗产; 文字; 保护与传承; 社会记忆; 古文明; 祝词; 押韵; 本土传承; 易经;

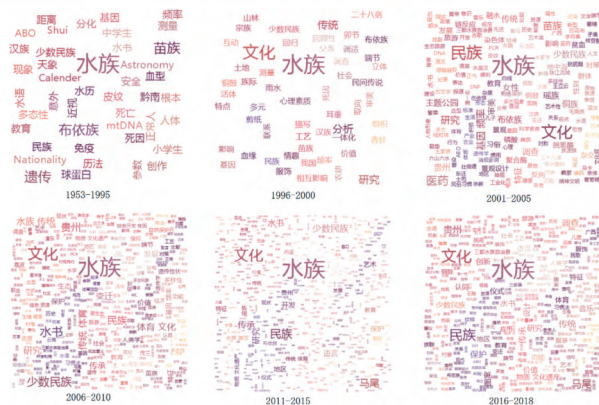


图 7 水族文献不同年份的主题关键词演化趋势

3 研究结论

本文采用计量分析方法和知识图谱方法研究中国知网的水族文献,涉及 1953–2018 年共 990 篇水族领域的学术成果。研究成果如下:

(1)通过对水族文献的高影响力期刊、报纸及机构计量分析发现:①民族类和人文社科类的期刊对水族文化研究做出了重要的贡献,如《黔南民族师范学院学报》《贵州民族研究》《人类学学报》等。②大学是水族领域研究机构的中坚力量,这些大学主要分布于贵州、云南、广西等省份,这些省份也是水族的聚居区;同时民族类和师范类大学也对水族文化有重要的贡献,如贵州民族大学、中央民族大学、华东师范大学等。③报纸是学术成果的重要组成部分,对水族文化和水族传

承具有一定的贡献,它能让社会大众接触到水族文化内涵、水书艺术和迁徙漂泊的历史,具有代表性的包括《贵州民族报》《黔南日报》《中国民族报》。

(2)高被引论文期刊和机构主要分布于民族类、师范类、人文社科专业较强以及西南水族聚居区所在的高校。被引频次前 10 位的学术论文的被引频次都是 20 次以上,是所有水族文献篇均被引用次数(2.15 次)的 10 倍,这些作者是水族文化研究的核心力量,这一定程度上反映出高校对我国民族文化研究的影响程度。

(3)通过水族文献时间跨度分析展示了水族文献研究的进程和不同年份的文献分布情况。最早一篇论文发表在 1953 年,是凌慰民老师发表在《中国金融》的“少数民族地区的农贷工作不能违反民族政策”。被引用 10 次及以上的文献共 45 篇,被引用 5 至 9 次的文献共 64 篇,被引用 2 至 4 次的文献共 151 篇,仅被引用 1 次的文献共 137 篇。2012 年被引用过的水族文献最高,共 45 篇;2006 年被引用的总频次最高,为 277 次,其中 10 篇文章被引用频次 10 次以上,为高被引论文;2004 年单篇均被引用频次最高,平均每篇论文被引用 10.69 次。

(4)高引论文的核心作者分析采用发文量、被引频次、篇均被引频次三个指标评估核心作者,并挖掘出 40 位核心作者,其中余跃生发表了水族相关的文章 23 篇,被引频次为 174 次,篇均被引频次为 7.57 次;顾晓艳发表了水族文献 16 篇,被引频次为 176 次,篇均被引频次 11 次;赵凌发表了水族领域的论文 14 篇,被引频次为 25 次,篇均被引频次 1.79 次。

(5)通过知识图谱构建了水族文献作者之间的合作关系,挖掘出水族文献的核心科研群体,形成了以余跃生、顾晓艳、赵凌、潘朝霖、何燕、吴昌学、刘世彬、单可人等学者为核心的学术研究团体。其中以余跃生为首的团队以水族基因、水族医学、遗传学方向为主;以顾晓艳和王亚琼为首的团队主要研究水族体育、水族文化传承;以何燕为首的团队研究方向是水族医学、水族基因、心血管疾病;以赵凌为首的团队主要研究水族音乐和水族乐器;以潘朝霖老师为首的团队主要研究水书、水族图腾、水族祝词;以张振江为首的团队以水族村落、水族祖灵和民族语言为主。这些团队主要来自于黔南民族医学高等专科学校、黔南民族师范学院、贵阳医学院、黔南民族师范学院、贵州民族大学、中山大学等机构。

(6)通过 LDA 模型识别出水族文献的五大主题(水族文化、水族体育、水族音乐、水族医学和水书文字)及关键词,并采用词云展示了 1953 年以来水族文献的主题

演化趋势,包括 1953–1995 年、1996–2000 年、2001–2005 年、2006–2010 年、2011–2015 年、2016–2018 年中不同年份的水族文献所研究的热点主题关键词。

总之,本文的研究成果具有重要的理论意义和实用价值。该结果展示了我国水族文献的发展历程、高引论文、核心科研群体的研究热点,指出了水族领域高

被引论文的主要特征及影响,识别出水族文化、体育、音乐、医学和水书五大主题的关键词及历年的主题演化趋势,对传承与弘扬水族传统文化具有重要意义,为水族文化的研究和进一步发展提供相关启示,为后续的水族引文分析、水族迁徙研究和水族文化传承提供有效支撑。

参考文献:

- [1]饶文谊,梁光华. 关于水族水字水书起源时代的学术思考[J]. 原生态民族化学刊,2009(4):90–93.
- [2]姜春林. 基于文献计量学时法引文的案例分析[J]. 现代情报,2005(10):140–145.
- [3]梁永霞,杨中楷,刘则渊. 基于 CSSCI 的中国引文分析的可视化研究[J]. 情报研究,2008:34–38.
- [4]黄晓斌,张欢庆. 我国情报学高被引论文分析[J]. 情报科学,2018,36(1):54–60.
- [5]蔡文伯,马杰. 我国民族教育研究文献态势的计量分析[J]. 民族教育研究,2014,25(2):138–144.
- [6]王宗水,赵红,刘宇,秦续忠. 社会网络研究范式的演化、发展与应用——基于 1998–2014 年中国社会科学引文数据分析[J]. 情报学报,2015,34(12):1235–1245.
- [7]徐庶睿,章成志,卢超. 利用引文内容进行主题级学科交叉类型分析[J]. 图书情报工作,2017,61(23):15–24.
- [8]罗双玲,张文琪,夏昊翔. 基于半积累引文网络社区发现的学科领域主题演化分析——以“合作演化”领域为例[J]. 情报学报,2017,36(1):100–110.
- [9]马文博,陈占明. 《经济研究》近十年载文的文献计量与知识图谱分析[J]. 现代情报,2018,38(2):148–156.
- [10]任晓松,孙天美,赵国浩. 中国碳排放研究热点演化知识图谱分析[J]. 科技管理研究,2018(10):235–243.

作者简介:

杨秀璋(1991–),男,贵州凯里人,硕士,研究方向为 Web 数据挖掘、知识图谱、图像识别、文献分析
 收稿日期:2018–11–08 修稿日期:2018–11–14

Econometric Analysis and Knowledge Graph Based on Shui Literature

YANG Xiu-zhang

(School of Information, Guizhou University of Finance and Economics, Guiyang 550025)

Abstract:

With the rapid development of science and technology, academic achievements have exploded, it is becoming increasingly important to mine the core authors, identify high-cited papers, and analyze the subject of literature. Through Python, grabs 990 Shui literature from 1953–2018 in CNKI, and based on the quantitative analysis method of literature, unearths high-impact journals, newspapers and institutions in China's Shui literature, analyzes high-cited papers in aquarium, hot time and core authors. Based on the knowledge graph analysis method, constructs the cooperative relationship between the core scientific research groups and authors of the Shui literature, and reveals the theme evolution trend of the Shui literature. The experimental results show the development history of China's Shui literature, core research groups and high-cited papers, identify the key themes of the five themes of culture, sports, music, medicine and character, and the theme evolution trends over the years, for the study of Shui culture and further development provides relevant inspiration.

Keywords:

Shui Literature; Econometric Analysis; Knowledge Graph; High-Cited Papers; CNKI