

我国图书情报学的认知结构及其演化^{*}

张 斌 贾 茜

摘 要 共词分析在学科认知结构分析中具有非常重要的作用。本文利用信息检索中的“前方一致”原则,从中国社会科学引文索引收录的 14 种图书情报学期刊的关键词数据集中找出 100 个频繁出现的词。在这 100 个词的基础上,利用层次聚类得到我国图书情报学的四个主要分支,利用层次聚类、多维尺度分析和热点图相结合的方法分析这 14 种期刊的内容特征,并研究它们的演化轨迹和使用趋势,从而揭示认知结构及其演化过程。研究发现:我国图书情报学在不同时间段有不同的关注点,它们也有着自身的生命周期,目前整个学科越来越将“信息”作为研究核心,热点偏重于技术应用与管理创新。图 12。表 4。参考文献 32。

关键词 图书情报学 认知结构 演化 共词分析 多维尺度分析 热点图

分类号 G350

The Cognitive Structure of Library and Information Science in China and Its Evolution

Zhang Bin & Jia Xi

ABSTRACT Co-word analysis plays a pivotal role in the cognitive structure analysis of discipline. This paper collects a dataset of keywords from fourteen key journals of library and information science (LIS) in the Chinese Social Sciences Citation Index (CSSCI) and extracts the most frequent 100 words based on the right truncation method. Based on the 100 words, four major branches are identified by hierarchical clustering, and content characters in fourteen journals are analyzed by hierarchical clustering, multidimensional scaling and heat map. Furthermore, evolution trajectory and usage trends of the 100 words can reveal the cognitive structure of LIS and its evolutionary process. This paper finds that the LIS in China has different concerns at different times, which have their own life-cycle. Currently, the LIS tends to treat “information” as its research core, with emphasis on technology application and management innovation. 12 figs. 4 tabs. 32 refs.

KEY WORDS LIS. Cognitive structure. Evolution. Co-word analysis. Multidimensional scaling analysis. Heat map.

20 世纪 60 年代末,由于计算机在文献加工处理中的应用,国外很多图书馆学院演变为图书馆与信息学院,或图书情报学院。在我国,情报学起源于文献学和图书馆学,并与之相互借鉴、融合^[1]。图书情报学(Library and Information

Science, LIS)作为科学共同体,其成员会自发形成并共同拥有一定的认知结构,使得成员能对研究方向和前沿领域形成正确的解释和预期,从而协调自己的行为以适应学科发展和其他成员的需求。文章关键词是研究者试图表征自我

^{*} 本文系国家自然科学基金项目“知识网络的形成机制及演化规律研究”(编号:71173249)的研究成果之一。

通讯作者:张斌,Email:zb0205@126.com

对文献研究内容心智理解的外在编码形式。以文章关键词为基础,探求研究者对研究内容的认知状态,进而勾画学科整体的认知结构,是一个可行而有意义的角度。本文借助中国社会科学引文索引(CSSCI)题录数据,对我国14种图书情报学期刊在1998—2011年间发表的文献中的关键词进行处理,并进行词频分析、共现分析、聚类分析、多维尺度分析和热点图呈现,这种系统性的回顾和整理有助于厘清学科的认知结构及其演化过程。

1 相关研究回顾

在阐述图书情报学这一学科概念时,不同学者给出了自己的解释。Webber^[2]在回顾英国图书情报学的发展历史时,使用Biglan学科分类去评判其学科性质,认为它属于一门应用性学科(applied discipline),其中有硬(hard)和软(soft)两种元素。Robinson和Bawden^[3]在分析英国图书情报学起源和特征时,认为它的主要驱动力来自文档运动、专业图书馆的建立以及现代科技信息的爆炸式发展,而它的主要特征可以总结为五个特点:以主题为导向,广泛的研究方法,具有职业素养,整合信息和技术,记忆保存机构的参与。特别是关于此学科兼具科学探究和专业领域实践这一点,得到了很多学者的认同^[4]。

此外,学者们也观察到了图书情报学的转向问题。如:Cronin^[5]重点讨论了图书情报学的社会学转向(sociological turn),认为认知观点(cognitive viewpoint)的介入会过于强调个人知识状态,这会导致我们忽视社会关系和社会结构;但我们毕竟是社会性动物,认知并不完全是个人现象,也不是很容易能从实验室中得到检验,相反,它具有多重组成成分且是字符上的一种互动。Nolin和Åström^[6]认为图书情报学处于一种支离破碎的危机中,缺少理论层面的发展,并列举了若干转向(认知转向、信息转向、以用户为中心、认识论转向、历史学转向、语用学转

向),鉴于“信息社会”使得信息在不同研究领域都成为热点,因此“信息转向”会使得我们在研究合作中处于强势地位。

总的来看,不管是学科性质还是学科转向,都可以归结到两个研究范式(paradigm)上来,即图书馆学范式和情报学范式^[7]。前者关注图书馆事业,把图书馆当成社会机构看待,研究传统更多地偏向社会科学;后者则关注信息及其交流情况,研究传统更多地偏向信息传播的数学理论和模型构建,比如信息生命周期^[8-10]。Bates^[11]认为情报学范式并不仅仅包括对选择、采集、组织、存取和检索信息的研究,它还有很多元素没有被表达出来。

随着计量对象和研究手段的丰富,很多学者把重点放在图书情报学中的一个特定的子领域进行实证研究。比如:选取若干种期刊,进行作者、文献和期刊的共被引分析^[12-15];抽取标题、摘要、文本中的词或索引词进行共现分析^[16-17];利用词—参考文献共现^[18]、LDA(Latent Dirichlet Allocation)模型^[19]等分析研究主题。早期的这类研究主要从事的是识别研究领域和预测研究趋势,较少从认知结构方面入手。此外,这些分析大多是共时的,将不同时期的数据放在一起处理,而不是随着时间的推移去探索潜在的变化。后续的研究则注意到了这点,并进行方法上的改进。比如:马瑞敏和倪超群^[20]整合“聚类分析”和“pathfinder算法”,基于作者共被引(ACA)分时段研究图书情报学的知识结构,并分析其发展演变情况;Milojević等^[21]抽取标题词(title words),利用大样本数据,将共词分析、聚类分析、多维尺度分析和趋势分析整合在一起研究认知结构;滕立等^[22]也注意到了共词知识网络中的认知结构问题,提出知识群中具有最高外部连接度的知识单元在知识群形成的知识链中具有知识基因的功能,并以石墨烯的发现及其演化为例进行了阐述;任红娟^[23]进一步阐述了将文献的内容和引用特征融合来进行知识结构划分的必要性,提出一种最大文献相似性函数并进行实证。显然,后续的研究

更加注重从多个角度、运用多种方法进行计量分析,也试图描述知识结构或认知结构。

对于知识结构,不同领域有不同的解释。心理学将知识结构定义为在人脑中形成的概念体系,采用概念地图^[24]进行表达;而在图书情报学领域,部分学者将知识结构理解为人类知识的分类,另一些学者如 Brookes^[25]则认为人类知识结构应分为客观知识结构和主观知识结构,前者可以理解为人类知识的分类,具体到学科分析上就表现为学科结构,后者则是指人脑中形成的知识状态,比如概念体系。Brookes 还提出用认知地图(cognitive map)去描述知识结构,认为它是“由概念间关系连接起来的一种概念结构”,这种概念联结网络可以表示学科认知结构,更偏向于心理学中的表达。虽然上述说法间存在细微差别,究其原因对于知识单元(knowledge unit)的理解不同^[26],但通常可以认为它们表达相近的内容,在一定条件下可以通用。本研究所强调的认知结构则偏向更为微观的语言层面。

从语言层面上看,一门学科的概念术语可以被视为控制该学科科学实践的基础器件,将自身从日常语言中区分出来,并能保持学者们在分享观点和实践经验时的稳定性^[27]。在科学实践中所留下的文本文档不仅是记录和传递知识的人工载体,其中的概念术语也体现学者们的认知状态。因此,许多学者将目光聚焦在科学共同体共享的概念系统上,通过在文档中使用的概念术语来辨识认知结构。

概念(concept)是对特征的独特组合而形成的知识单元,而术语(terminology)则是在特定学科领域用来表示概念的称谓的集合。本研究所说的概念术语是一种模糊指代,反映在学者们撰写文档时对标题词和文章关键词的选取上。但是,在期刊文章排版和结构上,不同期刊有不同的要求,文章关键词并不是必须的,比如著名期刊《美国情报科学技术学会会刊》(JASIST)所刊发的文章就没有关键词。此外,由于英文单词是由空格隔开,抽取单词相对容易,因此,国

外文献大多分析文档标题词以跟踪话语形成的过程、领域或学科的认知结构。中文分词技术的准确率一直有待提升,但国内大多数期刊对文章关键词是有要求的,因此,国内文献大多分析文章关键词以达到研究目的。

由文章关键词及其共现关系形成的共词网络为绘制“认知地图”提供了一种有效的工具^[28]。但对于中文来说,文章关键词更多的是以词组(短语)形式出现,并非最小知识单元(词是最小的能够独立运用的语言单位),因此用以分析更为微观的认知结构并不合适。针对这一问题,本研究将利用“前方一致”的原则进行知识单元的粒度划分,运用大样本数据集全面研究我国图书情报学的认知结构及其演化。

2 研究设计

在本研究中,为了探究图书情报学的认知结构,将文章关键词依据信息检索中的“前方一致”原则进行合并。研究基于如下假设:文章关键词表现为词组形式,一组特定的关键词在语义和认知关联上会更为紧密。比如:组成要素、组成因素、组成特征、组成信度、组成要素模型等,都是“组成”+“其它名词或词组”的形式,将这些关键词合并为“组成”,从宏观上讲,“组成”在语义表达上更粗一些,但从认知角度来看上述词组,“组成”描述了这些词组的共同特征和含义,是一个更细的认知表达,载荷了更多的信息。此外,合并还需要最大程度保留词的原意,如“图书”(book)与“图书馆”(library)为两个截然不同语义的词,关键词“图书馆事业”应归并至“图书馆”类,而非“图书”类,即可归并为“图书馆”类的词则不予归并至“图书”类。在这一过程中,一方面利用算法进行循环计算,另一方面在必要时进行人工干预。文章关键词经过这样处理之后,便可以和国外文献分析标题词保持一致,相关结论亦可以在相同粒度层面进行比较。在具体研究方法上,本研究采用词频分析、共现分析、聚类分析、多维尺度分析和热点

图相结合的方法分析我国图书情报学的认知结构和演化趋势。

2.1 研究对象的选取

如同科学计量学一样,本研究也要选择一组合适的期刊代表我国图书情报学领域,也就是说,研究基于如下假设:选定的期刊代表了整个学科领域的认知输出。显然,这一假设会有利于那些发表在选定期刊上的文章所研究的主题,从而带有明显的研究偏好。但由布拉德福定律可知,一门学科所发表论文成果中的大多数会集中在少数核心期刊上。因此,选择一组高排名的期刊(排名标准有影响因子、核心/非核心、索引来源期刊等)将有利于描述该学科领域的核心认知结构。需要说明的是,这种选择策略同样排除了发表在其它学科期刊上的少量文章,这样就会消除学科领域间的知识扩散(knowledge diffusion)。事实上,这些文章更多地

带有跨学科因素,研究的问题所代表的认知内容并非本学科领域的核心。

本研究选择中国社会科学引文索引(CSSCI)收录的14种图书馆、情报与文献学来源期刊(不含档案学期刊)作为研究对象(见表1)。CSSCI从1998年开始运行,其来源期刊会根据收录规则动态调整,从公布的来源期刊目录(2012—2013)来看,《图书馆理论与实践》曾在2008—2009年被降入扩展版,而《国家图书馆学报》在1998—2007年、《图书馆》在2000—2003年、《图书馆建设》在2000—2007年均未被收录,表明这四种期刊历史上有过不稳定状态,予以剔除,也就是说,该领域从建库至今一直收录的只有14种期刊。因此,本研究认为这14种期刊的载文(合计48,944篇)代表了我国图书情报学领域研究的主流,这些文章的关键词代表了作者对于图书情报学研究内容的认知。

表1 图书情报学期刊列表(1998—2011年)

	中文期刊名 (按音序排列)	英文期刊名	缩写	创刊时间 (年)	文章数量 (篇)
1	大学图书馆学报	<i>Journal of Academic Libraries</i>	JAL	1981	1,986
2	情报科学	<i>Information Science</i>	IS	1981	5,170
3	情报理论与实践	<i>Information Studies: Theory & Application</i>	ITA	1964	3,094
4	情报学报	<i>Journal of the China Society for Scientific and Technical Information</i>	JCSSTI	1982	1,717
5	情报杂志	<i>Journal of Intelligence</i>	JI	1982	7,384
6	情报资料工作	<i>Information and Documentation Services</i>	IDS	1980	2,118
7	图书馆工作与研究	<i>Library Work and Study</i>	LWS	1979	3,357
8	图书馆论坛	<i>Library Tribune</i>	LT	1981	4,701
9	图书馆杂志	<i>Library Journal</i>	LJ	1982	4,143
10	图书情报工作	<i>Library and Information Service</i>	LIS	1956	6,418
11	图书情报知识	<i>Document, Information & Knowledge</i>	DIK	1984	2,083
12	图书与情报	<i>Library and Information</i>	LI	1981	2,236
13	现代图书情报技术	<i>New Technology of Library and Information Service</i>	NTLIS	1985	2,725
14	中国图书馆学报	<i>Journal of Library Science in China</i>	LJSC	1957	1,812

2.2 数据获取与处理

本研究从 CSSCI 数据库中获取这 14 种期刊在 1998—2011 年(14 年)之间发表文章的题录

数据,使用 Python 编程语言、SPSS、R 语言等多种方式进行处理,具体过程如图 1 所示,主要为关键词处理与数据分析过程。

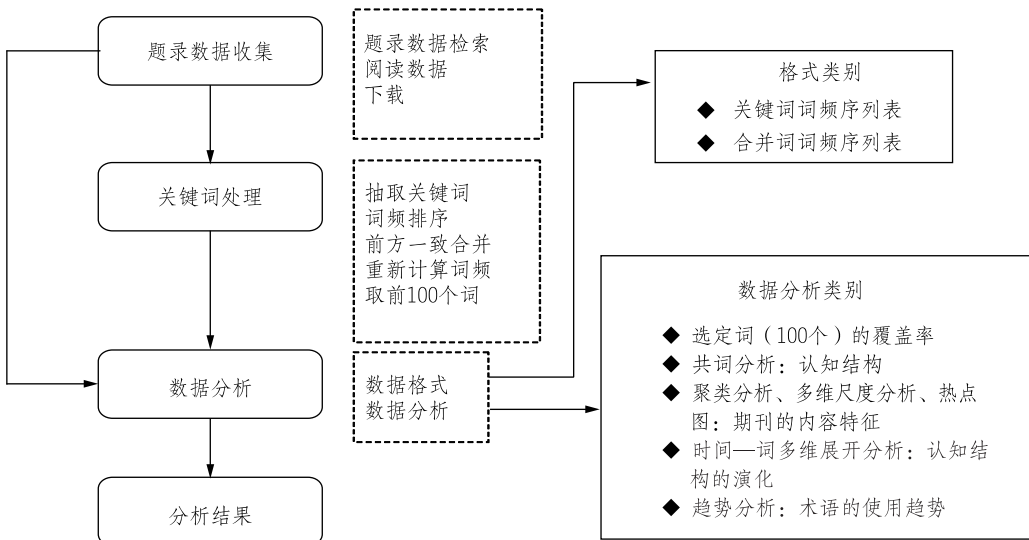


图 1 数据获取与处理

其中,关键词处理过程具体如下:

①获取样本期刊中所有文献的关键词,按照词频排序;

②按照“前方一致”的原则进行关键词合并;

③对归并后的词进行人工调整;

④对人工调整后的词重新计算词频;

词频是内容分析的一个重要方面,频繁出现的词反映了科学共同体所关心的问题,也反映了他们的认知。需要说明的是,人工调整不包括对同义词的合并(主要涉及中英文对应),如:“网络”和“Web”、“因特网”和“internet”。这主要是因为即便是同义词,其不同形式表达依然能反映出认知上的细微差异,特别是在加入时间和空间因素之后,可以很容易辨析出中国学者在概念上的接受进度及程度。本研究选取前 100 个词作为研究对象,如表 2 所示。

为了证明所选取的 100 个词可以代表大多

数文献,本研究计算了这些词在样本期刊文章中的覆盖率,如图 2 所示。其中,“信息”一词覆盖 24.31% 的文献,前四个词“信息”、“图书馆”、“网络”、“知识”覆盖 54.58% 的文献,超过半数,所选取的 100 个词能覆盖 93.80% 的文献。因此,这 100 个词能够很好地代表样本期刊的绝大多数文章,可作为图书情报学认知结构的研究基础。

随后,本研究以这 100 个词为基础探究图书情报学领域认知结构的特点和发展演化规律。主要步骤如下:①计算这 100 个词的共词矩阵;②利用 SPSS 绘制聚类图,以代表图书情报学领域认知结构的总体概况;③按照不同期刊对这 100 个词进行词频计算,绘出期刊聚类图、多维尺度图,并使用 R 语言绘出热点图;④按照时间(以年为单位)对这 100 个词进行词频计算,进行多维展开分析;⑤进一步总结这 100 个词的使用趋势。

表 2 样本期刊中排序前 100 的词

序号	词	词频	序号	词	词频	序号	词	词频	序号	词	词频
1	信息	11,898	26	学术	803	51	组织	447	76	影响	315
2	图书馆	10,534	27	研究	748	52	中文	446	77	人才	308
3	网络	4,980	28	科技	706	53	藏书	446	78	职业	306
4	知识	4,890	29	虚拟	674	54	因特网	442	79	智能	297
5	数字	4,155	30	个性	659	55	专利	431	80	XML	292
6	文献	3,363	31	评价	639	56	古籍	412	81	现代	291
7	高校	2,663	32	本体	619	57	自动	401	82	地方	288
8	电子	2,310	33	政府	601	58	比较	383	83	国家	284
9	情报	2,007	34	技术	598	59	版权	379	84	国际	280
10	数据	1,983	35	搜索	595	60	著作权	378	85	internet	276
11	公共	1,456	36	大学	583	61	教育	378	86	高等	275
12	企业	1,414	37	科学	578	62	专业	376	87	业务	264
13	资源	1,405	38	语义	576	63	阅读	369	88	内容	263
14	图书	1,367	39	计算机	562	64	文化	367	89	目录	262
15	服务	1,351	40	系统	555	65	网站	366	90	人文	252
16	竞争	1,126	41	核心	528	66	理论	366	91	全文	250
17	读者	1,098	42	发展	527	67	质量	363	92	法律	249
18	用户	1,081	43	引文	518	68	文本	346	93	指标	245
19	社会	1,075	44	元数据	504	69	美国	346	94	多媒体	244
20	Web	984	45	馆藏	502	70	网上	343	95	统计	243
21	中国	962	46	参考	499	71	经济	337	96	隐性	242
22	学科	948	47	开放	466	72	创新	335	97	咨询	239
23	管理	888	48	分类	458	73	市场	329	98	模糊	238
24	期刊	850	49	书目	452	74	特色	328	99	科研	237
25	检索	811	50	主题	449	75	教学	322	100	编目	236

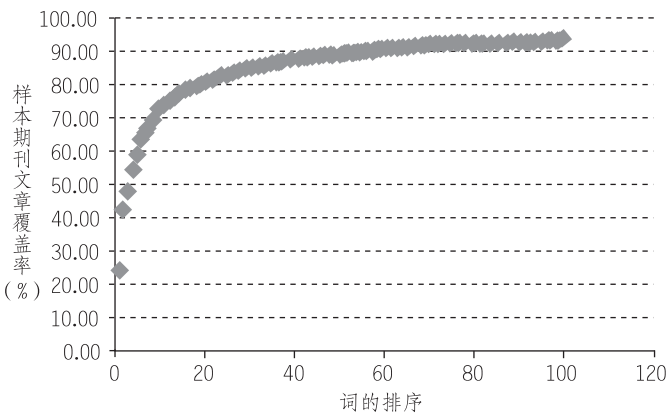


图 2 所选取的 100 个词在样本期刊文章的覆盖率

3 分析与结果

3.1 认知结构的概览

研究采用聚类分析中最常用的分层聚类法 (Hierarchical Clustering Method), 在 IBM SPSS Statistics 20 中完成, 选择离差平方和法 (Ward's method) 进行聚类, 数据类型为计数 (Counts), 测度方法为斐方方法 (Phi-square Measure)。这 100 个词聚类后, 可清晰地分为四大类, 将其命名为数字信息的组织与管理 (图 3)、图书馆学 (图 4)、信息计量与评价 (图 5) 和组织知识管理

(图 6)。可以进一步分析每个大类包含的内容及其子结构。

3.1.1 数字信息的组织与管理

这一类别从“本体”至“法律”, 是最大且最多样化的一类, 主要描述的是数字信息的组织与管理方面的内容, 归属于图书情报学的实践领域。数字信息是一个非常宽泛的概念, “数字”体现了当前的新环境, “信息”体现了研究核心依然是“信息—知识—情报 (智能)”的整条信息链^[29]。图 3 显示它可以分为若干小类, 包括语义网、信息检索、计算机、自动中文系统、内容数据、全文多媒体、分类主题文本、用户个性 Web、电子网上虚拟参考咨询、元数据 XML、数字版权、著作权、法律。

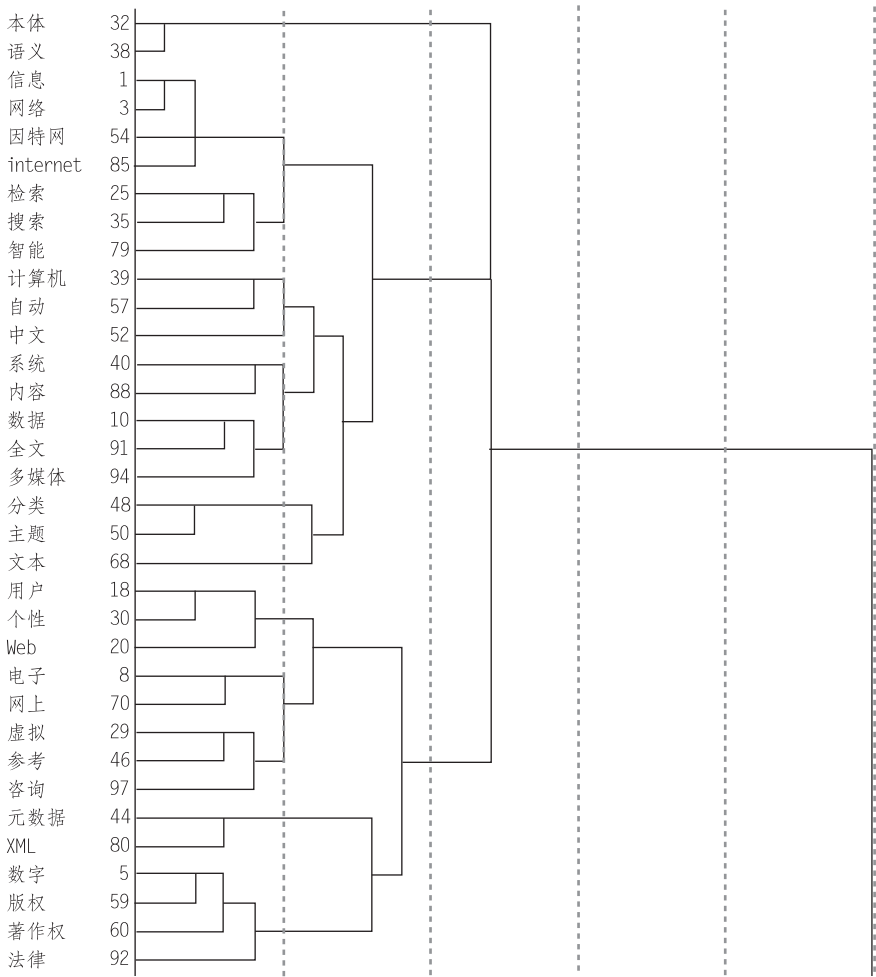


图 3 聚类一：数字信息的组织与管理

3.1.2 图书馆学

这一类别从“书目”至“阅读”,主要描述的是图书馆学的研究内容。图书馆学是研究图书馆的发生发展、组织管理,以及图书馆工作规律的科学,其目的是总结图书馆工作和图书馆事业的实践经验,建立科学的图书馆学理论体系,

以推动图书馆事业的发展,提高图书馆在人类社会中的地位 and 作用。图 4 显示它可以分为若干小类,包括书目工作、高校图书馆、特色馆藏、图书馆职业教育、读者服务与公共文化建

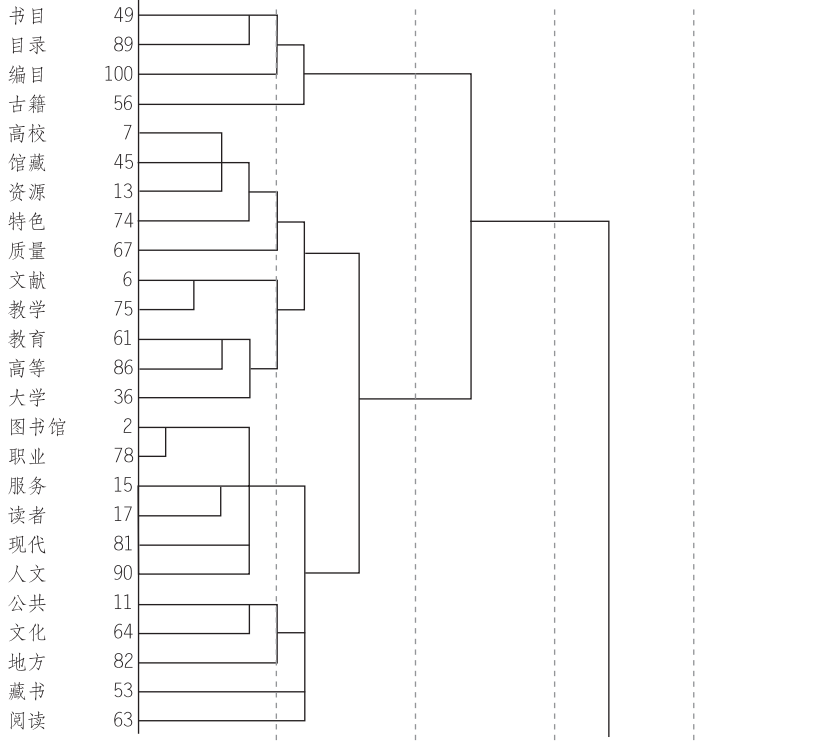


图 4 聚类二:图书馆学

3.1.3 信息计量与评价

这一类别从“期刊”至“国家”,主要描述的是信息计量与评价的研究内容,属于情报学领域,是定量研究中的一个重要分支。信息计量学的研究开始于 19 世纪末,目前可以再分为五个大的分支,分别为文献计量学(bibliometrics)、科学计量学(scientometrics)、信息计量学(informetrics)、网络计量学(webmetrics)和替代计量学(altmetrics)^[30-32]。图 5 显示它可以分为若干小类,包括计量对象与手段、情报学基础理论、国家及国际的比较研究等。

3.1.4 组织知识管理

这一类别从“评价”至“组织”,主要描述的是组织知识管理的研究内容。组织知识管理是把知识(信息)作为最重要的资源,把知识和知识活动作为企业的财富和核心,对信息的获取和传播、知识的学习和运用、知识的创新、知识交换,以及企业内部知识的共享和共享的结构、知识水平的提高进行管理,发挥员工和集体的智慧,在知识创新中谋求生存和发展。图 6 显示它可以分为若干小类,包括指标与评价、技术专利、市场经济、隐性知识、组织业务管理、人才创新等。

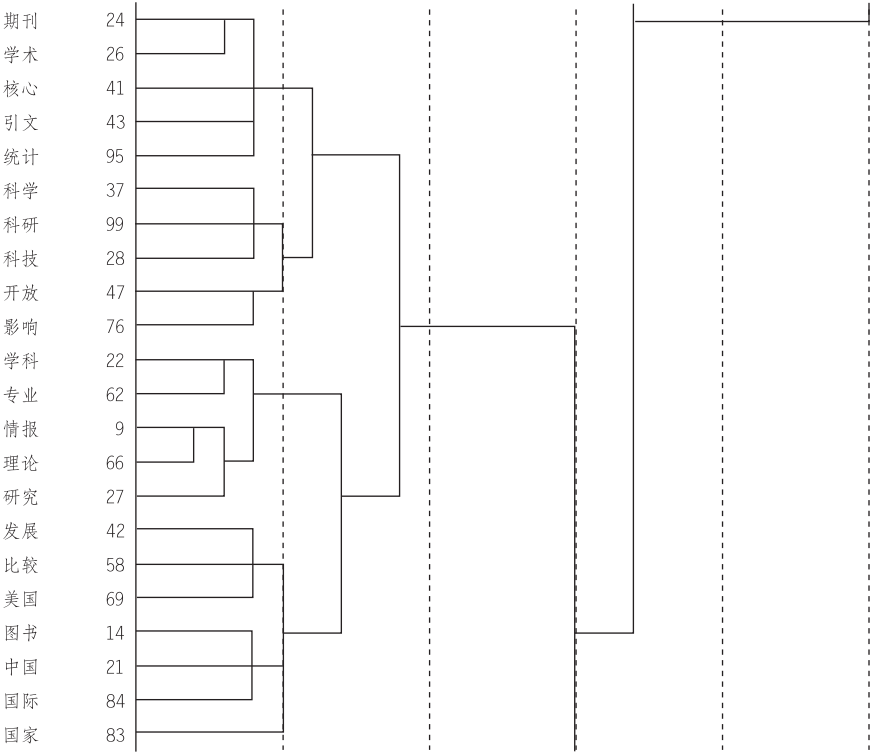


图 5 聚类三:信息计量与评价

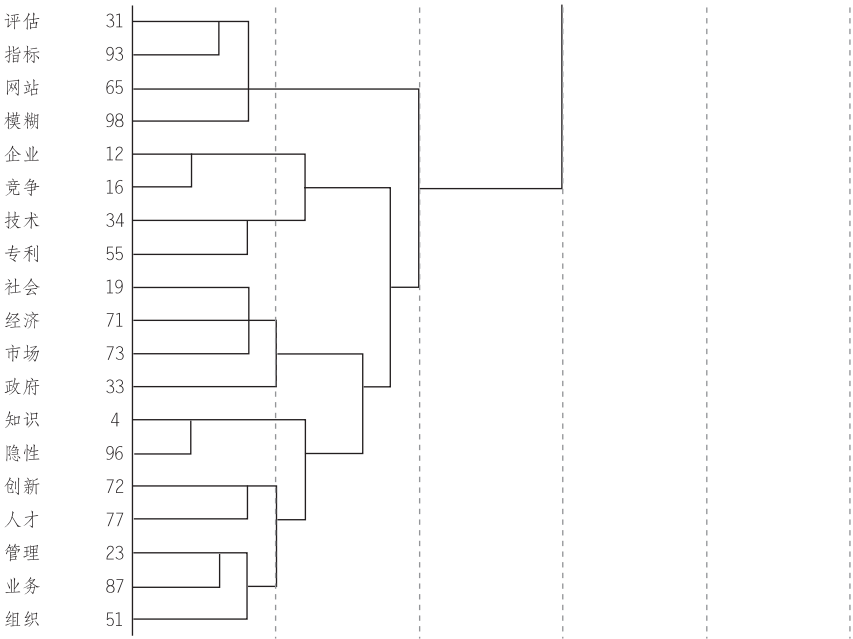


图 6 聚类四:组织知识管理

3.2 期刊的内容特征

上述聚类结果所展现的分布结构可以从宏观上勾勒出十几年来中国图书情报学领域的认知结构。为了使这一认知结构呈现得更加清晰,本研究将这 100 个词与样本期刊关联起来进行分析,研究发现这 100 个词在各个样本期刊中有着明显的分布特点。

利用 IBM SPSS Statistics 20 中的多维尺度分析(ALSCAL)方法,选择从数据创建距离,数据类型为计数,测度方法为斐方方法,测度水平为序数(ordinal),尺度模型为欧式距离(Euclidean distance),得到期刊多维尺度分析结

果,统计量 $\text{Stress} = 0.08462$, $\text{RSQ} = 0.96896$,表明取得了较好的拟合效果。图 7 展示了样本期刊的相似程度,位置相近的期刊其研究内容也比较一致,如《情报科学》与《情报杂志》两种期刊较为相似。维数 1 在 0 原点处的显示线可视为图书馆学与情报学的学科分界线,越靠近左侧其研究内容越靠近图书馆学,越靠近右侧其研究内容越靠近情报学。其中,情报学类期刊中《情报科学》、《情报杂志》和《情报理论与实践》的联系较为紧密,而《现代图书情报技术》和《情报学报》则由于其研究内容的独特性,相对于其它期刊来说更为孤立。

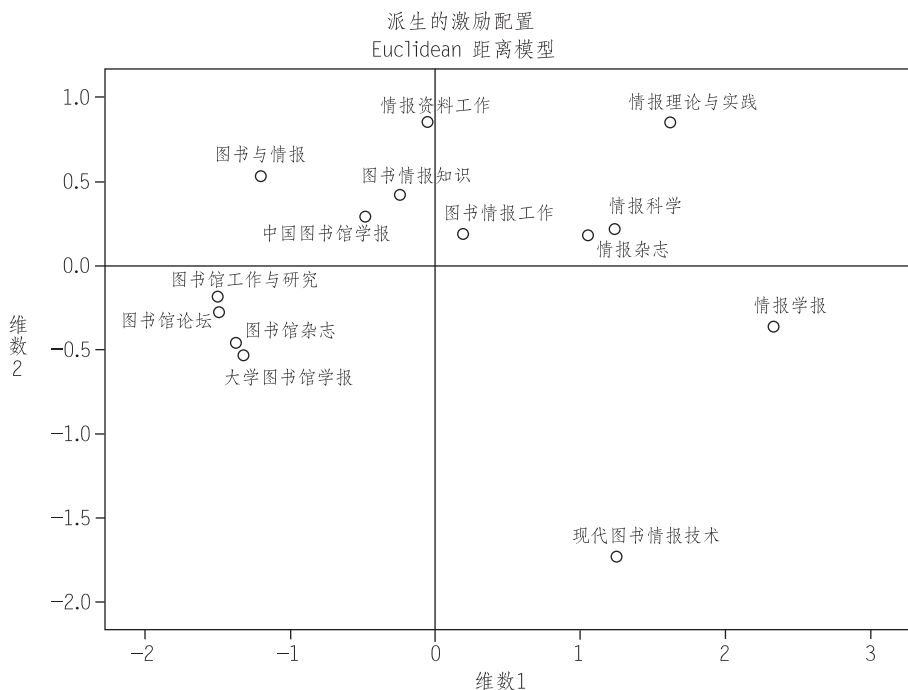


图 7 期刊多维尺度分析结果

为进一步研究不同期刊对这 100 个词的选择倾向,将其进行了归一化处理,尽量减小词间词频大小的纵向对比,更突出期刊间的横向对比。具体公式如下:

$$q_{ij} = n_{ij} / N_i \quad (\text{公式 1})$$

其中, n_{ij} 为词 i 在期刊 j 出现的频次, N_i 为词 i 出现的总频次, q_{ij} 为词 i 在期刊 j 出现频次

的归一化结果, $i = 1, 2, \dots, 100$, $j = 1, 2, \dots, 14$ 。

由 R 语言绘制了样本期刊与 100 个词之间的热点图,如图 8 所示。图中方格颜色由紫(暗色)变蓝变绿再变黄(亮色),亮色表明该期刊对于某些词具有偏好。从图中可明显看出,情报学期刊中的《情报科学》、《情报杂志》和《情报理论与实践》,明显偏好信息计量与评价和组织

知识管理方面的研究内容(区域1),而《情报学报》和《现代图书情报技术》更加偏好数字信息的组织与管理方面的研究内容(区域2)。图书馆学期刊中,《图书馆工作与研究》、《图书馆论坛》、《图书馆杂志》、《图书与情报》和《大学图

书馆学报》偏好图书馆学的研究内容(区域3)。另外四种期刊《图书情报工作》、《图书情报知识》、《情报资料工作》和《中国图书馆学报》则无特别明显的偏好,可归属于两栖刊物,这也和各个期刊的定位较为一致。

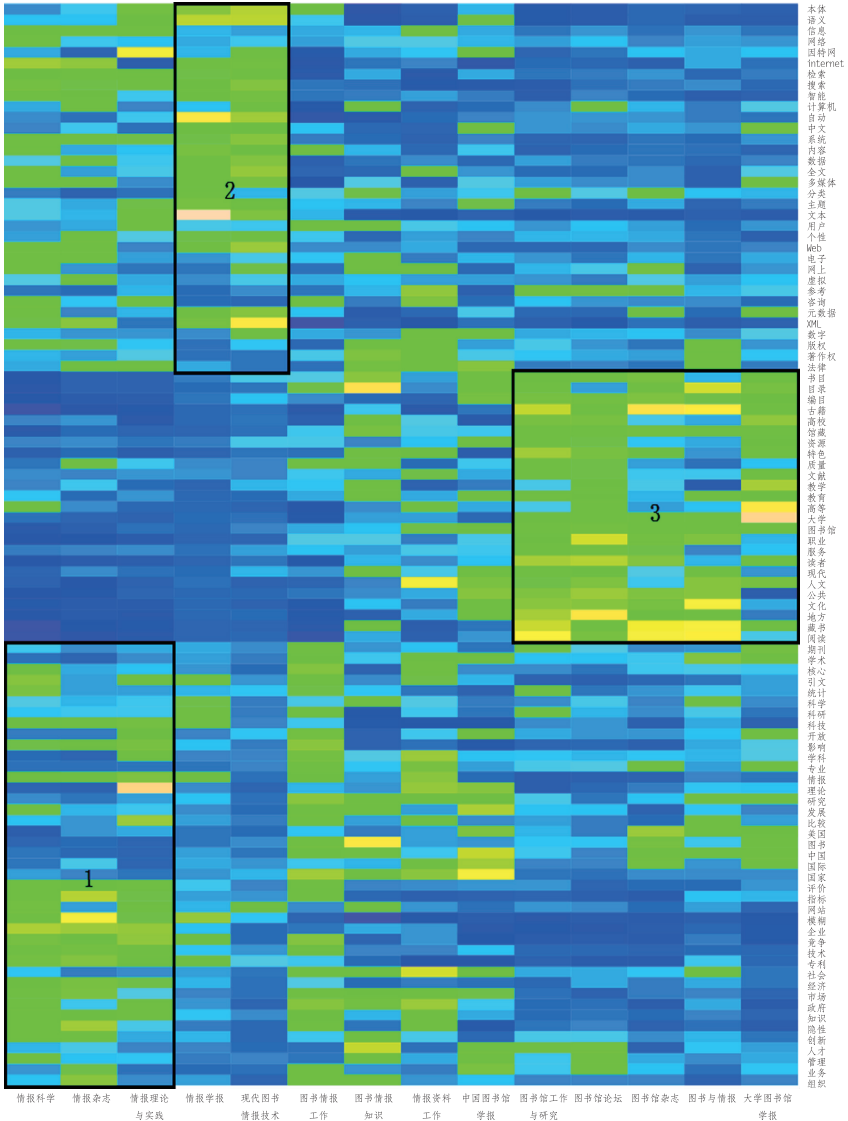


图8 “期刊—100 个词”热点图

由此可以明显看出《情报学报》和《现代图书情报技术》本身的特殊性。通过词频统计发现,《情报学报》偏好的词有文本、自动、语义等

(大于0.14),特别是“文本”一词所在的方格颜色为米黄色,表明研究主要集中在对文本的分析与处理上。《现代图书情报技术》偏好的词有

XML、语义、本体、自动、Web、全文、系统、元数据、搜索、数据、文本等(大于0.14),研究主要集中在网络环境下信息资源的自动处理上。此外,《情报学报》还被列为国家自然科学基金管理学部的重要期刊(A类),许多管理学领域的学者尤其是从事信息系统、商务智能和知识管理研究的学者会在该期刊上发表文章,也增加了它的特殊性。不妨这样推断,由于这两种期刊在研究内容上偏好信息技术应用,与其它期刊明显不同,因此被孤立出来。

需要说明的是,除《情报学报》的“文本”之外,图8中还有两个方块呈米黄色,即《情报理论

与实践》的“理论”和《大学图书馆学报》的“大学”。前者指出《情报理论与实践》的特色是研究探讨图书情报学的理论方法,后者指出《大学图书馆学报》的特色是研究探讨大学和高校图书馆相关理论与实践,有强烈的针对性。

利用 IBM SPSS Statistics 20 中的分层聚类法对期刊进行聚类分析,如图9所示。通过多维尺度分析所得出的期刊间相似关系被进一步肯定,同时树状图也提供了一些新的信息,即期刊在距离5以内可以明显地被划分为三类,而《情报学报》和《现代图书情报技术》依然非常特殊。图7、图8和图9所展现的内容可以相互印证。

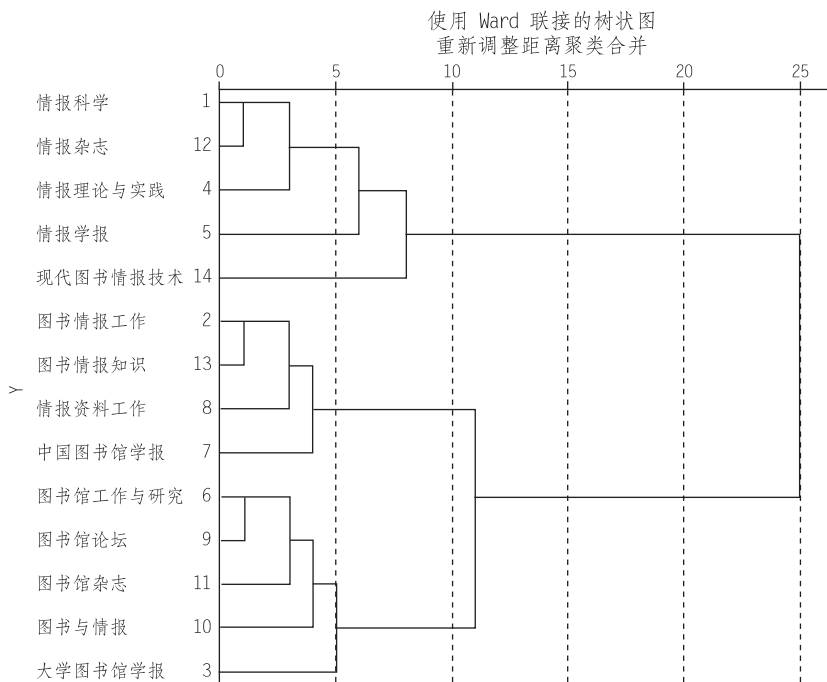


图9 期刊聚类分析结果

3.3 认知结构的演化

除期刊维度外,词在时间维度上的分布结构也有一定的规律和特点。为此也进行了归一化处理,尽量减小词间词频大小的纵向对比,更突出时间上的横向对比。具体的公式如下:

$$P_{iy} = n_{iy} / N_i \quad (\text{公式 } 2)$$

其中, n_{iy} 为词*i*在年份*y*出现的频次, N_i 为

词*i*出现的总频次, p_{iy} 为词*i*在年份*y*出现频次的归一化结果, $i = 1, 2, \dots, 100, y = 1, 2, \dots, 14$ 。

利用 IBM SPSS Statistics 20 中的多维展开分析(PREFSCAL)方法,尺度模型为加权欧几里得,近似值为相似性,初始配置为 Ross-Cliff,得到时间—100 个词多维展开结果,统计量 Normalized Stress = 0.0087552, Dispersion Accounted

for = 0.9912448,表明取得了较好的拟合效果,如图 10 所示。在这种可视化表达方式中,同一时间段内所使用到的词被聚集在一起,旁边显示了时间标签。换句话说,环绕在某时间标签周围的词代表该年份中这些词出现的次数相对

较多,即该方面的研究也相对较多。时间标签按逆时针方向均匀地形成了一个连续的“环”,这表明 1998—2011 年间认知一直在发生变化,由此可以展现出认知结构在时间上的演化轨迹。

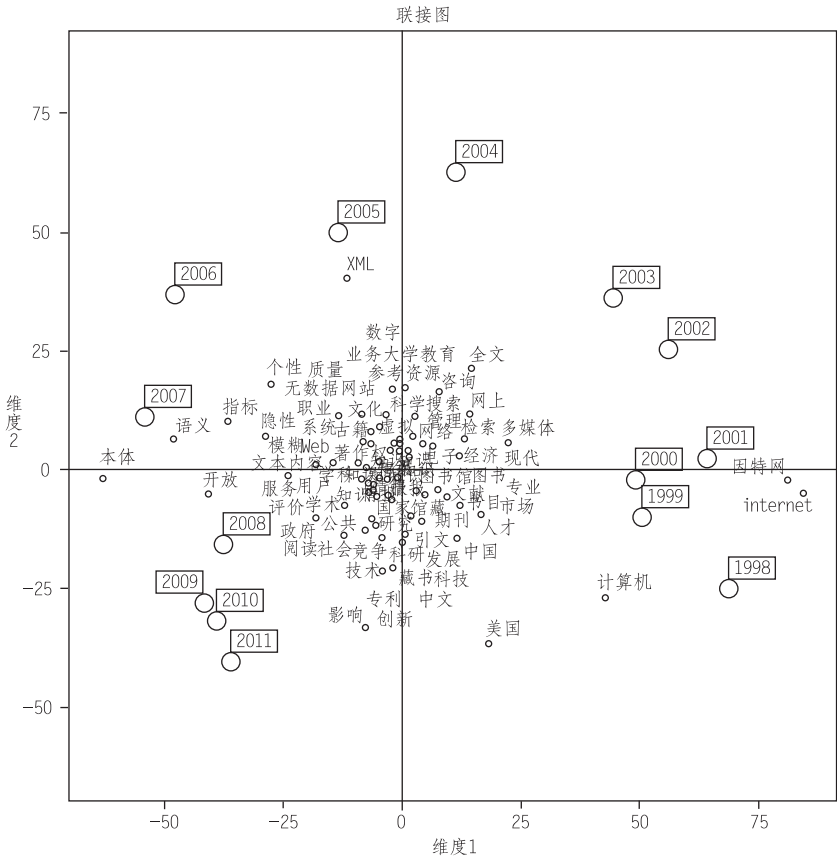


图 10 “时间—100 个词”多维展开分析

在图 10 中,高频次的词(如:信息)会集中在中心位置,反映了图书情报学领域的核心,这些词在各个年份中均有出现,研究较为持续。从演化轨迹上看,1998—2001 年,研究关注在 Internet、因特网、计算机等词上;2002—2003 年,研究转向全文、网上、检索等词上;2004—2005 年,研究显然已经转移到“XML”上;2006—2007 年,研究关注在“本体”和“语义”这两个词上,表明语义网成为热点;2008—2011 年,研究关注在开放、评价、用户、政府、公共、阅读、专利、影响、

创新等词上。显然,不同时间段的研究关注点是不一样的,虽然都有偏好的认知概念,但总体来说,每年的变化与跨度不大,认知结构基本稳定。

3.4 术语的使用趋势

科学共同体中的成员在自发形成并共同拥有一定的认知结构后,会对研究方向和前沿领域形成正确的解释和预期,并使用不同的概念术语进行表达,从而呈现出研究主题的总体演

化趋势。本研究定量分析了这 100 个词的使用变化,总结出术语的五种使用趋势。利用这些术语的使用趋势,可以探究图书情报学领域认知结构的演化,亦可以揭示研究主题的个体呈现趋势。

首先,笔者分别通过 1998—2011 年数据的发展变化热点图、线图,按照个人主观理解,将所有词的使用分为五类,包括上升趋势、下降趋势、先下降后上升趋势、先上升后下降趋势、平稳发展。但笔者将结果进行比对时发现,由于判断标准的不明确,导致对词的分类差异较大,且由于统计数据较少,波动较大,很难看出

真正的趋势。因此,研究进行了改进,分别求得 2011 年数据对于 1998 年数据的增长率(以下简称为总增长率),1998—2011 年最大值对于 1998 年数据的增长率(以下简称为最大值头增长率),1998—2011 年最大值对于 2011 年数据的增长率(以下简称最大值尾增长率),1998—2011 年最小值对于 1998 年数据的增长率(以下简称最小值头增长率),1998—2011 年最小值对于 2011 年数据的增长率(以下简称最小值尾增长率),通过对五个值的限定来定义使用趋势,具体标准如表 3 所示。

表 3 术语使用趋势的确定标准

趋势	总增长率	最大值 头增长率	最大值 尾增长率	最小值 头增长率	最小值 尾增长率
上升	>3	—	<2	>-0.5	—
下降	<0	<1	—	—	>-0.5
先升后降	>0	—	>2	>-0.5	—
	<0	>1	—	—	>-0.5
先降后升	>0	—	<2	<-0.5	—
	<0	<1	—	—	<-0.5
FALSE	其它,进行人工调整				

需要说明的是,阈值是根据经验确定的,参数的设置保证能够最大程度体现词的真正使用趋势,尽量将正常波动排除在外。将按照上述方法获得的词的类型与人工判断的词的类型进行比较,两者结果较为吻合。

术语使用趋势的计算结果如表 4 所示,表中列举了四个词聚类中除平稳发展之外的四种使用趋势,每种使用趋势中的词按变化程度从大到小依次排列。在“数字信息的组织与管理”中,一方面,因特网、Internet、计算机、多媒体、检索等词呈现下降趋势,而元数据、智能、XML、全文、网上、搜索、参考等词在上升一段时间后急速下降,这说明与信息技术相关的基础研究逐渐减少;另一方面,个性、本体、语义、文本、Web、内容、用户、虚拟、数字等词呈现上升趋势,这说

明数字信息的组织、管理与利用的研究逐渐增多,强调在新环境下应用新技术开展面向用户的信息服务。在“图书馆学”中,有变化的词较少,但值得注意的是,阅读、人文、特色、编目等词在下降一段时间后急速上升,凸显了图书馆学的人文导向;“服务”一直呈现上升趋势,“藏书”一直呈现下降趋势,这说明当前研究更加重视服务,对传统的藏书关注降低。在“信息计量与评价”中,开放、影响、国家、研究、发展、学科等词呈现上升趋势,比较、美国、中国等词在下降一段时间后急速上升,这说明信息计量与评价一直侧重于应用,重点关注学科的影响、发展与比较,特别是对中美两国的比较研究重新成为热点。在“组织知识管理”中,整体来看,呈现上升趋势的词有很多,这表明研究者们在近些

年开始转向这一分支,研究热点轮转很快。为更清晰地表明术语使用趋势的演化特点,本研究在四类趋势中选取一个特点明显的词,生成趋势图,如图 11 所示。

表 4 术语的使用趋势

使用趋势	数字信息的组织与管理	图书馆学	信息计量与评价	组织知识管理
上升	个性 本体 语义 文本 Web 内容 用户 虚拟 数字	服务	开放 影响 国家 研究 发展 学科	政府 组织 创新 网站 业务 知识 技术 隐性 评价 竞争 社会
下降	因特网 Internet 计算机 多媒体 检索	藏书		市场
先升后降	元数据 智能 XML 全文 网上 搜索 参考	现代 书目 目录	国际	
先降后升	咨询	阅读 人文 特色 编目	比较 美国 中国	指标 模糊

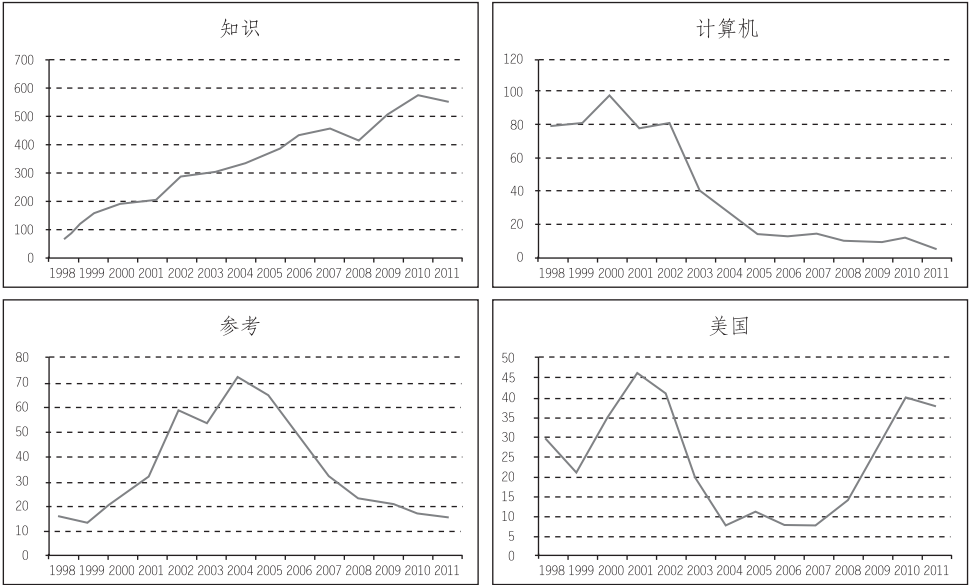


图 11 术语的四种使用趋势

利用这一方法,本研究又观察了同义词“网络”和“Web”、“因特网”和“internet”的使用趋势,如图 12 所示。显然,“因特网”和“internet”保持了高度一致,这可以理解为由于英汉的直译所致;而“网络”和“Web”在 2006 年之后出现

了反差,回顾“前方一致”原则,Web 中被合并的一部分关键词是 Web 2.0,其占比约为 25% (241/984),而这正是从 2006 年开始成为研究热点的,Web 2.0 无法很好地进行对应翻译,于是也自然成为一个常用的概念表达。

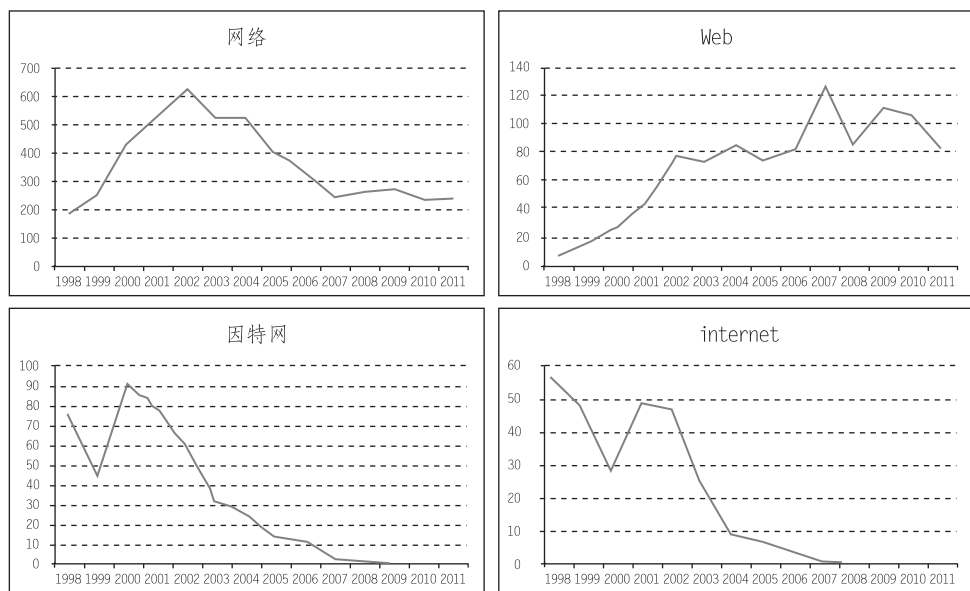


图 12 同义词的使用趋势

4 讨论与总结

本文回顾图书情报学的学科性质和学科转向,并总结描述学科认知结构的方法。一门学科中的概念术语体现了科学共同体的认知状态,并可由文章关键词进行表征。但文章关键词大多以词组(短语)形式出现,并非最小语言单位(词),因此用来分析学科认知结构并不合适。针对这一问题,利用“前方一致”原则对文章关键词进行合并处理,找出能代表图书情报学领域认知结构的 100 个频繁出现的词。在这 100 个词的基础上,分析它们之间的聚类分布,在 14 种样本期刊中的分布,以及在时间上的演化轨迹和使用趋势。

在对 100 个词进行聚类分析的过程中,本研究确定了中国图书情报学领域的四个主要分支,分别为数字信息的组织与管理、图书馆学、信息计量与评价、组织知识管理。需要强调的是,这一聚类结果在识别各个分支时,并不是基于论文、期刊或作者的数量“大小”,而是基于文章关键词中常用概念术语的清晰度。将这 100

个词与期刊关联起来分析,可以观察出不同期刊的认知偏好,也能得出期刊间的相似度。针对最常使用的概念“信息”和“图书馆”,可以明显观察到期刊的分类,图书馆类期刊关注图书馆,而情报类期刊关注信息。尽管有部分两栖刊物在期刊定位上不刻意强调认知偏好,但依然可以看出其关注点的分布情况。研究还具体给出了《现代图书情报技术》和《情报学报》这两种期刊的特殊偏好,后续研究可以进一步加以分析。

这 100 个词在时间上的演化轨迹和使用趋势可以清晰地表明概念术语的变化情况。不同时间段有不同的关注点,不同的概念术语有着自身的生命周期。当然,对这部分内容的解释需要和中国图书情报学的起源和特征相结合,同时还需要考虑图书情报学的转向问题。图 10 和表 4 显示了中国图书情报学作为一个学科整体在 14 年时间跨度上的整体概况,显然,中国图书情报学领域似乎越来越将“信息”作为它的研究核心,热点更偏重技术应用与管理创新。

本文主要针对图书情报学的认知结构,在方法论上补充了共词分析方法,相对于以往研

究,其特色是:在一个大样本基础上,通过“前方一致”原则将关键词聚合为关键词族,利用词的共现和聚类来揭示中国图书情报学领域在14年间的认知结构及其演化。在更广泛的层面上,本文所采用的“前方一致”原则,实际上是对关

键词进行压缩和聚合,将科学共同体共享的核心概念术语精炼出来,此时词频分析和共词分析就可以作为一个强有力的分析工具来分析学科认知结构,因此这一方法也能应用于对其它学科领域的分析。

参考文献

- [1] 马费成. 情报学发展的历史回顾及前沿课题[J]. 图书情报知识, 2013(2): 4-12. (Ma Feicheng. Historical review of the development of information science with proposing frontier topics[J]. Document, Information & Knowledge, 2013(2): 4-12.)
- [2] Webber S. Information science in 2003: a critique[J]. Journal of Information Science, 2003, 29(4): 311-330.
- [3] Robinson L, Bawden D. 'So wide and varied': the origins and character of British information science[J]. Journal of Information Science, 2013, 39(6): 754-763.
- [4] Saracevic T. Information science[J]. Journal of the American Society for Information Science, 1999, 50(12): 1051-1063.
- [5] Cronin B. The sociological turn in information science[J]. Journal of Information Science, 2008, 34(4): 465-475.
- [6] Nolin J, Åström F. Turning weakness into strength: strategies for future LIS[J]. Journal of Documentation, 2010, 66(1): 7-27.
- [7] Miksa F L. Machlup's categories of knowledge as a framework for viewing library and information science history[J]. Journal of library history, 1985, 20(2): 157-172.
- [8] Robinson L. Information science: communication chain and domain analysis[J]. Journal of Documentation, 2009, 65(4): 578-591.
- [9] 马费成, 望俊成. 信息生命周期研究述评(I)——价值视角[J]. 情报学报, 2010, 29(5): 939-947. (Ma Feicheng, Wang Juncheng. A literature review of studies on information lifecycle(I): the perspective of value[J]. Journal of the China Society for Scientific and Technical Information, 2010, 29(5): 939-947.)
- [10] 马费成, 望俊成. 信息生命周期研究述评(II)——管理视角[J]. 情报学报, 2010, 29(6): 1080-1086. (Ma Feicheng, Wang Juncheng. The review of studies on information lifecycle(II): the perspective of management[J]. Journal of the China Society for Scientific and Technical Information, 2010, 29(6): 1080-1086.)
- [11] Bates M J. The invisible substrate of information science[J]. Journal of the American Society for Information Science, 1999, 50(12): 1043-1050.
- [12] White H D, Griffith B C. Author cocitation: a literature measure of intellectual structure[J]. Journal of the American Society for information Science, 1981, 32(3): 163-171.
- [13] White H D, McCain K W. Visualizing a discipline: an author co-citation analysis of information science, 1972-1995[J]. Journal of the American Society for Information Science, 1998, 49(4): 327-355.
- [14] 马费成, 宋恩梅. 我国情报学研究分析: 以 ACA 为方法[J]. 情报学报, 2006, 25(3): 259-268. (Ma Feicheng, Song Enmei. An author co-citation analysis of information science in China[J]. Journal of the China Society for Scientific and Technical Information, 2006, 25(3): 259-268.)
- [15] Åström F. The visibility of information science and library science research in bibliometric mapping of the LIS field[J]. The Library Quarterly, 2010, 80(2): 143-159.
- [16] Ding Y, Chowdhury G G, Foo S. Bibliometric cartography of information retrieval research by using co-word analysis[J]. Information processing & management, 2001, 37(6): 817-842.
- [17] Janssens F, Leta J, Glänzel W, et al. Towards mapping library and information science[J]. Information Processing & Management, 2006, 42(6): 1614-1642.

- [18] Van Den Besselaar P, Heimeriks G. Mapping research topics using word-reference co-occurrences: a method and an exploratory case study[J]. *Scientometrics*, 2006, 68(3): 377-393.
- [19] Sugimoto C R, Li D, Russell T G, et al. The shifting sands of disciplinary development: analyzing north American library and information science dissertations using latent dirichlet allocation[J]. *Journal of the American Society for Information Science and Technology*, 2011, 62(1): 185-204.
- [20] 马瑞敏, 倪超群. 基于作者同被引分析的我国图书情报学知识结构及其演变研究[J]. *中国图书馆学报*, 2011, 37(6): 17-26. (Ma Ruimin, Ni Chaoqun. A study on intellectual structure and the evolution of library and information science in China based on author co-citation analysis[J]. *Journal of Library Science in China*, 2011, 37(6): 17-26.)
- [21] Milojević S, Sugimoto C R, Yan E, et al. The cognitive structure of library and information science: analysis of article title words[J]. *Journal of the American Society for Information Science and Technology*, 2011, 62(10): 1933-1953.
- [22] 滕立, 沈君, 高继平. 共词知识网络中的认知结构: 理论、方法与实证[J]. *情报学报*, 2013, 32(9): 976-989. (Teng Li, Shen Jun, Gao Jiping. Cognitive structure in co-word knowledge network: theory, method and a case study[J]. *Journal of the China Society for Scientific and Technical Information*, 2013, 32(9): 976-989.)
- [23] 任红娟. 一种内容和引用特征融合的知识结构划分方法研究[J]. *中国图书馆学报*, 2013, 39(5): 76-82. (Ren Hongjuan. A division method of intellectual structure of combining content and citation of documents [J]. *Journal of Library Science in China*, 2013, 39(5): 76-82.)
- [24] 马费成, 郝金星. 概念地图在知识表示和知识评价中的应用(I)——概念地图的基本内涵[J]. *中国图书馆学报*, 2006, 32(3): 5-9, 49. (Ma Feicheng, Hao Jinxing. Applications of concept maps in knowledge representation and knowledge evaluation(I)[J]. *Journal of Library Science in China*, 2006, 32(3): 5-9, 49.)
- [25] Brookes B C. The foundations of information science: part I. philosophical aspects[J]. *Journal of Information Science*, 1980, 2(3-4): 125-133.
- [26] 文庭孝, 罗贤春, 刘晓英, 等. 知识单元研究述评[J]. *中国图书馆学报*, 2011, 37(5): 75-86. (Wen Tingxiao. Luo Xianchun, Liu Xiaoying, et al. Knowledge unit research: review and comment[J]. *Journal of Library Science in China*, 2011, 37(5): 75-86.)
- [27] Albrechtsen H, Hjørland B. Understandings of language and cognition: implications for classification research [J]. *Advances in Classification Research Online*, 1994, 5(1): 1-16.
- [28] 王晓光. 科学知识网络的形成与演化(I): 共词网络方法的提出[J]. *情报学报*, 2009, 28(4): 599-605. (Wang Xiaoguang. Formation and evolution of science knowledge network(I): a new research method based on co-word network[J]. *Journal of the China Society for Scientific and Technical Information*, 2009, 28(4): 599-605.)
- [29] 梁战平. 情报学若干问题辨析[J]. *情报理论与实践*, 2003, 26(3): 193-198. (Liang Zhanping. An inquiry into some issues concerning information science[J]. *Information Studies: Theory & Application*, 2003, 26(3): 193-198.)
- [30] Priem J, Taraborelli D, Groth P, et al. Altmetrics: a manifesto [EB/OL]. [2010-10-26]. <http://altmetrics.org/manifesto>.
- [31] Piwowar H. Altmetrics: value all research products[J]. *Nature*, 2013, 493(7431): 159-159.
- [32] 邱均平, 余厚强. 替代计量学的提出过程与研究进展[J]. *图书情报工作*, 2013, 57(19): 5-12. (Qiu Junping, Yu Houqiang. The putting forward process and research progress of altmetrics[J]. *Library and Information Service*, 2013, 57(19): 5-12.)

张斌 武汉大学信息资源研究中心 2012 级管理科学与工程博士研究生。

通讯地址: 武汉市珞珈山。邮编: 430072。

贾茜 武汉大学信息资源研究中心 2012 级情报学硕士研究生。通讯地址同上。

(收稿日期: 2013-12-18)