

信息素养领域演进路径、研究热点与前沿的可视化分析¹

张士靖 杜建 周志超

华中科技大学同济医学院医药信息管理系 武汉 430030

[摘要] 梳理 30 年来信息素养领域研究的演进路径, 揭示研究热点与研究前沿。将 1983-2009 年 SCI、SCI、CPCI-S、CPCI-SSH 数据库中收录的主题为“信息素养”的全部论文数据作为研究对象, 应用 CiteSpace 知识可视化软件绘制文献共被引网络图谱, 析出信息素养演进过程中的关键节点文献; 并应用其关键词聚类 and 膨胀词探测功能分析研究热点与研究前沿。其中, 6 篇关键节点文献很好地展示了信息素养领域研究的演进路径, 2 个明显的聚类以及 5 个突现词表征了研究热点与研究前沿。因特网信息素养和健康素养是目前信息素养的两个主流研究领域, 定性研究与定量研究相结合是该领域的学科研究范式。老年人、公共卫生、循证决策和乳腺癌等对象或领域是信息素养的研究前沿与发展趋势。

[关键词] 信息素养 演进路径 研究热点 研究前沿 CiteSpace 知识图谱

[分类号] G350

1 问题与背景

信息素养 (Information Literacy, IL) 是近年来图书馆学、情报学、教育学等学科领域的一个研究热点, 国外有关信息素养的研究始于 20 世纪 70 年代并延续至今。自 1974 年美国信息产业协会主席 Zurkowski P 首次提出 “information Literacy” 这个概念以来^[1], 至今已有 35 年的历史。30 多年来, 信息素养的研究走过了怎样的发展轨迹? 期间有些什么样的关键事件促进了它的演进? 当前的研究热点和研究前沿是什么? 为此, 笔者借助知识可视化技术探讨信息素养领域研究的演进路径、研究热点与研究前沿, 以期国内学者在该领域的研究提供有益的启示。

2 样本与方法

本文采用以引文分析和信息可视化技术为基础的科学知识图谱的方法, 所用的工具选取美国德雷塞尔大学 (Drexel University) 信息科学与技术学院陈超美博士开发的 CiteSpace 软件 (目前已升级到 CiteSpace II), 形象地展示出信息素养领域研究的演进路径、研究热点与研究前沿。

2.1 样本选择

以 Web of Science 数据库中的四个子库 (SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH) 为数据来源, 检索策略为 “主题=(information literacy) AND 文献类型=(Article)”, 时间确定为 1975-2009 年, 共检索出 1 810 条论文记录 (检索时间为 2009 年 10 月 28 日)。考察该文献数据集的出版时间分布, 发现第 1 篇论文的发表时间为 1983 年, 故选取 1983-2009 年的文献数据做信息素养研究领域的演进路径分析。

2.2 可视化方法

“科学计量学之父”普赖斯 (Price) 在 1965 年注意到一个非常有趣的现象, 即在科学引文网络中, 频繁被引用最多是新近发表的文献。因此他提出了 “研究前沿” (Research Front) 的概念并以此来描述学科研究领域的过渡本质^[2]。1994 年, 瑞典科学计量学家皮尔逊 (Persson) 对研究前沿和知识基础作了界定: 从文献计量学来看, 引文 (施引文献) 形成了研究前沿, 被引文献组成了知识基础。知识基础由共引聚类来表示^[3]。2005 年, 陈超美 (Chaomei Chen) 对研究前沿和知识基础提出了新的观点^[4]: 研究前沿被定义为一组突现的动态概念和潜在的研究问题, 知识基础 (Intellectual Base) 是研究前沿在科学文献中的引文和共引轨迹, 即承载动态概念的引文及其参考文献形成的聚类。研究前沿总是沿时间推进的, 一个研究领域的知识基础将会同其研究前沿一起随着时间的变化而演进。在这种演进过程中, 必然会产生导致研究前沿演进的关键节点文献, 这些节点可能成为网络中由一个时间段向另一个时间段过渡的关键点。关键节点文献所起的过渡作用的大小用中心度 (也称中介中心性) 来度量。CiteSpace II 最突出的特点就是关键节点的计算测量, 并在可视化网络中用紫色的圆圈突

¹本文系教育部人文社会科学基金 “基于构建主义理论和 E-Learning 构建专业信息素质教育平台” (项目编号: 07JA870011) 研究成果之一。

出显示,明显的视图效果使用户很容易地看到那些具有较高中介中心性的节点。关键节点文献的意义有两点:①从知识领域的角度看,关键节点文献一般是提出重要的新理论或是具有重大理论创新的经典文献,也是最有可能形成科学研究前沿热点的文献;②对关键节点的分析能够帮助我们梳理信息素养领域研究的演进路径。

此外,CiteSpace II还有关键词聚类 and 膨胀词探测功能,以便可以确认信息素养领域的研究热点与研究前沿。它提供的“涌现”检测(burst detection)主要通过考察词频的时间分布,将那些频次变化率高、增长速度快的“涌现”词(burst terms)从大量的常用词中检测出来,用词频的变动趋势而不仅仅是词频的高低,来分析学科的前沿领域和发展趋势。

2.3 CiteSpace II 参数设置

将1983-2009年发表的信息素养主题的1810篇论文的全部题录数据输入CiteSpace软件中,这些题录数据主要包括:标题、关键词、摘要、参考文献等字段。时间分区设为每3年一个分区,共分为9个时间段。分析节点选择“cited reference”,调节阈值设定为(2, 1, 15), (3, 2, 18), (4, 3, 20)。运行CiteSpace软件,得到信息素养研究共被引网络分布图,析出关键节点文献。主题词分为名词短语(noun phrase)和词频突出增长的“涌现”词(burst term)两种,将其来源(term sources)选为标题(title)、摘要(abstract)、作者关键词(descriptors)、附加关键词(identifiers)。名词短语表达信息素养领域的研究热点,“涌现”词表达信息素养领域的研究前沿和发展趋势。

3 结果与分析

3.1 信息素养领域研究文献量的时间序列分布

图1所示信息素养的第1篇论文发表在1983年,从1983-1995年,为信息素养的起步阶段,发文量十分有限;1995-2000年为逐步上升阶段;从2000年开始至今则呈现了迅猛增长的态势。用Logistic曲线拟合文献量的时间序列分布,发现信息素养领域的研究还处于快速上升期,尚未出现成熟前的“拐点”。

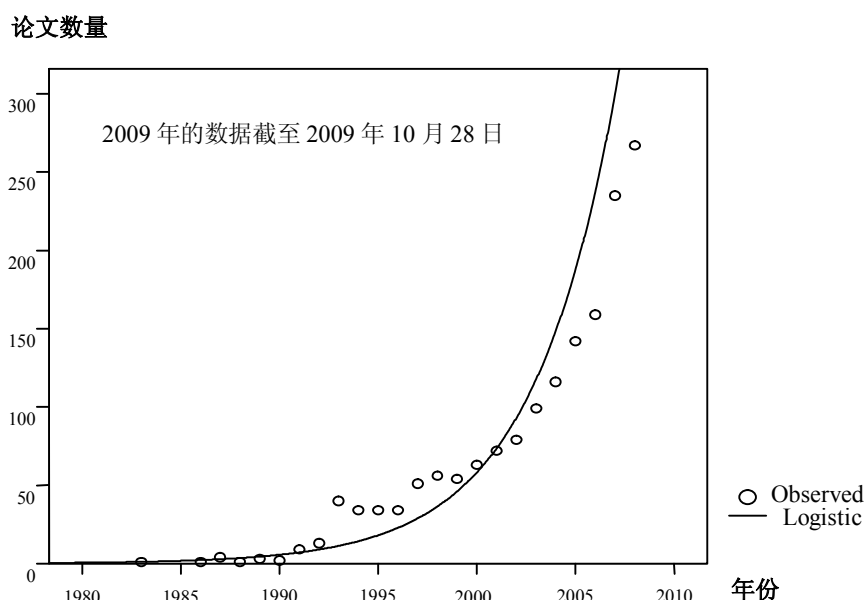


图1 SCI、SSCI、CPCI-S、CPCI-SSH收录主题为“Information Literacy”文献量的时间序列分布

3.2 信息素养领域研究的演进路径分析

CiteSpace II可视化软件对共引网络图谱的显示有两种不同的视图方式,即聚类视图(Cluster View)和时区视图(Time-zone View)。时区视图的显示方式突出共引网络节点随时间变化的结构关系。利用CiteSpace II软件运行得到信息素养领域文献共引网络节点的时区视图(见图2),中心性大于等于0.16的6篇关键节点文献的详细信息见表1。

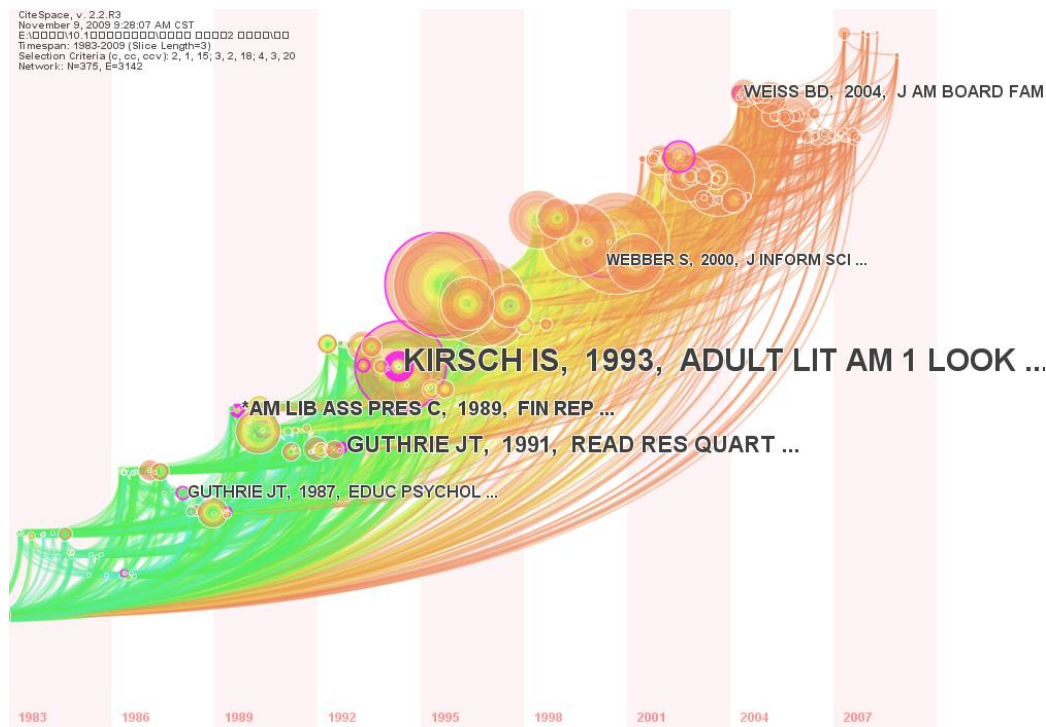


图 2 信息素养领域共引网络节点的时区

表 1 信息素养研究的 6 篇关键节点文献（按中心性排序）

关键节点文献	期刊/学科类别	被引频次（GS/本文数据集）	中心性
Kirsch IS, 1993. Adult Literacy in America: A First Look at the Findings of the National Adult Literacy Survey (美国成人素养：国家成人素养调查初步分析)	Report/调查报告	599/18	0.61
Guthrie JT, 1991. Relations of Document Literacy And Prose Literacy to Occupational&Societal Characteristics of Young Black-And-White Adults (年轻的白人或黑人的图表素养、文字素养与他们的职业和社会特征的关系)	Reading Research Quarterly (阅读研究季刊) /教育与教育研究	36/7	0.34
Am Lib Ass, 1989. Presidential Committee on Information Literacy, Final Report (信息素养总统委员会总报告)	Report/图书馆学	11135/11	0.25
Weiss BD, 2004. Relationship between health care costs and very low literacy skills in a medically needy and indigent Medicaid population (在医疗贫困与医疗救助贫乏的人群中, 低素养技能与医疗费用的关系)	Journal of the American Board of Family Practice (家庭医学美国版) /卫生与健康	42/10	0.21
Guthrie JT, 1987. Literacy as multidimensional: Locating information and reading comprehension (素养是多维的：定位信息与阅读理解)	Educational Psychologist (教育心理学家) /教育心理学	59/15	0.19
Webber S, 2000. Conceptions of information literacy: new perspectives and implications (信息素养的概念：新的观点与意义)	Journal of Information Science (情报科学杂志) /情报学与图书馆学	118/29	0.16

注：被引频次笔者分别检索了节点文献在 Google Scholar 和本文数据集中的被引频次，检索日期为 2009.11.14

综合图 2 和表 1，按照时间顺序分析这 6 篇关键节点文献，即可梳理出信息素养领域研究的演进路径。

第一篇文献是 1987 年，美国马里兰大学帕克分校从事语言学与阅读动机研究的教育心理学家 Guthrie J T 教授发表在《教育心理学家》上的“素养是多维的：定位信息与阅读理解”一文^[5]。该文将定位信息和阅读理解视为两种素养，并强调

后者包含着分析信息和理解信息两方面。这篇文章实际上指出了信息素养的基础素养,如果把信息素养作为一个学科的话,该文提出的观点可作为信息素养的“前学科时期”。

第二篇文献是1989年,美国图书馆学会(ALA)发表的著名的《信息素养总统委员会总报告》,提出了被普遍认同的信息素养定义:“作为具有信息素养能力的人,必须能够充分地认识到何时需要信息,并有能力去有效地发现、检索、评价和利用所需要的信息”^[6]。这一概念包含了信息意识和信息能力两个维度,是对信息素养内涵更准确的表述,论述了信息素养教育对个人、企业、国家的重要性,提出要进行信息素养教育。它标志人们对信息素养的本质把握到了一个新高度,信息素养概念发展到成熟的关键点,是美国信息素养运动正式启动的标志,直接推动美国国家信息素养论坛在当年成立。

第三篇是1991年Guthrie J T教授发表的另一篇著作,该文提出了两种素养,即Document Literacy(图表素养)和Prose Literacy(文字素养)^[7]。前者是指理解和利用从表格、图表、图片、日程表(schedules)中获取的信息的能力;后者则指理解和利用从报纸、杂志、小说、小册子、手册、传单(flyers)中获得信息的能力。该文还揭示了这两种素养与青年的职业地位和社会参与度的关联。结果发现在20世纪90年代的美国,黑人相对较低的职业地位要归因于他们相对低的信息素养水平。强调了文献信息素养的重要性,它可以决定年轻成年人的职业与社会特征。

第四篇文献发表于1993年,中心度最高(字体越大表明中心度越高)。该文对1993年美国全国成人素养调查研究结果分析发现^[8],有4千万美国成年人得分在五个等级中的最低水平(1级),还有5千万人的水平在2级,引起了很大轰动。这一结果与美国强调完成12年学校教育、国民普遍达到高中学历的政策所预期的效果大相径庭,突显了用受教育年限来代表实际文化程度(Literacy)的局限性。强调学校教育是不够的,必须要有终身学习的能力。首次提出了信息素养与终身学习的关系:信息素养是终身学习的必备条件与核心。该文的中心作用之所以如此显著,还因为该调查研究发现,有一半的成年人的健康素养技能很低,难以应付医疗环境,并由此引发了人们对信息素养在卫生和健康领域的关注。

第五篇是2000年英国谢菲尔德大学(Sheffield University)信息研究中心Webber S教授发表在《美国情报科学杂志》上的“信息素养的概念:新的观点与意义”一文^[9],该文在本数据集中的被引次数最高。从其题目和被引频次此可以推断该文对信息素养领域具有理论创新价值。该文的突出贡献是作者探究信息素养内涵时,在以往ALA强调的信息意识、信息能力的基础上特别增加了信息道德维度,强调了意识到在社会中合法使用信息的重要性。该文还首次提出了要对大学生的信息素养教育进行课程设计。重点探讨了两个问题:信息素养的合适的教育方法以及将信息素养本身作为一个学科。

最后一篇是2004年,Weiss B D发表于《美国家庭医学杂志》上的“在医疗贫困与医疗救助贫乏的人群中,低素养技能与医疗费用的关系”一文^[10]。该文的重要性体现在两方面:一是推翻了作者自己10年前,由于抽样欠科学而得出的“素养技能与医疗费用之间没有关联”的结论^[11];二是提示了健康素养的重要性。并由此引发了后人对健康素养的高度关注。美国国立医学图书馆认为,健康素养(Health Literacy, HL)是指个人能够获取、理解 and 处理基本的健康信息与服务,并以此做出合理的健康决策的能力^[12]。Nutbeam D在探究健康素养的内涵体系时指出健康素养不仅包括基本的阅读和书写技能,还包括高级认知和素养,即能够在日常生活中通过各种传播方式,积极寻求获得健康信息并采用批判思维分析健康信息从而应用新知识改变健康状况^[13]。Zarcadoolas C等认为健康素养是伴随公众一生的寻找、理解、评估、使用健康信息来进行健康决策、降低健康风险、提高生活质量的多维度技能^[14]。在信息素养和健康素养的基础上,健康信息素养(Health information Literacy, HIL)成为最近两年的一个研究热点^[15]。因此,该关键节点论文是信息素养在卫生与健康领域应用的一个重要标志。

通过上述分析,笔者把信息素养领域研究的演进路径表示,如图3所示:

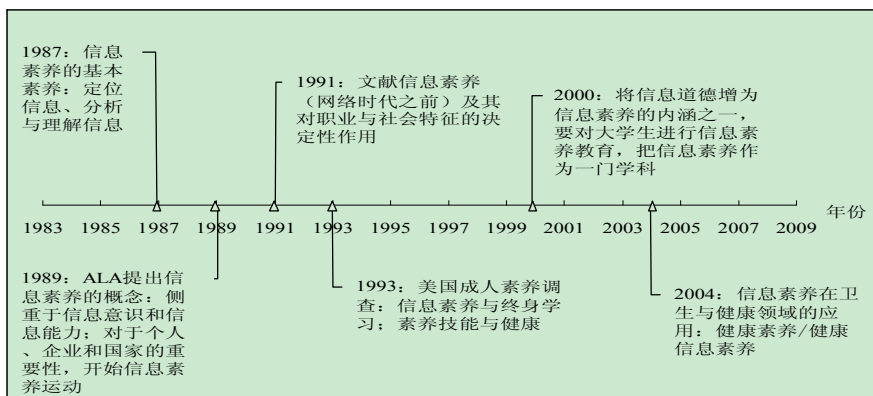


图3 6篇关键节点文献的核心观点展示的信息素养领域研究演进路径

3.3 信息素养领域研究热点分析

利用 CiteSpace 的术语自动聚类功能, 确定信息素养领域研究的主要研究领域和研究热点。在 CiteSpace2. 2 R1 及以后的版本中, 提供了 9 种聚类自动标识的方法。标识词的来源有 3 种, 分别是引文的题名、摘要、描述词 (descriptors), 然后每一种来源又可以对应三种算法, 分别是 tf*idf (词频-逆文档频率), Log Likelyhood Ratio (对数似然比)、mutual Information (交互信息)。所以共计 3*3=9 种聚类自动标识算法。在上述展示的关键节点文献知识图谱的基础上, 点击 “cluster”, 聚类标识来源选择 “label clusters with title words”, 聚类算法选择 “tf*idf”, 即利用词频-逆文档频率算法从引文的题名中提取聚类标识术语。结果形成了明显的两个聚类, 即#0 Internet Information Literacy (因特网信息素养) 和#1 Health Literacy (健康素养), 如图 4 所示:

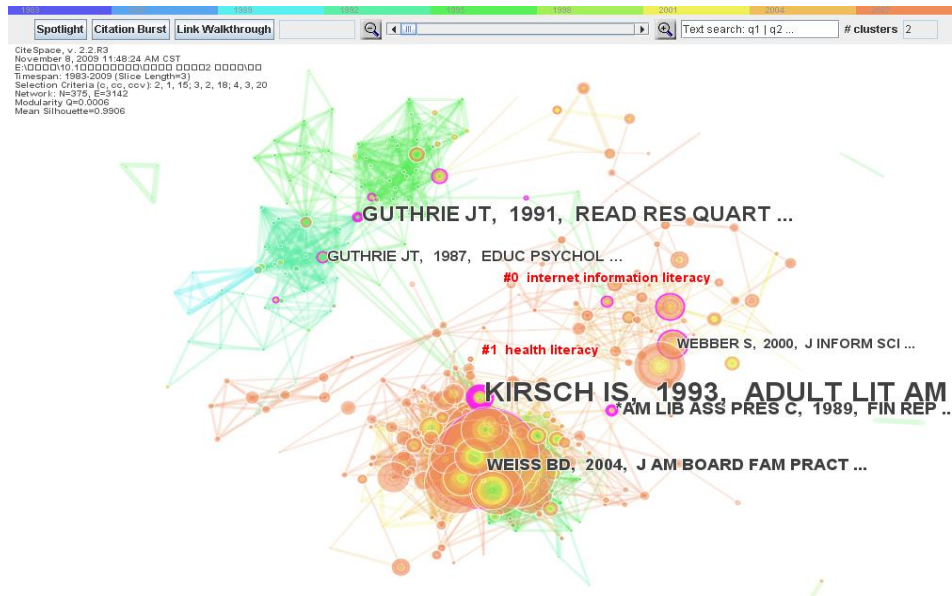


图 4 信息素养领域研究热点的自动聚类标识

两个聚类中的主要术语及其 tf*idf 加权值如表 2 所示:

表 2 两个聚类中的主要术语及其 tf*idf 加权值

	Top Terms (英文)	Top Terms (中文)	tf*idf 加权值
第一类: 网络信息素养	Internet Information Literacy	因特网信息素养	29.02
	Same difference	一致性差异	1.39
	Basic consideration	基本构思	1.39
第二类: 健康素养	Health Literacy	健康素养	84.56
	Health Information	健康信息	18.02
	Patient Education	患者教育	16.64

为了进一步验证这两个自动聚类标识, 笔者通过析出高频名词术语来进一步确定信息素养领域的研究热点。词项选择名词短语 (noun phrases), 选择 Pathfinder 算法以只显示网络中的关键结构并提高运行速度, 运行 CiteSpace 软件得到出现频次大于等于 44 次的 12 个名词术语, 如表 3 所示:

表 3 出现频次≥44 次的 12 个名词术语

序号	术语 (英文)	术语 (中文)	频次	序号	术语 (英文)	术语 (中文)	频次
1	Information Literacy	信息素养	251	7	Literacy skills	素养技能	49
2	Health Literacy	健康素养	136	8	Reading level	阅读水平	48
3	Health Information	健康信息	82	9	Significant difference	显著性差异	45
4	Information technology	信息技术	72	10	Literacy level	素养水平	45
5	Computer Literacy	计算机素养	65	11	Focus group	焦点小组	44
6	Health care	卫生保健	50	12	Communication technology	通信技术	44

表3可见,12个高频术语中,健康素养、健康信息和卫生保健3个术语可以表征“健康素养”这一聚类;而信息技术、计算机素养和通信技术可以证实“因特网信息素养”这个聚类。另外,笔者还发现,显著性差异和焦点小组这2个术语的出现频次也很高,前者说明信息素养领域的研究多采用统计分析方法,后者说明小组讨论(会议)法也是信息素养领域研究的最常用方法。这2个术语表明,信息素养领域研究多采用定量研究与定性研究相结合的方法。

3.3 信息素养领域研究前沿分析

陈超美博士认为,使用突现主题术语(surged topical terms)要比使用出现频次最高的主题词(title words)更适合探测学科发展的新兴趋势和突然变化^[16]。利用CiteSpace提供的膨胀词探测技术和算法,通过考察词频的时间分布,将其中心频变化率高的词(突现词,burst terms)从大量的主题词中探测出来,揭示信息素养领域的研究前沿。将词项选择为“burst terms”,运行CiteSpace软件,得到5个变化频次最高的突变词,如表4所示:

表4 5个变化频次最高的突现词

序号	突现词(英文)	突现词(中文)	变化频次
1	Older adults	老年人	65
2	Public Health	公共卫生	51
3	Evidence based	循证	45
4	Decision making	决策	45
5	Breast cancer	乳腺癌	33

由表4可见,“老年人”这一术语的变化频次最高,说明随着国外逐渐进入老龄化社会,老年人的信息素养问题是一个研究前沿,信息素养是他们参与民主社会、维护与增进健康的必要因素。“公共卫生”作为一个研究前沿提示我们,公众的健康素养/健康信息素养是一个非常值得关注的研究领域,信息素养的促进有助于提高预防保健服务的利用率、减少住院率、降低医疗费用并增强个体的疾病自我管理能力和患者和公众的健康信息素养的促进是对公共卫生领域的重要贡献。“循证”与“决策”两个术语的变化频次相同,笔者将其视为一个研究前沿,即“循证决策”,提示在未来各行各业的决策中,“循证”的思想表明了信息素养是一个必备因素。最后“乳腺癌”这一名词术语表明妇女这一群体的信息素养也是一个研究重点,乳腺癌作为女性最常见的恶性肿瘤之一,妇女对于乳腺癌及早期筛查的认知情况和相关信息的获取、理解和利用能力关系到自身的健康状况。

4 结 论

CiteSpace可视化应用软件在分析学科演化、热点与前沿领域等方面所表现出来的较强的探测和检验等辅助功能在以上信息素养领域的研究中得到了完美的体现,通过对Web of Science数据库中收录的所有信息素养主题的研究论文的可视化分析,可以得到以下结论:

● 信息素养涉及教育学、图书馆学、情报学等多学科的交叉领域,新世纪里在卫生与健康领域得到了广泛应用。

● 信息素养领域研究经历了这样一个演进路径:20世纪80年代中后期提出了信息素养的基本素养(定位信息、分析与理解信息),80年代末正式界定了信息素养的概念(侧重于信息意识与信息能力两个方面)并指出了信息素养对于个人、企业和国家的重要意义。90年代初强调了传统的文献信息素养对职业地位与社会参与度的决定作用;接着美国成人素养调查报告的发表指出了终身学习的重要性,使得信息素养的意义更为显著。世纪之交,对信息素养新的观点与意义的探讨将信息道德纳入信息素养的内涵,并指出应对大学生这一群体进行信息素养教育,最为重要的是提出将信息素养作为一门学科。在新世纪里,信息素养与健康成为一对须臾不可分离的概念,健康信息素养使得信息素养这一抽象概念在具体领域得到了实际应用。

● 信息素养领域的研究热点有两个,一是因特网信息素养;二是健康素养。研究方法采用定量研究与定性研究相结合的方法,逐渐形成了学科研究范式。

● 信息素养领域的研究前沿有4个:①老年人的信息素养问题;②信息素养在公共卫生领域的作为;③循证与决策;④妇女群体的信息素养,对乳腺癌的预防保健。

参考文献:

- [1] 王雯. 国外信息素养研究最新动态. 高校图书馆工作, 2009, 29 (4): 15-17.
- [2] Price, D. Networks of scientific papers. Science, 1965 (149): 510-515.
- [3] Persson, O. The intellectual base and research fronts of jasis 1986-1990. Journal of the American Society for Information Science, 1994, 45 (1): 31-38.
- [4] 刘则渊, 陈悦, 侯海燕. 科学知识图谱: 方法与应用. 北京: 人民出版社, 2008.
- [5] Guthrie, J. T., & Mosenthal, P. (1987). Literacy as multidimensional: Locating information and reading comprehension. Educational Psychologist, 22 (3 & 4), 279-297.
- [6] American Library Association (ALA). Presidential Committee on Information Literacy, Final Report, 1989.
- [7] Guthrie, John T. Relations of Document Literacy And Prose Literacy To Occupational And Societal Characteristics of Young Black-And-White Adults. Reading Research Quarterly, 1991, 26 (1): 30-48.
- [8] Kirsch, I. S., Jungeblut, A., Jenkins, L., et al. Adult Literacy in America: A First Look at the National Adult Literacy Survey[R]. Washington, D.C. Office of Educational Research and Improvement, U.S. Department of Education, 1993.
- [9] Webber S, Johnston B. Conceptions of information literacy: new perspectives and implications. Journal of Information Science, 2000, 26 (6): 381-397.
- [10] Weiss BD, Palmer R Relationship between health care costs and very low literacy skills in a medically needy and indigent Medicaid population. Journal of The American Board of Family Practice, 2004, 17 (1): 44-47.
- [11] Weiss BD, Blanchard JS, McGee DL, et al. Illiteracy among Medicaid recipients and its relationship to health care costs. J Health Care Poor Underserved. 1994; 5: 99 -111.
- [12] 孔燕, 沈菲飞. 健康素养内涵探析. 医学与哲学 (人文社会科学版), 2009, 30 (3): 17-19, 53.
- [13] Nutbeam. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promotion International. 2000, 15 (3): 259-267.
- [14] Zarcadoolas C, Pleasant A, Greer DS. Understanding health literacy: an expanded model. Health Promotion International. 2005, 20 (2): 195-203.
- [15] Shipman JP, Kurtz-Rossi S, Funk CJ. The health information literacy research project. Journal of Medical Library Association, 2009, 97 (4): 293-301.
- [16] 陈超美 (著). 陈悦, 侯剑华, 梁永霞 (译). CiteSpace II: 科学文献中新趋势与新动态的识别与可视化. 情报学报, 2009, 28 (3): 401-421.

[作者简介] 张士靖, 女, 1958-、副教授, 硕士生导师, 发文数 50 余篇。

杜建, 男, 1986 年生, 硕士研究生, 发文数 10 余篇;

周志超, 男, 1987 年生, 硕士研究生, 发文数 3 篇。