

doi:10.3772/j.issn.2095-915x.2015.05.007

基于社会网络分析的信息组织热点主题研究

杜二梅

(中国科学技术信息研究所 北京 100038)

摘要: 在 CNKI 中, 检索到 2010 年到 2014 年的期刊论文中主题为信息组织的记录 989 条, 并将每条记录中的关键词进行词频统计。利用文献题录信息统计分析工具 SATI 构建共词矩阵与相似矩阵, 然后利用社会网络分析软件 Ucinet 处理共词矩阵与相似矩阵, 并进行分析。研究结果表明, 信息组织领域的研究热点主题是数字图书馆、Web 2.0、知识组织和本体。

关键词: 社会网络分析, 信息组织, 热点主题, 共词分析

中图分类号: G350

A Study of Information Organization Hot Topics Based on Social Network Analysis

DU ErMei

(The Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract: This study obtains 989 records by retrieving journal papers published from 2010 to 2014 about information organization in CNKI, and then performs a keywords word frequency statistics. And the study builds and analyses co-word matrix and similarity matrix using SATI and Ucinet. The study indicates that the research hot topics in the field of information organization are the digital library, Web2.0, knowledge organization, etc.

Keywords: Social network analysis, information organization, hot topics, co-word analysis

作者简介: 杜二梅, 女, 硕士, 研究方向: 图书馆学。通讯方式: 13718908575。E-mail: 961748353@qq.com, duem2013@istic.ac.cn

1 引言

信息组织的兴起源于计算机技术的普遍应用和网络通讯技术的迅速普及^[1]。冷伏海曾在《信息组织概论》中给出这样的定义：“信息组织就是利用一定的科学规则和方法，通过对信息外在特征和内容特征的描述与序化，实现无序信息流向有序信息的转换，从而保证用户对信息的有效获取和利用及信息的有效流通和组合”^[2]。信息组织的目的是将杂乱无章的信息按某一规则进行排序，方便用户进行查找。其中信息流向明确和流速适度可以方便用户了解信息的脉络结构和高效获取信息。目前信息组织遵循的原理主要是语言学原理和系统论原理，信息组织方法主要是分类法和主题法^[1]。

近年来，信息组织成为我国图书情报领域研究的热点之一，涉及的范围非常广泛。焦丽于2008年对我国1997-2007年间信息组织领域的论文进行统计分析，发现信息组织的研究方法以分类法、主题法和书目控制为主，信息组织技术以搜索引擎、元数据和数据挖掘为主^[1]。付凯丽于2012年对Web2.0环境下的网络信息组织进行研究，指出信息的深入挖掘和互联是信息组织未来的发展方向^[3]。王知津和郑悦萍于2013年对信息组织的语义关系的概念、应用领域和关系类型进行探讨^[4]。杨雨霖于2014年对图情档领域的信息组织标准的现状进行研究^[5]。

从以上研究中可以发现，学者主要对信息组织进行文献综述研究、相关概念研究和标准研究，缺乏对该领域研究热点的关注。鉴于此，本文在前人的研究基础上，基于社会网络分析对信息组织领域文献中的关键词进行多角度的深入研究，以期从微观层面上揭示信息组织领域的热点主题，为今后的相关研究提供参考。

2 研究方法 with 数据处理

2.1 研究方法

论文以CNKI为来源数据库，检索条件设为：主题=“信息组织”；时间范围是2010年到2014年；来源类别是“全部期刊”，最终检索到五年间的相关论文共989篇。论文主要使用文献题录信息统计分析工具SATI、社会网络分析软件Ucinet等数据处理工具，采用词频统计分析、共词分析、中心性分析等分析方法对论文的关键词进行统计分析和社会网络分析，比较直观反映出近五年国内信息组织热点主题的分布情况。

2.2 数据处理

将检索到的结果转化为社会网络分析软件Ucinet可识别的格式。该论文中是将数据转化为矩阵形式。具体步骤如下：

(1) 将检索结果以EndNote方式导出，存为文本格式。

(2) 利用SATI^[6]（文献题录信息统计分析工具）处理导出的文本数据。将文本文件导入SATI中，统计关键词的词频，最终选出频次不小于10的关键词视为高频关键词，并将其作为处理对象，如表1所示。

(3) 利用SATI处理表1中的高频关键词：构建共词矩阵、相似矩阵。通过处理得到一个41行41列的共词矩阵部分数据如表2，相似矩阵部分数据如表3。

(4) 利用Ucinet处理高频关键词的相似矩阵。使用Network的Centrality的Multiple Measure功能，将相似矩阵导入，输出结果如表4。

表1 词频不小于10的高频关键词

序号	关键词	词频	序号	关键词	词频
1	信息组织	298	22	情报学	14
2	本体	62	23	信息管理	14
3	数字图书馆	52	24	叙词表	13
4	Web2.0	42	25	公共图书馆	13
5	信息构建	40	26	知识服务	13
6	图书馆	39	27	信息技术	12
7	信息服务	34	28	搜索引擎	12
8	知识组织	33	29	Web	12
9	网络信息资源	30	30	网络信息组织	12
10	信息检索	30	31	知识管理	11
11	高校图书馆	27	32	模式	11
12	信息	23	33	电子政务	11
13	元数据	20	34	可视化	11
14	数据库	19	35	多媒体	10
15	信息资源	19	36	Folksonomy	10
16	网络环境	18	37	电子商务	10
17	网络信息	18	38	用户	10
18	标签	18	39	文本分类	10
19	网络	17	40	XML	10
20	用户参与	15	41	信息资源建设	10
21	图书馆学	15			

表2 高频关键词的共词矩阵(部分矩阵)

	信息组织	本体	数字图书馆	Web2.0	信息构建	图书馆	信息服务	知识组织
信息组织	298	10	25	15	11	13	14	14
本体	10	62	3	0	0	0	2	4
数字图书馆	25	3	52	2	4	1	8	1
Web2.0	15	0	2	42	0	4	4	0
信息构建	11	0	4	0	40	2	0	1
图书馆	13	0	1	4	2	39	7	2
信息服务	14	2	8	4	0	7	34	0
知识组织	14	4	1	0	1	2	0	33

表3 高频关键词的相似矩阵(部分矩阵)

	信息组织	本体	数字图书馆	Web2.0	信息构建	图书馆	信息服务
信息组织	1	0.0054	0.0403	0.018	0.0102	0.0145	0.0193
本体	0.0054	1	0.0028	0	0	0	0.0019
数字图书馆	0.0403	0.0028	1	0.0018	0.0077	0.0005	0.0362
Web2.0	0.018	0	0.0018	1	0	0.0098	0.0112
信息构建	0.0102	0	0.0077	0	1	0.0026	0
图书馆	0.0145	0	0.0005	0.0098	0.0026	1	0.037
信息服务	0.0193	0.0019	0.0362	0.0112	0	0.037	1

表 4 高频关键词的中心度数据

序号	关键词	绝对度数中心度	绝对接近中心度	绝对中间中心度
1	信息组织	92.5	93.023	36.502
2	数字图书馆	50	66.667	4.813
3	信息服务	40	62.5	2.878
4	Web2.0	40	62.5	3.762
5	信息检索	37.5	60.606	3.532
6	本体	37.5	61.538	3.657
7	知识组织	35	60.606	3.044
8	图书馆	35	60.606	1.924
9	信息构建	35	60.606	2.592
10	高校图书馆	32.5	59.701	2.034
11	网络信息资源	32.5	59.701	1.69
12	网络环境	27.5	57.971	1.249
13	信息资源	27.5	57.971	0.74
14	元数据	27.5	57.143	1.54
15	Web	25	57.143	1.406
16	图书馆学	25	56.338	0.801
17	信息	25	54.054	1.424
18	知识管理	22.5	55.556	0.997
19	网络信息组织	22.5	51.282	1.113
20	网络	22.5	56.338	0.805
21	Folksonomy	20	54.795	0.418
22	信息技术	20	55.556	0.565
23	标签	20	54.795	0.526
24	网络信息	20	55.556	0.631
25	信息资源建设	17.5	54.795	0.292
26	模式	17.5	54.795	0.303
27	公共图书馆	17.5	54.795	0.114
28	数据库	17.5	54.795	0.329
29	多媒体	15	54.054	0.411
30	知识服务	15	54.054	0.141
31	叙词表	15	54.054	0.272
32	情报学	15	49.383	0.131
33	用户参与	15	54.054	0.247
34	XML	12.5	51.948	0.337
35	电子商务	12.5	53.333	0.136
36	搜索引擎	12.5	53.333	0.179
37	用户	10	51.948	0.043
38	信息管理	10	51.948	0.171
39	可视化	7.5	50.633	0
40	电子政务	7.5	50.633	0.043
41	文本分类	5	50	0
	平均值	24.268	56.856	1.995

3 研究结果分析

3.1 词频统计分析

文献关键词的组合在某种程度上可以表示该文献的研究内容,因此如果文献对某方面研究的越多,则相对应的关键词出现的频次也会越多^[7]。论文提取每篇文献中的关键词,去除与研究内容无关的一些关键词,如“研究”、“会议”等,得到3714个词。统计这3714个关键词的词频,可以看出词频不小于10的41词在此应视为高频词并作为论文的处理对象,如表1所示。通过观察可以看出,我国近五年来的研究热点主要集中在41个高频关键词所反映的主题上,其中以本体、数字图书馆、Web 2.0研究的最多,说明我国信息组织的研究多与本体、数字图书馆、Web 2.0结合起来。

3.2 共词矩阵分析与相似矩阵分析

共词分析主要是利用大量文献中共同出现的关键词来反映它们之间的关系强度,进而确定该文献集所代表的主题的研究热点^[8]。其做法是统计一组关键词在同一篇文献中共同出现的次数并加以汇总,然后再利用统计分析方法,将其共词关系简化,用数值、图形、表格等方式直观地表示出来^[9]。

如果两个关键词的共现频率越高,说明它们越容易结合起来共同研究^[10]。如果两个关键词的相似度越高,说明这两个关键词所代表的内容越接近。观察表2可知,共词矩阵与相似矩阵都是对称矩阵,共词矩阵的数值都是大于0的整数,而相似矩阵的数值是0到1之间的数。共词矩阵对角线的数值表示是该关键词总共出现在文献中的篇数;相似矩阵对角线数值的涵义是该词自身的相似度,很明显是1。共词矩阵非对角线的数

值表示相应的两个关键词共同出现在一篇文章中的次数;相似矩阵非对角线数值的表示两个关键词的共现强度。

观察表2,可得出以下结论。“信息组织”共出现了298次,是所有关键词中词频最高的。而“信息组织”与“数字图书馆”共现的频次是25,说明数字图书馆与信息组织同时出现在25篇文献中。“信息组织”与“Web 2.0”的共现频次是15,说明Web 2.0与信息组织同时出现在15篇文献中。数字图书馆、Web 2.0与信息组织共现的频次多,说明研究数字图书馆信息组织和Web 2.0信息组织的文献比较多,它们相继成为信息组织这一主题的研究热点。知识组织和数字图书馆、Web 2.0的共现频次分别是1与0,说明它们的联系目前还不紧密,结合研究的比较少,可能是由于技术手段不够成熟。

表3中的第*i*行与第*j*列交叉的值 Z_{ij} 表示对应两个关键词之间共现强度,0表示两个关键词没有同时出现。 Z_{ij} 越大说明这两个关键词的共现强度越大,它们之间的距离越近,相似度越大,反之距离越远,相似度越小。

3.3 中心性分析

社会网络分析(SNA)是20世纪70年代以来在社会学、人类学、数学、心理学、通讯科学等领域发展起来的一个研究分支^[11]。社会网络分析SNA的一个分析重点是中心性分析,通过对中心性的分析可以得出所分析对象在网络中的地位如何^[12]。社会网络分析中的中心度有6种,此论文中用到的是绝对度数中心度、绝对接近中心度、绝对中间中心度。

见表4,通过计算各高频关键词的中心度,各高频关键词的平均度数中心度是24.268,平均接近中心度是56.856,平均中间中心度是1.995。大于平均度数中心度的高频关键词有17个,表示

这 17 个关键词拥有最强的影响力。

3.3.1 度数中心度分析

度数中心度是反映行动者自身的交易能力,可分为绝对度数中心度,相对度数中心度^[13]。在此论文中代表的含义是某个关键词,如“情报学”,与其他关键词,“数字图书馆”,共同出现在同一篇文献中的可能性。关键词在网络中的交际能力越强,即它与多个关键词有联系,那么体现在论文中的是它的绝对度数中心度高。如“Web 2.0”的绝对度数中心度是 40,表示它和其他 40 个关键词共同出现在文献中。其中绝对度数中心度最高的是关键词“信息组织”,为 92.5,最低的是关键词“文本分类”,为 5。观察表 4,可知信息组织、数字图书馆、Web 2.0 等前 11 个关键词的绝对度数中心度明显大于其他的,说明这 11 个词是当前与信息组织相关性最高的,是当前研究的热点所在。

3.3.2 接近中心度分析

接近中心度是行动者在多大程度上不受网络中其他行动者控制的能力,分为绝对接近中心度、相对接近中心度,此处只用到绝对接近中心度^[14]。假如某关键词与其他关键词容易建立联系,在论文中体现在它与其他关键词容易出现在同一篇文章中,那么它的绝对接近中心度会越小。而通过 Ucinet 计算所得的接近中心度是真实接近中心度的倒数,所以在表 4 中,关键词的接近中心度越大,表明次关键词处在网络中的核心位置。网络中的绝对接近中心度的平均值是 56.856,有 15 个关键词的绝对接近中心度高于平均绝对接近中心度,说明这 15 个词比较不易受其他词的控制,处于相对核心的位置;“信息组织”的绝对接近中心度最大,为 93.023,表明“情报学”在网络中处于最核心的地位;“情报学”的接近中心度最小,为 49.383;“数字图书馆”、“Web 2.0”、“本体”等 15 个词处于网络的核心地位。

3.3.3 中间中心度分析

中间中心度是行动者在网络中的控制能力,分为绝对中间中心度、相对中间中心度,此处只用到绝对中间中心度。绝对中间中心度通常表示的是一个关键词影响其他关键词共现的大小。见表 4,“信息组织”的绝对中间中心度最大,表明它影响其他词是否共现的能力最强。“数字图书馆”、“Web 2.0”等 9 个词的绝对中间中心度明显大于其他的关键词,说明它们影响其它词是否共现的能力比较强;“可视化”与“文本分类”的绝对中间中心度为 0,表明它们没有影响其它词共现的能力;绝对中间中心度的平均值是 1.995,超过平均值的有 9 个关键词。

4 结论

根据上述对国内信息组织研究的分析,可以发现近五年来该领域的研究热点有以下主题。

(1) 数字图书馆

信息组织是图书馆的研究范围。随着网络和信息技术的发展,传统图书馆的资源不断数字化,服务方式不断网络化,数字图书馆应用而生。而数字图书馆的信息组织方式不同于传统图书馆的信息组织方式,要求更加适用于网络环境。其信息组织方式主要有元数据、数据库、叙词表等,这就需要更多的学者对其进行探索。所以目前图书馆领域的学者成为了研究信息组织的中坚力量,关于此方面的论文比较多,数字图书馆是信息组织的研究热点。

(2) Web 2.0

Web 2.0 是在 Web 1.0 的基础上,更加注重与用户的交互即用户参与,使用户不仅只是信息

的接受者,还成为信息的创建者和发布者。信息更加大众化、分散化和海量^[3]。因此 Web 2.0 环境下的信息组织与传统信息组织有所不同。其信息组织方式有 Folksonomy、知识组织、本体等;信息组织技术有 XML 技术;信息检索方式主要是搜索引擎。Web 2.0 是信息组织领域的一大研究热点。

(3) 知识组织

信息组织与知识组织的联系比较紧密。信息组织是信息的序化,而知识组织是将这些序化的信息从内容层次上连接起来,因此信息组织是知识组织的基础条件,即对知识进行组织必须先要将信息组织起来。目前论文中多是对两者的发展历程、理论基础和方法等方面进行比较分析。知识组织是信息组织的研究方向之一,是其领域的研究热点。

(4) 本体

本体是新的信息组织方法,具有传统信息组织方式所不具备的优势。本体是对领域知识的概念化、形式化的规范,是在语义和知识层面上描述信息系统的概念模型的建模工具^[3]。其基础是语义网的构建,语义网可以让机器按照信息组织的方式进行推理,实现一定程度的智能化,是智能检索的需要。因此本体是信息组织的研究热点。

综上所述,数字图书馆、Web 2.0、知识组织和本体是信息组织领域的研究热点。从这些研究热点中可以得知,目前国内学者关注重点已经从传统信息组织转变为网络环境下的信息组织,并取得了一定的成就。但是国内关于信息组织评价体系的研究比较少,未来要在此方面加大研究力度。同时要密切关注计算机领域的前沿技术并适当的加以引进和融合,并且努力将研究成果转化为实际应用。

参考文献

- [1] 焦丽.我国信息组织研究述评[J].农业图书情报学刊,2008,20(4):59-60.
- [2] 冷伏海.信息组织概论[M].北京:科学出版社,2004.
- [3] 付凯丽.Web2.0环境下网络信息组织研究综述[J].情报探索,2012(8):27-30.
- [4] 王知津,郑悦萍.信息组织中的语义关系概念及类型[J].图书馆工作与研究,2013(11):13-19.
- [5] 杨雨霖.我国图情档领域信息组织标准发展现状研究[J].图书馆学研究,2014(11):63-66.
- [6] 刘启元,叶鹰.文献题录信息挖掘技术方法及其软件 SATI 的实现-以中外图书情报学为例[J].信息资源管理学报,2012(1):50-58.
- [7] 叶平浩.基于社会网络分析的知识组织研究图谱[J].科技管理研究,2013,33(8):148-152.
- [8] 王红.近十年我国图书情报学科研究热点的共词分析[J].情报学报,2011,30(7):765-775.
- [9] [10] 王晓光.科学知识网络的形成与演化[J].情报学报,2009,28(4):599-605;2010,29(2):314-322.
- [11] 蒋春林,张立伟,刘胜博.图书情报学期刊“联锁编委”的社会网络分析[J].情报学报,2014,33(5):481-490.
- [12] 党洪莉,孙红霞.图书情报学博客的社会网络分析[J].情报杂志,2009(1):180-182.
- [13] 朱庆华,李亮.社会网络分析法及其在情报学中的应用[J].情报理论与实践,2008,31(2):179-183.
- [14] 裴雷,马费成.社会网络分析在情报学中的应用和发展[J].图书馆论坛,2006,26(6):40-45.