

我国数字图书馆研究论文 (2005-2009) 的统计分析 ——资源建设与资源组织研究*

□ 张鹏 王建冬 王继民 / 北京大学信息管理系 北京 100871

摘要: 基于2005—2009年国内数字图书馆资源建设和资源组织研究领域相关文献的计量分析,对数字图书馆资源领域近5年研究的总体情况、核心期刊发文量、热门关键词分布情况以及核心概念网络的分布情况进行统计分析。同时,采用社会网络分析和计量分析的方法,定义了领域关键词的专属度指标,并计算了国内数字图书馆资源建设和资源组织研究领域的向心度指标。

关键词: 数字图书馆, 资源建设, 资源组织, 文献计量, 专属度, 向心度

DOI: 10.3772/j.issn.1673-2286.2010.03.013

1 前言

数字图书馆建设的核心是数字信息资源建设和组织,目标是更好地对数字化信息资源进行开发利用。资源建设和资源组织包括理论、流程、方法、标准规范、系统框架、知识产权和技术问题等多方面的问题,只有将信息资源建设的这些要素有机地结合起来,从宏观上构建数字图书馆的信息资源建设和组织框架,形成科学的信息资源建设和组织流程,才能更好地指导资源建设实践^[1]。

20世纪90年代中后期,我国也相继开展了中国试验型数字图书馆工程、CALIS数字图书馆、全国文化信息资源共享工程以及中国科学院国家科学数字图书馆(CSDL)等项目,如今中国数字图书馆、上图数字图书馆、超星数字图书馆、书生之家数字图书馆也都投入运营。梁红^[2]指出,数字图书馆信息资源的种类和结构比较复杂,如何采取科学有效的方法和手段对各种信息资源进行筛选和整理,实现深层次的开发和整合,如何按照科学性、系统性、整体性的原则建设高质量的数字图书馆,如何更加合理地引进数字化信息

资源,是建设资源丰富、结构一体、共享型、开放式数字图书馆需要认真思考的问题。

本文拟从文献计量的角度,对近年来国内数字图书馆资源建设与资源组织研究领域的总体发展情况、核心期刊分布情况、领域分布情况等进行统计分析。

2 数据检索方法

本文所使用的数据是由万方数据股份有限公司收录的,包含国内图书馆学、情报学、档案学、编辑出版专业、新闻传播专业、计算机技术专业和综合类(各种学报等)的206种核心期刊的所有文献数据,共440594条记录,文献发表年份为1978-2009年,但前期数据收录不全。每条记录由标识号、题名、作者、机构、关键词、摘要、刊名、发表年、基金项目等字段组成。从以上文献数据中检索题名字段和关键词字段中包含“数字图书馆”的记录,并且设定发表年为2005-2009年的文献,得到记录1997条,作为国内数字图书馆研究的核心论文。

从国内数字图书馆研究核心论文中检索题名字

* 本文受北京大学前瞻性研究课题“基于网络使用的中文科技期刊及论文评价机制和方法研究”资助。

段中包含“资源建设”、“资源组织”、“数字资源”、“知识组织”或“信息组织”的记录,得到114条数据。对这些记录进行关键词词频统计,从高频关键词列表中手工筛选出以下与数字图书馆资源建设与资源组织相关的26个检索词:元数据、信息资源、资源整合、知识管理、长期保存、数字化资源、数字化馆藏、开放获取、馆藏建设、电子资源、文献资源、词汇列表、DC、组织模式、组织方法、资源集成、资源构成、知识集成、资源共享、资源共建、信息集成、文献保障、数据集成、数据交换、数据库、内容管理,并关键词字段进行扩展检索,得到738条数据。本文将作为国内数字图书馆资源建设和资源组织研究领域的主要文献。

3 总体情况

图1显示了2005-2009年5年间数字图书馆资源建设与资源组织研究论文占整个数字图书馆领域研究论文比例的历年变化情况:

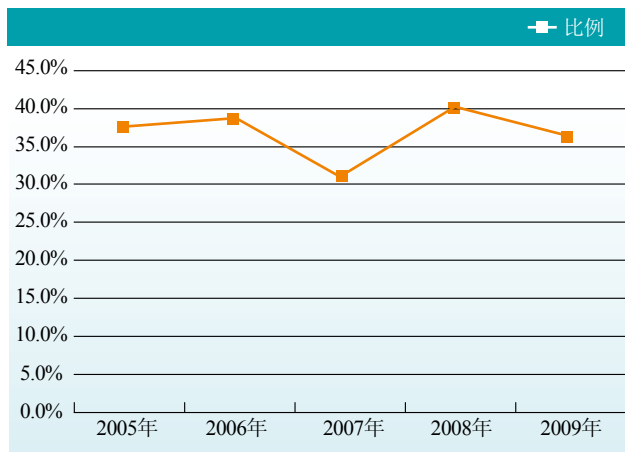


图1 数图资源研究论文的历年比例

2005-2008年数字图书馆资源建设与资源组织研究论文的年平均发文量为174篇^①,从图1可以看出,数字图书馆资源建设与资源组织研究论文占整个数字图书馆论文的比例基本稳定在30%到40%之间。除了2007年低于35%,数字图书馆资源建设与资源组织的研究论文在其余各年份中所占比例均大于35%,且上下浮动不大,这表明资源建设与资源组织研究在数字图书馆研究领域一直处于很重要位置。

^① 由于2009年数据尚不完整,因此在计算时并未考虑在内。

4 核心期刊发文量

该领域发文量最多的20种核心期刊与前文中曾提到的整个数字图书馆领域的20种核心期刊一致。20种核心期刊总发文量为705篇,占总数的95.5%。为进一步考察20种核心期刊在数字图书馆资源研究领域的偏好程度,我们计算每种期刊在数字图书馆资源研究领域的发文数量占其在整个数字图书馆领域的发文量的比重,将其作为考察期刊在数字图书馆资源研究领域的偏好程度的指标,计算结果如表1所示:

表1 期刊在数图资源领域的发文状况

期刊名	该领域 发文量	总发文量	比例
中国图书馆学报	46	81	56.8%
图书馆工作与amp;研究	38	89	42.7%
图书馆建设	26	61	42.6%
图书馆情报工作	82	198	41.4%
图书馆情报知识	11	28	39.3%
兰台世界	16	41	39.0%
情报科学	39	102	38.2%
情报杂志	56	149	37.6%
图书馆论坛	51	137	37.2%
情报学报	10	27	37.0%
现代图书情报技术	118	320	36.9%
情报资料工作	30	82	36.6%
情报理论与实践	19	54	35.2%
图书馆学研究	48	140	34.3%
大学图书馆学报	21	62	33.9%
图书馆杂志	23	70	32.9%
国家图书馆学刊	13	43	30.2%
图书与amp;情报	11	38	28.9%
图书馆理论与实践	34	124	27.4%
图书馆	13	49	26.5%

可以看出,20种核心期刊对数字图书馆资源研究领域的关注程度呈现较明显的不同,对该领域关注最多的几种期刊包括《中国图书馆学报》(56.8%)、

表2 数字图书馆资源研究高频词

排名	关键词	频次
1	数字图书馆	404
2	元数据	48
3	数字资源	38
4	图书馆	37
5	信息资源	29
6	信息组织	28
7	资源共享	27
8	知识管理	24
9	知识组织	22
10	信息资源建设	22
11	资源整合	19
12	资源建设	18
13	数据库	17
14	特色数据库	16
15	高校图书馆	16
16	读者服务	16
17	个性化服务	15
18	数字资源建设	14
19	本体	14
20	XML	14

从图2可以看出,目前国内数字图书馆资源领域研究中,对于资源建设和资源组织的研究涉及范围较广,主要有知识产权、信息组织、资源整合、信息技术、用户服务等方面。

为进一步考察国内数字图书馆资源建设和资源组



织研究领域的关键词从属于数字图书馆资源建设和资源组织研究领域的程度，我们设定了关键词的领域专属度指标。该指标的基本思想是，一个关键词在属于该领域的文献中出现的次数，与其在不属于该领域的文献中出现的次数的比值，再使用领域内外文献总数作为修正指标，得到关键词在数字图书馆资源建设和资源组织领域的专属度指标计算方法如下：

设关键词 K_i 在数字图书馆领域出现总词频为 F_i ，在数字图书馆资源建设和资源组织研究领域出现词频为 f_i 。所有关键词在数字图书馆其他领域出现总词频的平均值 $F = (\sum_{i=1}^n (F_i - f_i)) / n$ ，在数字图书馆资源建设和资源组织研究领域出现词频的平均值 $f = \sum_{i=1}^n f_i / n$ 。

则关键词 K_i 的专属度 $C_i = (f_i \times F) / ((F_i - f_i) \times f)$ 。

当 $C_i > 1$ 时，则可以认为该关键词“专属”于数字图书馆资源建设和资源组织研究领域。更进一步，还

可以通过计算专属度 >1 的关键词在分支领域所有关键词中所占的比例考察该领域的向心度。

结合上述计算结果，可以计算出数图资源研究领域的向心度 $=0.7549$ 。

7 结语

本文通过对国内数字图书馆资源建设和资源组织研究领域的相关文献的计量分析，从整体上考察了国内数字图书馆资源建设和资源组织研究领域的发展现状，目前，对于资源建设和资源组织的研究主要涉及知识产权、信息组织、资源整合、信息技术、用户服务等方面。希望本文的研究能够为领域内研究者的进一步探讨提供有益借鉴。

表3 关键词专属度

关键词	数图资源领域出现词频	数图领域出现词频	其他领域出现词频	专属度
数字图书馆	404	1148	744	0.39734
元数据	48	48	0	NA
数字资源	38	38	0	NA
图书馆	37	108	71	0.381327
信息资源	29	29	0	NA
信息组织	28	28	0	NA
资源共享	27	27	0	NA
知识管理	24	24	0	NA
知识组织	22	22	0	NA
信息资源建设	22	22	0	NA
资源整合	19	19	0	NA
资源建设	18	18	0	NA
数据库	17	17	0	NA
高校图书馆	16	32	16	0.731735
读者服务	16	18	2	5.853882
特色数据库	16	16	0	NA
个性化服务	15	52	37	0.296649
本体	14	49	35	0.292694
XML	14	22	8	1.280537
数字资源建设	14	14	0	NA

参考文献

- [1] 伏广红.论高校数字图书馆信息资源的建设与开发利用[J].现代情报,2007(6):74-76.
[2] 梁红.高校数字图书馆信息资源建设探析[J].图书馆工作与研究,2005(4):55-57.

作者简介

张鹏 (1984-), 北京大学信息管理系硕士研究生, 研究方向: 信息检索与分析。通讯地址: 北京大学信息管理系 100871。E-mail: pengzhang08@gmail.com

王建冬 (1982-), 北京大学信息管理系博士研究生, 研究方向: 信息服务与情报分析。通讯地址: 同上

王继民 (1966-), 北京大学信息管理系副教授, 博士, 研究方向: 搜索引擎与Web挖掘。通讯地址: 同上

A Statistical Analysis on China Digital Library Research Papers (2005-2009)--- Resource Development and Organization

Zhang Peng, Wang Jiandong, Wang Jimin / Department of Information Management, Peking University, Beijing, 100871

Abstract: Based on resource development and organization literature of digital library research from WangFang Date between 2005 and 2009, we carry out a statistical analysis of digital library management's 5-year study, which includes the overall situation of the study, the related articles of core journals, the distribution of popular keywords, as well as the core concept network. At the same time, using social network analysis and quantitative analysis methods, we define the exclusive degree of keyword and calculate the centripetal degree of the digital library resource development and organization field.

Keywords: Digital library, Resource development and organization, Bibliometric, Exclusive degree, Centripetal degree

(收稿日期: 2010-01-14)