

●缪其浩

观察国际图书馆学术前沿及其发展:内容分析

摘 要 通过对国际图联(IFLA)大会议题进行内容分析,可看出国际图书馆界对本学科学术前沿的关注及其变化。研究表明:从 1995~2001 年,图书馆学情报学界对数字化议题的关注逐年加强,对知识,特别是知识管理概念的关注程度急剧提高,而对信息技术、因特网等概念则出现兴趣消退迹象。表 3。图 2。参考文献 7。

关键词 国际图联 图书馆学研究 内容分析 知识管理

分类号 G250

ABSTRACT Through an analysis of the contents of topics in IFLA General Conferences, the author summarizes the following trends: during 1995-2001, there is a continual increase in the number of articles concerning digitization and knowledge management, and a decrease in the number of articles concerning IT and the Internet. 3 tabs. 2 figs. 7 refs.

KEY WORDS IFLA. Library science. Content analysis. Knowledge management.

CLASS NUMBER G250

1 背景:图书馆信息化潮流及其热点的变化

近年来随着信息技术的全球化发展,以计算机和网络应用为代表的信息化潮流不仅席卷发达国家的各类图书馆,而且开始波及广大发展中国家。信息化作为一个热门话题,自然也成为学术研究的重点出现在图书馆界各种会议的讨论中。然而同样明显的是,在信息化这个大题目下,图书馆界关注的具体热点是在变动的。例如“自动化”、“计算机”等曾经常出现在图书馆界学术论文标题和会议主题上,而在 2001 年国际图联大会上,600 多个论文标题或活动的说明文字中竟然没有一处提到这两个术语。这种变化对于我们观察学术动向是很重要的。

本文通过对国际图书馆界联合会(IFLA)年度大会的观察,来了解国际图书馆界对于信息化以及其他前沿课题中各个主题的起伏消长。由于国际图联是各国图书馆员和相关人员的共同论坛,其年度大会又汇集了数以百计的论文、发言和报告,并且具有良好的连续性,应当是合适的研究基础。就信息化而言,国际图联大会在专业深度上可能不如其他一些会议如美国图书馆协会(ALA)或美国信息科学技术协会(ASIST)的相应会议,然而国际图联的优势在于它具有广泛的代表性,特别是有不少发展中国家的代表参加,因而具有其独特的意义。

2 方法:内容分析法及其在图书馆研究中的应用

笔者在本文中所用的是“内容分析(Content Analysis)”方法。在社会科学领域,内容分析是一种研究通讯(沟通)的性质、隐含意义、动态过程以及沟通者行为的一种方法^[1]。所谓内容,可以是文本,也可以是数据、图像甚至行为姿态。近年来,随着信息技术的发展,利用专门编制的软件对数字化的内容进行分析逐渐成为研究方法的主流。在国外,内容分析法也广泛运用于图书馆和信息科学领域。1971 年出版的《图书馆与信息科学百科全书·第五卷》中“内容分析”条目长达 24 页^[2]。笔者查找了 1969~2001 年 6 月的《图书馆信息科学文摘(LISA)》,其中图书馆学研究采用内容分析方法的文章有 130 多篇,大多数都是用于图书馆行业的动向分析。在国内,笔者较早在信息研究领域对该方法进行过介绍和评述^[3],并做过一些实例^[4]。程焕文 1997 年的著作研究了我国图书馆学信息学期刊论文^[5],这是当时一个国际性研究的一部分,各国的研究都采用了基本统一的内容分析框架。

3 分析:框架设计、样本、关键词及质量控制

本次研究采用的具体方法又可称作“词频分析”,它通过文本型的分析对象中特定关键词的出现频度及互相关联进行分析、推断,有关该方法的具体描述可参见笔者文章^[6]。对国际图联大会内容进

行词频分析研究的依据是这样一些基本假设:论文作者和活动组织者往往将自认为最重要的概念列入标题或议题;在大会文献集合里某个关键词出现越是频繁,表明国际图书馆界对这个概念的关注程度越高;如果某关键词在数年内呈现连续递增,可以认为是出现了“热点”。

假设是需要检验的,后来在研究过程中发现,对某些假设需要作出修正。

(1) 样本:每年国际图联大会所产生的文献主要是程序册。从 1995 年起在网络上可以获得比较完整的程序册,同时也开始提供部分全文的超文本链接。程序册文字主要是英文,也有一小部分内容采用法文、德文和俄文译文,记载了各项大会活动的内容,包括组别、主题、论文题目、内容提要 and 作者等信息。本次研究使用的主要是标题,只是在深入分析时利用了少数全文。主要是因为网络版全文历年都很不完整。

本次研究以国际图联大会所有可以获得的网络版程序册为样本,即从第 61 届(1995 年)到最近的 67 届(2001 年)。所有样本都可以从国际图联网站下载。

(2) 关键词:本次研究的目的是观察信息化等图书馆研究前沿内容的热点。为此,笔者首先需要找出能够反映研究前沿的一些关键词。实际操作可以有两种做法:一是利用专门软件对样本的全部词汇进行词频统计,从中可以筛选出有关信息化的关键词;二是从研究者需要出发自行选择关键词,这样当然比较简单,缺点是容易忽略研究者事先不了解的新概念。由于笔者对这个领域比较熟悉,本文实际采用了第二种办法,初步选出了一些关键词及其下位词组(见表 1)。

表 1 初步选择的关键词及原文

关键词(组)	原文(含各种相关的词或词组)
自动化 *	Automation
因特网	Internet
计算机 *	Computer, personal computer, PC
数字化	digital, digitalization, digitizing, digitized
- 数字图书馆	digital library
信息 * *	Information
- 信息技术	Information technology, IT
知识	Knowledge
- 知识管理	Knowledge management, KM
元数据	Metadata, Dublin Core, DC

*“自动化”和“计算机”均从 1995 年起因词频太低而撤出了关键词表。

* * “信息”一词的含义太泛,其中许多不属于关键词。

(3) 计量方法:利用 IE 浏览器上“编辑”栏中的“查找”命令,逐项键入关键词,如只需键入词干 digit,就能够检出信息化的四种变形,再用人工方法剔除极少数误检词,即可获得简单计数。本次研究的计数规则是,任何活动议题或论文标题只要出现关键词,不论什么文种,也不管是否有重复,统计时只计为一次。关键词如果明显不反映内容含义(如“信息”一词常常如此)或出现在作者单位名称中,则不能计数。

(4) 基数调节:每年大会程序册的篇幅相差甚远,例如 67 届程序册的词汇数比 61 届要高出一倍以上,活动议题和论文标题总数也有一定差距,进行词频比较时,要扣除基数变化的因素。为此,统计了各届大会程序册的论文标题和活动议题总数,得出调节系数(见表 2)。图 1 和图 2 的关键词发展趋势就是经过计数调节后的数据。

表 2 基数的调节系数

	61 届	62 届	63 届	64 届	65 届	66 届	67 届
调节系数	1.00	1.02	1.04	1.10	0.86	1.16	1.35

(5) 文种问题:有极少数论文用法文写成,而且没有英文译文;还有个别文章在英文标题上不出现 digital 或其变型,但是在法文译文上却有 numerique 或变型。笔者补充了数字化的法文词干进行测试,结果表明,这方面影响微小。因此在具体分析时一般不加考虑。

(6) 增设“内容主题”。当需要对关键概念,如“数字化”进一步分析时,除了统计词频外,还需要对照程序册的文字,深入分析具体内容,必要时查询全文,用判断的方法提取该词涉及的不同内容主题。

(7) 误差分析。本研究误差的主要来源有两方面:一是数据本身,议题和标题采用什么词不完全是作者的随机行为,标题和实际内容的要点也可能不完全吻合。研究中发现有些词频起落可能是国际图联组织行为的结果,也不能排除其他我们不了解的原因对计数的影响,例如 65 届论文和议题数显著减少,原因有待分析。二是虽然本研究一般是以词频计数为基础的,但是分析结论的获得还要基于对专业的了解和从其他来源的信息进行定性的思考和推断。当然这个操作一方面是提高研究可靠性所必须的,同时也可能带来主观性和误判。但是无论软件技术如何发展,这个“两难”都不可避免,研究者的思考分析是无法替代的。

4 结论:趋势及其解读

限于篇幅,依照上述要求对文献对象进行统计分析的具体细节和大量数据表就不一一列出了。主要研究结果用图像来表达(见图1)。

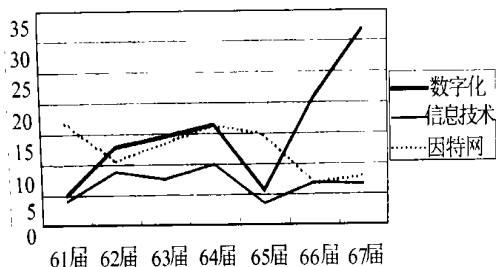


图1 数字化、信息技术、因特网三个关键词变动情况

从图1可以看出这样一些趋势:

(1) 因特网逐渐从热点淡出。因特网曾经是国际图书馆界的热门话题,在61届国际图联大会上谈得最多,然而词频分析表明,近年来图书馆界对其兴趣正在消退。当然,这仅仅表明它可能不再是一个学术研究的“中心话题”而已,在应用层面上,因特网无疑是正在稳步发展。再从具体内容主题来分析,61届国际图联大会时有一个研讨班(Workshop),从因特网和www的基本概念说起,还介绍如何利用Gopher和www查资料以及如何做一个网站;在64届成立的讨论组(Discussion Group)上,除了仍然有如何设计图书馆网站外,已经出现了利用因特网作为参考工具和www上版权等较为深入的问题;而到了67届已经全部是讨论诸如保存文化资源、在图书馆设立读者上网站和因特网内容审查等相当专门化的论题,实际上已经不再是一个十分专门的讨论组,其内容可以融入其他专题组。

(2) 数字化是目前最热门的话题。除65届大会之外,数字化呈现稳步的提高,到2001年67届大会时便遥遥领先于其它关键词,在此过程中,“信息技术”与“数字化”词频显示了相关性,只是在最近一两年才出现落伍。在数字化的各个内容主题中,资源的数字化以及数字化资源直接有关的管理、保存、评价、归档等具体内容仍占主要地位,在标题中特地提到资源具体类型的不多,其中制图和地图的数字化是一个例外,这说明数字化可以在复杂的图像资料中发挥特别作用。数字化服务虽然词频还很低,但是值得关注,在67届大会里出现的四次中,三次是

同参考服务有关的。笔者参加了该次大会,也亲身体会到数字化参考服务正在引起国际图书馆界极大的兴趣。为了验证这一点,笔者又对“参考(reference)”一词作了简单的词频分析,在最近的67届大会程序册中提到该词的有效次数17次,其中带有明显信息化修饰语的为9次,数量超过一半。这就有力地证明参考服务正在走向数字化。

67届国际图联大会有个引人注目的现象就是数字鸿沟和数字权利主题的突然浮现。这固然反映了国际图联对发展中国家面临问题的特别关注,然而对此仅从内容分析角度还难以判断。有意思的是,虽然数字化的概念有明显增长,词组“数字图书馆”并未同步增多,该词词频不仅一直不高,而且从66届起还出现下降迹象。在2000年66届的一次会议上曾经打算以后成立一个数字图书馆讨论组,还拟定了一些分组的论题,然而67届大会上并没有实现。这可能反映出对数字图书馆泛泛而谈的时候已经过去,现在图书馆界更关注的是其中的具体课题。

表3 三届国际图联大会关于数字化具体内容主题分布

内容主题归类	61届	64届	67届
数字化项目介绍	1	2	2
数字化资源	1	7	14
数字化服务	0	1	4
数字地理与地图	2	0	5
有关法律问题	0	1	3
数字权利与数字鸿沟	0	0	8
相关技术问题	0	2	0

(3) 对“知识”感兴趣是最新突发的亮点。简单计数表明,从64届到67届知识的词频出现了近10倍的增长,其中一个重要原因是67届大会的主旨中包含了一个短语“知识时代”。论文标题上“知识时代”这个短语可能因此而增加,例如本来作者只是讨论编目,可能就把题目改成了“知识时代的编目”。这些应该视作“噪声”,笔者均将其作非关键词处理。然而,尽管经过调整扣除了部分计数,最近一两年来“知识”概念的增长还是十分明显的,其中又以知识管理为主体(见图2)。经过提取内容主题步骤以后我们发现,对知识管理的讨论主要集中在图书信息学教育和各种机构的经验两个类别,后者主要包括国家、政府、公司等进行知识管理的实例,只有66届有一篇来自中国的论文讨论了图书馆未来的知识管

●张晓林

开放数字信息服务体系:概念、结构与技术^{*}

摘 要 开放数字信息服务体系,有其特定的概念、原则要求和功能框架。开放系统的开放描述基于扩展的元数据概念,信息系统通过开放语言和规范机制来实现对系统各层次内容的开放描述。分布服务机制源于分布对象技术,它将各种系统视为一个数字对象,对其界面、功能、数据流、传输协议等进行描述。开放体系的开放集成,包括横向和纵向开放集成,它们各有其技术线路。图 5。参考文献 29。

关键词 数字图书馆 开放体系 开放描述 分布服务 开放集成 互操作

分类号 G250.76

ABSTRACT An open digital information service infrastructure has its special concept, principle requirements and functional frameworks. The open description of an open system is based on the extended concept of metadata, and an information system realizes the open description of contents on various levels of the system through open languages and standard mechanisms. In this paper, the author also discusses distributed services, open integration of open systems, etc. 5 figs. 29 refs.

KEY WORDS Digital library. Open infrastructure. Open description. Distributed services.

Open integration. Interoperability.

CLASS NUMBER G250.76

开放数字信息服务体系(下称开放体系)指开放集成各种分布、异构和多样化数字信息资源和数字化

1 开放数字信息服务体系的概念与要求

理^[7]。这可能表明,虽然图书馆界认识到知识管理是有价值的,但至少目前还没有什么值得报告的成果出现。

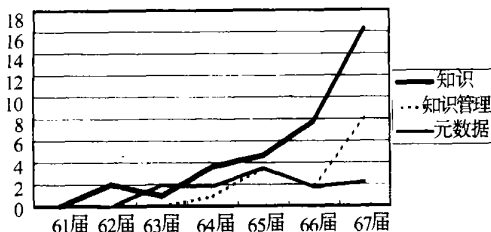


图2 知识、知识管理和元数据三个关键词的词频变化

在此要提一下元数据问题,如仅仅看词频的话,元数据还太低,不足以得出任何结论。然而元数据是数字化和知识管理这两大热门话题的共同焦点,2000年66届国际图联大会上开始筹备元数据讨论组,2001年元数据讨论组和知识管理讨论组一起成立。讨论组是国际图联大会的一级组织,而比研讨班或圆桌会议更为正式。它可能是一个“指示器”,表明图书馆信息化研究前沿出现了一个重要动向。

参考文献

- 1 Matilda W. Riley and Clarice S. Stoll: Content Analysis, in David L. Sills (ed): International Encyclopedia of the Social sciences, vol. 3, pp371-377, Collier-Macmillan Publishers, London, 1968
- 2 Patrick R. Penland: Content Analysis, in Allen Kent and Harrold Laucour (eds), Encyclopaedia of Library of Information Science, vol. 5, pp632 - 655, Maecel Dekker, Inc., New York, 1971
- 3 缪其浩.“内容分析”技术评述.情报科学,1986(1)
- 4 缪其浩,张左之.对JETRO海外技术情报活动的剖析.情报学报,1989(4)
- 5 程焕文.中国图书馆学信息学研究之文献计量研究.资讯传播学与图书馆学(台湾),1997(4)
- 6 缪其浩.内容分析法.见:包昌火主编.情报研究方法论.北京:科技文献出版社,1990:236~256
- 7 Tang, S. Knowledge Management in Libraries in the 21st Century, paper presented to the 66th IFLA General Conference, August 1999

缪其浩 上海图书馆研究员。通讯地址:上海市淮海中路1555号。邮编200031。

(来稿时间:2001-10-29)

^{*}本项研究得到“中国科学院百人计划”基金支持。