

●赵蓉英 张 洋 邱均平

网络信息计量学基本问题研究*

摘 要 “网络信息计量学”一词存在着多种定义,这些解释随学科的发展而不断变化。网络信息计量学是信息计量学的分支学科,也是一门交叉性的边缘应用性学科。“网络文献”所具有的三个层次的信息分别是:物理载体层次、元数据层次和信息单元层次。网络文献与传统文献具有相当大的相似性,传统文献计量研究方法可以运用于网络信息计量学的研究中。参考文献 23。

关键词 网络信息计量学 网络文献 计量学 文献计量学

分类号 G350

ABSTRACT There are various definitions for webometrics, which are changing with the development of the discipline. Webometrics is a branch of bibliometrics, and is also an applied cross-disciplinary science. Web documents have information in three levels, i. e. the levels of physical media, metadata and information unit. There are similarities between web documents and traditional documents, and traditional bibliometric research methods can be applied in webometric researches. 23 refs.

KEY WORDS Webometrics. Network documents. Statistical analysis. Bibliometrics.

CLASS NUMBER G350

在《网络信息计量学的兴起及其哲学思考》^[1]一文中,我们曾深入探讨了网络信息计量学的产生历程、产生背景以及该学科名称的由来。在此基础上,我们将在本文进一步分析网络信息计量学的学科定义、学科性质和研究对象等基本概念的发展演变过程,为探讨学科体系等其他基本问题打下基础。

1 网络信息计量学的定义

与其他新兴科学学科一样,学术界一直对“网络信息计量学”(Webometrics)一词有着各种各样的解释。而且这些解释还随着学科的不断发展和不断变化。

最早提出“Webometrics”的 T. C. Almind 等人将其定义为:“信息计量学方法在 WWW 上的应用。他们是将其视为一种全新的研究方法而提出的。因此,在最初的几年里,网络信息计量学一直被定义为一种新兴的研究方法或研究工具,例如, Moses A Boudourides 等人就明确指出,网络信息计量学是“一种应用于 WWW 上的方法”^[2]。李长忠曾将网络信息计量学定义为“应用文献计量学、科学计量学以及信息技术来分析各种信息媒介、信息交流的一种方法”^[3]。赵启民则认为,网络信息计量学是随着网络的发展和网络信息资源的激增而产生的一种网络计量工具,它综合采用文献计量、信息计量、统计学方法、计算机技术和网络技术对网络文献规律进行统计分析^[4]。此外,还有人将其定义为“基于 Web 的软件计量分析工具”,“集计算机技术、网络技术、计量学方法、统计学方法于一体(的方法),其应用范围覆盖了所有基于网络通信技术的信息测度”。

随着研究的深入,网络信息计量学逐步发展成一门独立

的科学学科,成为信息科学的新兴分支,这一术语的含义也相应地趋向于作为一门科学学科来描述。例如,李长忠等人后来就认为,“网络信息计量学是指在电子网络环境中,运用文献计量学、科学计量学、情报计量学的方法,对各种信息、信息媒介及其功能进行定量研究和分析(的一门学科)”^[5]。夏旭等人认为,“网络信息计量学是综合采用文献计量、信息计量、统计学方法、计算机技术和网络技术对网络文献规律进行统计分析的一门科学”^[6]。Peter Ingwersen 及其学生 Lennart Björneborn 则在近期的论文中将网络信息计量学定义为“运用文献计量学、信息计量学的方法对 Web 上信息资源、结构和技术的构成与应用进行定量研究的一门学科”^[7],这一定义被 Mike Thelwall 等人称之为“从情报科学的角度所下的定义”^[8]。此外,其他的定义还有:网络信息计量学是一门关于计算机软件设计的科学;是一门对网络文献规律进行分析的科学;是一门研究因特网上数据之间相互引用的科学。

邱均平指出,“从研究对象、方法、内容和目标等方面来看,网络信息计量学是采用数学、统计学等各种定量方法,对网上信息的组织、存贮、分布、传递、相互引证和开发利用等进行定量描述和统计分析,以便揭示其数量特征和内在规律的一门新兴分支学科。它主要是由网络技术、网络管理、信息资源管理与信息计量学等相互结合、交叉渗透而形成的一门交叉性边缘学科,也是信息计量学的一个新的发展方向 and 重要的研究领域”^[9]。这一定义自提出后就被学术界广泛引用,比较符合网络信息计量学的研究现状及其发展趋势。

需要说明的是,网络信息计量学作为一门发展中的新兴学科,其产生还不到 10 年,其定义还远没有定论。“网络信

* 本文系国家自然科学基金资助项目“网上学术信息的分布与变化规律研究及其应用”(70673071)的研究成果之一。

息计量学”这一术语所表示的含义的变化,实际上从一个角度反映出该学科由一种经验性的研究方法向一门严谨的科学学科发展转化的过程。可以预见,随着网络信息计量学研究的不断发展,这一学科的定义还将继续演化,直至发展成为一门成熟的为学术界所公认的科学学科。

2 网络信息计量学的学科性质

2.1 网络信息计量学是信息计量学的分支学科

20世纪60年代以来,在图书馆学、文献学、情报学和科学学领域相继出现了3个十分相似的定量性质的分支学科,即文献计量学(Bibliometrics)、科学计量学(Scientometrics)和信息计量学(Informetrics)(简称“三计学”)。

从某种意义上说,网络信息计量学就是“三计学”在网络上应用的一门学科,“三计学”构成了网络信息计量学的学科基础。首先,“三计学”中的很多理论、原理都可延伸到网络信息计量领域,成为网络信息计量学的理论基础,这其中既包括相似原理、耗散结构原理、马太效应、最小努力原则、随机过程原理等普遍原理,也包括信息交流理论、信息组织理论、信息系统理论等信息科学理论,还包括信息增长与老化规律、集中与分散规律、引文分布规律等经典规律。其次,在“三计学”中得到广泛应用的文献信息统计分析法、数学模型分析法、引文分析法、书目分析法、系统分析法、关键词统计分析法、关联数据分析法(包括聚类分析、共词分析、同域分析等)、计算机辅助文献信息计量分析法等定量方法同样成为网络信息计量研究的重要方法^[10]。第三,“三计学”在诸多应用领域,尤其是电子文献信息资料的统计分析等方面的研究成果,为网络信息计量学的形成奠定了基础,积累了经验。因此,近几年网络信息计量学的诞生,本质上说,乃是科学计量学、文献计量学、情报计量学和技术计量学在新的信息网络时代经过革命改造的结果,“三计学”构成了网络信息计量学的学科基础^[11]。

另一方面,“三计学”之间有着非常紧密的联系。从发展历史来看,信息计量学是在传统文献计量学和科学计量学的基础上扩展和演变而成,文献计量学和科学计量学是信息计量学的学科基础。从研究现状来看,尽管“三计学”仍然是相互独立的科学学科,但它们在研究对象、研究内容、研究方法等许多方面存在相似,甚至重合之处。从发展趋势来看,信息计量学的研究范围更加广泛,“三计学”将融合到信息计量学这一统一的学科体系之下。总之,“三计学”之间是相互交叉、相互促进的关系,“三计学”的融合已成为一种必然趋势^[12]。

因此,网络信息计量学最根本的性质是信息计量学的分支学科。

2.2 网络信息计量学是一门交叉性的边缘学科

和传统信息资源相比,网络信息在数量、结构、分布、传播范围、类型、载体形态、内涵、控制机制、传递手段等方面,都与传统的信息资源有很大差异,呈现出许多新特点,决定了网络信息计量学的研究过程具有许多区别于“三计学”的特征。例

如,网络信息的丰富性、多样性和分散性使统计数据很难做到准确而全面;网络信息的引用缺少统一规范,给网络资源的引用情况统计带来一定困难;网络信息的更新十分迅速,难以保存,统计工作难以保证真实性和准确性;网络信息自由度很大,使信息质量不稳定等等。“三计学”虽然构成了网络信息计量学的基础,但并不是网络信息计量学的全部,网络信息计量学必须不断汲取其他学科的成果,扩大自己的学科基础。网络信息计量学的这种“交叉”性质,一方面表现在信息科学内部分支学科之间的交叉、融合,包括信息技术、人工智能、计算机科学、信息资源管理、情报学、信息计量学等诸多分支学科;另一方面表现在与信息科学之外的其他学科交叉,包括数学、统计学、管理科学、社会学等更为广泛的科学领域。

2.3 网络信息计量学是一门应用性学科

科学史告诉我们,任何科学的产生和发展都由一定的科学背景和特定条件决定,社会实践需要是学科产生的根本动力。随着现代社会经济的发展和网络化的日益普及,加强网络信息资源的开发利用和网络管理已成为当务之急,而实施定量化管理则是其主要途径之一。网络信息计量学的研究成果必然会为网络管理的定量化和科学化提供理论指导及定量依据,而网络管理量化的实践需求又会促进网络信息计量学的全面发展。人类社会的这种客观实践需要,正是网络信息计量学产生和发展的根本动力。因此,网络信息计量学研究的根本目的主要是通过网上信息的计量研究,为网上信息的有序化组织和合理分布,为网络信息资源的优化配置和有效利用,为网络管理的规范化和科学化提供必要的定量依据,改善网络的组织管理和信息管理,提高管理水平,促进其经济效益和社会效益的充分发挥,推动社会信息化、网络化的健康发展。应用一直都是网络信息计量学的重要研究内容,虽然理论问题研究不可缺少,但更多的研究是结合实际情况,为解决网络管理工作中的具体问题展开,有明显的“应用性”特征。这一特征既符合所有计量科学的共性,也在一定程度上解释了目前网络信息计量学“应用”领先于“理论”的研究现状。

3 网络信息计量学的研究对象

研究对象是一门学科区别于其他学科的根本特点之一,也是一门学科之所以能够成为一门独立学科最根本的前提。对于网络信息计量学的研究对象,不同学者从不同角度做出了各自的总结和归纳。例如:徐久龄等人认为^[13],网络信息计量学的研究对象包括:(1)对网络信息计量学文献著者规律的研究;(2)对网络文献分散规律的研究;(3)对网络文献增长规律的研究;(4)对网络文献老化规律的研究;(5)对网络信息计量学文献引文的分析;(6)对网络上声音、图形、文字规律的研究。研究范围涉及网络信息计量学中的文献检索研究、引文分析、传统引文分析指标的改进(包括网络文献的引文数、平均引用数、自引数与自引率、被引用数、影响因素、快指标等)、老化规律的研究、累积优势原理与核心期刊效应等。李长忠等人则将网络信息计量学的研究对象归纳为^[14]:(1)万维网上的站点、主页、电子邮件、讨论组、新闻组等媒介和内

容;(2)万维网的研究测度与计量;(3)有关电子图书、期刊、图书馆和参考资源等电子网络信息资源。牟华等人认为^[15],网络信息计量学的研究对象包括网络技术、网页中文献计量学方法、万维网的电子数据库分析及网络结构、电子数据库资源的产出量、主页、电子引文分析、电子媒介和资源研究、数字图书馆、虚拟图书馆、电子信息系统、域名分布等。

我们认为,网络信息计量学的研究对象简单而言就是“网络信息”。问题的关键在于如何有效地描述“网络信息”。钟义信指出,“为了有效地描述信息,一定要对信息进行分类,分门别类地进行描述,建立分门别类的描述方法”^[16]。那么,如何对网络信息进行分类呢?

首先,网络信息归根到底属于信息的范畴,我们可以继承信息的分类,从不同角度将网络信息分为不同类型。例如按信息的性质,可分为语法信息、语义信息和语用信息;按信息的传递方向可分为纵向信息、横向信息和网状信息;按信息的内容,可以分为经济信息、科技信息、政治信息、文化信息、政策法规信息、娱乐信息等;按信息的流通渠道,可分为正式信息和非正式信息;按信息的记录方式,可分为语声信息、图像信息、文字信息和计算信息等;按照信息的加工程度,可以分为零次信息、一次信息、二次信息、三次信息等。其次,与传统信息相比,网络信息呈现出许多新特点。我们可以针对网络信息独有的特点,按照不同标准将它分成不同类型。例如按信息的时效性,可分为网上出版物、动态信息、联机馆藏书目数据库、国际联机数据库等;按发布形式,可分为书目信息、电子报刊、文本文档,以及网上电子信件、电子公告、专题讨论栏目等形式的非正式信息等;按信息存取方式,可划分为邮件型、电话型、揭示板型、广播型、图书馆型、书目型等;按具体的组织形式,可划分为文件、超媒体、数据库、网站、编目、学科信息门户、网络资源指南、搜索引擎、数字图书馆等。

对信息进行分类的目的是为了认识信息的性质和特征,以便描述信息和处理信息。信息究竟应该按照何种方式划分,没有固定标准,主要取决于人们分析问题的不同需要。例如,马费成从信息资源管理研究的角度出发,提出将信息资源分为记录型、实物型、智力型、零次型比较适宜^[17]。同样的道理,我们为了更好地进行网络信息计量研究,也必须从网络信息计量学的角度出发来对网络信息进行分类。上述分类方式虽然很多,但对网络信息计量学来说,大多都没有多少实际意义。

下面,我们从网络信息计量学的研究现状出发,综合目前网络信息计量研究的实际情况和发展趋势,按照研究方法和研究参与者的不同,将网络信息计量学的研究对象分为3个类型来论述,力求对网络信息计量学的研究对象有较为全面、清楚的认识。

3.1 网络特征信息的计量问题

网络作为日益重要的信息载体和交流工具,其客观特征可以通过诸多定量指标反映出来:既包括某个网站的点击量、网页数量、浏览次数等网络结构单元的个体特征信息,也包括网站数量、IP地址分布等反映网络整体特性的总体特

征信息;既包括计算机、作者、域名等反映网络内容提供者的相关信息,也包括年龄构成、文化程度、上网习惯等反映网络用户特征的相关信息;既包括人数、规模、构成、分布等静态信息,也包括流量、增长、发展趋势等动态信息。这类信息实际上反映了网络信息活动诸要素的总体统计特征。网络是一个复杂的自组织系统,在微观层面有不可预测性,只能在宏观层面上对网络信息活动的诸要素进行统计学意义的特征提取。这种经过加工形成的二次信息一般由专门的统计机构或研究机构在统计分析后形成统计报告。例如,中国互联网络信息中心每年发布的《中国互联网络发展状况统计报告》,就是对我国网络特征信息的权威统计。

3.2 网络拓扑信息的计量问题

从拓扑结构来看,所谓网络是由节点(node)和各节点之间的路径(path)构成。对万维网而言,网络节点是构成网络的结构单元,例如人们熟悉的站点、布告栏、聊天室、讨论组、电子邮件服务器等等,而路径就是连接各结构单元之间的超链接。在这个由“超链接”和“节点”构成的网络中,蕴含着大量的有用信息。例如,透过链接结构的拓扑学研究可以分析网络节点之间的联系程度和集中度,推论出网站群落,确认出权威网页,萃取出主题知识;通过对网路上大量、复杂但具结构化的网络节点进行网络结构的布局分析,可以合理配置资源,改进搜索引擎的算法等等。因此,网络链接结构信息也是网络信息计量学的重要研究对象之一。需要注意的是,这里所说的“网络结构单元”是一个分层次的概念,根据研究目的和视角不同,它可以是一个页面、一个文件夹、一个服务器、一个站点、一个机构,甚至是一个国家、一个地区。例如,当我们研究高校网站的访问者来源时,一般以网站作为基本的结构单元;而研究国家之间信息交流问题时,一般以国家作为基本的结构单元。有关这类信息的研究工作最为复杂。有关网络结构的研究,涉及到计算机、人工智能、拓扑学、社会学、图论等众多学科、领域的知识,一方面各个领域的研究者不得不吸收其他领域的研究方法,而另一方面又使不同领域的研究者都能够找到研究切入点。最后导致的结果就是这方面的研究者和研究方法的构成最为多样化,也是最容易引起争议的研究领域,成为各个相关学科的“必争之地”。

3.3 网络文献信息的计量问题

在传统文献计量学中,“文献”是指用文字、图形、符号、声频、视频等技术手段记录人类知识的一种载体,或称之为固化在一定物质载体上的知识。网络环境下,此概念延伸成为“网络文献”,被学术界广泛使用,但出于使用目的不同,其含义略有不同。例如,阎建国等人在探讨网络文献的收集与利用时,将它定义为“借助信息技术而存在于Internet上形式比较特殊的文献”^[18];姜仁珍在探讨网络文献的信息结构时将它定义为“以电子形式储存在光、磁等载体上,利用计算机技术、通信技术及多媒体技术在Internet上发布、传递,以节点为中心分布,并能在网络终端得以再现的文献信息单元或文献信息集合”^[19]。而黄晓斌在探讨网络文献的知识发现问题时将它定义为“在因特网上以数据形式存在

的各种文献信息内容的集合”^[20]。

以上有关“网络文献”的定义都不适合我们在此使用,其原因有二:第一,范围偏窄,以上定义实际上都没有超出“网络上传递的电子文献”的范围;第二,不够清晰,以上定义都使用了“文献”一词,这种递归式的定义没有反映出对象的本质特征。我们认为,要深刻掌握“网络文献”的含义,必须从“物质载体”和“知识内容”这两个构成文献的基本要素出发来探讨。首先,网络文献的载体是网络。虽然网络文献实际上是存储在网络服务器的存储设备上,但是“网络与各种服务器上的存储设备是一个有机的整体,而对用户来讲,网络比服务器上的存储设备更为直观”^[21]。因此,我们将网络作为载体看待。其次,网络文献必须记录有一定的知识内容。“知识”属于认识论的范畴,是“通过对信息的提炼和推理而获得的正确结论,是人的大脑通过思维重新组合的、系统化的信息集合”^[22],显然,无意义的文本、声音、图片都不属于此列。我们认为,任何满足以上两个要素的信息单元都属于“网络文献”的范畴。由此出发,我们将“网络文献(Web Document)”定义为“所有以网络为载体的包含知识内容的信息单元”。

按照我们的定义,网络文献不仅包括网络上传播的电子图书、电子期刊、电子报纸、专利信息、科技报告、学术会议论文、标准信息、学位论文、行政报告、会议资料等在网络上记录和传递的电子文献,也包括书目数据库、数字图书馆等结构化的信息集合,还包括日志文件、电子邮件、聊天室、电子留言板、个人主页等多种形式的信息单元,具有数量巨大、类型多样、变化频繁、结构复杂、质量层次参差不齐、用户差异大等特点。邱均平所描述的3个层次的网络信息计量学研究对象都属于“网络文献信息”的范畴,“网上信息本身”、“网上文献、文献信息及其相关特征信息”、“网络结构单元的信息”这3个组成部分实际上分别对应着“网络文献”所具有的3个层次的信息:物理载体层次、元数据层次和信息单元层次。

从本质上讲,“网络文献”仍然属于“文献”的范畴,这决定了网络文献与传统文献有相当大的相似性,因此,传统文献计量研究方法、原理能够十分有效地“借用”过来。我们从其产生历史也可以看出,网络信息计量学正是起源于对“网络文献”所进行的计量研究,而目前活跃在这一领域的研究者大多都是传统图书情报领域的文献计量学家,也就不足为奇。

参考文献

- 1,12 张洋,邱均平.网络信息计量学的兴起及其哲学思考.情报杂志,2005,24(1)
 - 2 Boundaries A M, Signist B, Alevizos P D. Webometrics and the Self-Organization of the European Information Society. Draft report on task 2. 1 of the SOEIS project, Rome meeting, June 17 ~ 19, 1999
 - 3 李长忠.网络计量学理论与实证研究.图书情报工作,2001(10)
 - 4 赵启民.网络计量学研究的构建和发展方向.吉林师范大学学报(自然科学版),2003(2)
 - 5 李长忠,李东洋,齐源.网络计量学的研究对象与方法.情报科学,2002(1)
 - 6 夏旭,李健康,葛驰.网络计量学研究:现状、问题与发展.图书馆论坛,2001(12)
 - 7 Björneborn L. Small-World Link Structures across an Academic Web Space: A Library and Information Science Approach. Ph. D. thesis from the Department of Informarion Studies, Royal School of Library and Information Science, Denmark, 2004
 - 8,14 Thelwall M, VauKhan L, Björneborn L. Webometrics. Annual Review of Information Science and Technology, 2005 39:81 - 135. by the American Society for Information Science and Technology. Information Today Inc. p82
 - 9,23 邱均平.信息计量学(一).信息计量学的兴起和发展.情报理论与实践,2000(1)
 - 10 邱均平,张洋.网络信息计量学综述.高校图书馆工作,2005(1)
 - 11 蒋国华.迎接科学计量学应用的新时代——第二届科研绩效定量评价国际学术会议暨第六次全国科学计量学与情报计量学年会.科学学研究,2001(6)
 - 13 徐久龄,刘春茂,刘亚轩.网络计量学的研究.见:张力治.情报学进展(第3卷).北京:航空工业出版社,1999:78~105
 - 15 牟华,周秀霞.网络信息计量学概论及应用探讨.情报资料工作,2002(6)
 - 16 钟义信.信息科学原理.北京:北京邮电大学出版社,2002:67
 - 17 马费成,李纲,查先进.信息资源管理.武汉:武汉大学出版社,2000:13
 - 18 阎建国,周立锦.网络文献的收集与利用.河北科技图苑,2005,15(1)
 - 19 姜仁珍.网络文献信息结构分析.现代情报,2004(4)
 - 20 黄晓斌.网络文献的知识发现研究.武汉大学博士论文,2002年4月,p39~40
 - 21 方卿.基于网络的科学信息交流载体整合与过程重构研究.武汉大学博士论文,2001年3月,p23.
 - 22 马费成,胡翠华,陈亮.信息管理学基础.武汉:武汉大学出版社,2002:8
- 赵蓉英 武汉大学信息资源研究中心.通讯地址:武汉.邮编430072.
- 张洋 中山大学资讯管理系.通讯地址:广州.邮编510275.
- 邱均平 武汉大学信息资源研究中心教授.通讯地址:武汉.邮编430072.
- (来稿时间:2007-02-25)