网络信息资源的计量与评价

WANGLUO XINXI ZIYUAN DE JILIANG YU PINGJIA

庞景安 著

科学技术文献出版社

WANGLUO XINXI ZIYUAN DE JILIANG YU PINGJIA

网络信息资源的计量与评价

WANGLUO XINXI ZIYUAN DE JILIANG YU PINGJIA

ISBN 978-7-5023-5759-7

9 787502 357597 >

定价：28.00元

2007

中信所重要科研项目预研基金资助项目 YY-200510

网络信息资源的计量与评价

庞景安 著

中国科学技术信息研究所

图书在版编目(CIP)数据

网络信息资源的计量与评价/庞景安著. -北京:科学技术文献出版社,2007.9
ISBN 978-7-5023-5759-7

Ⅰ.网… Ⅱ.庞… Ⅲ.①计算机网络-信息管理:资源管理-计量 ②计算机网络-信息管理:资源管理-评价 Ⅳ.TP393.07 G354.4

中国版本图书馆 CIP 数据核字(2007)第 134996 号

出　　版　　者　科学技术文献出版社
地　　　　址　北京市复兴路 15 号(中央电视台西侧)/100038
图书编务部电话　(010)51501739
图书发行部电话　(010)51501720,(010)68514035(传真)
邮 购 部 电 话　(010)51501729
网　　　　址　http://www.stdph.com
E-mail:stdph@istic.ac.cn
策 划 编 辑　周国臻
责 任 编 辑　杨　光
责 任 校 对　唐　炜
责 任 出 版　王杰馨
发　　行　　者　科学技术文献出版社发行　全国各地新华书店经销
印　　刷　　者　北京正豪彩色印刷有限责任公司
版 (印) 次　2007 年 9 月第 1 版第 1 次印刷
开　　　　本　787×960　16 开
字　　　　数　291 千
印　　　　张　16.25
印　　　　数　1～2000 册
定　　　　价　28.00 元

前　言

网络信息资源是一种新型的数字化资源和社会化信息。它利用超文本链接，构成立体网状文献信息链，把不同国家、不同地区、各种服务器、各种网站网页、各种不同信息资源通过结点链接起来，形成多渠道、互动交流的“非出版的数字化信息”。与传统信息资源相比，网络信息资源无论在数量类型、分布结构、传播范围、载体形态、交流机制、查寻手段等方面都存在显著的差异。

Web信息资源出现于20世纪80年代，是Internet上最重要的信息组织形式，构成网络信息资源的主体部分。Web采用超文本（hypertext）和超媒体（hypermedia）技术，以及超文本传输协议HTTP，集网上文字、图像、声音、动画等多媒体信息于一体，通过可视化界面向用户提供网络信息。利用Web浏览器，用户可以快速地浏览，查找和获取遍布全球所需的Web信息资源。同时，还可以轻松地访问Usenet、FTP、Gopher、WAIS等许多其他类型的网络信息资源。

随着数字化、网络化时代的到来，网络信息交流日益增加，虚拟图书馆与网络期刊成为科研工作者获取文献与信息的主要途径。这使得网络信息资

源的计量和评价变得越来越重要，传统的文献计量学与科学计量学的理论、方法、指标已不再适用于网络信息资源的测度与分析。正如著名信息计量学家 Egghe 教授指出：新的 Internet 虚拟世界对信息计量学分析提出了挑战，并且开辟了一个全新的时代。对于网络信息资源的开发利用，以及计量评价，已经成为图书情报部门与信息管理领域重要的研究方向和任务。

国际上从 1995 年开始重视并提出网络计量学的概念。最早的研究被认为是 Woodruff 等人关于网络文献特征（例如 HTML 文档大小与数目）的测度研究。1997 年 Alminel 和 Ingwersen 首次提出网络计量学（Webometrics）的术语，得到国际学术界的广泛认可和关注，并由此引发对网络计量学研究的热潮。2000 年以后，学术界开始认真思考和构建网络计量学的学科体系和理论框架。尽管目前国内外的研究尚缺乏整体性和系统性，但在很多方面都已取得了突破和进展。随着信息资源电子化、网络化程度的不断提高，以及电子文献信息资料统计分析技术的进一步发展，网络计量学将得到更快的发展，并对人类社会、经济、科技和文化等各个领域产生深刻的影响。

运用数字化信息技术的处理手段以及现代数学的理论和方法，面向用户的不同需求，对于各种类型的网络信息资源进行定量化的管理、利用、计量与评价，已经成为现代信息社会的重要标志，在未来的社会、经济、科技管理中发挥越来越大的作用。其内容涵盖文献计量学、科学计量学、信息计量学、网络计量学，以及信息组织管理、知识挖掘发现、决策支持系统等众多重要学科领域，构成与情报检索理论方法并驾齐驱的、现代情报学中最重要的两个分支学科，具有非常重要的理论研究意义和广阔的实际应用前景。

目前，国外学者对这一新兴研究领域非常重视，进行了广泛、深入地研

究，并已取得很好的结果。例如：网络信息资源采集、计量方法与技术研究；网络信息资源链接分析理论与方法研究；网络信息资源自动监测与评价研究；知识组织系统的集成及服务体系研究；网络信息资源计算机自动分类与标引研究；网络信息资源数据挖掘与知识发现理论与方法研究；元数据在网络信息资源管理评价中的应用研究；网络信息资源语义分析与概念检索研究；以及基于海量信息资源的分析技术与评价方法研究等。

我国开展这方面的研究还比较少，也不够深入，急需加强这方面理论基础与应用方法的研究，培养这方面高水平的自主创新型人才，大力开展该领域的基础理论与技术方法的研究，尽快接近国际的研究前沿。同时，建立健全我国科技信息资源的管理应用系统与计量评价体系，推动我国图书文献、情报学的学科发展，促进我国科技管理水平的不断提高。

为了顺应这一社会发展的需要，满足读者对数字化、网络化信息资源日益强烈的需求，中国科学技术信息研究所于2005年10月将“网络信息资源计量与评价研究”设立为重要科研项目预研基金资助项目。本书正是该研究项目的重要研究成果，主要针对网络信息资源的计量与评价的基础理论与技术方法进行了比较全面和深入的研究和梳理，并在此基础上，提出网络信息资源计量与评价研究发展的方向、趋势，以及关键问题。希望读者在阅读本书之后，能够对该学科领域的理论方法、学术发展，以及前沿研究有一个梗概的了解和认识。同时，能够循着前人研究的途径和轨迹，寻找确立自己的研究方向和课题，将有关网络信息资源计量与评价的研究不断开拓和持续发展下去。

如前所述，关于网络信息资源计量与评价的研究属于新兴的研究领域和

方向，对于许多理论、方法、概念、术语的认识与理解尚存在一定的分歧和争论；对于不少问题的研究也远非完善，还有待进一步深化和发展。由于著者水平和时间所限，书中一定存在不少的疏漏和错误，希望读者在阅读过程中不吝赐教，批评指正。

(京)新登字 130 号

内容简介

本书对网络信息资源计量与评价的基础理论与技术方法进行了比较全面和深入的研究和探讨，并在此基础上，提出网络信息资源计量与评价研究发展的方向、趋势，以及关键问题。具体内容包括：网络信息资源概述、Web 搜索引擎及其功能分析、网络环境用户查寻行为研究、网络信息数据的采集与计量、网络链接分析的理论与方法、超链接网络分析及其应用、Web 挖掘与知识发现、网络信息资源计算机自动处理工具、网络信息资源的元数据组织与管理、网络信息资源的评价方法及其应用等。

本书内容新颖，研究前沿，资料翔实，注重应用，可供各级科技管理部门、科研领导机构、图书情报部门，以及科学计量学、网络计量学研究者和大专院校师生学习参考。

目　录

第一章
网络信息资源概述

当前，Internet 正以前所未有的速度飞速发展，形成世界上覆盖面最大、用户数最多、内容最为丰富的互联网络。Internet 的影响已经渗透到社会、经济的各个方面，成为人类生活不可缺少的信息环境和网络文化。

网络信息资源是一种新型的数字化资源和社会化信息。它利用超文本链接，构成立体网状文献信息链，把不同国家、不同地区、各种服务器、各种网站网页、各种不同信息资源通过结点链接起来，形成多渠道、互动交流的“非出版的数字化信息”。与传统信息资源相比，网络信息资源无论在数量类型、分布结构、传播范围、载体形态、交流机制、查寻手段等方面都存在显著的差异。网络信息资源的开发利用，以及计量评价，已经成为图书情报部门与信息管理领域重要的研究课题。

1.1 Internet 及其信息资源

1.1.1 Internet 的由来与发展

Internet 是通过标准通信方式将世界各地的计算机网络连接起来，形成互联的网络体系。Internet 起源于美国国防部高级研究计划署在 20 世纪 60 年代建立的军用实验通信网 ARPANET(阿帕网)。最初建立了四个结点的连接，利用分组交换技术将斯坦福大

学、加州大学洛杉矶分校、加州大学圣芭芭拉分校和犹他大学连接起来。至 70 年代，ARPANET 已经连接了近 50 个网点，它们之间可以发送电子邮件(E-mail)、进行文件传输(FTP)和模拟远程终端使用远程计算机系统的信息资源(Telnet)。美国国防部为了使卫星通信网和无线分组通信网也能加入到 ARPANET 中，研制出了 TCP/IP 协议，使得原本独立并构的网络连接成为一个计算机网络的集合。

20 世纪 80 年代初，美国国防部按照是否包含军事内容将 ARPANET 分解为 MILNET 和 ARPANET Internet 两部分。80 年代中后期，美国国家科学基金会(NSF)将位于新泽西州、加州、伊利诺斯州、纽约州、密西根州和克罗拉多州的 6 个超级计算机中心连接起来建立了 NSFNET，并通过它将美国上百所大学、科研机构、政府部门的计算机网络连接起来，成为 Internet 的新主干。它对外全面开放，提供网络化信息环境，全世界各类计算机网络纷纷连入 Internet，形成世界范围的 Internet。

经过 30 年的发展，Internet 极大地促进了信息技术的革命。作为一种全新的数字化传播网络，突破了传统的媒体范围，具有海量信息、分布开放、媒体种类齐全、安全性好、实时性好等特点，对于社会、经济、科学教育、医疗卫生等方面产生深刻的影响，形成一种独特的计算机网络文化。

1.1.2 Internet 信息资源

Internet 信息资源是指以数字形式记录、以多媒体形式表达、存储在网络计算机磁介质、光介质，以及各类通信介质上的信息集合。Internet 上分布着丰富的信息资源，包括：为国家管理部门服务的政府信息资源，如各国政府发布的政策文件、法律法规、新闻报道、机构介绍、政府档案、统计信息，以及其他各种公开的官方信息；为社会公众服务的公共信息资源，如数据库联机信息服务系统、图书馆联机馆藏、科技信息资源、医疗信息资源、电子出版物(网络版报纸、期刊、图书)、电子公告、会议文献、广播电视、艺术作品、网上论坛等；为生产和消费提供的商用信息资源，如产品展示、服务介绍、商情广告、技术咨询、联机订购等。

Internet 信息资源主要以下列形式分布和交流。

(1)Web 信息资源

WWW 是 World Wide Web 的简称，也称为 Web，出现在 20 世纪 80 年代后期。Web 信息资源采用超文本(hypertext)和超媒体(hypermedia)技术，以及超文本传输协议 HT-