

Wangluo

Xinxi Zuzhi

网络信息组织

■ 陈耀盛 主编

科学技术文献出版社

网络信息组织

陈耀盛 主 编

科学技术文献出版社

Scientific and Technical Documents Publishing House

北 京

图书在版编目(CIP)数据

网络信息组织/陈耀盛主编.-北京:科学技术文献出版社,2004.7

ISBN 7-5023-4685-6

I. 网… II. 陈… III. 计算机网络-信息管理 IV. ①G203 ②TP393

中国版本图书馆 CIP 数据核字(2004)第 055441 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)68514027,(010)68537104(传真)

图书发行部电话 (010)68514035(传真),(010)68514009

邮 购 部 电 话 (010)68515381,(010)58882952

网 址 <http://www.stdph.com>

E-mail: stdph@istic.ac.cn

策 划 编 辑 科 文

责 任 编 辑 周 玲

责 任 校 对 唐 炜

责 任 出 版 王芳妮

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 富华印刷包装有限公司

版 (印) 次 2004 年 7 月第 1 版第 1 次印刷

开 本 787×1092 16 开

字 数 500 千

印 张 21.75

印 数 1~3000 册

定 价 38.00 元

© 版权所有 违法必究

购买本社图书,凡字迹不清、缺页、倒页、脱页者,本社发行部负责调换。

(京)新登字 130 号

内 容 简 介

本书是一本关于“网络信息组织”的图书,内容涉及网络信息组织的方方面面,包括网络信息组织体系、网络信息描述报道(目录)、网络信息序化控制(分类法、主题法)、网络信息浓缩增值(文摘法)、网络信息揭示开发(索引法)、网络信息数据库组织法、网络信息聚焦重组(译报评述)、网站信息组织等内容。适合广大网络服务业、网络信息加工处理、信息管理学、情报学、图书馆学、档案学专业人士和学生使用。

科学技术文献出版社是国家科学技术部系统唯一——
家中央级综合性科技出版机构,我们所有的努力都是为
了使您增长知识和才干。

前 言

20 世纪 90 年代以来,由于因特网的商业化,其发展速度以摩尔定律的速度爆炸般膨胀,网络信息量剧增,内容鱼龙混杂,流动杂乱无序,形式多样,变化频繁,结构复杂,信息污染严重,人们特定需要的适用信息、有价值信息、高质量信息则难以寻找。美国学者奈斯比特提出:“失去控制和组织的信息在信息社会里不再构成资源,相反,它成为信息工作的敌人。”因此,控制网络信息流向,把网络信息组织成为有序而优化的信息系统,是当前网络信息服务的主要业务之一,是解决庞大的网络信息混乱无序增长与用户有序个性化需求矛盾的主要措施。全面系统地研究网络信息组织,建立网络信息组织较为全面系统的体系,是网络信息服务业研究开发的必然趋势。

事物的结构与功能往往是对立统一的,网络信息组织方式方法体系的结构决定其功能,而其功能往往又是预先设定的,因而往往又要求其以相应的结构予以保证。网络信息组织方式方法体系有其功能的设定,它由层次要素依一定方式构成。我们认为,为了用户选择利用网络信息的方便,对庞大的网络信息资源,必须更系统、多层次地进行组织,应从网络信息组织的原理功能、方法体系、层次角度去认识理解。

为此,我们将承接的广东省教育厅人文社会科学研究项目《网络信息组织》(批准号:专项项目 01SJD870001)撰写成专著。本书综合运用有关原理方法,在调研的基础上,在著者们多年教学科研的基础上,系统地研究网络信息组织体系,即外部环境——网络,工作对象——网络信息,服务对象——用户及其需求,它既包括网络信息组织的技术、功能方法,也包括了网络信息组织的原理、原则和相关理

论。本书以第一章概述网络信息组织体系,以第二章至第九章分述各层次网络信息组织的功能方法:网络信息描述报道、网络信息序化控制、网络信息浓缩增值,网络信息揭示开发、网络信息聚焦重组、网络信息数据库组织法和网站信息组织。并以大量的实例阐述其原理和应用方法,使之既有较高的理论价值,又具有较好的实践应用价值。全书由陈耀盛同志主持和制定大纲,负责全书的统稿、定稿工作,编撰第一、二、三、四、八、九章;李惠珍同志编撰第五、六章;研究生丁宇同志编撰第七章第三节,研究生杨鹤林同志编撰第九章二、三节,廖番、张素芳两位老师曾参与部分章节初稿的编撰,研究生但碧霞同志提出了宝贵的修改意见。本书在编撰过程中,利用了不少同行的研究成果,在此表示衷心的感谢!本书的出版得到华南师范大学经济与管理学院的大力资助,得到学院领导的大力支持,责任编辑周玲同志为编辑本书付出了辛勤劳动,在此一并表示衷心的感谢!

网络信息组织体系的提出,在国内外尚是首次,由于收集资料、时间和研究水平有限,书中不妥之处,在所难免。我们诚恳地希望专家、学者、广大读者给予批评指教。

编著者

2004.3

图书在版编目(CIP)数据
 科学技术文献出版社. 北京: 科学出版社, 2004. 7
 ISBN 7-203-4682-6
 I. 科... II. 张... III. 计算机-信息-管理 IV. ①G203 ②TP393
 中国版本图书馆CIP数据核字(2004)第02441号
 科学技术文献出版社
 SCIENTIFIC AND TECHNICAL DOCUMENTS PUBLISHING HOUSE
 100044 Beijing, P.R. China Tel: 6848 4848 Fax: 6848 4849



科学技术文献出版社方位示意图

突必求道 育福殊福 ©
 对图黄页第行式排本, 告更通, 页图, 页端, 都不板半凡, 非图排本英图

目 录

第一章 网络信息组织体系	(1)
第一节 外部环境与工作对象:网络与网络信息资源	(3)
第二节 服务对象:网络用户及其需求	(11)
第三节 网络信息组织的原理体系	(15)
第四节 网络信息组织的功能方法体系	(26)
第二章 网络信息描述报道:目录	(35)
第一节 目录及其结构	(37)
第二节 元数据	(42)
第三节 网络信息描述工作	(54)
第三章 网络信息序化控制(一):分类法	(75)
第一节 载体信息分类	(77)
第二节 信息载体分类工作	(80)
第三节 体系分类法的宏观结构	(85)
第四节 体系分类表的微观结构	(94)
第五节 电子版《中图法》的特点与体系结构	(108)
第六节 四大门户网站分类引擎体系	(134)
第七节 机读数据分类标引	(151)
第四章 网络信息序化控制(二):主题法	(177)
第一节 主题法与叙词语言	(179)
第二节 叙词表的结构	(183)
第三节 机读数据的主题标引	(192)

第四节 关键词语言·····	(204)
第五章 网络信息浓缩增值:文摘法 ·····	(213)
第一节 文摘及其类型·····	(215)
第二节 文摘的结构与文摘要素·····	(218)
第三节 文摘的编写·····	(222)
第四节 自动文摘·····	(230)
第六章 网络信息揭示开发:索引法 ·····	(233)
第一节 索引和索引结构·····	(235)
第二节 索引的编制·····	(240)
第七章 网络信息数据库组织法·····	(249)
第一节 书目数据库·····	(251)
第二节 全文数据库·····	(253)
第三节 专题数据库·····	(255)
第八章 网络信息聚焦重组:译报评述 ·····	(271)
第一节 译报类文章的编写·····	(273)
第二节 评述类文章的编写·····	(277)
第九章 网站信息组织·····	(287)
第一节 网站与网络信息资源评价·····	(289)
第二节 网站策划·····	(302)
第三节 网页设计·····	(309)
附录 文摘编写规则·····	(331)
参考文献·····	(337)

第一章

网络信息组织体系

人类认识世界首先是从信息编码开始的。信息编码就是将信息符号依一定规则(称为码规)编排成信息符号系列以表达信息内容的活动。信息编码实质上是人们对外界感性信息进行组织,并把感性信息上升为理性信息的思维活动。

信息编码的过程,是对感性材料进行“组织”的过程。人在最初的感觉中所获得的对象信息,往往是零乱的、无序的。而在编码活动中,人用语言符号系统将它们“组织”起来,通过语言符号系统的有序性来再现出对象信息所表征的对象或事件之间结构的有序性。将个别、分散、随机的感觉结果组织成为和对象具有同构关系的结构,而只有这种经过系统符号组织起来的感性信息,才能输入大脑的“加工厂”进行制作,最后形成理性的认识。

认知心理学和创造学认为:个体信息知识的组合是智慧,信息知识的重新组合就是创造。信息组织是个体信息组织的群体化——社会分工的产物。据此,我们认为:网络信息组织应包括信息的序化和重组优化。

网络信息组织是将分散无序的信息,通过描述、揭示和报道信息特征,再现、重组网络信息联系,规范控制网络信息流向以使用户有效利用的过程。

网络信息组织体系离不开外部环境——网络,工作对象——网络信息,服务对象——用户及其需求。

第一节 外部环境与工作对象:网络与网络信息资源

一、中国互联网络发展的宏观概况

20世纪90年代以来,因特网以摩尔定律的速度爆炸般膨胀。我国的因特网也在高速发展。根据中国互联网络信息中心1997.10—2004.1的《中国Internet发展状况统计报告》的数据(<http://www.cnnic.net.cn/develst/report.shtml>),将中国互联网络发展的宏观概况整理如下,见表1.1.1。

表 1.1.1 中国 Internet 发展状况统计

统计日期	上网计算机	上网用户	CN下注册的域名	WWW站点数	国际线路的总容量
1997.10.31	29.9万台	62万	4 066	1 500个	25.408 Mbps
1998.6.30	54.2万台	117.5万	9 415	3 700个	84.64 Mbps
1998.12.31	74.7万台	210万	18 396	5 300个	143 M 256 K
1999.6.30	146万台	400万	29 045	9 906个	241 M
1999.12.31	350万台	890万	48 695	15 153个	351 M
2000.6.30	650万台	1 690万	99 734	27 289个	1 234 M
2000.12.31	892万台	2 250万	122 099	265 405个	2 799 M
2001.6.30	1 002万台	2 650万	128 362	242 739个	3 257 M

续表

统计日期	上网计算机	上网用户	CN 下注册的域名	WWW 站点数	国际线路的总容量
2001.12.31	1 254 万台	3 370 万	127 319	277 100 个	6 032 M
2002.7.22	1 613 万台	4 580 万	126 146	293 213 个	10 576.5 M
2002.12.31	2 083 万台	5 910 万	17.9 万	37.1 万	9 380 M
2003.6.30	2 572 万台	6 800 万	250 651	473 900 个	18 599 M
2003.12.31	3 089 万台	7 950 万	34 万	59.6 万	27 216 M

注:网站是指有独立域名的 Web 站点,其中包括 CN 和通用顶级域名下的 Web 站点。此处的独立域名指的是每个域名最多只对应一个网站“WWW. + 域名”,如:对域名 sina.com.cn 来说,它只有一个网站 www.sina.com.cn,并非它有 dailynews.sina.com.cn,mail.sina.com.cn 等多个网站。

1. 上网计算机数比较

截至 2003 年 12 月 31 日,我国上网计算机总数已达 3 089 万台,与半年前的第十二次调查结果相比,增加了 517 万台,增长率为 20.1%,和去年同期相比,增长率为 48.3%,增长了 1 006 万台。从报告的数据来看,我国的计算机产业依旧处于高速发展的阶段。

2. 上网用户人数比较

截至 2003 年 12 月 31 日,我国的互联网用户已经达到了 7 950 万,仍居世界第二位。和 2003 年 7 月发布的数据相比,半年内我国上网用户人数增加了 1 150 万,增长率为 16.9%;和 2002 年年底相比增加了 2 040 万人,增长率为 34.5%。数字显示我国互联网事业稳步上升,进入了快速发展的新阶段。虽然从 1998 年 7 月到 2000 年 7 月,上网用户人数每半年都以超过 50% 的比率增长,2000 年 1 月达到顶峰后开始下降,而且从 2002 年 1 月开始,又逐步回升,但是上网用户人数增长率总的发展趋势减缓。这可能和随着互联网的发展,网民的基数增大有一定的关系。不能忽略的是 4 580 万网民在我国将近 13 亿的总人口中仅占 3.6%,说明我国互联网的普及程度还很低,有非常大的发展空间。

3. CN 下注册的域名比较

截至 2003 年 12 月 31 日,我国 CN 下注册的域名数为 340 040 个,与半年前相比增加 89 389 个,增长率为 35.7%,与去年同期相比增长了 89.4%,同 1997 年 10 月第一次调查结果相比,域名总数已是当初 4 066 个的 83.6 倍。

整体而言,CN 下注册域名数有着较快的增长。CN 下注册域名数的这种变化趋势一方面和目前互联网产业的大气候好转有一定的关系,另一方面也反映了国家主管部门制定的新的域名管理政策(CN 顶级域名开放等)对 CN 域名注册服务的巨大推动作用。从 CN 下注册域名的地域分布可以看出,华北、华东、华南的 CN 下注册域名比例为 76.4%;而东北、西南、西北的 CN 下注册域名比例同以往调查结果相比虽略有增长,但所占比例仍非常小。这在一定程度上反映了我国地区之间的信息化水平的差距。

4. WWW 站点数比较

截至 2003 年 12 月 31 日,我国 WWW 站点数为 595 550 个,半年内增加了 121 650 个,增长率为 25.7%,和去年同期相比增长 60.3%。在 WWW 网站数的地域分布上,华北、华东、华南占了 85% 的比例,东北、西南、西北有所增加,我国互联网地区发展的差距还明显地存在。在互联网服务业务方面,网络教育、网上银行、在线交易、网络广告、网络新闻、网上视频服务、收费邮件服务、IP 电话、短信服务、网上人才服务、网络资讯服务、网络游戏等服务业务快速发展。从我国 WWW 站点数的变化趋势可以看出,尽管在 2001 年上半年,网站数下降了 2.3 万,但是从 2001 年下半年到现在,网站数又开始了增长。WWW 站点数近期的变化趋势在一定程度上说明了我国互联网产业在经历了一个低潮后,正呈现出进一步发展的迹象。

5. 我国国际出口带宽的总容量比较

截至 2003 年 12 月 31 日,我国国际出口带宽的总容量为 27 216 M,与半年前相比增加了 8 617 M,增长率为 46.3%,和去年同期相比增加了 190.1%,是 1997 年第一次调查结果 25.408 M 的 1 071 倍,可见我国国际出口带宽目前增长非常迅速,为互联网的发展提供了资源保障。

6. IP 地址数比较

截至 2003 年 12 月 31 日,中国大陆 IPv4 地址数已达 41 456 128 个,与半年前相比增加 9 371 648 个,增长率为 29.2%,和去年同期相比增长 42.9%。中国的 IP 地址资源近几年增长较快,在数量上达到了一定的规模,但是目前仍不能完全满足中国互联网络运营单位发展的需要,随着我国网民人数的大幅增加,网络应用的逐步加强,这种 IP 地址与我国互联网整体发展的不匹配现象会更加明显。

二、网站

1. 网站的一般类型

目前互联网上网站五花八门,包罗万象,但是我们仔细分析其内容主旨,不难发现它们其实可以分为生活、学习、服务和商业四大类型。

(1) 生活类网站

主要以新闻、娱乐、文艺和体育网站居多。网络用户进入这些站点的主要目的是休闲和个人爱好,大家在这类网站里浏览自己感兴趣的新闻和消息。

(2) 学习类网站

一般是指交流比较科学的知识,传授实用技术为主的网站,常见的有科学技术、文史知识、

计算机软硬件、学习交流等内容的网站。

(3) 服务类网站

以国家企事业单位网和政府电子政务网最常见,这类网站的目的就是为广大用户排忧解难,履行自身的社会职责。

(4) 商业类网站

主要是指各个企业在网上建立的网站,这种网站主要进行电子商务以及对自身产品、企业和发展规划的详细介绍。

2. 各行业网站分布状况

据中国互联网信息中心统计:截至 2003 年 12 月 31 日,全国网站数约 595 550 个。企业网站数的比例最大,占整个网站总体的 70.9%;其次为商业网站,占 8.2%;第三是个人网站,占 6.5%;随后依次为教育科研机构网站占 5.1%,其他非盈利机构网站占 5.0%,政府网站占 3.2%,其他类型占 1.1%。企业网站的行业分布中,制造业比例最高,为 35.3%;其次是 IT 业,占 18.5%;第三是批发零售业,占 6.8%;第四是建筑业,占 3.5%;第五是咨询服务业,占 3.1%。

全国网页总数为 311 864 590 个,其中静态网页数 226 725 557 个,动态网页数 85 139 033 个,静动态网页数比例为 2.66:1,平均每个网站的网页数 523.7 个。全国网页总字节数为 6 059 431 526 KB,每个网页平均字节数为 19.43 KB,平均每个网站的网页字节数为 10 174.51 KB。

截至 2004 年 2 月,全国在线数据库的总量为 169 867 个。在所有在线数据库中,企业网站拥有的在线数据库数量最多,占全部在线数据库的 54.3%;其次是商业网站拥有的在线数据库,占全部在线数据库的 14.1%;其他非赢利机构网站拥有的在线数据库排第三位,占全部在线数据库的 9.3%。在拥有在线数据库的网站中,56.7%的网站拥有产品信息数据库,16.5%的网站拥有报刊新闻数据库,14.2%的网站拥有企业名录数据库,14.0%的网站拥有科技信息数据库,8.2%的网站拥有政策法规数据库,8.1%的网站拥有数据库,2.2%的网站拥有金融股票信息数据库。

3. 网站提供的主要服务

(1) 政府网站提供的主要服务

政府网站所提供的主要信息服务包括:政府新闻、政府职能/业务介绍、统计数据/资料查询、法律法规/政策文件、办事指南/说明、办公/业务咨询、通知/公告、办事进程状态查询、企业/行业经济信息、便民生活/住行信息、表格下载等。此外还对信息是否面向对象分类、是否具有信息搜索功能、是否提供网站链接以及语种等进行调查。调查结果显示:关于信息服务,绝大多数政府网站提供政府职能/业务介绍(90.3%)、政府新闻(85.9%),大多数的政府网站提

供办事指南/说明(68%)、通知/公告(67.9%)、便民生活/住行信息(56.4%)、企业/行业经济信息(53.6%)。

(2) 商业网站提供的主要服务

商业网站提供的主要信息服务包括:新闻、在线数据库服务、网上社区、网上购物(B2C 电子商务)、免费电子信箱、搜索引擎、收费电子信箱、免费主页空间、软件下载、收费主页空间、网上教育、B2B 电子商务、在线聊天室、在线游戏、邮件订阅、网上酒店预订、网上股票交易、网上订票、拍卖/集体议价等。调查显示,新闻和在线数据库服务是所有商业网站中提供最多的服务,在所调查的所有商业网站中,有 46.8% 的网站提供新闻服务,有 40.1% 的网站提供在线数据库服务。第 3~10 的服务种类分别是网上社区、网上购物(B2C 电子商务)、免费电子信箱、搜索引擎、收费电子信箱、免费主页空间、软件下载、收费主页空间。其他服务中比例最高的为除新闻外其他各种信息服务(如产品信息、商贸信息、科技信息、生活信息等),另外还有域名注册、技术咨询、网上短信、在线视听等服务。

在各种信息服务中,提供产品信息的商业网站比例最高,为 68.5%;其次是提供新闻的网站,为 46.8%;第三是提供商贸信息的网站,为 45.3%;第四、第五是提供生活财经信息和教育信息的网站,均为 30.1%。

(3) 企业网站提供的主要服务

企业网站所提供的主要信息服务包括:企业介绍、产品/服务介绍、行业新闻、企业动态、售后服务/技术支持、在线招聘信息、用户在线咨询/投诉、网上采购招标、网上销售(针对最终用户,即 B2C)、网上销售(针对代理商、经销商,即 B2B)等。绝大部分企业网站提供企业介绍(98.3%)和产品/服务介绍(95.6%),超过半数的企业网站提供企业动态(56.6%),接近半数的企业网站提供在线招聘信息,40% 左右的企业网站提供用户在线咨询/投诉(40.4%)和行业新闻(38.8%);而发生网上交易行为的网站比例很小,仅有 8.1% 的企业网站发生网上采购招标行为,7.5% 的企业网站发生 B2C 行为,7.3% 的企业网站发生 B2B 行为。

企业网站的行业分布中,制造业比例最高,为 58.8%;其次是邮电通信业,占 8.38%;第三是公关、咨询、广告和市场研究等服务业,占 6.58%;第四是零售批发贸易,占 6.12%;第五是社会服务业,占 5.84%。

三、网络信息资源

中国互联网信息中心将中国互联网络信息资源定义为:中国互联网络上公开发布的网页和在线数据库的总和。

20 世纪 90 年代以来,由于商业动作大量进入因特网,有人估计,因特网上每 24 小时信息流量达万亿比特,每天发布 14 万件新信息。据中国互联网信息中心统计(<http://www.cnnic.net.cn/index/OE/00/12/index.htm>)至 2003 年 12 月 31 日全国共有网页数 311 864 590 个,平均每个网站的网页数 523.7 个。全国网页字节数 6 059 431 526 KB 平均每

个网页的字节数 10 174.51 KB。全国在线数据库的总量为 169 867 个。

1. 网络信息资源的类型

对网络信息资源的类型,许多人从不同角度进行不同的划分。如孟广均等主编《国外图书馆学情报学研究进展》(北京图书馆出版社,1999)第 171~172 页所介绍的:“美国联机计算机图书馆(OCLC)的‘因特网资源计划’根据文件类型将网络信息资源划分为:政府出版物、百科全书、电子论坛、书目、索引和摘要、图书馆目录、字词典、公告、导览、宣传资料等几十种”。国际数据集团(IDG)1996 年出版的《万维网指南》依学科领域将网络信息资源划分为:农业、人类学、天文学和商业等 57 个类目。学者董小英在《网络环境下的信息资源管理与信息服务》(中国对外翻译出版公司,2000)从出版发布方式将网络信息资源分为:非正式出版的网络信息、正式出版的电子出版物、半正式出版的网络信息。我们认为可从信息交流的方式、载体、信息源、信息内容和信息存取方式来划分网络信息资源的种类。

(1) 从信息交流的方式分

从信息交流的方式可将网络信息资源分为:

● 正式交流的网络信息

正式交流是通过中介的网络信息流通。它包括网络信息编辑出版发行工作(如各种网上大众传播机构、联机数据库以及各种数据库、网上书店、网络出版社、杂志社、报社等),网络信息检索工作(如各种搜索引擎机构、各种联机检索中心、文摘社、索引社),网络信息工作(如网站、网络评价机构、各类网络信息中心和网络公司、档案馆等)和数字图书馆工作作为中介的信息流通。正式交流的网络信息多是正式出版发行的电子出版物,包括电子图书报刊、各种数据库(包括联机数据库)、电子工具书、专利信息、教育资源、数字图书馆数字资源等。

● 非正式交流的网络信息

非正式交流流通的是原始信息,流通形式是信息生产者与用户直接交流,多是非正式、半正式出版发行的网络信息。如软件与娱乐游戏、网上教育培训信息、文艺活动信息、旅游景点介绍、交通信息、天气预报、股票行情、BBS、广告、网络新闻等动态信息,也有电子邮件、专题讨论组、论坛、电子会议、电子布告新闻、文件报告、统计数据、法规条例等等。

(2) 从载体分

从载体分,有各种指示性信息,各种信息单元,也有各种电子信息载体;有自由文本,也有规范文本;有文本文件,也有软件与娱乐游戏、声像文件;有网上教育培训信息、文艺活动信息、旅游景点介绍、交通信息、天气预报、股票行情、BBS、广告、网络新闻等动态信息,也有电子邮件、专题讨论组、论坛、电子会议、电子布告新闻、文件报告、统计数据、法规条例等等。

(3) 从信息源分

从信息源分,有政府机构、各种学术团体和教育机构、研究机构、公司企业和商业部门、行业协会、国际组织,也有社会团体、大学以及个人。可从域名知道,通过网址或网页可以查到。