

网络信息资源重组理论与实践

郭振安 主编

兵器工业出版社

PDG

网络信息资源重组理论与实践

郭振安 主编

兵器工业出版社

图书在版编目 (CIP) 数据

网络信息资源重组理论与实践/郭振安主编. —北京: 兵器工业出版社, 2004.4

ISBN 7 - 80132 - 251 - 7

I. 网… II. 郭… III. 网络信息资源开发 IV. G25

出版发行: 兵器工业出版社

封面设计: 王同建

责任编辑: 张小洁

责任印刷: 王京华

社 址: 100089 北京市海淀区车道沟 10 号 开 本: 850 × 1168 1/32

印 刷: 军械工程学院印刷厂印装 印 张: 9.8125

版 次: 2004 年 4 月第 1 版第 1 次印刷 字 数: 235 千字

印 数: 1—2000 册 定 价: 26.00 元

(版权所有 翻印必究 印装有误 负责调换)

内 容 简 介

本书以网络信息资源开发利用为主线，结合近几年来作者从事因特网信息资源重组研究和实践的知识与经验，介绍了网络信息资源重组的基础理论和网络信息资源的分析、采集、加工标引；结合具体事例，介绍了网络信息的分类体系、网络信息资源的管理与发布；推荐了部分国内外军事、学术网站网址。

本书作为国内系统介绍网络信息资源重组的专著，内容全面、新颖，适用于广大从事网络信息资源开发与利用的人员，也可文献情报单位的管理人员提供参

主 编 郭振安

副主编（按姓氏笔划排序）

尹振东 冯 锐 车建国 邢 峰 李 雷

宋英超 张静平 周 力 侯东陵 胡 晖

前 言

因特网上信息资源浩如烟海、十分丰富,但其海量的、未经系统组织和加工的、无序的、自生自灭的零散信息给人们获取带来很大困难。人们迫切需要经过分析、处理、重组后的有序化信息,需要新型的信息服务体系。所以,网络信息资源重组是开发、利用因特网信息资源的有效途径,也是知识创新的一个重要途径;网络信息资源重组,是文献情报单位数字化信息资源体系的重要组成部分,是其文献信息资源共享的重要措施,也是文献情报单位开展信息集成服务的一个新方向。开展这方面的研究具有重要的现实意义。

机械工程学院图书馆在深入研究了网上信息资源组织特点的情况下,认真分析了传统数据库和信息资源建设中存在的问题,总结了近年来全文数据库建设的经验,开展了网络信息资源重组的研究与实践工作,探索出了网络环境下快速组织信息资源的新路子。《网络信息资源重组理论与实践》一书是该研究成果的重要体现,通过网络信息资源重组在文献情报单位建设中的作用,系统地对网络信息资源重组的理论、方法、实施过程进行了有益的探讨。

本书分七章,第一、二章,主要研究了网络信息资源重组在文献情报单位建设中的地位和作用、网络信息资源重组的指导思想、原则、建设内容和平台以及服务模式等重要问题,是网络信息资源重组的理论性研究。第三至第七章,主要论述了网络信息资源分析、采集、回采后信息的加工标引、信息资源体系结构的建立、信息资源的管理与发布等问题,是网络信息资源重组的实践性成果。

本书在论述网络信息资源重组的理论与实践过程中，力求体现“创新”思想、“信息集成服务”理念；力求使“技术问题业务化，复杂问题简单化”，实现快速组织数字化信息资源。

“创新”思想是网络信息资源开发、利用过程的灵魂。网络信息资源重组的提出，本身就是“创新”思想的体现，是以信息集成服务为理念的进一步深化。书中体现了以技术创新为手段，全新构建网络环境下多功能信息资源重组平台；以内容组织创新为主线，系统建立网络环境下特色化的信息资源分类体系；以服务内容创新为根本，适当确立网络环境下多元化的服务模式等一系列带有创新意识的思想。

“信息集成服务”是信息服务工作在当前网络信息无序的状态下，提供有效服务所应持有的基本思路。“信息集成服务”的核心，是以信息资源作为大系统，采取技术手段，进行集成整合，实行“一站到位”式的服务，从而实现信息资源共享。书中体现了信息集成服务的理念，全面探讨了网络环境下信息资源建设与利用的新模式：即以网络资源为主的信息获取模式，以特色数据库为主的资源建设模式，以网络支持为主的成果发布服务模式。

快速组织数字化信息资源是文献情报单位面对的一个重要问题，在网络信息资源重组的过程中也是如此。书中体现了信息资源组织由传统的建库模式向快速资源组织模式跨越的思想：即将多格式全文文献的存储管理与信息资源的快速组织融为一体、将信息资源目录体系的自动建立与个性化的信息服务融为一体、将信息资源的管理与发布系统紧集成融为一体、将全文检索与数据库技术融为一体，使之成为满足网络信息资源重组需求的、动态发展的数字化信息资源系统。

本书的大纲和全书的统稿以及各章节的修改，均由郭振安负责；胡晖、侯京陵、张静平参加了统稿、校稿工作；胡晖负责了

书稿的编辑、排版工作。

各章节的编写具体分工如下：

第一章由侯京陵、郭振安编写，第二章由郭振安编写，第三章由周力编写，第四章由邢峰编写，第五章由胡晖编写，第六章由冯锐、郭振安编写，第七章由牟建国编写，附录由张静平编写。

本书在编写过程中参阅和利用了国内同行的有关文献资料、研究成果，在此表示衷心的感谢。

由于编著者理论实践水平有限，疏漏和不当之处在所难免，诚望军内外同行学者和广大读者不吝指教。

郭振安

2003年9月30日

目 录

第一章 概 论	(1)
第一节 网络信息资源利用概述	(1)
一、因特网环境下的用户信息需求分析	(1)
二、因特网资源利用中的问题分析	(3)
三、文献情报单位在网络信息资源开发利用中的作用分析	(8)
第二节 网络信息资源重组的作用与意义	(9)
一、网络信息资源重组的作用	(9)
二、网络信息资源重组的意义	(12)
三、网络信息资源重组的指导思想	(15)
第二章 网络信息资源重组的基础理论	(17)
第一节 网络信息资源重组概述	(17)
一、网络信息资源重组的相关概念	(17)
二、网络信息资源重组的理论基础	(20)
第二节 网络信息资源重组的基本问题	(26)
一、网络信息资源重组的基本原则	(26)
二、网络信息资源重组的基本方式	(29)
三、网络信息资源重组的基本方法	(34)
第三节 网络信息资源重组的实施	(41)
一、网络信息资源重组的基本模式	(41)
二、网络信息资源重组的工作内容	(44)
三、网络信息资源重组的工作流程	(47)
四、网络信息资源重组的工作平台	(49)
第三章 网络信息资源分析	(53)
第一节 网络信息资源概述	(53)
一、网络信息资源的特点	(53)

二、网络信息资源的分类	(54)
三、网络信息资源的分布	(59)
第二节 网络学术信息资源	(64)
一、网络学术信息资源的类型	(64)
二、国外主要学术信息资源	(73)
三、国内主要学术信息资源	(84)
第三节 网络军事信息资源	(90)
一、网络军事信息资源的概念	(90)
二、网络军事信息资源的特点	(91)
三、国外网络军事信息资源的分布	(92)
四、美国 AD、PB、NASA、DOE 报告	(94)
第四章 网络信息资源采集	(98)
第一节 网络信息资源检索	(98)
一、网络信息检索工具概述	(98)
二、网络信息检索技术	(102)
三、常用网络信息检索工具	(108)
第二节 网络信息资源下载	(119)
一、网络信息资源下载概述	(119)
二、常用下载软件简介	(122)
第五章 网络信息加工标引	(142)
第一节 网络信息加工	(142)
一、网络信息加工概述	(142)
二、网络信息加工方法	(145)
三、PDF 文档加工方法	(154)
第二节 网络信息标引	(166)
一、网络信息标引概述	(166)
二、网络信息标引方法	(170)

第六章 网络信息分类体系	(178)
第一节 网络信息分类体系概述	(178)
一、网络信息分类体系建立原则.....	(178)
二、网络信息分类体系结构设计要点.....	(180)
三、网络信息分类体系结构类型.....	(182)
第二节 网络信息分类体系结构典型应用	(184)
一、装备相关信息资源体系结构.....	(184)
二、军事相关信息资源体系结构.....	(197)
三、科学教育相关信息资源体系结构.....	(200)
第七章 网络信息资源的管理与发布	(206)
第一节 网络数据库	(206)
一、网络数据库概述.....	(206)
二、网络数据库技术.....	(212)
三、网络数据库体系结构.....	(219)
第二节 网络数据库发布	(222)
一、WWW 服务中的 B/S 模式及其实现.....	(222)
二、网络数据库 Web 发布方法.....	(223)
三、网络数据库 Web 发布技术.....	(225)
四、发布数据的安全保护机制.....	(236)
五、网络数据库 Web 发布系统的远程维护.....	(237)
第三节 网络数据库管理与发布系统的典型应用	(238)
一、iBASE Index System 文档索引系统简介.....	(238)
二、信息资源的管理与发布实施过程.....	(242)
附录 相关网站	(267)
一、世界军事网站.....	(267)
二、学术网站.....	(287)
主要参考文献	(296)

第一章 概 论

第一节 网络信息资源利用概述

因特网(Internet)起源于美国,其前身是始建于1969年的美国国防部资助建成的Arpanet网络。Arpanet网络实现了信息的远程传送和广域分布式处理,并比较好地解决了异地网络互联的技术问题,为因特网的诞生和以后的发展奠定了基础。随后不久推出的TCP/IP协议,扫清了计算机互联的主要技术障碍,从根本上解决了不同类型计算机系统之间通信的问题。此后,网络进入了一个大发展时期。今天,因特网已成为世界上覆盖面最广、规模最大、信息资源最丰富的计算机网络,已成为人们走向世界、了解世界、与世界沟通的重要窗口。

一、因特网环境下的用户信息需求分析

(一)信息资源需求的全方位与综合化

用户信息资源需求的全方位与综合化主要是指用户对信息资源内容的综合和信息需求的全方位发展。这两方面的需求集中体现在21世纪知识经济时代用户愈来愈需要为其提供内容全面、类型完整、形式多样、来源广泛的网络信息,而这些信息又必须与用户的工作、学习和科研等各方面相适应,其形式和内容随着工作需求的变化而变化,因而需求又是全方位的。

(二)信息资源需求的开放化与社会化

用户信息需求开放化的直接原因是社会信息化发展中信息活

动贯穿于工作、学习和科研等过程的始终，其各个环节必然存在基本的信息需求。随着工作、学习和科研等社会交往范围的扩大，信息交流日益广泛，致使用户从面向部门的信息需求向面向社会的信息服务模式转变，最终导致传统服务走向开放化的网络信息服务。用户信息需求的开放化从客观上提出了信息资源的社会化共享，从而形成了网络信息资源的开放化、社会化发展趋势。

(三) 信息资源需求的专业化与网络化

用户信息资源需求的专业化是随着知识经济发展中知识密集型行业的增多和社会职业活动的知识化，各类用户的信息需求范围不断扩大、内容不断深化、专业化程度不断加强而体现的。用户信息资源需求的网络化则是由于简单的电子化信息处理已不能满足人们获取知识信息的需要，尤其是系统地获取专业有效知识信息的需要。因而，网络化信息服务开始崛起与发展，使用户信息需求结构由传统型为主向专业化与网络化方向发展。这一趋势反映在信息服务需求上便是信息的专业化组织和迅速实现网络化服务的要求。

(四) 信息资源需求的集成化与高效化

用户信息资源需求的集成化是随着科技进步及其对经济发展推动作用的增强，尤其是从事高科技领域研究与开发的用户已不再满足于文献情报单位为其提供一般服务而发展的，要求通过网络信息资源共享将分散在本领域及其相关领域的专门知识信息加以集中组织，进行文献信息的内涵知识二次开发和知识重组，从中提炼出对用户的研究、开发与管理创新思路形成的“知识基因”，供其使用。

用户对信息需求的高效化主要表现在：工作节奏的加快从客观上要求迅速满足工作中的信息需求；信息处理和利用状态的优

化要求有快速、高效的信息服务作保障；网络信息组织与传递方式的变化使用户逐步适应利用新技术处理信息和进行信息交流与利用的环境，从而进一步激发了用户对高效化信息服务的需求。以因特网为基础的电子信息载体技术的不断进步，为信息需求的高效化提供了信息高速组织与传递的条件，而信息服务费用的降低则进一步创造了高效化信息需求得以更加充分满足的经济条件，由此构成了网络信息需求的新机制。

二、因特网信息资源利用中的问题分析

因特网上丰富的信息资源为文献情报单位开展信息服务提供了极为有利的条件，但必须清楚地认识到，网上信息并不是完备有序的，其利用现状仍存在很多问题，主要有下述诸方面。

（一）基于信息源方面的问题

1. 网络信息资源缺乏有效的组织

在网络环境下，信息资源多以数据库、超文本、多媒体的形式存在，数字化信息资源占据主导地位。为适应网络环境下信息组织的对象多样、种类繁多、成分复杂、传递高速的要求，传统的以文献为单元的信息组织形式逐渐失去了优势，需要创造全新的知识信息组织方法。这种方法不仅仅是对信息特征的简单描述，而是要实现从信息层次到知识层次的根本性转变。

迄今为止，人们还没有找到知识信息的最佳表达方式、信息生产及流通的标准和规范，网络信息资源的流动、加工、更新等没有一定的规律和秩序，达不到网络信息资源开发社会化和管理有序化的要求，使网络信息资源的有效性下降，从而增大了网络信息资源开发利用的难度。与网络信息缺乏有效的组织与控制机制形成强烈反差的是网络信息资源的迅速扩充，各个组织及个人借助信息技术和手段，将大量信息“堆积”到网络上，使网络成

为一个“没有书架和目录的图书馆”。

2. 网络信息资源缺乏深层次的开发

在网络环境下,用户及其信息需求日益个人化和专门化,由选择检索点转向最相关信息;从以文献、信息利用为中心转为以知识利用为中心。虽然因特网上提供了网络信息检索工具,但只能停留在提供浏览功能来查询信息的阶段,尚未深入到对网络信息资源在信息单元、知识单元层面的开发利用。为此,美国图书情报学家们形象而尖锐地指出:“因特网上的信息如果有一英里宽,则只有一英寸深。”因此,对网络信息资源进行宏观组织与微观管理已势在必行。

3. 网络信息“垃圾”缺乏有效的控制

由于因特网的开放性等特点,使得网上信息发布无法控制,信息污染相当严重,出现许多的信息“垃圾”。如网上虚假信息、失真信息、黄色淫秽信息、巫术信息、政治反动信息、暴力犯罪信息和陈旧无用信息等“污染源”大量充斥整个网络。这些信息“垃圾”不仅给网络造成混乱,也给文献情报单位开发利用网络信息资源增加了难度。

另外,由于因特网的开放性和自由性,“电脑病毒”和“黑客”等不安全因素也经常在网上随意流动,相互传染。这不仅威胁各国网络信息系统的安全,同时也给文献情报单位开展网络信息服务造成了一定障碍。

(二) 基于网络检索工具方面的问题

网络检索工具,主要是指因特网上检索各类信息资源的工具。随着因特网的迅速发展以及网上信息资源的快速增长,各种网络检索工具相继出现。但是由于各方面的原因,这些检索工具总的来说存在许多缺点,这给文献情报单位开展网络信息服务造成了一定的障碍。具体包括:

1. 网络检索工具的覆盖范围有限

现在使用的各种检索工具中,即使是功能较完善的搜索引擎,也只能搜索到 Web 上约三分之二的网页,如 Google 搜索引擎可检索 33 亿个网页,Openfind 包含 35 亿个网页,而因特网上可搜索的网页已经超过 50 亿个,可以说,迄今为止,还没有出现一种能覆盖整个因特网信息资源的检索工具。

2. 网络检索工具的功能不够完善

网络检索工具的功能不够完善主要表现为检索点太少,不能从多个方面对检索问题进行限制,只能就某一或某几个关键词进行笼统检索。同时,许多检索工具还不能对多媒体信息进行检索,而因特网上的信息除了文本信息外,还有图形、图表、图像、声音和动画等。这就影响到网络信息检索的质量和水平。

3. 网络检索工具对网络信息资源的检准率较低

如今还没有一种网络检索工具能按照所提供的关键词检出与之密切相关的信息,而对于一个较广义的检索词,往往会检索出大量无关的信息。同时,网络检索工具本身也无法判断所检出信息的真假,更不可能将那些无关紧要的不具备增值可能性的信息过滤掉。这对文献情报单位开展网络信息服务增加了一定的难度。

4. 网络检索工具对信息的组织和标引缺乏规范控制

各种检索工具都有自己的信息收集方式、检索算法和结果排序方式;信息组织没有统一的规格,索引方式也各不相同;大多数检索工具都用自然语言来标引和检索,其结果是同义词和近义词得不到控制,词间相互关系得不到揭示。

5. 网络检索工具检索专业信息的功能较差

目前的检索工具多是通用型的,专指性较差,无法满足从某一学科专业的角度进行信息检索的要求。这也影响到网络信息资源利用的效率。

(三) 基于文献情报单位本身的问题

1. 文献情报单位现代化设备方面的缺陷

要开展网络信息服务,文献情报单位必须具备一定的计算机软硬件系统以及网络环境。然而,目前许多文献情报单位的计算机软硬件系统及相关的软件系统等配置不佳,自动化、网络化和数字化程度不高,网络系统的带宽、速度等也不够理想。这些都对查询、检索网络信息资源产生一定影响。

2. 文献情报单位的“复合”型人才不足

由于因特网起源于美国等西方发达国家,英语又是世界上地理分布最广、传播范围最大的媒体语言,随着计算机技术、多媒体技术和网络技术的发展,英语作为科学技术先导语言的地位相当突出。因特网上的信息涉及各种学科专业知识,且90%以上的信息资源是用英文传播的,而中文信息资源在网上相当匮乏。在整个因特网信息输入与输出量上,我国分别只占万分之一和万分之五。同时,我国信息资源的数字化率尚不到10%。而目前我国文献情报单位的信息服务人员中,既精通英语又擅长计算机和网络操作,同时掌握有关学科专业知识的“复合型”人才还不多,特别是有关新兴学科或交叉学科方面的知识比较缺乏。这必然会影响到网络信息资源的有效开发和利用,严重制约文献情报单位网络信息服务的规模 and 水平。

要开展高层次的网络信息服务,文献情报单位必须拥有一支多层次、全方位的“复合”型人才队伍,他们既要具备一定的外语水平,还要具备相关的计算机和网络知识;同时还必须掌握相关的学科专业知识。要有较高的科学文化水平,能对网络信息进行有效的评价、辨析、浓缩、重组和深加工;能编制专业网络检索工具,为用户提供网络信息服务。而实际上,目前文献情报单位这方面的“复合”型人才比较缺乏,另外许多文献情报单位的