



文献信息资源开发与研究丛书

主编 张昌山

# 网络信息资源检索与利用

张志东 编著



中国社会科学出版社



文献信息资源开发与研究丛书

主编 张昌山

# 网络信息资源检索与利用

张志东 编著



中国社会科学出版社

## 图书在版编目 (CIP) 数据

网络信息资源检索与利用 / 张志东编著. —北京: 中国社会科学出版社,  
2015. 5

ISBN 978 - 7 - 5161 - 5990 - 3

I. ①网… II. ①张… III. ①网络检索—基本知识 IV. ①G354. 4

中国版本图书馆 CIP 数据核字 (2015) 第 081332 号

---

出版人 赵剑英  
责任编辑 孔继萍  
责任校对 张依婧  
责任印制 何 艳

---

出 版 中国社会科学出版社  
社 址 北京鼓楼西大街甲 158 号  
邮 编 100720  
网 址 <http://www.csspw.cn>  
发 行 部 010 - 84083685  
门 市 部 010 - 84029450  
经 销 新华书店及其他书店

---

印刷装订 北京市兴怀印刷厂  
版 次 2015 年 5 月第 1 版  
印 次 2015 年 5 月第 1 次印刷

---

开 本 710 × 1000 1/16  
印 张 15.5  
插 页 2  
字 数 262 千字  
定 价 52.00 元

---

凡购买中国社会科学出版社图书, 如有质量问题请与本社联系调换  
电话: 010 - 84083683

版权所有 侵权必究

# 《文献信息资源开发与研究丛书》

## 编委会

主 编	张昌山				
副主编	郑 文	万永林	华 林	周 铭	
编 委	杨 勇	罗茂斌	陈子丹	杨 毅	
	胡立耘	马自坤			
编 务	侯明昌	陈海玉	甘友庆		

# 总 序

文献是文明的结晶，也是文明的载体。人类创造文献，积累文献，开发利用文献信息资源，不断推动文明的进程。

中国乃人类文明古国。中华文明史也可以说是文献的历史。其文献类型之多样，内容之丰赡，数量之巨大，正所谓“浩如烟海”“汗牛充栋”。而今，由于新技术的广泛使用，信息时代的迅速来临，文献已呈“爆炸”之势；尤其是互联网的出现，使人类社会又进入了一个崭新的时代。

与此同时，我们亦面临诸多的问题与挑战，比如，传统文献怎样才能得到更好的开发利用，现代文献在充分传播利用的同时又怎样才能得以有效传承，人类文献资源如何才能更好地实现共享，而对于按几何级数增长的文献，怎样才能真正管好用好，怎样才能处理好文献的多样性与一体化的关系，等等。这些都需要文献学者和实际工作者进行更广泛、更深入的调查与研究，努力探求并切实遵循“文献之道”，以文献积淀文化，以文献创新文化，进而实现中华文化的伟大复兴。

云南地处祖国的西南边疆，中原文化很早就在这里传播，大量的汉文典籍源源不断地传入并积累，成为云南文化的主流与传统。而其地方、民族与边疆诸特色亦在云南文献中得以彰显。以地方特色而言，编史修志从来都是文化胜业，且成绩斐然。专著、文集不断被创制和保存，民国年间辑刻的《云南丛书》，“初编”“二编”即达二百零五种一千六百三十一卷及不分卷的五十册。其后更有数千种图书文献问世。从民族特色来说，云南民族众多，文化多元，民族典籍文化源远流长，傣族的贝叶文献、彝族的毕摩文献、纳西族的东巴文献、藏族文献、白族文献等，早已产生了国际性影响，是人类共同的文化财富。就边疆特色来看，记载或论述边地、边境、边界、边民、边防及边贸等内容的边疆文献，种类多，价值

高,历来都受重视;尤其是在西部大开发、中国面向西南开放重要“桥头堡”的建设及文化强省的建设中,云南文献的研究与建设被赋予了神圣的文化使命,推进到了新的发展阶段。

云南大学是西南地区建立最早的综合性大学之一。其前身东陆大学奠基于云南贡院故地,承续着悠久深厚的历史文化,而今日云南大学的步伐正向着她的第一个百年迈进。建校之初即已呈现出“西学”“国学”以及“滇学”并重的特质,而文献的研究、整理与开发利用,一直被视为学术的根基,袁嘉谷、方树梅、刘文典、钱穆、顾颉刚、姜亮夫、吴晗、向达、白寿彝、徐嘉瑞、方国瑜、江应樑、谢国桢、李埏等诸多学术泰斗、文献大家都曾执教于斯,他们的学术著述多已成为传世之作,他们的文献学成就早已名垂青史。特别是自20世纪80年代以来,相继建立了档案学、图书馆学、信息资源管理等专业,经过一代代学者的努力,继承传统,开拓创新,文献学科获得长足发展。培养出的学士、硕士和博士,多已成为业务骨干;培育出的学术成果,已显现出自身的优势和特色。更可喜的是,资深学者不断推出自己的力作,学术新秀正在脱颖而出。

为促进文献学科的发展,繁荣学术文化,进一步做好文献工作,在云南大学的大力支持下,我们组织出版这套《文献信息资源开发与研究丛书》。我们的初衷,概而言之就是求真求新、继承创造。求真是科学研究的本义,无论是文献学学术研究,还是文献资源的整理实践,定当以求真为基本原则,否则,绝无科学可言。求新是我们的学术追求,没有独到的见解,没有新意,不足以言学术贡献。继承是学术创造的源泉,创造是学者永恒的使命。这是我们的学术心愿。古人说,取法乎上,仅得其中,虽不能至,心向往之。我们秉承这种精神,朝着这个方向,努力前行。

是为序。

张昌山

2013年9月于云南大学会泽院

# 前 言

在网络化和数字化时代，传统的信息检索方式和手段已经不能满足人们对信息资源的利用需求。如何有效利用计算机和互联网获取各种信息资源，已成为广大师生、科研人员所必须具备的基本技能之一。为帮助读者更好地适应不断涌现的新网络检索工具、掌握新的方法策略、获取新的知识和技能，同时结合作者在信息检索教学工作中的经验而编写了本书。

全书共六章，内容包括网络信息资源检索的有关基础知识、网络搜索引擎和中外文检索工具的利用。其中第一、二章为基本理论，由徐国英老师编写，主要对网络信息资源、网络信息检索、网络信息检索工具进行了概述，讲解了网络信息资源检索的方法、技术与策略；第三章至第六章着重介绍综合性检索工具的使用，包括各种网络搜索引擎、中外文数据库、特种文献数据等，由张志东老师编写。本书收录了一些有代表性和特色的检索工具，描述也是详略各异，有的只介绍了功能、特点，没有进一步展开，有的则详细讲述了检索技巧、检索式的构造，并辅以必要的界面和实例。希望读者能够亲身实践、熟练掌握，并借助这些典型的检索工具，触类旁通、举一反三，从而达到学会使用其他的网络检索工具的目的。

本书在编写过程中参考、借鉴了大量相关的文献资料，得益于许多前辈、专家的优秀著作和教材的启发，在此对参考过的文献资料、著作、教材的作者表示衷心的感谢！

由于作者能力和水平有限，加之对许多新知识、新观点、新内容的掌握、消化吸收还尚有欠缺，所以对于书中存在的疏漏和不足之处，恳请各位读者朋友批评指正。

编 者

2014年4月于昆明

# 目 录

<b>第一章 网络信息资源检索概述</b> .....	(1)
<b>第一节 网络信息资源</b> .....	(1)
一 网络信息资源的定义 .....	(1)
二 网络信息资源的特点 .....	(2)
三 网络信息资源的类型 .....	(3)
四 网络信息资源的评价 .....	(9)
<b>第二节 网络信息检索</b> .....	(11)
一 网络信息检索的定义 .....	(11)
二 网络信息检索的特点 .....	(11)
三 网络信息检索的发展 .....	(13)
<b>第三节 网络信息检索工具</b> .....	(17)
一 网络信息检索工具的定义 .....	(17)
二 网络信息检索工具的构成 .....	(17)
三 网络信息检索工具的运行机理 .....	(18)
四 网络信息检索工具的类型 .....	(18)
五 网络信息检索工具的评价 .....	(20)
<b>第二章 网络信息资源检索的方法技术与策略</b> .....	(22)
<b>第一节 网络信息资源检索的一般方法</b> .....	(22)
一 基于超文本的信息检索 (浏览) .....	(22)
二 基于目录的信息检索 .....	(22)
三 基于搜索引擎的信息检索 .....	(23)



四 基于网络数据库的信息检索 .....	(24)
第二节 网络信息资源检索的基本技术 .....	(24)
一 逻辑控制技术 .....	(24)
二 词形控制技术 .....	(27)
三 位置控制技术 .....	(28)
四 其他技术 .....	(31)
第三节 网络信息资源检索的策略 .....	(34)
一 分析课题 .....	(34)
二 选择检索工具 .....	(34)
三 选择检索途径 .....	(34)
四 确定检索词 .....	(35)
五 构造检索式 .....	(35)
六 调整检索策略 .....	(35)
第三章 网络搜索引擎 .....	(37)
第一节 搜索引擎概述 .....	(37)
一 搜索引擎的定义 .....	(37)
二 搜索引擎的类型 .....	(37)
三 搜索引擎的工作原理 .....	(39)
第二节 综合性独立搜索引擎 .....	(46)
一 Google .....	(47)
二 Yahoo! .....	(51)
三 百度 .....	(54)
四 Bing .....	(60)
第三节 垂直型独立搜索引擎 .....	(62)
一 学科(专业)搜索引擎 .....	(62)
二 购物搜索引擎 .....	(65)
三 博客搜索引擎 .....	(67)
四 电子地图搜索引擎 .....	(68)
第四节 元搜索引擎 .....	(71)
一 Dogpile .....	(71)
二 Vivisimo .....	(74)

三	Ixquick .....	(76)
<b>第四章</b>	<b>中文数据库信息资源检索与利用 .....</b>	<b>(77)</b>
第一节	CNKI 与中国期刊全文数据库 .....	(77)
一	CNKI 概述 .....	(77)
二	中国期刊全文数据库概述 .....	(78)
三	中国期刊全文数据库的检索 .....	(79)
四	检索示例 .....	(83)
第二节	维普资讯系统与中文科技期刊全文数据库 .....	(84)
一	维普资讯系统概述 .....	(84)
二	中文科技期刊全文数据库概述 .....	(84)
三	中文科技期刊全文数据库的检索 .....	(84)
四	检索示例 .....	(91)
第三节	万方数据资源系统 .....	(92)
一	万方数据资源系统概述 .....	(92)
二	万方数据库资源 .....	(92)
三	万方数据库的检索 .....	(94)
四	检索示例 .....	(96)
第四节	中国高等教育文献保障系统 (CALIS) .....	(99)
一	CALIS 概述 .....	(99)
二	CALIS 数据库资源 .....	(101)
三	CALIS 数据库的检索 .....	(106)
四	检索示例 .....	(107)
第五节	联机公共检索目录系统 (OPAC) .....	(107)
一	OPAC 概述 .....	(107)
二	OPAC 的检索 .....	(109)
三	检索示例 .....	(109)
<b>第五章</b>	<b>外文数据库信息资源检索与利用 .....</b>	<b>(111)</b>
第一节	Dialog 联机检索系统 .....	(111)
一	Dialog 系统概述 .....	(111)
二	Dialog 数据库资源 .....	(113)

三	Dialog 检索技术 .....	(115)
四	DialogWeb 系统的检索 .....	(119)
五	检索示例 .....	(121)
第二节	OCLC FirstSearch 联机检索系统 .....	(123)
一	OCLC FirstSearch 系统概述 .....	(123)
二	FirstSearch 数据库资源 .....	(125)
三	FirstSearch 数据库的检索 .....	(127)
四	检索示例 .....	(131)
第三节	ISI Web of Knowledge 数据库 .....	(132)
一	ISI 三大引文索引概述 .....	(132)
二	ISI Web of Knowledge 平台的数据库资源 .....	(133)
三	ISI Web of Science 数据库的检索 .....	(135)
四	检索示例 .....	(143)
第四节	工程索引 (Ei Engineering Village) .....	(145)
一	EI 概述 .....	(145)
二	Ei Village 数据库资源 .....	(145)
三	Ei Engineering Village 的检索 .....	(147)
四	检索示例 .....	(154)
第五节	EBSCOhost 全文数据库 .....	(155)
一	EBSCOhost 数据库资源 .....	(156)
二	EBSCOhost 检索技术 .....	(157)
三	EBSCOhost 数据库的检索 .....	(158)
四	检索示例 .....	(168)
第六节	ProQuest 全文数据库 .....	(169)
一	ProQuest 数据库资源 .....	(170)
二	ProQuest 检索技术 .....	(173)
三	ProQuest 数据库的检索 .....	(174)
四	检索示例 .....	(179)
第七节	Elsevier Science Direct 电子期刊全文数据库 .....	(179)
一	Elsevier Science Direct 数据库资源 .....	(180)
二	Elsevier Science Direct 检索技术 .....	(180)
三	Elsevier Science Direct 数据库的检索 .....	(181)

---

四 检索示例 .....	(186)
<b>第六章 特种文献信息资源检索与利用 .....</b>	<b>(188)</b>
<b>第一节 专利文献 .....</b>	<b>(188)</b>
一 专利文献概述 .....	(188)
二 国内专利文献检索 .....	(189)
三 国外专利文献检索 .....	(192)
<b>第二节 标准文献 .....</b>	<b>(197)</b>
一 标准文献概述 .....	(197)
二 国内标准文献检索 .....	(198)
三 国外标准文献检索 .....	(200)
<b>第三节 学位论文 .....</b>	<b>(205)</b>
一 学位论文概述 .....	(205)
二 国内外学位论文检索 .....	(205)
<b>第四节 会议文献 .....</b>	<b>(209)</b>
一 会议文献概述 .....	(209)
二 会议文献的网站检索 .....	(209)
三 会议文献的数据库检索 .....	(212)
<b>第五节 科技报告 .....</b>	<b>(215)</b>
一 科技报告概述 .....	(215)
二 国内科技报告检索 .....	(217)
三 国外科技报告检索 .....	(218)
<b>第六节 政府出版物 .....</b>	<b>(220)</b>
一 政府出版物概述 .....	(220)
二 国内政府出版物检索 .....	(221)
三 国外政府出版物检索 .....	(222)
<b>参考文献 .....</b>	<b>(225)</b>
<b>后记 .....</b>	<b>(232)</b>

# 第一章

## 网络信息资源检索概述

### 第一节 网络信息资源

随着计算机技术和网络通信技术的发展, Internet 已经发展成为世界上规模最大、资源最丰富的网络互联系统, 为全球范围内快速传递信息提供了有效手段, 也为信息检索提供了广阔的发展平台。与传统的信息资源相比, 网络信息资源作为一种新的资源类型, 既继承了一些传统的信息组织方式, 又在网络技术的支撑下呈现出许多与传统信息资源显著不同的独特之处。因此, 了解信息资源的特点、类型、组织形式等方面的信息, 对有效利用网络信息资源检索工具, 实施网络信息资源检索具有重要的作用。

#### 一 网络信息资源的定义

随着互联网发展进程的加快, 信息资源网络化成为一大潮流, 与传统的信息资源相比, 网络信息资源在数量、结构、分布、传播范围、载体形态和传递手段等方面都显示出新的特点, 这些新特点赋予了网络信息资源新的内涵。

网络信息资源, 指以数字化形式记录的, 以多种媒体形式表达的, 分布式存储在互联网不同主机上的, 并通过计算机网络通信方式进行传递的信息资源集合, 是计算机技术、通信技术、多媒体技术相互融合而形成的, 互联网上可查找、利用到的资源。

## 二 网络信息资源的特点

网络信息资源由于依托 Internet 平台, 与印刷型、联机和光盘电子信息资源相比较, 在数量、结构、分布、控制机制、传递手段等方面存在显著的差异, 呈现出许多新特点。

### (一) 信息数量巨大、内容丰富

Internet 是个开放的信息传播平台, 是一个集各种信息资源为一体的信息资源网, 由于政府、机构、企业、个人都可以在网上发布信息, 这些文献几乎覆盖生产、生活以及其他社会实践产生的各种信息, 内容包括学术信息、商业信息、政府信息、个人信息。因此成为无所不有的庞杂信息资源, 并且有跨地区、分布广、多语种、高度共享的特点。

### (二) 信息类型复杂、形式多样

网络信息资源以数字化的形式存储在全球范围内的不同服务器上, 是多媒体、多类型、多语种、集成式的信息混合体。从信息的类型来看, 有文本、图表、图像以及多媒体信息; 从存在的形式看, 有文件、数据库、超文本和超媒体等, 从而能够满足网络用户的各种信息需求。

### (三) 信息动态性高、不稳定性强

网络信息发布的自由性和随意性使得任何机构、任何人都可能既是信息接受者又是信息提供者。网络信息地址、信息链接、信息内容处于经常性的变动之中, 信息资源的更迭消亡无法预测。网络信息随时都处于增长和变化之中, 新的信息不断出现, 旧的信息或消失, 或失效, 或更改, 使网络信息呈现动态性和不稳定性。

### (四) 信息质量参差不齐、价值差异性大

网络上大部分资源并不像图书和期刊那样经过编辑和出版部门的权威审核, 而且网站信息资源的提供不受任何组织或制度的控制, 导致了网上信息资源质量良莠不齐, 使得大量垃圾信息混于高质量信息中, 增加了有效信息获取的不便, 影响了检索效率。

### （五）信息分散无序、关联度高

网络信息资源以分布式数据库的形式存放在不同国家、不同地区的各种服务器上，导致了信息资源的分布相对无序和分散；但利用超文本链接按知识单元及其关系建立的知识结构网络，又使内容之间具有较高的关联程度，从而通过各种搜索引擎、检索系统使信息检索变得方便快捷。

### （六）信息共享度高、检索便捷

由于信息存储形式及数据结构具有通用性、开放性和标准化的特点，它在网络环境下，时间和空间范围得到了最大程度的延伸和扩展，利用以各类搜索引擎为代表的检索工具进行检索，使用户不需排队等候就可以共享同一份信息资源。

### （七）信息交互性强、平台多样化

Internet 是交互性的，用户不仅可以从中获取信息，也可以向网上发布信息。Internet 为用户提供了方便的讨论、交流渠道，如 E-mail、BBS 论坛、网上问答互动平台等。

## 三 网络信息资源的类型

网络信息资源类型繁多、形式多样，由于划分依据不同，类型也就不同，可以从以下方面划分。

### （一）按照网络信息资源的组织形式划分

#### 1. 文件方式

文件方式采用主题组织法的思想，以文件名标识信息内容，用文件夹组织信息资源，并通过网络共享实现信息传播。文件服务器（FTP）即以此方式组织网络信息资源（多为非结构化信息），用户界面就是人们熟悉的文件夹窗口，浏览和下载信息操作简便。文件方式是网络信息资源组织的简单方式，但随着网络信息资源利用的不断普及和信息量的不断增多，以文件为单位共享和传输信息会使网络负载越来越大。同时文件方式对结构化信息的管理则显得力不从心，文件系统只涉及信息的简单逻辑结构，

当信息结构较为复杂时,就难以实现有效的控制和管理,从而降低了信息组织的效率。因此文件方式只能是网络信息资源管理的辅助形式或者是作为信息单位成为其他信息组织方式的管理对象。

## 2. 数据库方式

所谓数据库组织模式,就是将所有已获得的网络信息资源以固定的记录格式存储,用户通过关键词及其组配查询,就可以找到所需要的信息线索。

利用数据库技术进行网络信息资源的管理具有强大的优势:能高速处理大量结构化的数据,从而大大提高了信息管理的效率;数据库以信息项作为数据的最小单位,可根据用户需求灵活地改变查询结果集的大小,从而大大降低网络数据传输的负载。因此,数据库方式是当前普遍使用的网络信息资源的组织方式。但数据库方式也存在不足之处:对非结构化信息的处理困难较大;不能提供数据信息之间的知识关联;无法有效处理结构日益复杂的信息单元;缺乏直观性和人机交互性。

## 3. 主题树方式

主题树组织模式就是将确定范畴所获得的信息资源按照某种事先确定的概念体系,分门别类地逐层加以组织,用户通过浏览的方式逐层加以选择,层层遍历,直至找到所需的信息线索(即相关站点链接),再通过信息线索连接到相应的网络信息资源。主题树方式能较好地满足用户族性检索的要求,同时它也允许用户根据兴趣,从任意等级类目入手选择不同深度的浏览检索范围。主题树方式具有严密的系统性和良好的扩充性,但它不适合建立大型的综合性网络资源系统,只有在建立专业性或示范性的网络信息资源体系时,才能体现其结构清晰、使用方便的优点。同时,其检索专指性较差。

## 4. 超文本/超媒体方式

超文本、超媒体方式是将网络信息按照相互关系非线性存储在许多的结点上,结点间以链路相连,形成一个可任意连接的、有层次的、复杂的网状结构。超文本方式以线性和静态的文本信息为处理对象,超媒体方式是超文本和多媒体技术的结合,将文字、图表、声音、图像、视频等多媒体信息以超文本方式组织。超文本/超媒体方式体现了信息的层次关系,用户既可以根据链路的指向进行检索,也可以根据自己的需要任意选择链路进行信息的检索。



### 5. 自由文本方式

自由文本方式主要用于全文数据库的信息组织,是针对非结构化的文本信息进行组织和处理的一种方式。基于全文数据库的全文检索可以将任意字符作为检索标识,这样,用户用自然语言即可直接检索未经标引的一次文献。

### 6. 主页方式

主页方式类似于档案全宗的组织方式,它将有关机构或个人的各种信息集中组织在一起,是对某机构或人物等各类对象进行全面介绍,如热门站点和相关站点的推荐。目前单位和个人上网均采用建立主页的方式,详略程度由其自行决定。

### 7. 非结构化数据库组织方式

非结构化数据库可根据半结构化和非结构化数据的特殊性进行存储和管理,在管理机制上超越了传统结构化关系数据库对于信息采取定长和结构化定义的局限,与 Web 技术结合可实现非结构化数据库型信息资源的直接上网发布与全文检索。非结构化数据库突破了关系数据库结构定义不易改变和数据定长的限制,支持重复字段、子字段以及变长字段并实现了对变长数据和重复字段进行处理和数据项的生长存储管理,在处理连续信息(包括全文信息)和非结构信息(重复数据和变长数据)方面,有着传统关系型数据库所无法比拟的优势。

### 8. 镜像站点方式

当一个单位有了自己的 www 服务站点时,对于网上最重要和本单位最常用的信息,可将这些信息移植到自己的服务器上,供本单位使用。这种站点为镜像信息站点,它对信息的更新及时跟踪,保证拷贝过来的信息与源站点一致。这样用户经常使用的信息不需要每次都从远程调用,既节省通信的开销,又可加快信息的传递。

## (二) 按照采用的网络传输协议划分

### 1. WWW 信息资源

WWW 信息资源是指建立在超文本、超媒体技术基础上,集文字、图形、图像、声音等为一体,以网页的形式存在,以直观的图形用户界面展现和提供信息,采用超文本传输协议(Hypertext Transfer Protocol, HTTP)进行传输的一类信息资源形式。www 信息资源是 Internet 信息资源的最主