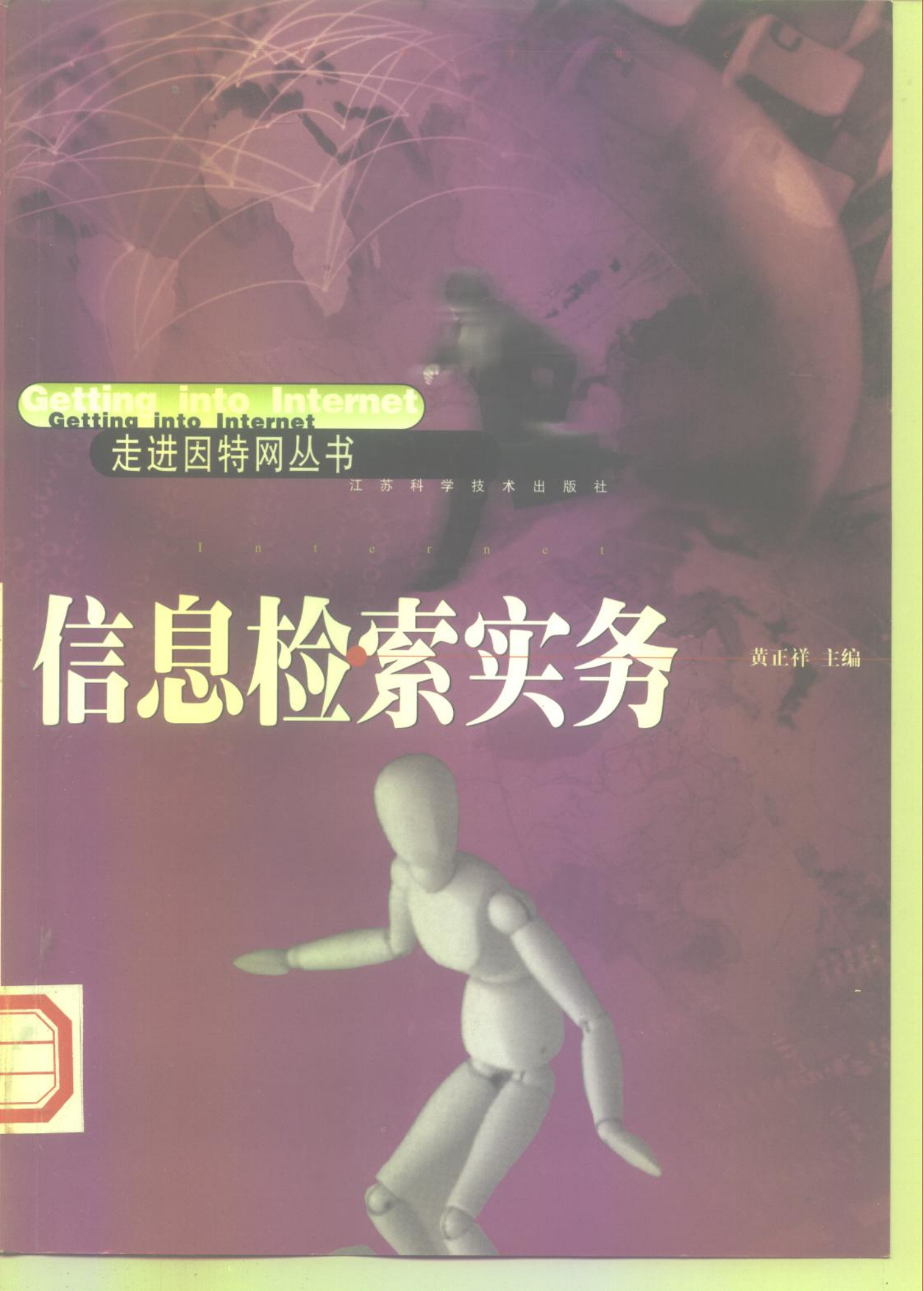




Getting into Internet

Getting into Internet

走进因特网丛书

江苏科学技术出版社

I n t e r n e t

信息检索实务

黄正祥 主编



TP393.4
H950

走进因特网丛书

465113

Getting into **Internet**
信息检索实务

黄正祥 主编

江苏科学技术出版社

图书在版编目(CIP)数据

信息检索实务/黄正祥主编;徐伟弘,官程编. - 南京:江苏科学技术出版社,1999.9

(走进因特网丛书/楼佩煌,黄正祥主编)

ISBN 7-5345-2741-4

I. 信… II. ①黄… ②徐… ③官… III. 因特网-情报检索 IV. TP393.4

中国版本图书馆 CIP 数据核字(1999)第 40746 号

走进因特网丛书

信息检索实务

主 编 黄正祥
责任编辑 宋 平

出版发行 江苏科学技术出版社
(南京市中央路 165 号,邮编: 210009)

经 销 江苏省新华书店
照 排 南京展望照排印刷有限公司
印 刷 通州市印刷总厂

开 本 787mm × 1092mm 1/16
印 张 8
字 数 138 000
版 次 1999 年 9 月第 1 版
印 次 2000 年 4 月第 2 次印刷
印 数 5 001—10 000 册

标准书号 ISBN 7-5345-2741-4/TP·90
定 价 15.00 元

图书如有印装质量问题,可随时向我社出版科调换

《走进因特网》丛书

编 委 会

主 编 楼佩煌 黄正祥 JS232/25
编 委 马跃勇 刘 立 宋 平 陈丹伟
徐伟弘 黄正祥 龚 翔 龚福亚
楼佩煌 戴正农

本书编写人员 黄正祥 徐伟弘 官 程

责任编辑 宋 平

封面及版式设计 赵 清

出版说明

几年前,人们似乎还觉得因特网很遥远,但在今天这个信息时代,它已延伸到地球的每一个角落,从根本上改变了信息的传播方式,正影响着我们的生活、工作、学习和交流环境,人们可以方便地通过因特网实现信息交流、资料查找、网上购物、网上求学、浏览报刊、阅读小说、欣赏音乐等。

为了让读者对因特网世界有个粗线条的了解并快速学会感兴趣的实际应用,我们精心设计了这套《走进因特网》丛书,讲述上网之前的准备工作并快速浏览因特网上的迷人世界,介绍因特网常用软件工具,介绍如何使用电子邮件、因特网电话、网络传真、网络寻呼、网络会议、电子贺卡等省钱省时的因特网通信手段,讲述如何使用 BBS、Web 上的信息交流、Usenet 新闻组、mIRC 实时交谈等多种网络聊天工具参与聊天、“张贴”新闻公告、获得帮助、与他人交流思想、分享数据资源、结识素未谋面的朋友,介绍在网上查找产品、政治、商贸、科技、专利、标准等专业资料的搜索策略和实用要点,讲述如何就读网上知名中学、国内外名牌大学、职业培训学校和课余如何进入网上图书馆、网上书店,告诉您网上订阅电子报刊、浏览电子图书、点播音频视频节目、观看实况转播、参与跨时空的网络游戏等休闲新时尚,另外还介绍了设计制作有特色的网页、在因特网上安“家”的基本方法。

这套图文并茂的丛书内容相互联系,使用彼此独立,为读者提供了有重点、有选择的学习途径。我们衷心期望丛书成为您学习的朋友,引导您顺利进入信息高速公路,奔向灿烂的 21 世纪。

前 言

在今天这个知识经济时代,因特网已经进入到人们学习、工作和日常生活的方方面面,网上信息已成为重要的信息来源。全世界成千上万的计算机通过因特网互联在一起,传播与分享着各种各样的信息资源,这些信息种类繁多,内容广泛,语言多样,更新频繁,常常使遨游于网海中搜寻信息的朋友感到无所适从。如何在这浩瀚的信息海洋中快速而准确地查找到您所需要的信息资料呢?本书特向广大读者介绍这些工具和方法。

本书在概略介绍因特网信息资源基本知识的基础上,描述了国内外著名的英文搜索引擎、简体中文搜索引擎、繁体中文搜索引擎、区域搜索引擎的特色和应用范围,列举详实丰富的例证说明查找各种网上信息的检索策略和使用技巧。其中介绍的多内码转换工具解决了中文多内码系统并存问题,避免了在查找香港、台湾地区的繁体中文资源时可能遇到的麻烦;特种搜索工具介绍了如何在因特网上找软件、找工作、寻访亲友、查询电话号码及电子邮件地址等;专题信息查找为从事科研、信息工作者给出各个专业领域的查找例证,以便读者举一反三,掌握专题查找技巧。

我们衷心希望,当广大朋友漫游在因特网这个信息高速公路上时,本书能够充当路标,帮助大家一矢中的,迅速在网上查找到所需的信息资料,让因特网真正成为每个人的信息宝库。

编 者

1999年6月 南京

目 录

网上查资料必备知识	1
-----------------	---

1 Internet 网上资源	1
------------------------------	---

2 搜索引擎	3
---------------------	---

■ 搜索引擎的基本概念	4
-------------------	---

■ 搜索引擎的分类	4
-----------------	---

■ 搜索引擎的收录范围	7
-------------------	---

■ 搜索引擎的语言版本	9
-------------------	---

3 搜索引擎查询基本方法	10
---------------------------	----

■ 明确搜索目标	10
----------------	----

■ 使用“交”和“差”	11
-------------------	----

■ 使用短语搜索	11
----------------	----

英文搜索引擎	13
--------------	----

1 英文搜索引擎及其搜索技巧	13
-----------------------------	----

■ AltaVista	13
-------------------	----

■ Lycos	18
---------------	----

■ Excite	19
----------------	----

■ WebCrawler	20
--------------------	----

■ Infoseek	22
------------------	----

■ Yahoo!	24
----------------	----

■ 其它英文搜索引擎	25
------------------	----

2 图形、图像的检索	27
-------------------------	----

■ 网路罗盘	27
--------------	----

■ Lycos	28
---------------	----

■ AltaVista	29
-------------------	----

■ Smithsonian	30
---------------------	----

中文搜索引擎.....32

1 简体中文搜索引擎.....32

■ 新浪网.....32

■ Yahoo! 中文.....37

■ 中文 Excite.....38

■ 搜狐.....40

■ 中经搜索.....40

■ 天网.....41

■ 若比邻.....42

■ 软件速递搜索引擎.....44

■ 其它简体中文搜索引擎.....44

2 常用多内码转换工具.....46

■ 使用 IE.....47

■ 使用简繁体转换软件.....51

3 繁体中文搜索引擎.....58

■ 哇塞.....58

■ 悠游.....59

■ 八爪鱼.....60

■ SEEDNet 搜索引擎.....61

■ 其它繁体中文搜索引擎.....63

特种检索工具.....65

1 因特网上找工作.....65

■ 直接登录单位站点.....65

■ 访问专门从事招聘的网站.....66

■ 找工作搜索引擎.....73

2 网上寻人.....73

■ 网上寻人搜索引擎.....74

■ 电话号码搜索器——PhoneFerret.....76

■ 电子邮件地址搜索器——EmailFerret.....81

3 软件搜索利器.....83

■ 软件搜索.....84

■ 处理搜索结果.....86

■ 下载文件.....86

■ 查看 Web 页.....87

■ 处理搜索结果的显示	87
■ 设定搜索引擎配置	88
■ FileFerret 操作技巧	89
4 其它特种搜索引擎	90
■ 多面手搜索引擎——Tile	90
■ 幽默搜寻器——Humor	93
■ 169 访问专用引擎——Top88	94

新闻搜索.....95

1 因特网网络新闻	95
■ BBS 电子公告栏	95
■ Usenet 新闻论坛	95
2 Deja News 新闻搜索.....	96
■ Deja News 的一般搜索	96
■ 阅读新闻和回复新闻	97
■ Deja News 的高级搜索	99
3 用 NewsFerret 搜索新闻.....	101
■ NewsFerret 的下载和安装	101
■ 搜索新闻	103
■ 阅读新闻	104
■ NewsFerret 的软件设置	105

区域性搜索引擎.....106

1 医学信息查询——中国生物医学文献数据库	106
2 科技资料检索——万方数据资源系统.....	107
3 计算机信息资料的查找.....	109
■ 查找程序	109
■ 查找协议和标准	111
4 科技工程资料数据的查找.....	115
5 法律检索——中国法律检索系统.....	117
6 专利检索——中国专利数据库检索系统.....	117
7 商业信息查询——中国大黄页.....	119
■ 快速搜索	119
■ 高级搜索	120

网上查资料必备知识

1 Internet 网上资源

近几年来,因特网在全世界迅速走红,成了报刊、广播电视、街头巷尾议论的话题,这表明我们已经进入了一个崭新的信息时代。因特网这个全世界最大的计算机网络极大地改变了我们与世界的联系,使我们的信息变得丰富多彩,来源更加广泛。那么,因特网究竟能做什么?我们又能获得什么样的信息?下面,就因特网的常用功能作一简单介绍,以便读者对因特网有个大致了解。

● WWW(World Wide Web)环球信息网

WWW 是因特网上的环球信息网,它提供广域超媒体信息服务,可以在同一画面中同时显示文字、图像、声音、动画,这就是我们常听到的超文本(HTML)。WWW 的基本结构是由一个浏览器和提供信息的网上服务器组成,两者分别架设在网络的一端,使用者所下达的指令通过浏览器分析后发往服务器。当服务器接收到使用请求后,就会把储存在服务器上的相关 HTML 文件回传给用户端,然后再由浏览器显示到屏幕上。使用者的浏览器不但可以连接本地网络的 WWW 服务器、而且也可以通过因特网连接至远端的 WWW 服务器,从而实现全球信息的交流。可以说用于浏览 WWW 主页的用户端浏览器是 WWW 世界的窗口,我们通过它就能够浏览多姿多彩的 WWW 世界。

● Gopher

在因特网中,Gopher 给网络用户提供各式各样的信息服务。Gopher 是一个世界性的客户/服务器(Client/Server)方式的信息检索工具,常用作图书馆数据查询、生活信息检索、气象预测等。早期的 Gopher 仅仅设计成文本检索工具,后来逐步发展成为一个多媒体信息检索工具,现在它不但可以检索一般文本,还可以检索格式化的文本、图像、声音和全动态视频图像。通常,Gopher 采用客户/服务器工作方式,即在用户计算机中运行一个客户程序,该程序提供与 Gopher 服务器的连接,而 Gopher 服务器在广泛的主题范围内为许多信息提供菜单形式的人机界面,以方便用户获取所需的相关资料。

● FTP 文件传输

File Transfer Protocol(FTP)是实现在计算机之间传送文件的最古老的方法,它允许用户进入另一个计算机系统,获取该系统中的文件。尽管最早的 FTP 说明书出现在 1961 年 4 月,但在今天,FTP 仍然是因特网的主要工作软件,并且还会在很长的时间内保持其地位。FTP 提供了在两个完全不同的计算机之间传送文件及提供公用文件共享,这两个决定性因素使得它非常有用。由于网上存在无数的匿名(Anonymous)FTP 服务器,这些服务器中有大量共享资源,人们通过访问匿名 FTP 可以得到各自所需的公用软件和其它有用信息,因此,后一个特点最具普遍意义。

● Archie 文件搜索

上面提到的匿名 FTP 服务器是最有用的因特网新资源。我们可以从因特网上得到大量信息和软件,现在的主要问题就是寻找所需要的信息了,而这正是用得着 Archie 的地方。

Archie 服务站点收集了许多资源发掘工具,这些工具为在因特网环境中查找信息提供电子目录服务,使用它们就可以跟踪许多匿名 FTP 档案节点的内容。现在,Archie 服务已经扩展到因特网的许多方面,包括多种其它的联机目录和信息资源清单,其最基本的用途是在因特网中查找那些通过匿名 FTP 可以得到的特殊文件。在 Archie 数据库中有大量常用的匿名 FTP 地址索引,这些地址是可以扩展的。Archie 服务已遍布世界每一个角落,并且从任何地方都可以在因特网中访问它。

● Telnet 远程登录

Telnet 是一个独立程序,可以在本地机上运行它来访问远程计算机。使用 Telnet 远程登录系统,能够使用户进入一个超出个人计算机的新环境,即可以使用因特网登录进入远程计算机(宿主机)中,并在该宿主计算机中运行程序,具有像在本地机上一样的操作形式。

Telnet 程序通常被用来访问因特网上的所有共享类型数据。在远程主机上有许多可利用的信息,通过 Telnet 便可登录进入这些远程系统进行访问以获得相关资料。

● 新闻论坛

新闻论坛是网络上一个没有言论限制的讨论区,其中有各式各样的讨论主题,这些主题涵盖范围之广、参与人数之多、辩论之激烈难以用笔墨形容。只要你感兴趣,保证看也看不完。

● BBS 电子公告栏

BBS 有点像新闻论坛,不过规模较小,大多数是区域性的,提供一个信息交流、经验交换的园地。

- Email 电子邮件

Email 是因特网中用得最多的一种工具, 它让远隔两地的用户高效率、低成本地相互传递电子文本、图像、表格和声音等形式的电子邮件。

- IRC 网络聊天

这是一个国际性的多人聊天系统, 提供成千上万个不同主题的“频道”, 进入感兴趣的频道, 你就可以参与来自世界各地的网迷们的聊天。

- 网络电话

这是最近刚刚兴起的因特网应用之一, 可以极低的费用拨打国际、国内电话。

综合来说, 因特网上资源极其丰富, 取之不尽, 整个网络就像一个前所未有的超级数据库, 包容了从军事、政治、文化、经济、科学、工程、医学等各个方面。现在全球有上百万家 WWW 站点, 构建着一个虚拟的、没有国界的信息世界, 并且以极高的速度迅速膨胀。但是, 要能在众多的网络资源中找到满足自己需要的资源并非易事, 必须先知道网络上的什么地方有哪些资源。以当前 WWW 全球信息网来说, 最常用到的搜索工具有以下几种:

- WWW 网页搜索
- FTP 文件搜索
- Gopher 文章查询
- Usenet 和 BBS 文章查询
- 针对 WWW 里的多媒体对象的搜索等

2 搜索引擎

就因特网的使用而言可以归为两大用途: 一个是信息交流, 如上面提到的 Email、Usenet、BBS 等; 另一个是信息查询。在网上的主机纷繁复杂, 每台主机都产生大量的或有用或无用的信息。如何在浩如烟海的信息空间里快速找到所需的信息, 显得越来越重要起来。

在因特网世界里, 信息的获取是通过对网上主机的访问来实现的。我们这里的问题是如何才能找到含有所需信息的主机。通过对网上每一台主机的逐次访问显然既不经济, 也不现实。因此, 我们需要一种导航工具, 它能引领我们在网海中到达正确的地方。这种工具就是网上搜索引擎!

这里所说的搜索引擎主要指以 Web 站点、网址、网页、文章搜索及综合服务的 Search Engine。随着因特网的急速发展, 许多搜索引擎已开始提供全面的查询服务, 通过这个搜索站点我们即可完成过去几种查询工具才能完成的查

询任务。

目前,网上搜索引擎站点多达几十个,搜索方式、搜索范围均不相同。选择一个好的搜索引擎站点去上网寻宝,可以节省大量网络通道的空间和您宝贵的上网经费与时间。

下面,我们主要介绍搜索引擎如何工作、搜索引擎的类型、搜索引擎的收录范围、搜索引擎的分类以及搜索引擎的查询技巧等内容。

■ 搜索引擎的基本概念

每一个搜索引擎都是一个万维网网站。与普通网站不同的是,搜索引擎网站的主要资源是它的索引数据库。搜索网上信息后,主页显示所得到的满足要求的各种链接资源信息,提供人们寻找所需信息的途径。大多数搜索引擎搜索到的信息资源以万维网资源为主,也有少数提供电子邮件地址、新闻论坛文章、FTP 匿名服务器等站点资源。

要想用好搜索引擎,要先了解搜索引擎的工作原理、分类方式和基本查询规则。

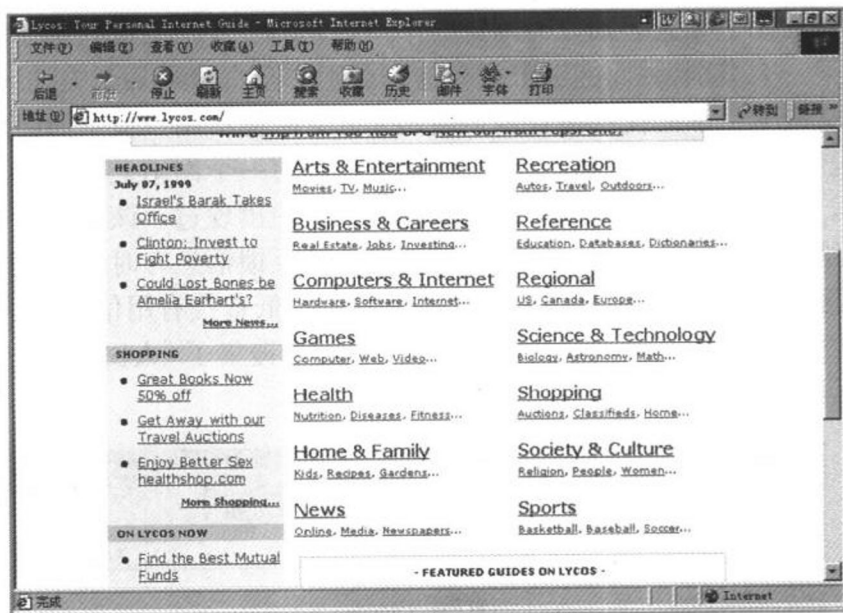
■ 搜索引擎的分类

从组织信息的角度出发,可以把搜索引擎区分为分类目录式和全文式两种:

分类目录式搜索引擎

分类目录式搜索引擎提供按类别编排的因特网网站目录。在各类别下面,就像一本电话号码簿一样,排列着这一类别网站的站名和网址链接,不同的是有些搜索引擎还提供了各个网站的内容简介,往往还伴有网站查询功能。查询网站时,在搜索引擎网站的主页上提供一个文本输入框和一个按钮,在文本框中输入的要查找的字、词或短语称为关键词(Keywords)。输入关键词后单击搜索按钮(按钮名称一般为“查找”、“搜索”、“开始”,或英文“Start”、“Search”、“Submit”、“Find”、“Go”等),搜索引擎便会查找相关的站名、网址和内容提要,将查找到的内容馈送回来。

分类目录式搜索引擎提供的是网站简介,并不将网站上的所有文章和信息都收录进去,仅仅将网站划分成各个分类,再记录一些有关网站的摘要信息(Abstract),用户提出的搜索关键词就在网站的简介中搜索。分类搜索引擎的优点是将信息分门别类,用户可以清晰方便地查找到某一大类信息,符合传统的信息查找方式,尤其适合那些仅希望了解某一方面、某个范围内的信息,并不严格局限于查询关键字的用户。分类式搜索引擎的缺点是它的搜索范围较



全文搜索引擎要小得多。

如图所示的分类式搜索引擎站点是 Lycos，具有代表性的还有 Yahoo! 等。

全文式(网页级)搜索引擎

这类搜索引擎看起来与前一类搜索引擎的网站查询非常相似，虽然也提供一个文本输入框和搜索按钮，使用方法相同，而且有些全文搜索引擎也提供分类目录，但是两者却有本质上的区别。

在分类式搜索引擎的资料库中，搜集保存的是因特网上各个网站的站名、网址和内容提要；而在全文检索搜索引擎的资料库中，搜集保存的则是因特网上各网站的每一个网页的全部内容，涉及范围要大得多。因此，全文检索搜索引擎所查到的结果不仅仅是站名、网址和内容提要，而是与输入的关键词相关的一个个具体网页的地址和该网页第一小段的文字。这段文字可能没有输入的那个关键词，甚至只是一段无法看懂的标记，但在这个网页中一定有所输入的那个关键词，或者是相关的词汇。

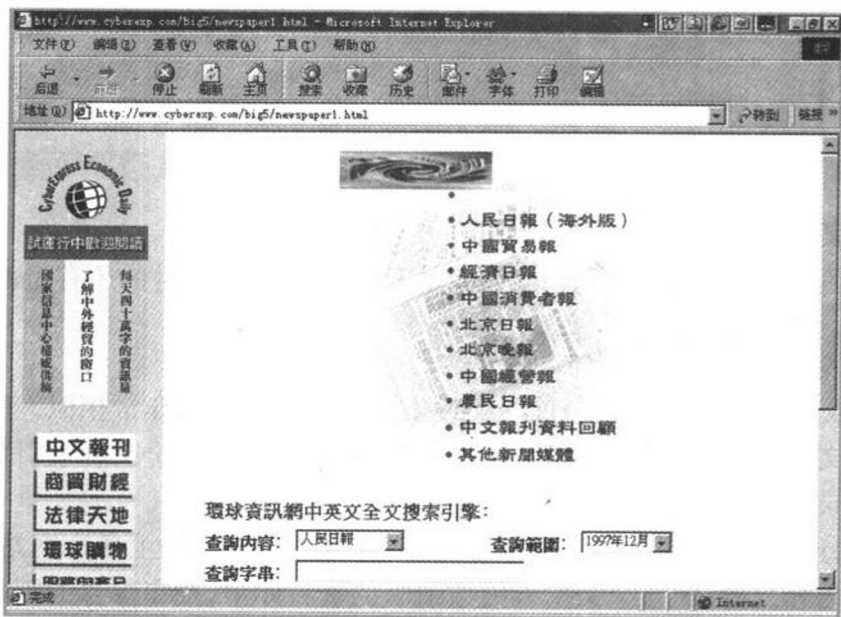
打个形象的比方，分类式搜索引擎可以查到网上有哪些报纸，如《文汇报》、《大公报》，而全文检索则可以查到网上这些报纸上的每一篇文章中的词汇。这两种类型的搜索引擎各有所长，不可替代。

有些全文检索搜索引擎也提供分类目录，但这种目录不是网站的分类目录，而只是网页的分类目录。也就是说，在其各类目录下所排列的不是网站站名、地址，而是大量的属于这一类别的网页地址。由于网页数目非常庞大，所

以这种目录几乎无法起到分类浏览的作用。它的主要功能也只是让使用者能够进入某一大类别,因而能够限定在这一类别中全文检索某个关键词。

全文搜索引擎的优点是查询全面而充分,它能给用户最全面而广泛的搜索信息。全文搜索引擎的数据库会将涉及到的每一个网站上所有网页通通保留下来。当用户查询的关键词在数据库的主页中出现过,该主页就会作为一项搜索结果返回到搜索结果的页面上。广泛的反馈信息有时也会成为全文搜索方式搜索引擎的缺点,因为信息太多反而会降低查找有用信息的精确度。

下图是中文全文检索系统网站主页,该网站搜集了《人民日报》、《中国贸易报》、《经济日报》等中文资料。



该类搜索引擎的代表站点还有: Aitavista, Hotbot, Infoseek, Excite 等。从搜索引擎的搜索对象来看,又可分为独立式搜索引擎和超级式搜索引擎。

☞ 独立式搜索引擎

采用主动或被动方式处理网上站点资源,整理网上信息,登录到自己的网站中用网站自身的数据库向查询用户提供服务,这就是独立式的搜索引擎。许多比较有名气的搜索引擎网站均为独立式搜索引擎,如 Yahoo!, Altavista, Lycos 等。

☞ 超级式搜索引擎

由于任何一个搜索引擎都不可能在网上海罗万象,各个搜索引擎均有其各自的侧重点,各自的搜索方式往往会有一些不同。所以,对同一主题作网上

查询,若只通过一个搜索引擎去完成,难免有挂一漏万之嫌。为此,网上又出现了一种特殊的站点,它虽也提供查询服务,但除了自身数据库的查找以外,主要将用户的查询词递交给网上多个搜索引擎同时进行查询,得到的查询结果会先进行相关度排序,然后再列出。这样所得的信息,由于信息源的扩大,质量较单一引擎的搜索结果要好,查询效率大大提高。要同时处理多个引擎上的信息,带来了处理查询时间长的缺点。

超级搜索引擎分为两类:串行处理引擎和并行处理引擎。超级搜索引擎又被称为多搜索引擎、元搜索引擎、集搜索引擎等,至今尚无统一的称呼,在后续的章节中,我们将作些较为详细的介绍。

■ 搜索引擎的收录范围

就像我们查工具书要先知道这本书是干什么的,是字典、词典,还是成语词典一样,使用一个搜索引擎应该先了解它收录的范围是什么,由此才能知道在这里是否能查到自己需要的东西。搜索引擎的收录范围主要有以下几方面:

● 地域范围

大型搜索引擎的资料库往往面向全球,收录全球因特网的网站或网页,而且在网站中还专门建立一些地域范围的检索。



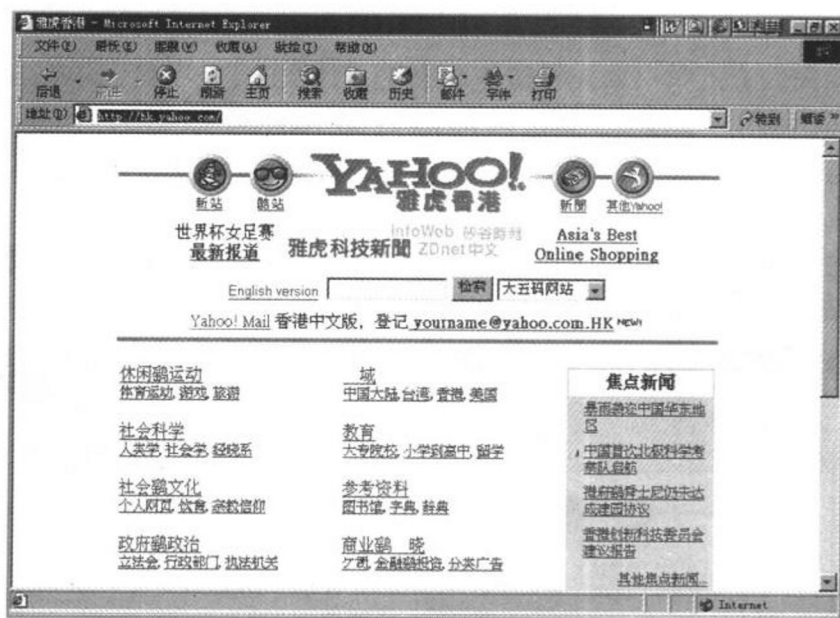
如图, Yahoo! 就是这样的站点,它一方面收录全球的网站,同时专门开辟有收录亚洲地区网站的分站。

一些中小型的搜索引擎则专门收录某一地区的信息,如美洲、欧洲、亚洲地区,甚至小至纽约、东京、香港等城市。全球性搜索引擎收录的范围虽然广泛,但对某一局部而言,却很可能不如区域性搜索引擎内容来得丰富和完备。

● 语言范围

世界上大的搜索引擎都有不同的语言版本。就因特网上的中文网站来说,往往使用不同的汉语内码来编写,常用的有 GB 简体中文(中国大陆地区、新加坡常用)、Big5 繁体中文(港台地区常用),还有 HZ 码、图形方式等。

因此,有些搜索引擎主要收录简体中文网站的信息(特别是中国大陆地区的搜索引擎),有些搜索引擎则主要收录繁体中文的信息(特别是港台地区的搜索引擎);还有些搜索引擎则兼收各种中文网站信息。



在因特网上,还有一些英文网站也发布有关中国和华人信息,因此也有一些专门查询这一类网站住处信息的搜索引擎。

● 网站类型范围

在所有国际互联网的网站中,最常用的是页面浏览、超文本链接的网站,即 WWW 网站。除此之外,也还有其它多种形式的网站,如 Ftp、Gopher、BBS、Usenet 等。

大型搜索引擎往往开辟专门的选项,能够搜索包括 Ftp、Gopher、BBS、Usenet 在内的各种形式的网站信息,但是相当多的搜索引擎却只能搜索 WWW 网站,无法查询其它形式网站的信息。如图是一个 Gopher 网站。