

Internet 搜索引擎指南

陆吉林 杨建芳 编著

把世界缩小到一个屏幕，
将地球掌握在键盘中。

Internet 搜索引擎

将带领你胸有成竹地走进Internet的世界。

复旦大学出版社

图书在版编目(CIP)数据

Internet 搜索引擎指南/陆吉林,杨建芳编著. —上海:
复旦大学出版社,1999. 10
ISBN 7-309-02356-0

I. I… II. ①陆…②杨… 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 39277 号

出版发行 复旦大学出版社

上海市国权路 579 号 200433

86-21-65102941(发行部) 86-21-65642892(编辑部)

fupnet@fudanpress.com <http://www.fudanpress.com>

经销 新华书店上海发行所

印刷 江苏大丰市印刷二厂

开本 787×1092 1/16

印张 7.75

字数 190 千

版次 1999 年 10 月第一版 1999 年 10 月第一次印刷

印数 1—6 000

定价 12.00 元

如有印装质量问题,请向复旦大学出版社发行部调换。

版权所有 侵权必究

内 容 提 要

搜索引擎是用户在网上浏览与查找信息的基本工具。

本书主要内容有:搜索引擎基础知识,介绍搜索引擎的基本类型、简单检索与进阶检索、检索策略与技巧;常用搜索引擎评介,对 18 个著名搜索引擎的主要特色和使用方法作了评论和介绍;搜索引擎分类指南,按 14 个类别列出 1 300 个搜索引擎的网上地址;中文搜索引擎使用指南,介绍中文搜索引擎的现状与特点,并对全球范围的 174 个中文搜索引擎作了简要描述。

本书适合于各种层次的 Internet 爱好者在网上浏览与查找信息时使用。

前 言

搜索引擎作为网络之门,是用户在网上浏览与查找信息的基本工具。Yahoo! 的成功和华裔学生杨致远白手起家到亿万富翁的传奇故事,刺激了众多 Internet 网上淘金者,形形色色的搜索引擎网站层出不穷,给广大网络用户带来了莫大便利。

然而,初次上网的用户,一般只是对简单易用的分类目录索引感兴趣,经常上网的用户,又习惯于挑选一些自己熟悉和经常使用的搜索引擎,这实在辜负了众多搜索引擎设计者的良苦用心,也使得用户自己与很多美好的东西失之交臂。

本书共分 4 章,旨在使读者对精彩纷呈、各树一帜的诸多搜索引擎有一个比较全面、基本的了解。第一章介绍了搜索引擎的基本知识和一些常用的检索策略与技巧;第二章是本书的重点,对 18 个著名搜索引擎的主要特色和使用方法作了评论和介绍,一般读者只要掌握其中几种搜索引擎的使用方法,再通过它们就可以将整个 Internet 展现在自己眼前;第三章按 14 个类别列出了 1 300 个搜索引擎的网上地址,为读者挑选适合于自己使用的搜索引擎提供了便利;第四章针对国内部分上网用户外语水平欠佳的情况,着重介绍了中文搜索引擎的现状与特点,并对全球范围的 174 个中文搜索引擎作了简要描述。

工欲善其事,必先利其器。用户在上网之前,先了解一些有关搜索引擎的基本知识,掌握一点网上查找信息的策略与技巧,熟悉几个最有影响的搜索引擎的使用方法并记住它们的网址,则在上网时就可以事半功倍、满载而归。这也正是笔者所期望的。

本书引导读者如何利用搜索引擎在网上浏览和查找信息,但用户在网上搜索目标的过程漫长而又遥远,展现在用户面前的是一个浩瀚无际的海洋。这其中既蕴藏着无穷的知识宝库,也充斥着大量的信息垃圾,甚至还有插满鲜花的“美丽”陷阱。请不要为这些虚拟的世界所迷惑,也不要中途停留,只要沿着自己最初确定的目标和方向坚定不移地走下去:“蓦然回首,那人却在,灯火阑珊处”。

中山大学数学与计算科学学院的林云寰研究员对本书的策划提出了很好的建议,南京大学图书馆副馆长马金川先生、中国科学技术大学数学图书分馆负责人叶林秀女士、四川大学白苏华研究员在本书的编写过程中提供了许多有益的启示和帮助,复旦大学图书馆金环女士负责全书的资料收集和电脑文字处理,复旦大学出版社孙未未先生从本书的选题、策划到材料的落实等许多方面给予了具体指导和帮助,复旦大学出版社严晴燕小姐作为本书的责任编辑,对本书的审校和出版付出了辛勤劳动,我们谨向为此

书出版作出贡献和给予帮助的所有朋友表示衷心感谢。

由于水平有限,书中难免有不当之处,恭请广大读者和用户不吝赐教,谢谢!

编著者

1999年6月

目 录

第一章 搜索引擎基础知识	1
1.1 搜索引擎概述	1
1.2 搜索引擎基本类型	1
1.3 搜索引擎使用方法	3
1.4 如何选择搜索引擎	8
1.5 检索策略与技巧	9
第二章 常用搜索引擎评介	12
2.1 AltaVista	12
2.2 Excite	15
2.3 HotBot	16
2.4 Infoseek	18
2.5 Lycos	19
2.6 Northern Light	20
2.7 WebCrawler	21
2.8 Yahoo!	22
2.9 蕃薯藤(Yam)	24
2.10 盖世引擎(GAIS)	26
2.11 华页指南	28
2.12 茉莉之窗	29
2.13 搜狐(Sohu)	31
2.14 天网(WebGather)	32
2.15 添达	33
2.16 网擎(Openfind)	34
2.17 悠游(Goyoyo)	38
2.18 哇塞!(Whatsite)	40
第三章 搜索引擎分类指南	42
3.1 一般搜索引擎	42
3.2 多元搜索引擎	44
3.3 教育类搜索引擎	49
3.4 社会、政治、法律类搜索引擎	51
3.5 语言、文化、出版类搜索引擎	55
3.6 广播影视、音乐艺术类搜索引擎	63
3.7 商业、就业指导类搜索引擎	66
3.8 医药、保健类搜索引擎	75

3.9	休闲、娱乐类搜索引擎	79
3.10	自然科学、技术类搜索引擎	81
3.11	计算机、网络类搜索引擎	84
3.12	新闻、气象、交通类搜索引擎	90
3.13	E-mail 地址、个人主页搜索引擎	91
3.14	其他	92
第四章	中文搜索引擎使用指南	94
4.1	汉字编码及其转换	94
4.2	中文搜索引擎的现状与特点	96
4.3	中国大陆搜索引擎一览	98
4.4	香港地区中文搜索引擎一览	103
4.5	台湾地区中文搜索引擎一览	104
4.6	其他中文搜索引擎一览	111
附 录	网上参考资料	114

第一章 搜索引擎基础知识

1.1 搜索引擎概述

Internet 的迅速发展和信息量的急剧增长,导致各种网络信息检索工具应运而生,被称为网络之门(Web Portal)的搜索引擎(Search Engine)就是其中之一。Yahoo! 的成功和华裔学生杨致远的传奇故事,使得形形色色的搜索引擎网站层出不穷,现已发展成为 Internet 的一个新兴产业。

搜索引擎通常由搜索器(又称作网络蜘蛛——Web Spider,爬行者——Crawlers,机器人——Robots 等)和检索程序两部分组成。网络蜘蛛在网络间连续不断地来回交叉搜索,对所有发现的网站和网页进行判断和标记;检索程序则负责对搜索到的信息进行整理、分类和索引,并对用户提出的各种检索要求作出响应。搜索引擎的运行速度一般都非常快,每次检索只要几秒钟即可完成。每一个搜索引擎实际上就是一个庞大的网上信息数据库,而资金与技术则是影响每一个搜索引擎知名度的主要因素。

为了扩大影响、争取用户,许多搜索引擎站点如今已不再单纯地提供网络搜索服务,而是发展成为包括网络搜索、网络新闻、网络游戏、网上聊天、免费电子信箱等诸多内容在内的综合性网络服务媒体。

由于不同的搜索引擎具有不同的检索功能,同一种检索方式或同一个检索主题对不同的搜索引擎通常会得到不同的检索结果,因此,根据不同的检索要求,选择合适的搜索引擎,制定切实可行的检索策略,再辅之以必要的检索技巧,可以使用户以较少的时间去获得较为满意的检索结果,从中领略到 Internet 的浩瀚无际和博大精深。一次好的检索,可以是一次令人心旷神怡的知识海洋的遨游。

1.2 搜索引擎基本类型

目前,Internet 中的搜索引擎主要有三种基本类型:以分类目录为主的搜索引擎,以全文搜索为主的搜索引擎和建立在独立搜索引擎之上的多元搜索引擎(Meta-Search Engine)。其中,前两类搜索引擎的功能是相互借鉴和互为补充的,两者之间的界限也渐趋模糊。

一、以分类目录为主的搜索引擎

这类搜索引擎又称为目录服务(Director Service),著名的 Yahoo! 就是其代表。检

索系统将搜索到的 Internet 中的所有资源按其主题分成若干大类,每一大类下面又分设若干二级类目、三级类目等。有的搜索引擎常细分至十几级类目,每个类目下面列出属于这一类目的网站名称和网址链接(有的搜索引擎为防止本系统资源的白白流失,只提供网站镜像而没有直接链接)。有的搜索引擎还对各个网站进行简要介绍和说明。

以分类目录为主的搜索引擎通常还设有网站查询功能,通过在查询框内输入用户感兴趣的词组,即通常所说的关键词(keywords),即可获得与之相关的网站目录和信息。

以分类目录为主的搜索引擎的特点是由系统先将搜索到的网络信息分别归类。这一工作有的系统由人工操作,有的则由机器自动完成,用户只要遵循该系统的分类体系按图索骥、层层深入即可。这与图书馆中传统的分类索引的功能十分相似,用户在使用时十分方便,是一般用户浏览 Internet 的良好导游。

以分类目录为主的搜索引擎的不足之处是搜索范围通常比以全文检索为主的搜索引擎的范围要小很多,加之各个系统没有统一的分类体系,操作系统和用户对类目的判断和选择将直接影响到检索的效果;而类目之间的交叉,又导致了许多内容的重复;此外,有的搜索引擎分类太细,也使得用户无所适从。

二、以全文检索为主的搜索引擎

这类搜索引擎通常称为索引服务(Indexing Service),它们与分类目录为主的搜索引擎中的网站查询功能看上去十分相似,但却有着本质的区别,著名的 AltaVista 就是这方面的代表。

以全文检索为主的搜索引擎,它们的搜索对象是 Internet 所有网站中每个网页的全部内容。用户得到的检索结果,通常是一个个网页的地址和一段段的文字,这里面也许并没有用户在查询框中输入的词组,但在反馈信息所指明的那一个个网页所包含的内容中,一定有用户输入的词组或与之相关的内容。

有些以全文检索为主的搜索引擎也提供分类目录,但这是网页的分类目录而不是网站的分类目录。由于网页数目巨大,即使类目分得再细,用户也很难从浏览中得到明确的启发和提示,它们的作用更多的是体现在先让用户进入某一个类目,然后限定在此类目中对给定的词组进行全文检索,这样可以有效地压缩搜索范围。

以全文检索为主的搜索引擎的特点是信息量特别大,在理论上用户可以对 Internet 所有网站的每一页内容(即可以对每篇文章中的每个单词)进行检索,当用户想对某一特定的主题进行深入细致的研究时,不妨借助于这类搜索引擎。

以全文检索为主的搜索引擎的不足之处则是每次常规意义下的普通检索,反馈的信息往往太多、太滥,以致用户很难直接从中筛选出自己真正感兴趣的内容,要想达到理想的检索效果,往往要借助于进阶检索(Advanced Search)中必要的语法规则和限制符号,而这一点又恰恰是一般用户最感陌生的。

三、多元搜索引擎

Internet 庞大的信息,使得任何孤立的搜索引擎都无法单独覆盖,而建立在搜索引擎之上的搜索引擎,通常称之为多元搜索引擎。由于它可以调用其他搜索引擎,从而在

一定程度上满足了用户更多、更快地获得网络信息的要求。

当用户向一个多元搜索引擎发出检索要求之后,该引擎会将检索要求迅速发送给其他独立的搜索引擎,并将它们反馈的结果集中进行处理之后提供给用户。多元搜索引擎具有同时查询多个数据库的优势,至于其本身是否拥有数据库则无关紧要。

多元搜索引擎根据其工作方式,又可具体分为串行处理和并行处理两种。串行处理是将检索要求先发送给某一个搜索引擎,然后将其检索结果经处理后,传递给下一个搜索引擎,依次进行下去,直至最终将结果反馈给系统。串行处理的搜索引擎准确率高,但速度慢。并行处理则是将检索要求同时发送给所有要调用的搜索引擎。并行处理的搜索引擎速度快,但重复的内容较多。

日前,比较有影响的多元搜索引擎有 Meta Crawler, Savvy Search, Search Spaniel 等。

四、All-in-One 集成搜索

All-in-One 集成搜索是指用户只要通过一个网站,即可选择多个搜索引擎依次进行搜索。这与多元搜索引擎的工作方式有些相似,只是多元搜索引擎只要一次性输入检索要求,就可以同时让多个搜索引擎一起或分别进行搜索,并对搜索结果进行分析整理;而 All-in-One 通常是逐一输入检索要求(可以相同,也可以不相同),然后从搜索引擎列表中每次选择一个进行检索,一般不对搜索结果进行处理。在很多场合,多元搜索引擎被视为 All-in-One 集成搜索的一部分。

All-in-One 搜索公司的“All-in-One Search Page”可以看做是 All-in-One 集成搜索的代表。该网站汇集了 Internet 中 400 多个大型的搜索引擎、数据库、索引和分类目录,其数据库容量、系统响应速度和用户界面等多项指标在同类产品中均居领先地位。

五、特殊用途的搜索引擎

一般人们所说的搜索引擎,总是指综合性搜索引擎,它们通常涵盖了各个学科的各个方面。但也有些搜索引擎,它们是专为某一特殊用途或某一特定目标而设立的,人们将其称之为特殊用途的搜索引擎,如全球华人寻人搜索引擎——Look 4U,软件搜索引擎——Softseek 和 Download Power Search,图形、图像搜索引擎——Image Surfer,支持自然语言的搜索引擎——Ask Jeeves 等等。

这些专业的搜索引擎,由于其鲜明的特色和对专业的精通,往往更为用户所欢迎。

1.3 搜索引擎使用方法

搜索引擎作为网络之门,是用户上网的必经之路,尽管搜索引擎的种类繁多,但其基本使用方法却很简单。一般搜索引擎通常由分类目录和关键词检索两部分组成,以分类目录为主的搜索引擎几乎都有一个关键词检索框,而以全文检索为主的搜索引擎,除极个别(如 AltaVista)外,也都有一个分类目录索引(有的称做频道,有的称做文件夹)。

一、分类目录的使用

根据有关机构对上网用户的调查分析,大部分用户使用搜索引擎的目的,只是想大致了解一下某个方面的信息,一般并不准备对此作进一步深入、细致的调查研究。这些用户最适合使用分类目录,因为分类目录可以使用户仅仅通过浏览的方式就可以集中地查找到某一类的信息,这与传统的信息查找方法十分相似。用户进入某一个搜索引擎之后,只要按照该系统的分类目录体系,按图索骥,层层深入即可。

雅虎中文(Yahoo!)是以分类目录为主的中文搜索引擎中的佼佼者,搜狐(Sohu)是国内搜索引擎中分类目录做得比较出色的一个。现以这两个搜索引擎作为工具,以了解我国大专院校图书馆上网情况为检索课题,简单介绍一下分类目录的使用。

先从雅虎中文入手,进入该站主页之后,就可以在显著位置看到一个分类目录索引。共有14个一级类目可供选择,其中有一个类目名为“参考资料(图书馆、字典、辞典)”;选中它并进入该类目,有目录、百科全书、图书馆等22个二级子类目可供选择;选择“图书馆”并进入该子类目,这时又有大学图书馆、公共图书馆、国家图书馆等10个三级子类目可供选择;选择“大学图书馆”并进入该子类目后,一共列有73个大专院校图书馆的网址与简要介绍,其中使用简体中文(GB码)的有33个,使用繁体中文(Big5码)的有40个。而使用搜狐,进入主页之后,分类目录索引中共有18个一级类目可供选择,其中有“综合参考(图书馆、地图)”;选中它并进入该类目,有图书馆、辞书参考、地图等12个二级子类目可供选择;选择“图书馆”并进入该子类目,这时又有大学图书馆、公共图书馆等6个三级子类目可供选择;选择“大学图书馆”并进入该子类目后,一共列有123个大专院校图书馆的网址与简要介绍,其中使用简体中文(GB码)的有64个,使用繁体中文(Big5码)的有59个。此时,用户如果对这些大专院校图书馆有兴趣,就可以逐一或有选择地访问它们。

二、关键词检索

关键词检索是搜索引擎的基本组成部分,一个搜索引擎可以没有分类目录,但却很少没有关键词检索,只不过有的搜索引擎关键词检索的对象是网页,有的则是网站而已。

用户进入某一个搜索引擎之后,一般总是在比较显著的位置就可以看到关键词检索框。用户只要在检索框内输入一个检索表达式,然后按一下边上的检索键,检索系统就会开始工作,并很快将检索结果返回。用户输入的检索表达式,可以是一个单词、几个单词、几个单词的组合,甚至可以是一句普通的问话等等。不同的搜索引擎对检索表达式的构成有着不同的要求,同一个检索表达式运用于不同的搜索引擎,返回的检索结果也不尽相同,因此,构造一个目的明确、针对性强,且能被系统识别的检索表达式,是用户使用关键词检索能否取得良好效果的关键。

天网中英文搜索引擎(WebGather)是北京大学计算机系网络研究室设计开发的国内为数不多的支持中英文全文检索的搜索引擎,且只支持关键词检索而不提供分类目录。现将该搜索引擎作为工具,以有关中文搜索引擎的使用为检索课题,简单介绍一下

关键词检索的使用过程。

用户进入天网中英文搜索引擎之后,在它的查询框中输入“搜索引擎”,然后按查询键。系统很快就返回检索结果,共搜索到有关的文档 6 060 篇,并显示前 10 篇文档的内容摘要。由于返回的文档数量太多,故需要缩小检索范围,这时重新在查询框中输入“中文搜索引擎”,则返回的结果显示共找到有关的文档 855 篇。仍然太多,再在查询框中输入“中文搜索引擎使用”,则返回的结果显示共找到有关的文档 19 篇,这时的目标已经相对集中了。如果再改输入“中文搜索引擎使用指南”,则返回的结果显示只找到有关的文档 8 篇,但这 8 篇文档与检索主题的接近程度则是前面的那些文档所无法与之相比的,用户此时就可以逐一访问并阅读这些文档,从而获得有关中文搜索引擎使用的基本知识。

用户在使用搜索引擎的分类目录时,有时也会需要关键词检索的帮助。如果用户要了解的内容在分类目录索引中一下子找不到,就可以通过在检索框中输入相应的关键词予以协助,系统有可能很快就找到相关的类目,也有可能找不到相关的类目而仅仅提供一些系统认为还是比较接近的类目和相关的站点或网页。现仍以雅虎中文和搜狐这两个搜索引擎作为工具,以了解我国大学出版社上网情况为检索课题予以说明。

先进入雅虎中文,由于在分类目录索引中一下子找不到有关“大学出版社”的类目,故只好在关键词检索框中输入“大学出版社”,然后按搜索键开始检索,系统返回的结果显示,雅虎中文的各级类目中均没有以“大学出版社”为标题的类目,只有一个六级子类目“出版”与之最为接近,而与大学出版社相关的站点共有 27 个。剔除重复部分,尚有清华大学出版社、复旦大学出版社、武汉大学出版社、中央广播电视大学出版社等 4 条记录。再使用搜狐,搜狐的分类目录索引中一下子也找不到有关“大学出版社”的类目,同样在关键词检索框中输入“大学出版社”,则系统返回的结果显示,搜狐的各级类目中也没有以“大学出版社”为标题的类目,只有一个四级子类目“出版”与之最为接近,而与大学出版社相关的站点也是 27 个。剔除重复部分,尚有中国科学技术大学出版社、中国地质大学出版社、中国矿业大学出版社、南京大学出版社、南京师范大学出版社、东南大学出版社、河海大学出版社、苏州大学出版社、重庆大学出版社和西南师范大学出版社等 10 条记录。

三、简单检索与进阶检索

关键词检索通常分为简单检索与进阶检索(Advanced Search)两种。简单检索就是在关键词检索框中输入一个或几个关键词,然后递交系统进行检索。简单检索的结果往往不尽如人意,因为如果这些单词的使用频率较高,则反馈的信息就有可能多得让人无法应付;而如果这些单词过于“冷僻”,则反馈的信息又往往令人大失所望。单靠一个或几个单词毕竟很难完整地表达用户的检索要求。

为此,许多搜索引擎就纷纷提供一些不太复杂的新方法和新途径来对用户的检索要求作些小小的限制(如逻辑条件限制、特殊操作符限制等),力图使搜索系统能更好地领会用户的意图。这些方法和技巧通常就称为进阶检索,也有的搜索引擎将其称之为高级检索或复杂检索。

一般而言,简单检索与进阶检索之间并没有十分明确的界限。有的搜索引擎整体水平较高,不但其进阶检索具备了许多复杂的功能,就连其简单检索所包含的一般功能都超过了普通搜索引擎的进阶检索功能。像专门为科技工作者设计的 Northern Light,它的简单检索表达式就允许使用逻辑条件限制。有的搜索引擎的进阶检索的功能则十分简单,如台湾盖世引擎(GAIS)的进阶检索,仅比一般的简单检索多了一个精确匹配(Exact Matching)与模糊匹配(Fuzzy Matching)的选择框而已。

不同的搜索引擎,提供的进阶检索的方法和功能不尽相同,但在很多方面,还是有许多共同之处的,现将一些比较常见的进阶检索的方法简要介绍如下。

1. 逻辑检索

逻辑检索也称布尔检索。常用的逻辑检索有“AND”,“OR”和“NOT”。

AND 表示逻辑“与”,有的搜索引擎也常用“&”,“+”,“,”和空格来表示。AND 用于检索两个以上关键词的情形,检索的结果应该与这几个关键词都有关系,如“经济 AND 文化”,就表示既包括经济又包括文化。

OR 表示逻辑“或”,有的搜索引擎用“|”来表示。检索的结果只要求与若干个关键词中的一个有关系即可,如“体操 OR 游泳”,就表示可以包括体操,也可以包括游泳,使用 OR 通常会得到许多无关紧要的信息,一般应慎重使用。

NOT 表示逻辑“非”,有的搜索引擎用“!”表示。NOT 检索的结果将完全排除与 NOT 后面的关键词有关的信息,如“水果 NOT 苹果”,就表示可以包括水果但其中不能有苹果。

一般能提供进阶检索的搜索引擎,都支持逻辑检索,但有的是“完全支持”,如 Excite,Infoseek 等;有的则是“部分支持”,如 Yahoo! 就只支持“AND”和“OR”;有的在其进阶检索中“完全支持”而在其简单检索中则“部分支持”,如 HotBot, Lycos 等。中文搜索引擎一般不直接支持“AND”,“OR”和“NOT”,而是通过使用特殊的操作符如“+”,“-”,“|”,“!”等来达到同样的目的。

2. 使用特殊操作符

一般的搜索引擎都支持通过一些特殊的符号来改进和限制检索的结果,这些常见的特殊符号有:

+, 关键词前加“+”,要求该词组应出现在检索结果中,如“+引擎”就表示检索结果中一定要有引擎字样。

-, 关键词前加“-”,要求在检索结果中将该词组排除在外,如“-引擎”就表示检索结果中一定不能有引擎字样。

*, 表示通配符,一般在检索表达式的中间或尾部使用,用以代替任意的字母组合。如“中央 * 政府”,既可以包括中央人民政府,也可以包括中央和地方各级政府等;又如“文化 *”,既可以表示文化生活,也可以表示文化娱乐,还可以表示文化氛围等,熟悉计算机 DOS 命令的用户对此不应陌生。

., 作用类似于逻辑检索中的 OR。

(), 作用与数学公式中括号的作用相似。

“, 在多个关键词组合中,通过加双引号,来查询完全符合关键词组合的信息,如

“Search Engine”将只检索 Search Engine,而忽略仅包含 Search 或 Engine 的有关信息。

t:, 在关键词前加上“t:”,搜索引擎将只检索网站或网页名称,如“t:引擎”,则搜索引擎将只检索网站或网页名称中含有引擎字样的对象。

u:, 在关键词前加“u:”,搜索引擎将只检索网址(URL),如“u:intel”,则搜索引擎将只检索网址中含有 intel 字样的对象。

NEAR, 要求两个关键词在句子或文中的间隔不应超过限定的字数,如 Search NEAR/10 Engine,表示要求检索 Search 和 Engine 之间间隔不超过 10 个字的有关文档。

以上提供的特殊操作符,也可以组合使用。有的搜索引擎可能只支持部分特殊操作符,有的搜索引擎的表达方式可能有所不同,但每个搜索引擎对此都会有详细的说明。用户只要在每一个搜索引擎的主页上仔细找一下,通常在主页的上端或关键词检索框的边上就可以发现一个“在线帮助”(Online Help)或“常见问题解答”(FAQ)的标签或按钮。用户在使用一个新的搜索引擎之前,不妨先花些时间访问一下它们,从而熟悉和了解每个搜索引擎的进阶检索方法和其独到之处。

尽管很多用户不使用、也不喜欢进阶检索,他们使用搜索引擎的目的只是走马观花,仅仅满足于对某个方面的信息浏览而已,但对于大多数专业人员而言,若要深入到某个主题并从中挖掘出一些真正有价值的资料,熟悉几种搜索引擎并掌握它们提供的进阶检索方法,是绝对有必要的。

Northern Light 是一个功能比较齐全、适合于科技工作者使用的搜索引擎。现以该搜索引擎作为工具,以了解数学在生命保险方面的应用课题,简单介绍一下进阶检索的基本使用方法。

进入 Northern Light 主页之后,首先在它的查询框中输入“Mathematics”(数学),系统返回的检索结果高达 3 464 250 条记录,根本无法阅读。运用该搜索引擎提供的进阶检索功能,选择“Industry”(行业)标签并激活“Insurance”(保险),系统返回的结果为 1 296 100 条,仍然无法阅读。重新在查询框中输入逻辑检索表达式“Mathematics AND Insurance”(数学 AND 保险),系统返回的结果下降到 3 323 条。此时运用该搜索引擎提供的另一进阶检索功能“Custom Search Folders”(用户检索文件夹),单击“Insurance Industry”(保险行业),系统返回的结果为 2 450 条;继续运用“Custom Search Folders”,单击“Life Insurance”(生命保险),系统返回的结果为 40 条,效果已经相当不错了。为了使目标更集中些,继续发挥进阶检索的作用,将文档的“Select Date Range”(选择日期范围)限制在从 1998 年 1 月 1 日到最近一天,将结果的显示按文档的发表日期(Sort Results by Date)排列,将文档的类型(Limit Documents to)限制在正式出版物范围之内,则此时系统返回的记录仅为 22 条,用户可以逐一浏览,从中获取自己感兴趣的内容。

四、一般检索步骤

用户在使用搜索引擎之前,首先要根据自己的具体要求选择合适的搜索引擎。如果

用户仅仅是在网上浏览一般的信息,则可以选择以分类目录为主的搜索引擎,浏览中文信息的用户可以选择雅虎中文、搜狐、悠游等,浏览外文信息的用户可以选择 Yahoo! 等。如果用户要对某一课题进行比较深入的研究,则应该选择能支持全文检索的搜索引擎,在这方面做得比较好的搜索引擎有 AltaVista, HotBot, Northern Light 等,而能支持中英文全文检索的搜索引擎则有天网(WebGather)、网擎(Openfind)等,雅虎中文原来不支持中文全文检索,自从与网擎合作之后,如今也能支持中文全文检索,用户也可以选择它。

如果用户选择的是以分类目录为主的搜索引擎,则只要遵循该系统的分类体系,层层深入即可。假如用户接连访问了几个类目均未发现有自己感兴趣的内容,此时最好更换一个新的搜索引擎重新寻找。

如果用户使用的是关键词检索,则通常应从简单检索开始(一开始没有必要去构造什么复杂的检索表达式),用户可以在检索框中直接输入一个或几个单词,如果系统找不到有关的内容或返回的结果与用户期望的相去甚远,则仍可以用简单检索,更换几个新的单词接着再试。一般来讲,用户使用简单检索不会找不到有关信息,就怕返回的结果太多,反而使用户无所适从。如果碰到这种情况,就应该求助于该系统的进阶检索或寻找更高级、更专业的搜索引擎了。

1.4 如何选择搜索引擎

经常上网的用户一般都有自己喜爱的搜索引擎,一旦熟悉了其使用方法轻易不会再改用其他的搜索引擎,而新上网的用户因为对任何搜索引擎都没有成见,他们可以有充分的选择余地。

每一种搜索引擎,都有其自身的特色和长处,但也有其难以弥补的不足,通常在选择搜索引擎时,以下几个方面应该引起注意。

1. 收录范围

综合性搜索引擎通常以全球的网络资源为目标,而一些中、小型的搜索引擎则致力于某一区域或某一专业的资料信息。综合性搜索引擎的搜索范围虽然广泛,但就某一局部、某个专业而言,内容却未必有中、小型搜索引擎来得丰富和完备。

综合性搜索引擎通常支持多语种,有的综合型搜索引擎支持的语言竟高达 25 种之多,有特殊需要的用户,应该首先选用相应语种的搜索引擎。

目前在 Internet 中,使用得最多的当属万维网(即 WWW)资源,现有的绝大多数搜索引擎就是以 WWW 为目标的,但在 Internet 中还有许多其他有用的网络资源,如 BBS, FTP, Gopher, Newsgroup, UseNet 等,有的搜索引擎将这些统统拒之于门外,有的搜索引擎则专门致力于这些网络资源的开发和利用。

即使在 WWW 中,不同的搜索引擎,其功能也不尽相同。例如,有的搜索引擎只搜索网络,有的只搜索网址,还有的只搜索网页,用户对每个搜索引擎的收录范围应该有所了解。

2. 数据库容量

不同的搜索引擎,其数据库的容量彼此间相差得十分悬殊,现今搜索引擎界的巨无霸 AltaVista,其数据库容量号称已达到 1.5 亿个网页。一般而言,数据库容量较大的搜索引擎,比较容易获得满意的检索结果。

3. 用户界面

用户界面的简单易用与搜索引擎的功能齐全始终是一对矛盾,但是却存在一些搜索引擎,其网络搜索的技术不怎么样,一进入主页,迎接用户的首先是一大堆令人眼花缭乱的广告,这样的搜索引擎应该坚决摒弃。

4. 响应速度

通常各个搜索引擎的系统响应速度都很快。有专业机构曾对此作过测试,结论是彼此间的差异微乎其微,完全可以忽略不计。让用户对速度无法忍受的真正原因,往往不是系统本身造成的,而是与用户本地网络的传输速度或与用户选择的上网时间有关。

5. 更新周期

Internet 始终处在不断变化、发展之中,一个好的搜索引擎,除了内容丰富、查找迅速之外,还应该对数据库中已有的内容经常进行审核、更新,及时剔除死链、坏链。不同的搜索引擎更新的速度不同,有的搜索引擎的更新周期是以天计算,有的以周计算,有的则以月计算,孰优孰劣自见高低。

6. 准确性

通常用户总是希望从搜索引擎中返回的结果都是他们所需要的,而实际结果离用户的希望往往相去甚远。为此,有的搜索引擎在反馈的信息中加入了诸如相关程度(或关联系数)之类的指标,将所反馈的信息按与检索词组的相关程度从高到低排列,以期弥补这方面的不足。

7. 全面性

当用户在使用搜索引擎时,总是希望他所感兴趣的信息通过搜索引擎都能查找出来,但就目前的搜索技术而言,准确性与全面性是相互制约的,鱼与熊掌不可兼得,用户对此不必过于苛求。

1.5 检索策略与技巧

1. 选择合适的搜索引擎

每个搜索引擎都有其自身的特点,不同的搜索引擎适合于完成不同类型的搜索任务,同一个检索关键词,在两个搜索引擎中并不会产生完全相同的检索结果。因此,用户应该根据具体的检索要求和对各个搜索引擎特征的了解,选择适合于自己检索要求的搜索引擎。具体来说,当用户仅仅满足于在网上进行一般的浏览时,可以选择以分类目录为主的搜索引擎,如 Yahoo!、蕃薯藤等,也可以使用其他搜索引擎提供的分类目录索引;当用户要获得比较全面、丰富的资料时,就要选择以全文检索为主的搜索引擎,如 AltaVista, HotBot 等;如果用户想对某个专业或主题进行比较深入、专门的研究,则应该选择一些具有鲜明特色的专业搜索引擎,如适合于科技工作者使用的 Northern Light,全球找工作的 JobEngine,全球华人寻人引擎 Look 4U 等。就某一个具体的专业

和主题而言,综合性搜索引擎未必比专业搜索引擎占有优势。

2. 发挥进阶检索的功能

用户在使用搜索引擎时,最令他们头疼的往往不是因为找不到有关资料,而是由于找到的资料太多、太滥。当系统找不到用户所需要的资料或返回的检索结果不理想时,用户只要更换一个新的搜索引擎继续寻找,最终总会有所收获。可是,当用户一下子要面对由系统返回的成千上万个检索结果时,如何从中迅速挑选出自己真正感兴趣的内容,这就要看用户驾驭搜索引擎的能力了。

通常情况下,用户在选择好一个搜索引擎之后,可以先用简单检索,即在检索框中输入若干个关键词进行初步尝试,如果系统返回的检索结果比较多,则就需要通过进阶检索来进一步提高检索的准确度、缩小搜索范围和过滤无关紧要的内容。一般而言,利用普通搜索引擎都能提供的进阶检索功能(Advanced Search),选择一般用户不愿或不熟悉的进阶检索技术,通过灵活运用正确的检索表达式和特殊的操作符、限定词,往往可以收到事半功倍的效果。

3. 同时打开多个检索窗口

利用 Windows 平台的多任务性,用户在检索时可以同时打开多个窗口,分别启动不同的搜索引擎同时进行工作,也可以让不同的窗口分别用来作为搜索界面,或传输数据。

4. 使用 All-in-One 集成搜索

当用户不满足于一般的浏览,对检索结果有较高的要求时,可以使用 All-in-One 集成搜索,或多元搜索引擎。只要你有足够的时间和耐心,通过多个搜索引擎的通力合作,总可以使你达到目的。

5. 利用搜索引擎的特色服务

为了吸引用户,许多搜索引擎增添了一些诸如新闻、天气预报、交通信息、E-mail 地址查询之类具有鲜明个性的特色服务,通过它们去直接查询这方面的信息,比通过常规检索要快许多,结果也更令人满意些。

6. 注意对检索结果的甄别

通常在搜索引擎反馈的检索结果中,与检索要求关联程度较高的内容总是被安排在最前面,当用户来不及对检索结果全部浏览时,可先选择前面的内容阅读。

在搜索引擎反馈的结果中,经常会出现一些令用户感到陌生的站点,请不要忽视这些名不见经传的网站和网页,它们当中很可能包含着一些专门从事某一领域或某一主题研究的专业网站,其专业资料之丰富往往令你目不暇接。

当搜索引擎反馈的结果或链接指针中有你感兴趣的网站或能使你为之一振的信息时,请及时将它们加入到浏览器的书签(Bookmarks)或将它们马上打印出来,否则,在下次检索过程中你可能再也“遇”不到它们了。

当搜索引擎反馈的结果中,有些网站或网页的地址(URL)很长,并伴有“~”,“%”,“?”之类的特殊符号时,你很可能是发现了一个“寄生”在其他网站上的个人主页,除非此人是某个方面的权威人士,否则,请赶快转移目标。

7. 不要在检索表达式中只使用一个关键词