

INTERNET JINGJI XINXI ZIYUAN JIANSUO YU LIYONG
INTERNET JINGJI XINXI ZIYUAN JIANSUO YU LIYONG
INTERNET JINGJI XINXI ZIYUAN JIANSUO YU LIYONG
INTERNET JINGJI XINXI ZIYUAN JIANSUO YU LIYONG

INTERNET JINGJI XINXI ZIYUAN JIANSUO YU LIYONG

Internet经济信息资源 检索与利用

◎主 编 李树青

◎副主编 徐美凤 王维民 蒋伟伟



南京大学出版社

Internet经济信息资源

检索与利用

主 编 李树青

副主编 徐美凤 王维民 蒋伟伟

INTERNET JINGJI XINXI ZIYUAN JIANSUO YULIYONG
INTERNET JINGJI XINXI ZIYUAN JIANSUO YULIYONG
INTERNET JINGJI XINXI ZIYUAN JIANSUO YULIYONG

南京大学出版社

图书在版编目(CIP)数据

Internet 经济信息资源检索与利用 / 李树青主编

— 南京 : 南京大学出版社, 2010. 11

ISBN 978-7-305-07689-3

I. ①I… II. ①李… III. ①经济信息—互联网络—
情报检索 IV. ①G354.4

中国版本图书馆 CIP 数据核字(2010)第 205270 号

出版发行 南京大学出版社

社 址 南京市汉口路 22 号

网 址 <http://www.NjupCo.com>

出版人 左 健

书 名 Internet 经济信息资源检索与利用

主 编 李树青

责任编辑 孙 辉 李 琳 编辑热线 025-83592146

照 排 南京南琳图文制作有限公司

印 刷 南京人文印刷厂

开 本 787×1092 1/16 印张 21.5 字数 550 千

版 次 2010 年 11 月第 1 版 2010 年 11 月第 1 次印刷

ISBN 978-7-305-07689-3

定 价 39.00 元

发行热线 025-83594756

电子邮箱 Press@NjupCo.com

Sales@NjupCo.com(市场部)

* 版权所有,侵权必究

* 凡购买南大版图书,如有印装质量问题,请与所购
图书销售部门联系调换

序

很长一段时间,一直想写一本专注于 Internet 信息资源检索的书。作为一名专职教师,深感此门课程的实用性。Internet 信息资源检索是一门非常实用的工具类课程,它主要帮助用户掌握对各种 Internet 信息资源进行检索、获取和利用的方法。纵观目前市场,此类图书琳琅满目,种类很多。那么,这本书应该具有什么特点呢?我为什么要写这本书呢?

它具有以下三个主要特点:

一是在内容上完整,涵盖 Internet 信息资源检索的各种方法。通过对该书的学习,用户应该掌握两个基本能力:一是用户应该能够熟练地使用搜索引擎在内的各种 Internet 信息检索工具,一旦使用得当,你会发现它们的功能非常强大;二是用户应该了解很多常见的信息资源站点,知道哪些站点在哪方面的信息资源上比较丰富和权威,我们可以直接访问此类站点来获取最有价值和最完整的信息资源。于是,这就构成了本书的两大主要内容。

二是在主题上集中于经济类信息资源检索。经济类信息资源是一种较为常见和使用频率较高的信息资源,与我们的生活和工作密切相关。本书在谈论这些信息资源检索方法时,多围绕着经济类资源获取这个主题,同时还专门介绍了搜索引擎中经济信息的获取与分析、常见经济信息资源检索这两个重要章节的内容。当然,各种 Internet 信息检索方法本身具有通用性,因此这些所述方法也完全适用于对其他类型信息资源的检索。

三是在风格上侧重于实用性和易读性。我不想写一般的理论书籍,事实上,恐怕读者也不太关心“什么是”或者“特点有什么”之类的问题,更想知道一些“如何去做”的问题。因此,对于大部分内容来说,除非必要(即必须知道这些概念上的区别才能更好的操作),在大部分情况下,本书都不去做太多的概念解释。同时,考虑到信息资源获取的方便性,本书还着力介绍很多相关免费资源的获取方法。

另外,对于操作步骤,本书尽可能介绍的具体和详细,以增加易读性。这体现在以下几个方面:一是所有操作都在 Internet 上予以实践,因为 Internet 发展变化非常快,所以这些操作都是在最近的时间完成的,时效性比较强;二是把丰富的截

图演示添加在书中,用户只需阅读本书,即可根据操作说明和截图来对各种方法的使用和效果获得比较清晰的认识。

下面对本书的结构做一说明。如前所述,本书主要有两大内容,即搜索引擎和 Internet 信息资源站点,具体可以分为四个部分:

第一部分为基本概念,主要介绍信息检索和信息资源的基本概念和类型。它们对于理解和更好地掌握 Internet 信息资源检索方法都是至关重要的。

第二部分为搜索引擎,主要介绍搜索引擎的检索方法和搜索引擎中经济信息的获取与分析。值得注意的是,利用搜索引擎获取的信息多为免费信息,所以该部分内容的实用性很大。

第三部分为常见电子文献信息的检索,主要介绍网络图书资源检索、网络电子论文资源检索、网络百科知识资源检索和其他常见电子文献资源的检索,其中对经济类相关资源有专门说明和重点介绍。

第四部分为常见经济信息资源的检索,主要介绍商贸信息检索、名录信息检索、产品价格信息检索、金融信息检索、综合经济信息检索、经济术语信息检索、经济人物信息检索、经济类院校和研究机构的检索等。

这本书的网络辅助学习站点可以提供各种与该书相关的信息资源服务,如电子文档、讲义和一些常见软件资源的下载等。这些站点资源可以及时更新,以补充些最新的相关知识内容。网址为 <http://www.njcie.com/eir/>。另外,作者的联系邮箱是:leeshuqing@163.com。

事实上,本书的写作也得益于 Internet 所提供的各种信息资源,现代 Internet 网络确实能够为我们提供很多非常有用的帮助。同时,本书写作还得到了南京财经大学信息工程学院的大力支持,南京大学出版社也给予了很大的帮助。在此一并表示感谢!

限于个人能力,本书还有很多不足之处,敬请各位读者和专家批评指正。

目 录

第一章 信息检索与信息资源	1
1.1 信息检索	1
1.1.1 信息检索的必要性	1
1.1.2 信息检索的基本过程	7
1.1.3 信息检索效果的衡量指标	8
1.1.4 信息检索系统的发展历史	9
1.2 信息资源	11
1.2.1 一般信息资源的常见类型	11
1.2.2 学术信息资源的常见类型	12
第二章 搜索引擎的检索方法	13
2.1 基于关键词的基本检索方法	13
2.1.1 布尔检索	13
2.1.2 词组检索	19
2.1.3 模糊检索	24
2.1.4 字段检索	26
2.2 基于 Web 目录的分类检索方法	37
2.3 检索策略与典型案例	40
2.3.1 合理选择检索关键词	40
2.3.2 综合使用各种检索方法	48
2.3.3 间接获取方法	50
2.3.4 其他方法	55
第三章 搜索引擎中经济信息检索与分析	57
3.1 经济信息获取	57
3.1.1 财经新闻信息检索	57
3.1.2 股票信息检索	61
3.1.3 统计信息检索	65
3.1.4 商品信息检索	67
3.1.5 生活信息检索	70
3.2 信息分析与决策	73
3.2.1 百度的信息分析决策功能	74

3.2.2 Google 的信息分析决策功能	78
第四章 经济类网络图书资源检索	90
4.1 经济类网络书目检索	90
4.1.1 图书馆书目系统	90
4.1.2 图书网站的网络书目	100
4.1.3 网络书目数据库	105
4.1.4 利用搜索引擎检索书目	108
4.2 经济类网络电子书全文数据库	110
4.2.1 全文数据库	110
4.2.2 利用搜索引擎获取全文数据库	119
4.2.3 其他图书资源下载站点	122
4.3 经济类网络图书的其他相关信息检索	123
4.3.1 网络图书的文摘检索	123
4.3.2 网络图书的书评检索	125
第五章 经济类网络电子论文资源检索	127
5.1 经济类网络电子论文检索	127
5.1.1 电子论文数据库检索	127
5.1.2 搜索引擎的电子论文检索	144
5.2 利用引文信息获取相关学术研究资源的基本方法	154
5.2.1 参考文献检索与管理	155
5.2.2 引文索引检索	158
5.2.3 结合参考文献和引文索引获取相关学术研究资源的基本方法	166
第六章 经济类网络百科知识资源检索	171
6.1 网络专业百科全书的检索	171
6.1.1 网络英文百科全书的检索	171
6.1.2 网络中文百科全书的检索	178
6.2 百科知识网络站点的检索	182
6.2.1 专业百科知识网站的检索	183
6.2.2 搜索引擎的百科知识服务	186
6.3 经济类网络专题百科知识检索	193
6.4 经济类网络年鉴检索	194
6.4.1 年鉴资源站点的检索	195
6.4.2 网络年鉴数据库的检索	199
6.4.3 搜索引擎的年鉴信息检索	210
第七章 其他经济类电子文献资源的检索	211
7.1 经济类电子报纸信息检索	211

7.1.1 纸质报纸的网络资源检索	211
7.1.2 网络电子报纸的检索	223
7.2 经济类专利信息检索	224
7.2.1 从专利主管机构来检索	224
7.2.2 专利数据库检索	227
7.2.3 搜索引擎中的专利信息检索	230
第八章 经济信息资源的检索	234
8.1 商贸信息检索	234
8.1.1 商品供求信息检索	234
8.1.2 市场信息检索	248
8.1.3 会展信息检索	250
8.2 名录信息检索	251
8.2.1 企业名录信息检索	252
8.2.2 产品名录信息检索	260
8.3 产品价格信息检索	277
8.3.1 专业的产品价格数据库检索	277
8.3.2 产品价格信息的服务站点检索	279
8.4 金融信息检索	285
8.4.1 专业的金融信息数据库检索	286
8.4.2 相关金融站点检索	295
8.5 经济分析信息检索	297
8.5.1 国内经济分析信息数据库	297
8.5.2 国外经济分析信息数据库	299
8.6 经济类专业学习资源检索	300
8.6.1 中国经济学教育科研网	300
8.6.2 人大经济论坛	301
8.6.3 ProQuest 信息学习数据库	301
8.7 综合经济信息的检索	302
8.7.1 中国经济信息网	302
8.7.2 中国国家经济信息系统	304
第九章 其他经济相关信息资源的检索	306
9.1 经济术语信息的检索	306
9.1.1 利用词典站点来检索	306
9.1.2 利用搜索引擎字典服务来检索	311
9.1.3 利用专业术语检索站点来检索	314
9.1.4 专业术语的中英文互译检索	315
9.2 经济人物信息的检索	317
9.2.1 中国经济人物信息的检索	318

9.2.2 国外经济人物信息的检索	325
9.3 经济类院校和研究机构的检索	327
9.3.1 中国经济类院校和研究机构的检索	327
9.3.2 国外经济类院校和研究机构的检索	330

参考文献	336
------------	-----

第一章 信息检索与信息资源

本书主要介绍在 Internet 上检索各种信息资源的方法,其中对经济类信息检索有专门的说明。不过,在正式介绍之前,我们需要首先了解一些比较基本的概念和原理,以方便后续书中内容的学习,同时也有助于那些对信息检索原理不是很了解的读者,使其掌握一些信息检索的基本理论内容,从而加深对各种实践操作的理解。本章主要介绍信息检索和信息资源的一些基本概念。

1.1 信息检索

我们先对信息检索这个概念做个解释。所谓信息检索,有时也被称为信息获取、信息查询等。它们的意思相差无几,都表示用户利用现代信息检索系统来获取所需信息内容的过程。这种信息检索系统通常有两类:一类是搜索引擎,另一类就是各种信息资源站点。它们对于信息检索来说,都是重要的信息来源体。本书也主要围绕这两类信息检索系统来介绍。

让我们从一个简单的问题出发:为什么要学习这本书?

1.1.1 信息检索的必要性

1.1.1.1 表面的理解

为什么要进行信息检索活动呢?我想更多的时候,读者只是偶然翻开这本书,或者是刚开始学习这门课,因此对这个问题可能并没有十分清楚的认识。

我们先来看一个例子!这是《暮光之城》电影的一个截图,吸血鬼对那个女孩说了一句话,如图 1.1 所示:



图 1.1 Google 已经成为一个动词(截图来自于电影《暮光之城》)

在这里你会发现 Google 已经成为一个很常用的动词。这并非杜撰,在 2006 年 6 月 15 日,著名的牛津英语大辞典(The Oxford English Dictionary, OED)把 Google 收录成一个动词。这可以在牛津词典的网站上看到,如图 1.2 所示:

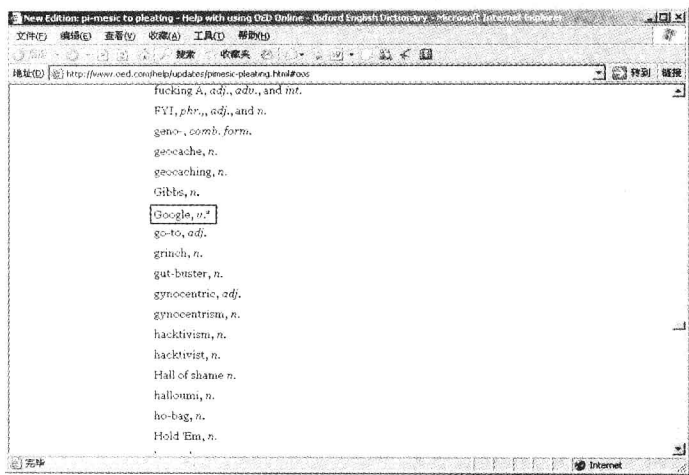


图 1.2 牛津英语大辞典网站收录的“Google”词语说明(截取于 2010-4)^①

事实上,Google 是一家软件公司的名称,这家公司在 1998 年才成立,发展至今已经成为能和微软等软件公司竞争的大型公司。然而,它只是一家专门从事 Internet 信息检索服务的搜索引擎公司!

当然,这样的公司远远不止一家,国外还有雅虎等搜索引擎,甚至微软后来也开始从事搜索引擎的开发和运营。中国国内也有很多,比较著名的有百度和搜狗等。它们不仅广告出镜多,而且经常会在各种媒体上展示自己。

这些充分说明搜索引擎厂家已经成为现代 Internet 企业中举足轻重的力量,造成这种现象的一个根本原因就是现代社会已经越来越多地需要利用搜索引擎等信息检索服务来完成日常工作和生活等各个方面的活动,因此促进了此类搜索引擎的快速发展。

不过,你也许会想,虽然搜索引擎厂商重要,但对于个人而言,信息检索又有什么用呢?2006 年曾经有一本畅销书,名称叫《搜商:人类的第三种能力》^②。这里并不想对此书做什么介绍,只是关注这本书的名称。什么是搜商?专业些说,搜商(Search Quotient, SQ)是一种与智商、情商并列的人类智力因素,也就是人类通过某种手段获取新知识的能力,其本质就是检索信息和搜索信息的能力。

很多人都经常说“做功课”。比如,旅游出行前,需要对旅游目的地做必要的了解,防止一些完全可以避免的事情出现,常见的方法就是检索 Internet 上的搜索引擎。前些年有朋友去外地旅游买手表,回来感觉不对劲,网上查一查,果然有问题,而且很多人都有同样的遭遇!如果事先借助搜索引擎查查各种旅游攻略、注意事项和别人旅游回来后发表的各种心得,完全可以避免这种事情发生。常见的检索结果如图 1.3 所示:

^① 牛津英语大辞典网站收录“Google”词语的说明网址为: <http://www.oed.com/help/updates/pimesic-pleating.html#oos>。

^② 陈沛. 搜商:人类的第三种能力[M]. 北京:清华大学出版社,2006

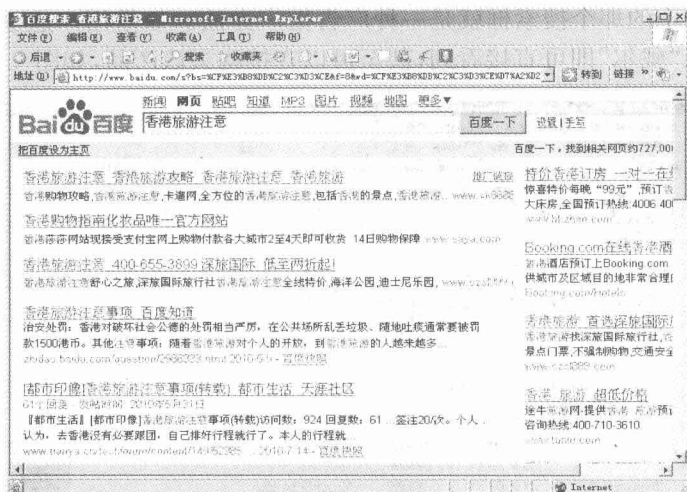


图 1.3 在百度中检索“香港旅游注意”的结果页面(截取于 2010-7)

再如,在生活和工作中遇到问题怎么办?过去似乎手边总是需要一本如“百事通”之类的手册,但今天只需打开 Internet 网络,检索一下即可。例如,前段时间有同事的惠普打印机出了问题,总是在液晶屏上显示一个“Printer Mispick”的提示,同事一时没想起来,就找了半天说明手册。其实,面前就是一个电脑,在百度上输入“Printer Mispick”,就解决了问题,结果如图 1.4 所示:

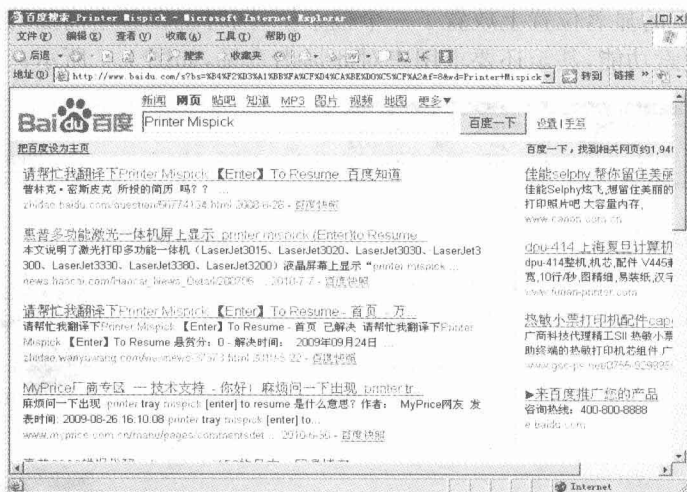


图 1.4 在百度中检索“Printer Mispick”的结果页面(截取于 2010-7)

在与经济有关的各个领域,搜索引擎和信息检索更是无处不在。比如,利用搜索引擎进行广告推广已经成为一种常见的市场营销策略,用户只需在搜索引擎中输入一些检索词,搜索引擎就会把相应的广告有效地推送给用户浏览。《读者》杂志上的 Dell 电脑广告如图 1.5 所示:



图 1.5 《读者》杂志上的 Dell 电脑广告

Dell 电脑广告中的那个搜索框就是一种广告推广方法。如果用户想要了解该产品,只需在搜索引擎中输入“戴尔”即可直接看到,甚至可以直接购买该产品,如图 1.6 所示:

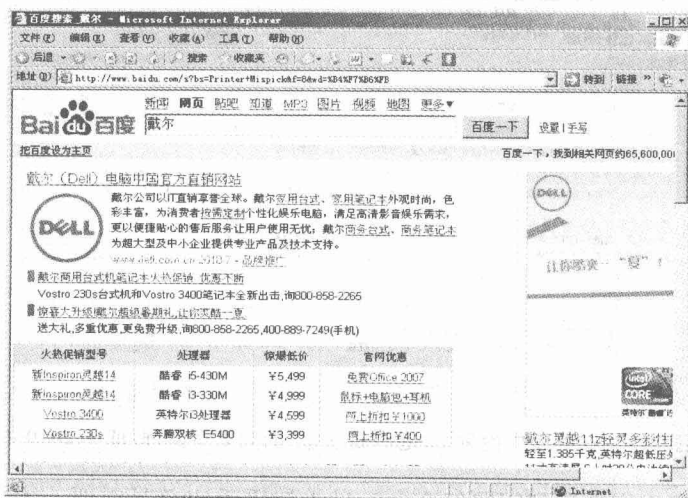


图 1.6 在百度中检索“戴尔”的结果页面(截取于 2010-7)

近几年,电子商务得到了大力发展,如淘宝之类的网络购物站点也如雨后的春笋般陆续出现。其实,网络购物的关键就在信息检索,这也是影响用户使用感受的一个最为明显的因素。如果这种系统不能很好地帮助用户找到自己所需的商品,恐怕用户就不愿意再使用它们。为此,淘宝在自己主页的显著位置上放置了一个搜索框,同时也在各个商品的浏览页面中集成了各种方便用户的检索功能,甚至还要求用户对此提出意见,如图 1.7 所示:

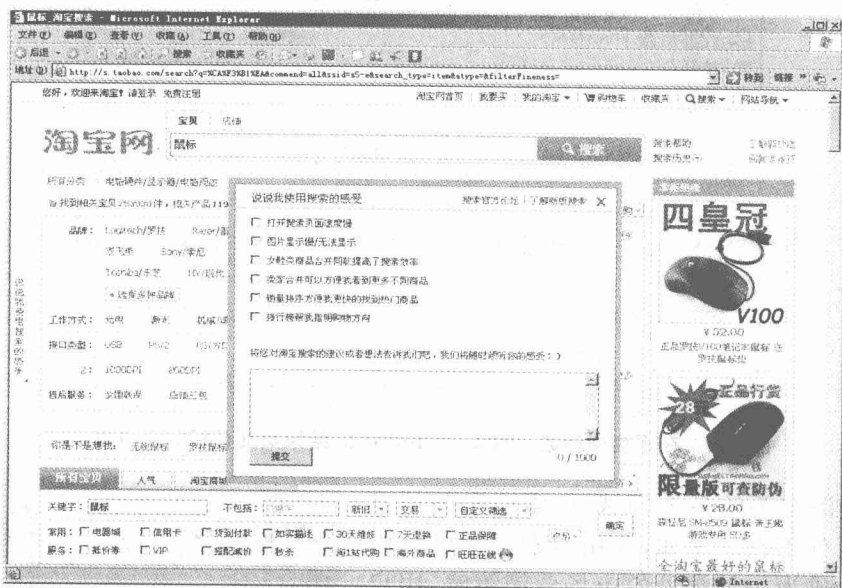


图 1.7 在淘宝网的商品页面中查看和信息检索有关的各种内容(截取于 2010-7)

不妨把信息检索的必要性再说大一些,人生的几件大事可能都与信息检索有关!高中毕业选择高校和专业,仅仅查阅那个小小的高校专业介绍手册显然不够。看看大家

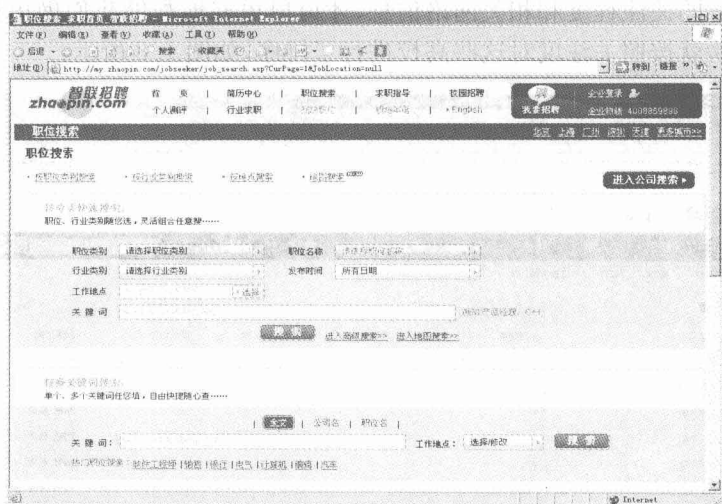


图 1.10 智联招聘网提供的职位搜索界面(截取于 2010-7)

甚至连找对象这种事情,也都成为现代 Internet 信息检索服务产业中一个很大的市场,如各种婚恋介绍站点等,如图 1.11 所示。

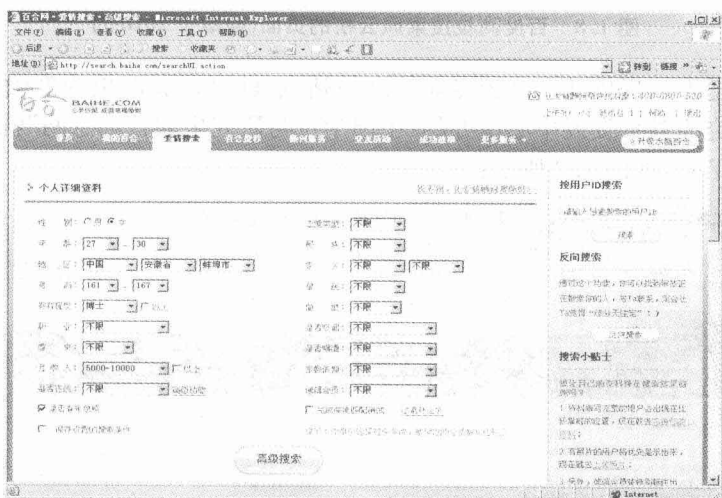


图 1.11 婚恋社交站点百合网提供的爱情搜索界面(截取于 2010-7)

更不要说买房、买车之类的事情了。到此,我们可以对信息检索的必要性有一个感性的认识。在现代社会中,信息检索已经成为一种重要的用户行为,和我们日常工作生活密切相关。因此,我们有必要学习如何更好地使用各种诸如搜索引擎在内的信息检索系统,同时也应该了解一些常见的信息资源站点,知道从哪些站点可以更为方便地获取所需信息,从而为我们提供更多的便利。

1.1.1.2 深层的理解

我们可以站在更为专业的角度来对信息检索的必要性做一说明。

如果我们把 Internet 信息系统抽象成一个巨大的人类知识体,那么就会发现信息检索活

动是一种人类认识和获取知识的基本活动过程。在这种场景中,这种巨大的知识体既可以包括 Internet 网络,也可以将图书、报纸等各种传统媒体资源包含进来。因此,用户必须掌握与这种知识体交互的方法,即信息检索方法,才能更好地使用它们。

那么,我们是如何使用这些知识的呢?

其实,这包括两个重要条件:一是要存在这样的知识体。事实上,不管是图书、报纸等传统纸质文献,还是 Internet 存储的电子资源,它们都是一种知识体的具体存在形式,因此知识体是客观存在的。我们这里主要探讨如何在 Internet 信息资源这个知识体中获取所需信息。二是用户能够表达需要什么样的知识。相对于第一个条件而言,似乎这个条件更为简单,然而对于用户来说,这才是需要着力掌握的技能之一。这其实也就是一种信息检索的能力,越能有效地掌握获取和检索知识的能力,用户就越有可能在今天海量的信息世界中找到自己所需的内容。

事实上,今天 Internet 上的信息量已经呈现出一种爆炸性增长的态势。据报道,2008 年 7 月,Google 搜索引擎能够检索到的 Internet 网页就高达 1 万亿,而且每天增加的网页数量也在十亿数量级上。面对着海量的信息资源,人们却难以有效地获取所需知识,主要原因在于这种信息资源的增长速度远远超出了人们能够处理它们的能力。约翰·奈斯比特(John Naisbitt)在《大趋势》一书中是这样形容人们目前所处的困境:“信息是丰富的,而我们正在渴求知识”^①。

当然,通过改进搜索引擎等信息检索系统的功能可以解决这一问题,然而我们也注意到,在更多的情况下,用户检索技能方面的改进空间更大一些,甚至可以说,如果用户不能很好地掌握信息检索方法,再好的信息检索系统也难以发挥它们的威力。事实上,在很多情况下,我们都会看到用户遇到问题时会使用搜索引擎进行检索,但是结果往往不能让他们很满意。

搜狗实验室在 2007 年曾经发表过一篇研究论文^②。文中指出,在对搜狗搜索引擎一个月内近 5 000 万条检索日志进行分析处理后,发现长度不超过 3 个词的检索占总检索数的 93.15%,平均长度为 1.85 个词,这说明用户输入的检索通常都比较短。而我们都能想象出,检索词语越短就越难以有效地表达用户的准确信息需求。更为吃惊的现象是,只有约 0.73% 的检索含有用于高级检索功能的符号,也就是说目前大多数中文检索用户只是通过输入很少的几个关键词就开始检索。其实,这些包括布尔检索在内的高级检索往往都能更为有效地表达用户的信息需求。

这些都充分说明我们需要掌握现代 Internet 信息资源的检索方法,提高我们使用 Internet 信息资源的能力,利用这些宝贵的信息资源为我们提供更好的帮助。

1.1.2 信息检索的基本过程

不管用户使用的是搜索引擎,还是其他信息检索系统来检索知识体,一般而言,信息检索的基本过程主要分为两个步骤:一个步骤是用户发出对信息的检索请求;另一个步骤是信息检索系统响应用户的请求,并返回请求的检索结果。不过,这种理解过于简单。事实上,用户和知识体并不能直接交流,中间存在两个主要的转换环节:

① [美]J. 奈斯比特. 大趋势[M]. 北京:新华出版社,1998

② 余慧佳,刘奕群,张敏,茹立云,马少平. 基于大规模日志分析的搜索引擎用户行为分析[J]. 中文信息学报,2007,21(1)

一是用户需要将自己的信息需求转换为信息检索。这既需要用户掌握一些信息检索的技能,同时也需要检索系统提供一个良好的界面以方便用户表达信息需求和使用信息检索功能。事实上,我们也能看到很多搜索引擎的界面差别正是体现了它们对用户检索感受的不同理解和考虑。

二是信息检索系统要能够在知识体中找到用户所需的信息内容。这就需要信息检索系统对这些知识体的信息内容做必要的处理,以保证在较短的时间内找到最为相关的结果。对于如此巨大的 Internet 网络来说,这并非一件很简单的事情。因此,大多数搜索引擎都是由一些技术先进的大公司来运作和维护。

用户在信息检索系统中检索信息的完整过程如图 1.12 所示:

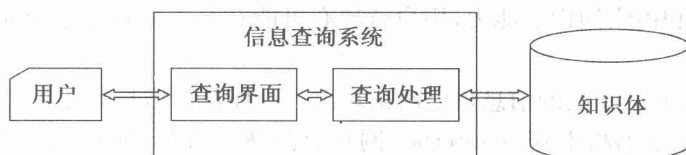


图 1.12 信息检索系统中用户检索活动的示意图

这种信息检索活动是一种常见的行为,尤其在 Internet 网络上。一般而言,使用搜索引擎就是一种典型的信息检索行为。那么有人会说:我通常上网并不是使用搜索引擎,只是看看网页,这也是信息检索行为吗?比如用户打开网易主页,看到了主页上的体育新闻,很快点击该超链,在弹出的新页面中看到了更多的体育新闻。由于该用户是个篮球迷,于是在这个网页中又连续点击看到很多关于 NBA 联赛的消息。

这种操作看起来并不像是信息检索,其实它具有信息检索活动的全部特点,即用户有比较明显的信息需求,同时也在不停地获取满足这种需求的各类信息资源。具体来看,用户为什么要点击关于体育的新闻呢?又为什么继续点击 NBA 的消息呢?这些都能反映出用户的一种个性化信息需求,也就是说,正是因为该用户想了解这方面的信息,才会通过点击超链来获取相应的信息内容。不过和一般搜索引擎检索不一样,这种检索过程没有显式的检索词语输入,而且用户的信息需求是逐渐在浏览操作中被启发而形成的,但它确实是一种信息检索过程,我们通常称之为基于超链访问行为的信息检索。

1.1.3 信息检索效果的衡量指标

在大多数情况下,信息检索结果的好坏是一个非常主观的判断结果。不同的人,甚至同一个人在不同的时间,都可能会对同一个检索结果做出完全不一样的判断。因此,为了能够清楚和客观地测度信息检索系统的检索效果,经常使用的两个经典指标,分别是查全率和查准率。

所谓查全率,英文名称为“Recall”,也常被唤作“召回率”,是指在所有满足用户检索需求的相关记录中,检索系统实际返回的相关记录比重。使用公式表示为:

$$\text{查全率} = \frac{\text{检索出来的相关记录}}{\text{所有的相关记录(包括没有检索出来的相关记录)}}$$

所谓查准率,英文名称为“Precision”,是指在检索出来的所有记录结果中,真正相关记录的返回比重。使用公式表示为:

$$\text{查准率} = \frac{\text{检索出来的相关记录}}{\text{检索出来的所有记录(包括不相关的记录)}}$$