

# Internet 信息资源及其检索探讨

狄文凯

(皖南医学院图书馆,安徽 芜湖 241002)

[摘要] 分析了 Internet 信息资源的类型和特点,介绍了 Internet 信息资源检索方法,针对影响 Internet 信息资源检索的因素,提出了提高 Internet 信息资源检索效率的对策。

[关键词] Internet 信息资源;信息检索;网络检索

[中图分类号] G354.4 [文献标识码] B [文章编号] 1003-725X(2011)03-0108-03

## 1 Internet 信息资源及其特点

Internet 信息资源是指以数字化形式存储在 Internet 上的各种信息资源的总和。按信息来源划分,Internet 信息资源可分为政府信息资源、公众信息资源、商用信息资源;按信息时效划分,可分为电子邮件型、图书馆目录、书目与索引、全文资料及电子出版物、数据库等信息资源;按网络传输协议划分,可分为 WWW、Telnet、FTP、用户服务组、Gopher 等信息资源。Internet 信息资源具有以下特点:(1)内容多样性。Internet 是个开放的信息传播平台,任何机构、任何人都可以将自己拥有的且愿意与他人共享的信息传递到网络上。如公共图书馆、网络信息服务商、传统媒体、高等院校、科研机构、各类商业公司等是 Internet 信息供应源。Internet 信息资源包含科学技术领域信息、历史档案信息、知识性和教育性的信息、传媒信息、学术文化信息、经济信息;(2)信息表现形式多样。Internet 是一个集声音、图像、文字、照片、图形、动画、电影、音乐为一体的综合性信息系统;(3)Internet 信息资源集信息关联性、信息开放性、信息时效性于一体。Internet 的信息组织是基于超文本的,有关联的信息之间通过链接形成一个相互联系的信息渠道。Internet 具有很强的时效性,能很快地将信息传播到世界各地;(4)信息交互性强。用户不仅可以在 Internet 上获取信息,而且也可以

在 Internet 上发布信息。用户通过 Internet 可以与科学家、工程技术专家、医生、律师、教育家进行交流,同时也可发表个人的见解。

## 2 Internet 信息资源检索方法

用户要了解和利用 Internet 信息资源,必须借助检索工具对 Internet 信息资源进行检索。而搜索引擎(Search Engine)作为 Internet 的信息检索系统,已成为用户普遍使用的信息检索工具。它以一定的策略在 Internet 上搜集、发现信息,对信息进行理解、提取、组织和处理,并为用户提供检索服务。搜索引擎一般有搜索器、索引器、检索器、用户接口等组成,搜索引擎有两个主要的功能:第一个功能是收集信息并建立索引数据库,自动跟踪信息源的变动,不断更新索引记录,定期维护数据库;第二个功能(最主要的功能)是提供网络的信息导航与检索服务。搜索引擎的查询方式有:(1)简单检索。直接输入一个关键词,提交搜索引擎查询,这是最基本的查询方式;(2)词组检索。输入两个单词以上的词组(短语),当做一个独立运算单元提交搜索引擎查询。这种方式也叫短语检索;(3)语句检索。输入一个多词的任意语句,提交搜索引擎查询。这种方式也叫任意查询;(4)高级检索。用布尔逻辑组配方式查询。

基于搜索引擎的 Internet 信息资源检索方法主要分为基于超文本的信息查询、基于目录的信

息查询、基于元搜索引擎的查询三种。

### 2.1 基于超文本的信息查询

该类搜索引擎由检索器根据用户的查询输入,按照关键词检索索引数据库。这种方式是大多数搜索引擎最主要的功能。在主页上有一个检索框,用户在检索框中输入要查询的关键词,单击“检索”(或“搜索”、“search”、“go”等)按钮,搜索引擎就会在自己的信息库中搜索含有输入的关键词的信息条目。用户可以通过分析选择所需的网页链接,直接访问要找的网页。此类搜索引擎主要有:(1)天网(<http://e.pku.edu.cn>)。天网提供全文检索、新闻组检索、FTP检索(北京大学、中科院等FTP站点)。目前大约收集了100万个WWW页面(国内)和14万篇Newsgroup(新闻组)文章。支持简体中文、繁体中文、英文关键词搜索,不支持数字关键词和URL名检索;(2)百度(<http://www.baidu.com/>)。百度是目前全球最大的中文搜索引擎,除提供网页搜索外,还提供MP3、图片、视频、地图等多样化的搜索服务,给用户提供更加完善的搜索体验;(3)AltaVista(<http://www.altavista.com>)。该种搜索引擎提供全文检索功能,并有较细致的分类目录。提供检索新闻、讨论组、图形、MP3/音频、视频等检索服务以及进入频道区(zones),对诸如健康、新闻、旅游等类进行专题检索;(4)Infoseek(<http://www.infoseek.com>)。Infoseek提供包括通过电子函件发送新闻、外国语搜索、按地理区域的搜索以及个人的金融文件夹等。提供全文检索功能,并有较细致的分类目录,还可搜索图像。网页收录极其丰富,以西文为主,支持简体和繁体中文检索。

### 2.2 基于目录的信息查询

目录分类式(网站级)搜索引擎的数据库是依靠专职编辑人员建立的。当用户提出检索要求时,搜索引擎只在网站的简介中搜索。用户只要用鼠标单击这些分类链接就可以一级一级地深入这个目录,最终搜索到所需的网页。所收录的网络资源经过专业人员的鉴别、选择和组织,保证了检索工具的质量,提高了检索的准确率。常见的目录分类式搜索引擎主要有以下几种:(1)搜狐(<http://www.sohu.com>)。搜狐为用户提供查找网站、网页、新闻、网址、软件、黄页等信息;(2)新浪(<http://www.sina.com.cn>)。新浪将网络资源分为18个大类,收录网

站20万个。提供网站、中文网页、英文网页、新闻、汉英辞典等多种资源的查询;(3)网易(<http://www.163.com>)。网易新一代开放式目录管理系统,为用户创建了一个拥有超过1万个类目的信息;(4)Yahoo(<http://www.yahoo.com>)。Yahoo拥有第一流的Web目录和最佳的新闻链接以及许多附加服务。有10余种语言版本,提供类目、网站及全文检索功能;(5)Looksmart(<http://www.looksmart.com>)。LookSmart是人工目录集合网站,向其他搜索引擎提供搜索结果。目前,LookSmart已建成含有25亿URL,11亿索引文档的网络索引目录,集合了400万个网站。

### 2.3 基于元搜索引擎的查询

元搜索引擎是一种调用其他独立搜索引擎的引擎,对多个独立搜索引擎进行整合、调用、控制和优化利用。检索时,元搜索引擎根据用户提交的检索请求,调用源搜索引擎进行搜索,对搜索结果进行汇集、筛选、删并等优化处理后,以统一的格式在同一界面集中显示。常见的元搜索引擎有:(1)Dogpile(<http://www.dogpile.com>)。它可以同时调用25个独立搜索引擎进行信息检索;(2)Mamma(<http://www.mamma.com>)。Mamma是并行式元搜索引擎,可以同时调用AltaVista、Excite、Infoseek、Lycos、WebCrawler、Yahoo等独立搜索引擎,并且可以查新闻组、商业黄页等;(3)AskJeeves(<http://www.askjeeves.com>)。AskJeeves提供同时搜索AltaVista、Excite、Yahoo、Infoseek、Lycos和WebCrawler的功能,此外还能同时搜索自己独立的数据库。

## 3 影响Internet信息资源检索的因素

### 3.1 信息资源质量对信息检索的影响

第一,Internet信息资源的不完整、不系统、不科学,导致信息检索必须多次进行,造成人力、物力和时间上的浪费。第二,Internet信息资源加工处理不规范、不标准,使信息检索的查全率、查准率下降。第三,Internet信息资源分散、无序、时常更换,用户无法判断网上有多少信息同自己需求有关,检索评价标准无法确定。第四,信息资源版权和知识产权问题,也给信息检索带来麻烦。第五,信息的语言障碍问题。目前Internet上80%以上的信息是以英语形式发布的,英语水平低和不懂英语的用户很难利用Internet上庞大的信息资源。对中国用户来说,虽然网上中文信息剧增,但

还是需要查询西方国家先进科技信息,由于缺乏汉化软件、自动翻译系统尚未成熟,因此,语言障碍也影响了广大用户对网上信息资源的开发与应用。

### 3.2 检索软件对信息检索的影响

用户必须掌握各种网络信息检索工具,才能检索到自己所需要的网络信息资源。但由于 Internet 信息组织的特殊性和目前检索工具自身存在的一些问题,信息检索不是一件轻而易举的事情。第一,Internet 上的信息存放地址会频繁转换和更名,根据检索工具检索的结果并不一定就能获得相应的内容。第二,基于一个较广定义的检索项,往往会获得数以千万计的检索结果,而使用户难于选择真正所需的信息。第三,每种检索工具虽然仅收集各自范围内的信息资源,但也难免使各种检索工具的信息资源出现交叉重复现象。

### 3.3 用户技术水平对信息检索的影响

在 Internet 这个开放式的信息检索系统中,用户不仅要检索信息资源,同时还进行信息资源的收集、整理、存储工作。因此,Internet 用户的信息获取与检索能力对信息检索有着直接的影响。用户对信息检索需求的理解和检索策略的制定关系到信息检索的质量,用户的计算机操作能力及网络相关知识的掌握程度影响着信息检索的效率,用户对网络信息检索工具的应用熟练程度影响着信息检索的效果,用户的外语水平影响着信息检索的广度。

## 4 提高 Internet 信息资源检索效率的对策

### 4.1 选择合适的网络信息检索工具,采取切实可行的检索策略

网络信息资源检索工具数量众多,各有千秋,不同的检索工具,其索引规模、搜索范围、索引组织、查询的表示形式、特征项的选择、输出结果的形式、检索功能等各不相同。如果用户能选择合适的检索工具,对取得检索成功有很大帮助。在选择网络信息资源检索工具时,用户要明确不同类型网络信息资源检索工具的适用范围,要了解主要网络信息资源检索工具的特点与功能,重视网络信息资源检索工具的分类浏览功能,注重多种网络信息资源检索工具的组合使用。

提高 Internet 信息资源检索效率,除了选择合适的检索工具以外,还要制定切实可行的检索策

略。首先,要对检索课题的主题进行分析,提炼出正确的关键词。其次,要选择适当的词语,确定检索项。检索词可以是规范词,也可以是自由词。再次,要编制既能表达检索课题需求又能为计算机识别的检索提问式。用户可关闭图像提高下载速度,打开多个检索窗口减少等待时间,以提高检索效率;降低检索词的专指度,使用同义词、近义词,使用布尔逻辑或 or 等扩大检索范围,以提高查全率;提高检索词的专指度、使用布尔逻辑与 and、利用高级(进阶)检索、使用词组检索、使用字段限制检索、使用完全字符串检索、使用大小写检索等缩小检索范围以提高查准率。

### 4.2 加强用户网络信息资源检索能力培训

加强用户培训,是提高网络信息检索效率的最有效途径之一。图书馆可通过开设文献检索或计算机检索课程培养用户的信息资源检索能力。根据不同类型的用户采取不同的培训方式,如网上自助式培训、定期不定期举办讲座或培训班、编印网络信息资源检索方法手册等,对用户开展网络基础知识、各类网络信息资源介绍及网络检索工具使用方法等内容的培训,提高用户获取网上信息资源的技能。

### 【参考文献】

- [1] 邹永利,王春强.影响网络信息检索效率的用户因素[J].情报理论与实践,2008(3):374-376.
- [2] 聂建霞.提高网络信息检索效率探讨[J].情报探索,2010(6):95-96.
- [3] 张友梅.网络信息资源检索问题研究[J].科技情报开发与经济,2010(1):92-94.
- [4] 王新卫.Internet 信息资源检索机制研究[J].西安文理学院学报(自然科学版),2007(2):100-102.
- [5] 吴江.搜索引擎关键词和目录检索系统合并提高检索率[J].图书馆论坛,2008(4):80-82.
- [6] 袁津生.搜索引擎与信息检索教程[M].北京:中国水利水电出版社,2008.
- [7] 克罗夫特.搜索引擎:信息检索实践[M].刘挺,译.北京:机械工业出版社,2010.
- [8] 搜索引擎 [EB/OL].[2010-12-25]. <http://baike.baidu.com/view/1154.htm#sub1154>.

【作者简介】 狄文凯(1953-),男,馆员,大专,研究方向:图书馆学。