

●王进孝 (中科院文献情报中心 北京 100080)

搜索引擎与网络信息资源检索研究*

Abstract: Based on a discussion on the basic theory and the evaluation indexes of search engines, this article points out the major problems in the development of search engines. Meanwhile, it introduces several important directions for the development of search engines for searching information resources on the Internet. The research on searching information resources on the Internet is also discussed.

Keywords: Internet; information resources; search engine; development study

网络环境和全文检索为网络信息资源检索带来了一场革命性的变化。搜索引擎的智能化发展,将从根本上改变现有的信息资源检索方式和信息服务模式,全面实现自动化、网络化、信息化。

1 搜索引擎的基本理论

信息检索的研究与因特网的发展是同步进行、互相推进的,在因特网刚刚诞生的时候,还没有搜索引擎(Search Engine)的概念。随着因特网的信息呈几何级数迅速增长,最初的搜索引擎开始出现。所谓面向因特网的搜索引擎,是一种可以从各类网络资源中浏览和检索信息的工具。这些网络资源包括:Web、FTP文档、新闻组、Gopher、Email以及多媒体信息等。早期的搜索引擎系统大部分基于字符界面,虽然以免费和开放著称,但由于应用不便难以普及,其代表产品包括 Archie、Gopher、Whois、Agora、Knowbot、分布式文本搜索系统 Wais 等。

1994年初,第一代真正基于WWW的搜索引擎Lycos诞生。到1995年,商业化的搜索引擎开始大规模开发,其第一代产品的代表厂商包括Yahoo、Excite、Infoseek、AltaVista等,并从典型的目录式分类结构(如Yahoo)发展到全文搜索引擎(如AltaVista)。目前,搜索引擎的使用已成为排在收发电子邮件之后的第二大互联网应用技术。

1.1 搜索引擎的类别

按照搜索引擎提供的功能和使用的技术来分,目前因特网上的搜索引擎大致可以分成以下5类。

1.1.1 普通搜索引擎 此类搜索引擎是利用网络流浪者(Web Crawlers)、网络蜘蛛(Spider)、机器人(Robot)对因特网资源进行索引,一般无须人工干预。所谓的网络蜘蛛

是一个程序,通过自动读取一篇文档遍历其超链接结构,从而递归获得被引用的所有文档。不同的搜索引擎搜索的内容不尽相同:有些着重于站点搜索;有些搜索包括Gopher、新闻组、Email等。搜索引擎的性能主要取决于索引数据库的容量、存放内容、更新速度、搜索速度、用户界面的友好程度以及是否易用等。这类搜索引擎国外的有:AltaVista (<http://www.altavista.com>), Excite (<http://www.excite.com>), Infoseek (<http://www.go.com>), Lycos (<http://www.lycos.com>), Google (<http://www.google.com>)等;国内的有:中文Excite (<http://chinese.excite.com>), 中文雅虎 (<http://www.cn.yahoo.com>), 搜狐 (<http://www.sohu.com>), 悠游 (<http://goyoyo.com>), 中文Google (<http://www.google.com>)等。

1.1.2 元搜索引擎(Meta-Search Engine) 这类搜索引擎实际是搜索引擎之上的搜索引擎,也称为集成搜索引擎、多搜索引擎、索引搜索引擎等。这类搜索引擎可以接受用户的一个搜索请求,然后将该请求转交给其他若干个搜索引擎同时处理,最后对多个引擎的搜索结果进行整合处理后返回给查询者。整合处理包括消除重复,对多个引擎的结果进行排序等,这类搜索引擎的优点是返回结果的信息量更大、更全,能够在尽可能短的时间内提供相对全面、准确的信息,而且即使不能完全满足用户需求,仍可以作为相对可靠的参考源进行扩展搜索,因此成为倍受推崇的首选检索入口。例如All4One (<http://all4one.com/>), WebCrawler (<http://www.webcrawler.com>), InfoMarket (<http://www.looksmart.com>)和FastFind (<http://www.symantec.com/iff/>)等就是这样的搜索引擎。

搜索引擎与元搜索引擎的主要区别在于搜索引擎拥有独立的网络资源采集索引机制和相应的数据库,而元搜索引擎一般没有自己独立的数据库,却更多地是提供统一联结界面(或进一步提供统一检索方式和结果整理),形成一个由多个分布的具有独立功能的搜索引擎构成的虚拟整

* 本文为国家自然科学基金资助项目“信息资源管理与信息资源管理思想史”研究分报告之一,项目号:79870013。

体, 用户通过元搜索引擎的功能实现对这个虚拟整体中各独立搜索引擎数据库的查询显示等一切操作。元搜索引擎中各独立搜索引擎被称为目标搜索引擎或者成员搜索引擎, 它们各自保持其原来的局部数据模式和自己的检索指令; 元搜索引擎给出一个全局外部模式, 用以接受用户检索输入和结果输出。

1.1.3 智能搜索引擎 (Smart Hunter) 智能搜索引擎是根据目前搜索引擎的发展趋势, 除提供传统的全网快速检索、相关度排序等功能外, 还提供用户角色登记、用户兴趣自动识别、内容的语义理解、智能化信息过滤和推送等功能, 为用户提供了一个真正个性化、智能化的网络工具。智能搜索引擎服务端由多台分布式服务器组成, 查询信息响应时间小于 1 秒, 查全率指标相对较高, 查准率接近人工分类的搜索引擎的准确率。客户端通过专用浏览器或插件, 自动识别用户兴趣, 通过智能搜索引擎按用户兴趣搜索、过滤、推送信息, 并可实现网上兴趣相近用户查找。搜索引擎缺乏知识处理能力和理解能力, 对要检索的信息仅仅采用机械的关键词匹配来实现。智能搜索引擎能把目前基于关键词层面检索提高到基于知识 (或概念) 层面。目前国内较先进的智能搜索引擎有: 世纪永联 (<http://www.softhouse.com.cn>)、百度 (<http://www.baidu.com>)、尤里卡 (<http://www.ulika.com>) 等。

1.1.4 个性化的网上搜索引擎代理^[1] 此类搜索引擎最主要的特征是具有学习功能, 能够在信息交互中获得用户的信息, 包括用户的兴趣、爱好和思维方式。在此前提下, 系统可以主动、定期地为用户查找信息, 并根据用户搜索信息的变化调整“知识库”中的通用字和关键词, 使之能够有效适用专门领域的信息搜索。系统的本地信息库还可以对搜索到的信息进行分类存储和管理, 并具有与其他系统的协作功能, 扩展灵活, 维护方便。目前此类搜索引擎在搜索引擎中所占比例很小, 但其发展代表着网络信息资源检索的一个方向, 已经得到业界的重视, 其主要代表为 Agent (www.agent.com)。

1.1.5 专用搜索引擎 如果用户想查找电话、人名、电子邮件、地址等资料, 就可以使用专用搜索引擎。这些搜索引擎专门搜集某一类的信息资源, 内容丰富, 数据量大, 能帮助用户迅速找到一些专门的信息。譬如人物搜索、旅行路线搜索和产品搜索等, 这些搜索都依赖于具体的数据库。例如专门提供有关域名注册信息、域名有效性测试等服务的域名搜索引擎 Domain Notes (<http://wd-vl.互联网.com/互联网/domains>); 专门搜索网上免费资源的搜索引擎 AAATreeStuff (<http://www.laaafreestuff.com>), AbsolutelyFreebies (<http://www.absolutelyfreebies.com>), Allfreesites (<http://www.allfreesites.com>); 地图搜索引擎 Map-

blast (<http://www.mapblast.com>), 图行天下 (<http://www.go2map.com>), 城市交通旅游电子地图 (<http://www.afmap.com.cn>) 等。

搜索引擎的其他分类方法还有^[2]: 按照自动化程度分为人工与自动引擎; 按照是否具有智能功能分为智能与非智能引擎; 按照搜索内容分为文本搜索引擎、语音搜索引擎、图形搜索引擎、视频搜索引擎等。

1.2 搜索引擎的搜索原理

搜索引擎主要由 4 部分组成: 搜索器、索引器、检索器、用户接口。搜索器的功能是在因特网中发现和搜集信息。它还通过网络流浪者、网络蜘蛛、机器人等相关技术, 根据网页链接进行搜索, 在网上自动抓取和分析被它找到的网页信息, 并将其加入到索引数据库中。它要尽可能快、尽可能多地搜集各种类型的新信息, 同时还要定期更新已有信息, 避免死链接。索引器的功能是理解搜索器所搜索的信息, 从中抽取索引项, 用于表示文档以及生成文档库的索引表, 建立起自己的物理索引数据库。一个搜索引擎有效性在很大程度上取决于索引的质量。检索器的功能是根据用户的查询在索引库中快速检出文档, 进行文档与查询的相关度评价, 对将要输出的结果进行排序, 并实现某种用户相关性反馈机制。用户接口的作用是输入用户查询、显示查询结果, 提供用户相关性反馈机制。

搜索引擎的检索思想起源于传统的信息全文检索理论, 即计算机程序通过扫描每一篇文章中的每一个词, 建立以词为单位的倒排文件, 检索程度根据检索词在每一篇文章中出现的频率和每一个检索词在一篇文章中出现的概率, 对包含这些检索词的文章进行排序, 最后输出排序的结果。全文检索技术是搜索引擎的核心支撑技术。

目前的搜索引擎一般使用两种技术来实现信息检索: 一是使用网站分类技术, 即把网站进行树状归类, 登录的网站至少属于一个类别, 对每个站点都有简略的描述, 雅虎即是如此。二是使用全文检索技术, 处理的对象是文本, 它能够对大量文档 (这里是大量网页数据) 建立由字 (词) 到文档的倒排索引, 在此基础上, 用户使用关键词来对文档 (网页) 进行查询时, 系统将给用户返回含该关键词的网页。

一般说来, 无论哪一种搜索引擎都是通过某种界面跟用户交互, 接受用户查询特定信息的请求, 然后对用户查询请求进行分析。譬如将查询请求分解成若干关键词, 在分析用户请求之后, 在索引数据库中不断进行匹配, 挑出符合条件的信息, 同时按照匹配程度的高低对结果进行排序, 最后将排序后的结果返回给用户。因为网络信息时刻变动, 所以搜索引擎在后台通过网络流浪者、网络蜘蛛、机器人等漫游因特网, 收集网络信息, 自动对收集到的信

息进行分析,并按照一定的格式,将收集到的信息保存到本地索引数据库中。

用户最关心的是搜索结果能否满足自己的需要,尤其是在搜索引擎可以获得的信息资源非常多的情况下。目前,搜索引擎仍不能很好地理解人的查询请求,所以,现在的搜索引擎采用的常见策略是将用户的查询请求分解成若干关键词,然后再根据这些关键词计算 Web 文档与用户请求的匹配程度,从而挑出若干匹配的文档。匹配程度的衡量准则很多:一种是根据关键词在文档中出现的频率确定它对用户请求的匹配程度,另一种是计算关键词出现次数和页面总词数之比。不同的搜索引擎采取的匹配策略是不同的。Google 是一个表现出色的搜索引擎,在匹配用户请求的过程中,它考虑到了关键词的频率、位置甚至格式等信息以衡量文档对用户请求的匹配程度。

但是基于关键词匹配的搜索技术有较大的局限性。首先,它不能区分同形异义;其次,不能联想到关键词的同义词。目前有一种有生命力的搜索引擎技术就是基于内容的搜索引擎。基于内容的搜索不是根据字形,而是试图理解用户的请求,同时根据文档的内容选出符合用户要求的文档。这里所谓的理解包括用户查询的理解和文档内容的理解,它允许用户以非常自然的形式提出查询请求,采取语义网络、汉语分词技术等分析用户的请求,了解用户真正的需求。

1.3 搜索引擎与传统信息资源检索的比较

基于因特网的搜索引擎与传统的信息资源检索相比,有如下特点。

1.3.1 信息服务的综合性 由于采用了知识库导航,搜索引擎将面貌一新,给用户提供更全面、更综合的信息服务,在这里,信息检索只是信息服务的一部分。

1.3.2 信息服务的智能性 所谓“智能来自知识”,有综合知识库作为背景,信息检索和导航服务将更智能化。因特网中的语言层面知识有助于解决前面提到的“表达差异”问题。例如,只要定义“计算机”、“电子计算机”、“电脑”是同义关系就可以消除用户由于使用不同的词表达同一概念而带来的检索困难。另一方面,根据因特网的常识性和本体论层次知识对用户的查询进行相关性联想,提供引导用户进行下一步查询的线索,这样一步一步地在与用户的交互过程中诱导用户“表达”出他真正想找的东西,从而实现查询的智能导航。这种逐步求精的策略解决了信息检索“忠实表达”的难题。至于多语种信息检索问题,也可在语言层面来解决。

1.3.3 信息服务的个性化 因特网的知识库可以存放与具体用户相关的知识(用户的专业兴趣、购买力等),搜索引擎将利用这些知识来为用户提供个性化的信息服务。

1.3.4 具有支持代理(Agent)的能力 由于 Web 服务器端有综合性知识库,为智能代理的活动提供了基础。例如,活动在客户端的智能代理可对用户正在浏览的网页进行主动观察、分析内容,根据服务器端的知识库来推荐内容相近的其他网页供用户参考。

2 衡量搜索引擎的基本指标

经过几年的发展,搜索引擎的检索从最初的字符串匹配程序已经演进到能对超大文本、语音、图像、活动影像等非结构化数据进行综合管理的大型软件,衡量搜索引擎的基本指标也逐渐形成规范。

1) 查全率是系统在进行某一检索时,检索出的相关资料量与系统资料库中相关资料总量的比率。目前,查全率的概念在逐渐弱化,这是因特网信息资源在量上激增所致。

2) 查准率是系统在进行某一检索时,检索出的有用资料数量与检索出资料总量的比率,即搜索得到的信息与用户所要求的信息相关性的程度,查准率直接影响到搜索的速度和费用。为提高搜索的准确度,大多数的英文搜索引擎都支持逻辑查询。用户可以用多个单词,加上适当的逻辑字符来缩小搜索范围,从而显著提高搜索结果的准确度。

3) 检索速度或者说响应时间是提高工作效率的保障,指的是从提交检索命令到查出资料结果所需的时间。最基本的检索速度是应该达到“千万汉字,秒级响应”。

4) 收录范围是指搜索引擎所查找的范围。造成搜索引擎覆盖面差距的原因是多方面的,如网络带宽、磁盘容量、计算能力等。当然,搜索引擎收录范围的大小仍然是衡量其功能的一个重要指标。

5) 死链接。普通搜索引擎总有些搜索结果是点不进去的,少到 1%~2%,多到 8%~9%,这也常被用作评测条件之一。著名的搜索引擎 Google 使用了网页快照功能解决死链接问题,就算搜索结果中的那个网站已关闭,用户还是可以看到 Google 自己储存的网页。

6) 用户负担是指用户在检索过程中付出精力的总和,任何妨碍和延迟用户到达最终搜索结果的形式都属于用户负担的范畴。包括搜索界面的简洁、搜索结果描述、搜索结果描述显示。

其他评价标准还有,是否支持本目录下搜索,索引数据库更新时间长短,搜索引擎的稳定性,对高级搜索的支持能力强弱等。

3 当前搜索引擎发展面临的主要问题

当前搜索引擎发展面临的主要问题包括以下几个方

面。

1) 网络信息量迅猛增加,人工无法对它们进行有效的分类、索引和利用。劳伦斯博士(S. Lawrence)和贾尔斯(G. L. Giles)曾对 AltaVista、Excite、HotBot、InfoSeek、Lycos、Northern Light、EuroSeek、Google、Microsoft、Snap、Yahoo 等搜索引擎收集的大量试验数据,利用统计方法进行比较、分析和评价。研究结果宣称,1997年公众搜索引擎包含至少3.2亿个网页,最大的搜索引擎覆盖率估计为网页总量的1/3^[3]。1999年他们利用随机抽样估计出网络服务器总数为1.6亿个^[4],而公众搜索引擎则为8亿页,两年中翻了2.6倍。他们还估计出因特网上的影像资料数量约为 3^{1012} 字节,公众搜索页面上有1.8亿个图像。同时搜索引擎的搜索结果有重叠部分,引擎的不同覆盖组合之和为3.35亿页,是估计总页数的42%。显然,不可能有一个包罗万象的搜索引擎将因特网上的信息资源全部覆盖。

2) 网络信息组织的多样性发展。因特网用户面对的是非常多的随机的未组织的信息,其无序程度是非常大的。从如此庞杂的信息海洋中取出对用户最有用的信息是搜索引擎面临的一项挑战,而信息的有序化组织也是搜索引擎高效工作的前提。

3) 信息资源评价困难。网络信息资源的开发水平差距颇大,这与信息发布者的素质有很大的关系。因此,采取什么样的评价标准来衡量网络信息资源的价值,是目前网络信息资源检索所要面对的一个大问题。一些网站在网页中大量重复某些关键词,使得容易被某些著名的搜索引擎选中,以期借此提高网站的地位,但事实上却可能没有提供任何对用户有价值的信息,这更加深了评价信息有用性的难度。

4) 网络信息资源的更新速度迅速。用户在查询信息时,总是期望挑出最新的信息,然而网络信息时刻变动,实时搜索几乎不可能,就是刚刚浏览过的网页也随时都有更新、过期、删除的可能。主要的引擎供应商都声称他们的引擎能够跟上整个因特网的复杂和膨胀速度,但是,因特网是分布式的、动态的、快速增长的信息资源,这对于传统的信息收集技术来讲有着不可克服的困难。传统的信息收集技术是为不同环境设计,并代表性地用来索引一个静态的且可以直接访问的文档。而因特网的发展为信息检索带来的重要问题就是,搜索引擎的集中化架构是否能跟上文档的扩张速度;搜索引擎是否能规律地及时更新他们的数据库,以便检测那些已修改过的、已删除的、已重新定位的信息。

5) 信息媒体多样化。迄今为止,虽然图像、声音、气味的检索技术都已经出现,但是在搜索引擎中几乎没有

任何应用,网络信息资源检索的主要对象仍然是文本。多媒体技术的发展对搜索引擎提出了更多的要求,人们期望引擎不仅能挑出自己需要的文章,还能挑出自己所关心的图片、电影、音乐等。

6) 带宽等其他因素的制约^[5]。搜索引擎的关键问题之一就是如何进行网络信息的收集与整理,也就是如何将网络信息有序化。为此,搜索引擎需要定期不断地访问网络资源。然而,遍历如此庞杂的网络本身就是一件非常困难的事情。目前网络带宽不足,网络速度不够理想,使得搜索引擎搜索网络资源的速度较慢。

因此,传统的搜索引擎不能适应网络信息资源检索的高速发展,新一代智能化搜索引擎作为一种高效搜索引擎技术,在网络信息时代将日益引起人们的关注。

4 网络信息资源检索中搜索引擎的发展方向

新一代智能搜索引擎设计追求的目标是根据用户的要求,从可以获得的网络信息资源中检索出对用户最有价值的信息。目前的理论研究认为,针对网络信息资源检索的要求,新一代的智能搜索引擎将向以下几个主要方向发展。

1) 注重网站的重要性。所谓重要性是指网站内容的丰富性和准确性,并能够满足用户的需求。搜索引擎根据网站重要性的等级进行排序,并向用户返送这些信息。一般的搜索引擎是根据一个关键词在主页中的位置和出现频率来计分和排序,而智能搜索引擎则是根据以往用户实际访问一个网站并在该网站上所花费的时间来确定一个网站的重要性,或者根据一个网站被其他网站链接的数量来确定网站的重要性等级。这种根据用户忠诚度的评判方法更具客观性,因而用户所获得的信息也就更准确。智能搜索引擎是根据用户访问情况来认定一个网站重要性的。

2) 为特定用户提供相关信息。智能搜索引擎能通过观察用户的行为,了解用户的兴趣爱好,另外能通过不断的训练学习增长智能。每次用户对引擎返回的信息进行评价,智能引擎根据用户的评价调整自己的行为,智能搜索引擎还能对搜索结果进行合理解释。智能搜索引擎具有主动性,可以在任何特定的时候(如用户最关心的信息发生了某种变化的时候)用各种方法与用户取得联系,这些方法包括电子邮件、电话、传真、寻呼机、移动电话等。搜索引擎还可根据用户特定时刻的位置信息,选择恰当的方法跟用户通信。

3) 专用(业)搜索引擎在资源检索中的地位逐渐重要^[6]。近年来专用搜索引擎发展很快,其作用和功能是综合性搜索引擎所不能替代的。专用搜索引擎针对性强、目

(下转第320页)

在所有场合都能奏效。尽管科斯定理有一定的局限性，但它指出了产权界定与外部效应的关系，指导人们通过明确界定产权来促进资源配置的“帕累托最优”。在图书市场，由于图书具有明显的“效果外溢”，所以版权制度较好地界定和保护了知识产权，促进了图书市场的出版效率。

3) 罚款和征税。政府的政策能够解决由外部效应而导致的失灵问题。在负外部效应存在的情况下，如果政府向有关企业或个人课以相当于他所造成的边际外部成本的罚款或税收，外部成本就成了有关当事人的内部成本，他们在决策时就会考虑到这些成本，避免效率损失。以盗版软件市场为例，由于盗版软件强加了外部成本给正版软件生产厂商，于是产生了负外部效应。政府对盗版软件生产厂商总是实行没收或罚以重款。这使得盗版软件供给曲线上移到 S_1 ，从而减少盗版软件生产的市场均衡产量（见图6）。

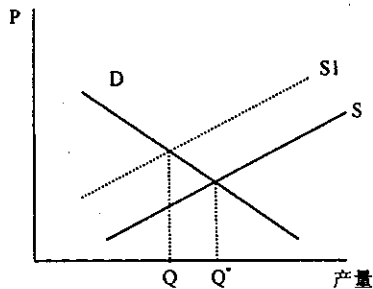


图6 对负外部效应的纠正——罚款

(上接第292页)

标明确、查准率高，是从事专业信息咨询人员的重要工具，也是用户在因特网上进行专业信息查询的好帮手。

4) 爬行软件的智能化。爬行软件（如网络蜘蛛、机器人等）通过启发式学习并采取最有效的搜索策略，选择最佳时机获取从因特网上自动收集、整理的信息。众所周知，信息动态更替无时无刻不在进行，即使在搜索过程中文档会被添加、删除、改变。因此，智能引擎将通过智能化的爬行软件，自动完成在线信息的索引。

5) 支持跨平台多语言的文档处理。智能搜索引擎应具有跨平台工作和处理多种混合文档结构的能力。譬如既能处理 HTML，又能处理 SGML 和 XML 文档以及其他类型的文档，如 Word、WPS 等。

6) 多媒体和图像搜索引擎。严格地说，多媒体和图像搜索引擎亦属于专用搜索范畴，只是由于所检索对象的特殊性，所以其技术实现要复杂得多。多媒体搜索引擎可以检索因特网上的音乐、电影和电台播音节目。最近出现的图像搜索引擎则在检索技术方面有较大的改善，提高了图像查准率，同时还开发了一些新的功能，方便用户使用。

7) 搜索引擎人机接口的智能化^[7]。智能搜索引擎可

4) 补贴。在正外部效应存在的情况下，为了扩大实际消费量，政府一方面可以直接提供准公共产品，另一方面可以给私人企业提供补贴，降低企业的边际生产成本，从而使企业在一定的价格下扩大供给量，达到纠正市场失灵的目的。我们继续以《钢》为例。由于《钢》具有正外部效应，所以如果政府对每本书补贴9元，也即要求该书定价为6元，则将使该书的需求曲线向上移至 D_1 ，从而使市场均衡产量达到帕累托最优的 Q 点（见图7）。

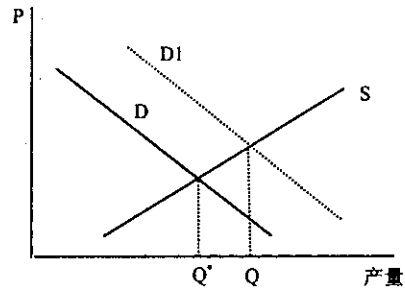


图7 对正外部效应的纠正——补贴

参考文献

- 1 金格马. 信息经济学. 马费成, 袁红译. 太原: 山西经济出版社, 1999
- 2 马费成. 信息经济学. 武汉: 武汉大学出版社, 1997
- 3 周惠中. 微观经济学. 上海: 上海人民出版社, 1997
- 4 李纲. 市场信息学. 武汉: 武汉大学出版社, 1996
- 5 樊勇明, 杜莉. 公共经济学. 上海: 复旦大学出版社, 2001
- 6 斯蒂格利茨. 经济学上册. 第二版. 北京: 中国人民大学出版社, 2001

以通过自然语言与用户交互。它采取诸如语义网络等智能技术，通过汉语分词、句法分析以及统计理论有效地理解用户的请求，甚至能体会出用户的弦外之音，最大程度地了解用户的需求。

综上所述，在网络信息资源检索的研究中，搜索引擎的研究具有非常重要的意义，随着信息技术和新一代因特网的发展，搜索引擎的研究和应用也将迈上一个新的台阶，我们期待着搜索引擎的发展为网络信息资源检索带来一个美好的明天。□

参考文献

- 1 潘登登. 各类搜索引擎说明. 电脑报, 2000 (49)
- 2 张蕊. 发展中的搜索引擎模式. Available from: <http://itsearch.ccidnet.com/info/info.htm>
- 3 Lawrence S, Giles G L. Science, 1998, 280
- 4 Lawrence S, Giles G L. Nature, 1999, 400
- 5 王卫国, 万跃华. 搜索引擎的六大发展趋势. Available from: <http://sanf.3322.net/seek/seekgengduo.htm>
- 6 李爱红. 网络搜索引擎的比较研究. 中国信息导报, 1999 (1)
- 7 郑茵. 下一代因特网研究概述. Available from: <http://www.jl.jl.cn/jlweb/school/index.html>

作者简介: 王进孝, 男, 1968年生, 博士生。发表论文数篇。
收稿日期: 2001-08-06