高等院校信息技术规划教材

搜索引擎基础教程

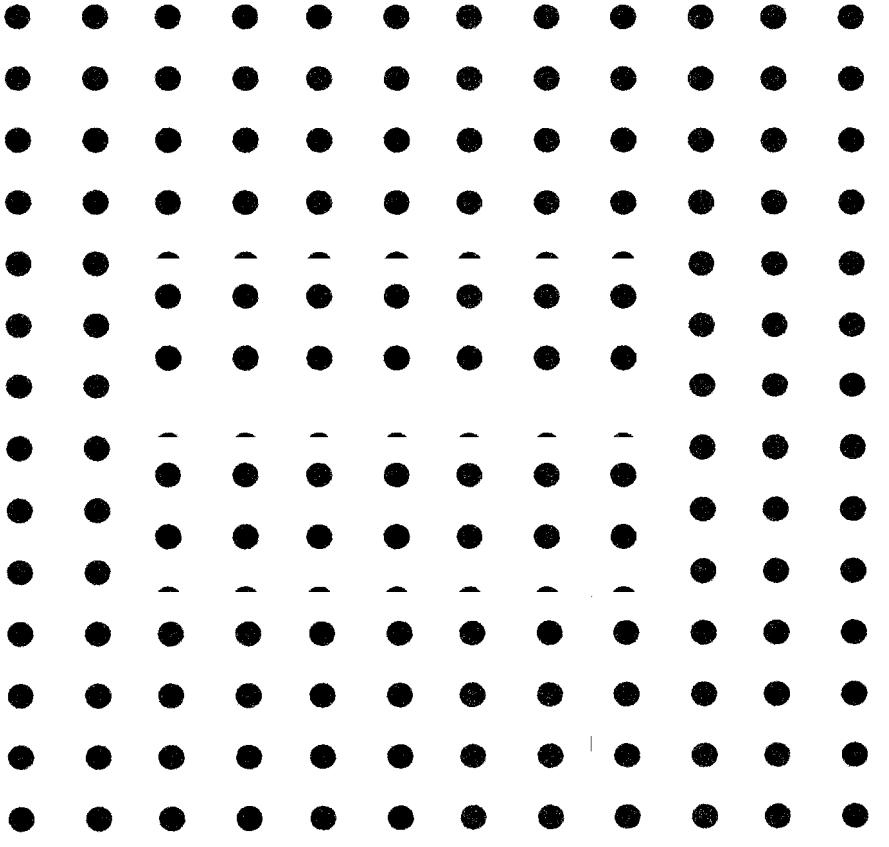
袁津生 李群 主编

清华大学出版社

高等院校信息技术规划教材

搜索引擎基础教程

袁津生 李群 主编



清华大学出版社
北京

内 容 简 介

本书从教学的角度出发，对搜索引擎的原理及开发技术进行了全面的介绍，内容包括搜索引擎的基本原理、网页抓取技术、信息预处理技术、信息索引技术、信息查询技术和多媒体信息检索技术。另外，本书还对搜索引擎开发技术进行了详细的讨论。

本书适合高等院校计算机科学与技术专业及相关专业的高年级学生和研究生阅读参考，也适合相关领域的工程技术人员参阅。

图书在版编目（CIP）数据

搜索引擎基础教程 / 袁津生，李群主编．—北京：清华大学出版社，2010.7
（高等院校信息技术规划教材）
ISBN 978-7-302-22049-7

Ⅰ．①搜…　Ⅱ．①袁…②李…　Ⅲ．①互联网络－情报检索－高等学校－教材
Ⅳ．①G354.4

中国版本图书馆 CIP 数据核字(2010)第 026117 号

责任编辑：战晓雷　徐跃进
责任校对：时翠兰
责任印制：杨　艳

出版发行：清华大学出版社　　**地　　址**：北京清华大学学研大厦 A 座
http://www.tup.com.cn　　**邮　　编**：100084
社　总　机：010-62770175　　**邮　购**：010-62786544
投稿与读者服务：010-62795954，jsjjc@tup.tsinghua.edu.cn
质　量　反　馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：三河市春园印刷有限公司
经　　销：全国新华书店
开　　本：185×260　　**印　　张**：21　　**字　　数**：505 千字
版　　次：2010 年 7 月第 1 版　　**印　　次**：2010 年 7 月第 1 次印刷
印　　数：1～3000
定　　价：29.50 元

产品编号：033609-01

编者的话
EDITOR'S NOTE

网络改变了人们的思维，搜索改变了人们的生活。面对浩如烟海的网络资源，搜索引擎就好像是航船的指南针，引领着人们在网络中寻找自己想要的信息。不论是办公室工作人员、在校学生，还是科学研究人员，使用搜索引擎查询信息几乎成为每日必做的一件事情，搜索引擎已经成为人们的一项新的生活内容。

为了适应目前形势的发展，各个高校先后都开设了搜索引擎这门课程。我们编写这本书的目的就是系统地讨论和研究搜索引擎的基本理论，学会构建自己的搜索引擎。

全书较为系统地阐述了搜索引擎的基本概念以及相关的技术，总共分为 9 章。第 1 章全面地介绍了搜索引擎的概念、搜索引擎的发展、分类及建立搜索引擎的关键技术；第 2 章讨论了搜索引擎的体系结构、工作原理、搜索引擎的数据结构、元搜索引擎以及职能搜索引擎的概念；第 3 章介绍了网页抓取技术，主要内容包括搜索引擎爬虫的工作原理、爬虫使用的关键技术和 Robots 协议；第 4 章介绍了网页信息预处理技术，主要内容有网页信息结构化、文本处理技术、中文分词技术和 PageRank 算法；第 5 章介绍了信息索引技术，主要包括顺排检索、倒排索引、后缀数组索引和文本压缩技术；第 6 章介绍了信息查询与评价技术，主要包括信息检索的模型、常用的检索方法、查询服务以及相关性的评价和查全率和查准率等内容；第 7 章介绍了多媒体信息检索的基本概念，主要内容有多媒体的基本概念、多媒体数据的压缩、多媒体内容的理解以及多媒体信息检索的关键技术；第 8 章介绍了基于 Lucene 的搜索引擎开发技术，主要内容有搜索引擎开发实例简介、环境的搭建与配置、网页搜集技术、网页预处理技术和查询服务；第 9 章介绍了基于 Nutch 的搜索引擎开发技术，主要内容有 Nutch 简介、环境的搭建与配置、Nutch 的初始配置及运行、开发自己的搜索引擎平台。

希望本书的出版能够对搜索引擎的设计者、Web 站点的管理员以及广大用户有所帮助，也希望它成为搜索引擎和信息检索有关领域的学生学习的参考书。

本书是作者在多年教学的基础上，参考若干资料整理而成的。本书对基本概念、基础知识的介绍力求简明扼要；各章相互配合并附有小结和习题，同时还有相关的实验。建议本课程为 40 学时，其中讲课 30 学时，实验 10 学时。

本书由袁津生、李群、蔡岳、程超然和张帆共同编写，其中，蔡岳和张帆编写了本书的第 8 章、程超然编写了本书的第 9 章、李群编写了本书的第 7 章并校阅了全部书稿。由于作者水平有限，书中难免有许多错误和不当之处，请读者批评指正。

编者

2010 年 3 月

目录 CONTENTS

第1章 搜索引擎概述

随着互联网的飞速发展，人们越来越依靠网络来查找他们所需要的信息。但是，由于网上的信息源数不胜数，所以如何有效地去发现我们所需要的信息，就成为一个很关键的问题。为了解决这个问题，搜索引擎就随之诞生。搜索引擎自从出现就创造了一个个发展奇迹。搜索引擎虽然只有10多年的历史，但是在Web上已经有了不可或缺的地位。在近些年来搜索引擎发展尤为迅猛，百度2005年在纳斯达克成功上市，Google在全球市场突飞猛进。搜索引擎的开发爱好者也形成了浩大的队伍，仅在开源社区SourceForge上，搜索引擎的项目就有将近10 000项。搜索引擎得到了前所未有的关注。

搜索引擎并不是一个完全创新的系统，而是借鉴了以往全文检索系统和网络软件系统开发而成的。搜索引擎采用了以往产品的很多技术和思路，尤其是继承了很多信息检索系统的技术和方法。互联网搜索引擎在继承历史技术的同时，针对互联网信息处理的特点，开发出了互联网信息查找工具。

本章主要介绍搜索引擎的概念、搜索引擎的发展史、搜索引擎的分类以及一些著名的搜索引擎。

1.1 搜索引擎的概念、原理及历史与发展

搜索引擎是指根据一定的策略、运用特定的计算机程序搜集互联网上的信息，在对信息进行组织和处理后，并将处理后的信息显示给用户的为用户提供检索服务的系统。

1.1.1 搜索引擎的概念

从使用者的角度看，搜索引擎提供一个包含搜索框的页面，在搜索框输入词语，通过浏览器提交给搜索引擎后，搜索引擎就会返回和用户输入的内容相关的信息列表。

互联网发展早期，以雅虎为代表的网站分类目录查询非常流行。网站分类目录由人工整理维护，精选互联网上的优秀网站，并简要描述，分类放置到不同目录下。用户查询时，通过一层层的点击来查找自己想找的网站。也有人把这种基于目录的检索服务网站称为搜索引擎，但从严格意义上讲，它并不是搜索引擎。

搜索引擎并不真正搜索互联网，它搜索的实际上是预先整理好的网页索引数据库。真正意义上的搜索引擎，通常指的是收集了互联网上几千万到几十亿个网页并对网页中的每

一个词(即关键词)进行索引,建立索引数据库的全文搜索引擎。当用户查找某个关键词的时候,所有在页面内容中包含了该关键词的网页都将作为搜索结果被搜出来。在经过复杂的算法进行排序后,这些结果将按照与搜索关键词的相关度高低,依次排列。

现在的搜索引擎已普遍使用超链分析技术,除了分析索引网页本身的内容,还分析索引所有指向该网页的链接的URL、锚文本,甚至链接周围的文字。所以,有时候,即使某个网页 A 中并没有某个词,比如"信息检索",但如果有网页 B 用链接"信息检索"指向这个网页 A,那么用户搜索"信息检索"时也能找到网页 A。而且,如果有越多网页的"信息检索"链接指向网页 A,那么网页 A 在用户搜索"信息检索"时也会被认为更相关,排序也会越靠前。

1.1.2 搜索引擎的原理

搜索引擎的原理,可以分为4步:从互联网上抓取网页、建立索引数据库、在索引数据库中搜索排序、对搜索结果进行处理和排序。

1. 从互联网上抓取网页

利用能够从互联网上自动收集网页的Spider系统程序,自动访问互联网,并沿着任何网页中的所有URL爬到其他网页,重复这过程,并把爬过的所有网页收集回来。

2. 建立索引数据库

由分析索引系统程序对收集回来的网页进行分析,提取相关网页信息(包括网页所在URL、编码类型、页面内容包含的关键词、关键词位置、生成时间、大小、与其他网页的链接关系等),根据一定的相关度算法进行大量复杂计算,得到每一个网页针对页面内容中及超链中每一个关键词的相关度(或重要性),然后用这些相关信息建立网页索引数据库。

3. 在索引数据库中搜索排序

当用户输入关键词搜索后,由搜索系统程序从网页索引数据库中找到符合该关键词的所有相关网页。因为所有相关网页针对该关键词的相关度早已计算好,所以只需按照现成的相关度数值排序,相关度越高,排名越靠前。最后,由页面生成系统将搜索结果的链接地址和页面内容摘要等内容组织起来返回给用户。

4. 对搜索结果进行处理和排序

所有相关网页针对该关键词的相关信息在索引库中都有记录,只需综合相关信息和网页级别形成相关度数值,然后进行排序,相关度越高,排名越靠前。最后由页面生成系统将搜索结果的链接地址和页面内容摘要等内容组织起来返回给用户。

1.2 搜索引擎的历史与发展趋势

搜索引擎至今已经经历了三代发展阶段:

第一代搜索引擎出现于1994年,主要特征为集中式检索。这类搜索引擎一般都索引少于1百万个网页,极少重新搜集网页并去刷新索引,而且其检索速度非常慢,一般都要等待

数10秒甚至更长的时间。在实现技术上也基本沿用较为成熟的信息检索、网络、数据库等技术，相当于利用一些已有技术实现的一个WWW上的应用。

第二代搜索引擎系统大约出现在1996年，大多采用分布式检索方案，即多个微型计算机协同工作来提高数据规模、响应速度和用户数量。它们一般都保持一个大约5千万网页的索引数据库，每天能够响应1千万次用户检索请求。

第三代搜索引擎系统出现在1998年到2000年间，这一时期是搜索引擎空前繁荣的时期。第三代搜索引擎的发展有以下几个特点：

(1) 索引数据库的规模继续增大，一般的商业搜索引擎都保持在几千万甚至上亿个网页。

(2) 除了一般意义上的搜索以外，开始出现主题搜索和地域搜索，很多小型的垂直门户站点开始使用该技术。

(3) 由于搜索返回数据量过大，检索结果相关度评价成为研究的焦点。

1.2.1 搜索引擎的发展史

在互联网发展初期，网站相对较少，信息查找比较容易。然而伴随互联网爆炸性的发展，普通网络用户想找到所需的资料简直如同大海捞针，这时为满足大众信息检索需求的专业搜索网站便应运而生了。

现代意义上的搜索引擎的祖先，是1990年由蒙特利尔大学学生Alan Emtage发明的Archie。虽然当时World Wide Web还未出现，但是网络中文件的传输还是相当频繁的，由于大量的文件散布在各个分散的FTP主机中，查询起来非常不便，因此Alan Emtage想到了开发一个可以文件名查找文件的系统，于是便有了Archie。Archie是第一个自动索引互联网上匿名FTP网站文件的程序，但它还不是真正的搜索引擎。Archie是一个可搜索的FTP文件名列表，用户必须输入精确的文件名搜索，然后Archie会告诉用户哪一个FTP地址可以下载该文件。

由于Archie深受欢迎，受其启发，Nevada System Computing Services大学于1993年开发了一个Gopher(Gopher FAQ)搜索工具Veronica(Veronica FAQ)。Jughead是后来另一个Gopher搜索工具。

Robot(机器人)一词对编程者有特殊的意义。Computer Robot是指某个能以人类无法达到的速度不断重复执行某项任务的自动程序。由于专门用于检索信息的Robot程序像蜘蛛(Spider)一样在网络间爬来爬去，因此，搜索引擎的Robot程序被称为Spider(Spider FAQ)程序。世界上第一个Spider程序，是MIT Matthew Gray的World Wide Web Wanderer，它用于追踪互联网发展规模。刚开始它只用来统计互联网上的服务器数量，后来则发展为也能够捕获网址(URL)。

与Wanderer相对应，1993年10月Martijn Koster创建了ALIWEB(Martijn Koster Announces the Availability of Aliweb)，它相当于Archie的HTTP版本。ALIWEB不使用网络搜寻Robot，如果网站主管们希望自己的网页被ALIWEB收录，需要自己提交每一个网页的简介索引信息，类似于后来大家熟知的雅虎。

随着互联网的迅速发展，使得检索所有新出现的网页变得越来越困难，因此，在Wanderer基础上，一些编程者将传统的Spider程序工作原理做了些改进。其设想是，既然

所有网页都可能有连向其他网站的链接，那么从一个网站开始，跟踪所有网页上的所有链接，就有可能检索整个互联网。到 1993 年底，一些基于此原理的搜索引擎开始纷纷涌现，其中最负盛名的 3 个是：Scotland 的 JumpStation、Colorado 大学 Oliver McBryan 的 WWW Worm(First Mention of McBryan's World Wide Web Worm)、NASA 的 Repository-Based Software Engineering (RBSE) Spider。JumpStation 和 WWW Worm 只是以搜索工具在数据库中找到匹配信息的先后次序排列搜索结果，因此毫无信息关联度可言。而 RBSE 是第一个索引 HTML 文件正文的搜索引擎，也是第一个在搜索结果排列中引入关键字串匹配程度概念的引擎。

Excite 的历史可以上溯到 1993 年 2 月，6 个 Stanford 大学的学生的想法是分析字词关系，以对互联网上的大量信息进行更有效的检索。到 1993 年中，这已是一个完全投资项目 Architext，他们还发布了一个供网站管理员在自己网站上使用的搜索软件版本，后来被叫做 Excite for Web Servers。Excite 后来曾以概念搜索闻名，2002 年 5 月，被 Infospace 收购的 Excite 停止自己的搜索引擎，改用元搜索引擎 Dogpile。

1994 年初，Washington 大学的学生 Brian Pinkerton 开始了他的小项目 WebCrawler (Brian Pinkerton Announces the Availability of WebCrawler)。1994 年 4 月 20 日，WebCrawler 正式亮相时仅包含来自 6000 个服务器的内容。WebCrawler 是互联网上第一个支持搜索文件全部文字的全文搜索引擎，在它之前，用户只能通过 URL 和摘要搜索，摘要一般来自人工评论或程序自动取正文的前 100 个字。后来 WebCrawler 陆续被 AOL 和 Excite 收购，现在和 excite 一样改用元搜索引擎 Dogpile。

1994 年 1 月，第一个既可搜索又可浏览的分类目录 EINet Galaxy(Tradewave Galaxy) 上线。除了网站搜索，它还支持 Gopher 和 Telnet 搜索。

1994 年 4 月，Stanford University 的两名博士生，美籍华人 Jerry Yang(杨致远)和 David Filo 共同创办了雅虎(Yahoo!)。随着访问量和收录链接数的增长，雅虎目录开始支持简单的数据库搜索。因为雅虎的数据是手工输入的，所以不能真正被归为搜索引擎，事实上只是一个可搜索的目录。Wanderer 只抓取 URL，但 URL 信息含量太小，很多信息难以单靠 URL 说清楚，搜索效率很低。雅虎中收录的网站，因为都附有简介信息，所以搜索效率明显提高。雅虎以后陆续使用 Altavista、Inktomi、Google 提供搜索引擎服务；2002 年 10 月 9 日，雅虎放弃自己的网站目录默认搜索，改为默认谷歌(Google)的搜索结果，成为一个真正的搜索引擎。1999 年 9 月，雅虎中国网站(www. yahoo. com. cn)正式开通，继承了雅虎全球的分类目录搜索的基因，为中国互联网用户提供了强大的搜索功能。

Lycos(Carnegie Mellon University Center for Machine Translation Announces Lycos) 是搜索引擎史上又一个重要的进步。Carnegie Mellon University 的 Michael Mauldin 将 John Leavitt 的 Spider 程序接入到其索引程序中，创建了 Lycos。1994 年 7 月 20 日，数据量为 54 000 的 Lycos 正式发布。除了相关性排序外，Lycos 还提供了前缀匹配和字符相近限制，Lycos 第一个在搜索结果中使用了网页自动摘要，而最大的优势还是它远胜过其他搜索引擎的数据量，1994 年 8 月已搜集了 394 000 个文档；1995 年 1 月搜集了 150 万个文档；1996 年 11 月已超过 6000 万个文档。1999 年 4 月，Lycos 停止自己的 Spider，改由 Fast 提供搜索引擎服务。

Infoseek(Steve Kirsch Announces Free Demos Of the Infoseek Search Engine)是另一

个重要的搜索引擎。Infoseek 沿袭 Yahoo!和 Lycos 的概念，它具有友善的用户界面和大量的附加服务，而使它成为一个强势搜索引擎。当用户单击 Netscape 浏览器上的搜索按钮时，弹出 Infoseek 的搜索服务，而此前由 Yahoo!提供该服务。Infoseek 后来曾以相关性闻名，2001 年 2 月，Infoseek 停止了自己的搜索引擎，开始改用 Overture 的搜索结果。

1995 年，一种新的搜索引擎形式元搜索引擎(A Meta Search Engine Roundup)出现了。用户只需提交一次搜索请求，由元搜索引擎负责转换处理后提交给多个预先选定的独立搜索引擎，并将从各独立搜索引擎返回的所有查询结果，集中起来处理后再返回给用户。第一个元搜索引擎是 Washington 大学硕士生 Eric Selberg 和 Oren Etzioni 的 Metacrawler。元搜索引擎概念上好听，但搜索效果始终不理想，所以没有哪个元搜索引擎有过强势地位。

1995 年 12 月 DEC 的 AltaVista 登场亮相，大量的创新功能使它迅速到达当时搜索引擎的顶峰。AltaVista 是第一个支持自然语言搜索的搜索引擎，AltaVista 是第一个实现高级搜索语法的搜索引擎，如 AND、OR、NOT 等。用户可以用 AltaVista 搜索新闻组(Newsgroups)的内容并从互联网上获得文章，还可以搜索图片名称中的文字、搜索 Titles、搜索 Java applets、搜索 ActiveX objects。AltaVista 是第一个支持用户自己向网页索引库提交或删除 URL 的搜索引擎，并能在 24 小时内上线。在面向用户的界面上，AltaVista 也做了大量革新。在搜索框下放了 tips 以帮助用户更好地表达搜索式，这些小 tip 经常更新，这样，在搜索过几次以后，用户会看到很多他们可能从来不知道的有趣功能。这系列功能，逐渐被其他搜索引擎广泛采用。1997 年，AltaVista 发布了一个图形演示系统 LiveTopics，帮助用户从成千上万的搜索结果中找到想要的。2003 年 2 月 18 日，AltaVista 被 Overture 收购。

1995 年 9 月 26 日，加州伯克利分校 CS 助教 Eric Brewer、博士生 Paul Gauthier 创立了 Inktomi(UC Berkeley Announces Inktomi)，1996 年 5 月 20 日，Inktomi 公司成立，强大的 HotBot 出现在世人面前。声称每天能抓取索引 1000 万个网页，所以有远超过其他搜索引擎的新内容。Inktomi 于 2002 年 12 月 23 日被 Yahoo!收购。

1998 年 10 月之前，Google 只是 Stanford 大学的一个小项目 BackRub。1995 年博士生 Larry Page 开始学习搜索引擎设计，于 1997 年 9 月 15 日注册了 google.com 的域名，1997 年底，在 Sergey Brin、Scott Hassan 和 Alan Steremberg 的共同参与下，BachRub 开始提供 Demo。1999 年 2 月，Google 完成了从 Alpha 版到 Beta 版的蜕变。Google 公司则把 1998 年 9 月 27 日认作自己的生日。Google 在 Pagerank、动态摘要、网页快照、DailyRefresh、多文档格式支持、地图股票词典寻人等集成搜索、多语言支持、用户界面等功能上的革新，像 AltaVista 一样，再一次永远改变了搜索引擎的定义。在 2000 年以前，Google 虽然以搜索准确性备受赞誉，但因为数据库不如其他搜索引擎大，缺乏高级搜索语法，所以推广并不快。直到 2000 年数据库升级后，又借着被 Yahoo!选作搜索引擎的东风，才名声大振。Google 自 2000 年开始提供中文搜索服务。2006 年 4 月，Google 宣布其中文名称“谷歌”，这是 Google 第一个在非英语国家起的名字。

1999 年 5 月，挪威科技大学的 Fast 公司发布了自己的搜索引擎 AllTheWeb。Fast 创立的目标是做世界上最大和最快的搜索引擎，Fast(Alltheweb)的网页搜索可利用 ODP 自

动分类，支持 Flash 和 pdf 搜索，支持多语言搜索，还提供新闻搜索、图像搜索、视频、MP3 和 FTP 搜索，拥有极其强大的高级搜索功能。2003 年 2 月 25 日，Fast 的互联网搜索部门被 Overture 收购。

1996 年 8 月，sohu 公司成立，制作中文网站分类目录，曾有“出门找地图，上网找搜狐”的美誉。随着互联网网站的急剧增加，这种人工编辑的分类目录已经不适应。sohu 于 2004 年 8 月组建独立域名的搜索网站“搜狗”，自称“第三代搜索引擎”。

Teoma 起源于 1998 年 Rutgers 大学的一个项目。Apostolos Gerasoulis 教授带领华裔 Tao Yang 教授等人创立 Teoma 于新泽西 Piscataway，2001 年春初次登场，2001 年 9 月被提问式搜索引擎 Ask Jeeves 收购，2002 年 4 月再次发布。Teoma 的数据库目前仍偏小，但有两个出色的功能：支持类似自动分类的 Refine，同时提供专业链接目录的 Resources。

Wisenut 由韩裔 Yeogirl Yun 创立。2001 年春季发布 Beta 版，2001 年 9 月 5 日发布正式版，2002 年 4 月被分类目录提供商 looksmart 收购。Wisenut 也有两个出色的功能：包含类似自动分类和相关检索词的 WiseGuide、预览搜索结果的 Sneak-a-Peek。

Openfind 创立于 1998 年 1 月，其技术源自台湾中正大学吴升教授所领导的 GAIS 实验室。Openfind 起先只做中文搜索引擎，鼎盛时期同时为三大著名门户（新浪、奇摩、雅虎）提供中文搜索引擎，但 2000 年后市场逐渐被 Baidu 和 Google 瓜分。2002 年 6 月，Openfind 重新发布基于 GAIS30 Project 的 Openfind 搜索引擎 Beta 版，推出多元排序（PolyRankTM），宣布累计抓取网页 35 亿，开始进入英文搜索领域，此后技术升级明显加快。

北大天网是国家“九五”重点科技攻关项目“中文编码和分布式中英文信息发现”的研究成果，由北大计算机系网络与分布式系统研究室开发，于 1997 年 10 月 29 日正式在 CERnet 上提供服务。2000 年初成立天网搜索引擎新课题组，由国家 973 重点基础研究发展规划项目基金资助开发，收录网页约 6000 万，利用教育网的优势，有强大的 FTP 搜索功能。

2000 年 1 月，两位北大校友，超链分析专利发明人、前 Infoseek 资深工程师李彦宏与好友徐勇（加州伯克利分校博士后）在北京中关村创立了百度（Baidu）公司。2001 年 8 月发布 Baidu.com 搜索引擎 Beta 版（此前 Baidu 只为其他门户网站如搜狐、新浪 Tom 等提供搜索引擎），2001 年 10 月 22 日正式发布 Baidu 搜索引擎，专注于中文搜索。Baidu 搜索引擎的其他特色包括百度快照、网页预览、预览全部网页、相关搜索词、错别字纠正提示、MP3 搜索、Flash 搜索。2002 年 3 月闪电计划（Blitzen Project）开始后，技术升级明显加快。

2003 年 12 月 23 日，原慧聪搜索正式独立运作，成立了中国搜索。2004 年 2 月，中国搜索发布桌面搜索引擎网络猪 1.0，2006 年 3 月中搜将网络猪更名为 IG（Internet Gateway）。

2005 年 6 月，新浪正式推出自主研发的搜索引擎“爱问”。2007 年起，新浪爱问使用 Google 搜索引擎。

2007 年 7 月 1 日，网易全面采用自主研发的有道搜索技术，并且合并了原来的综合搜索和网页搜索。有道网页搜索、图片搜索和博客搜索为网易搜索提供服务。其中网页搜索使用了其自主研发的自然语言处理、分布式存储及计算技术；图片搜索首创根据拍摄像机品牌、型号，甚至季节等高级搜索功能；博客搜索相比同类产品具有抓取全面、更新及时的优势，提供“文章预览”、“博客档案”等创新功能。