

搜索引擎 Google 与百度的比较分析

二十一整理于 [兰州SEO](#)

摘要：随着网上信息量的海量增加，搜索引擎的发展也非常迅速，搜索引擎成为广大用户上网不可或缺的得力助手。本文从功能服务和排序技术两个方面对目前中国使用率最高的两大搜索引擎——Google 和百度进行比较分析，以期为更好的掌握和使用搜索引擎提供借鉴意义。

关键词：搜索引擎 Google 百度 比较分析

1、概述

随着计算机、网络技术的不断发展，网上信息资源海量增加。如何检索和查询这些信息资源已迫在眉睫，在这样的背景下搜索引擎的发展应运而生。自从 1994 年网络搜索引擎被 yahoo 公司成功开发并应用于网络信息检索以来，搜索引擎的研究与发展非常迅速。据统计，目前全世界共有搜索引擎 10000 余种，搜索引擎的核心技术和相关性排序技术也逐步趋于成熟，几乎所有上网的人都会用到搜索引擎来查找信息。因此，对广大用户来说，搜索引擎已成为大家查找网络信息不可或缺的得力助手。机器人搜索引擎一般由搜索软件、索引软件和检索软件三部分组成。搜索软件通常称为机器人(Robot)、爬虫(Crawlers)或蜘蛛(Spiders)，它们可以运行在互联网上，是能够沿着网站的链接从一个页面跨越到另一个页面，自动追寻和发掘网上的各种文献信息资源，采集新出现的信息，确认网页之间的链接是否有效并剔除死链的一种软件。索引软件将采集的网页信息进行自动标引，建立索引数据库。查询软件决定搜索引擎的检索功能和返回结果的相关性。在检索过程中，该软件还会利用特殊的计算机算法对文献与检索词的相关性进行计算和评估。目前，常用的网络搜索引擎主要有独立搜索引擎 (Single search Engine)、元搜索引擎 (Meta search Engine)、网络搜索软件 (Net search software)。

2005 年 8 月 29 日下午，中国互联网络信息中心(CN-NIC)对外发布了北京、上海、广州三地的《2005 年中国搜索引擎市场调查报告》。报告显示，中国搜索市场目前是两强多极竞争的格局，百度和 Google 市场份额大大高于其他公司，二者合计在北京和上海占有超过 80% 的市场份额，在广州合计占有率也接近达到 75%。在这样的背景下，我们有必要对在中国使用率最高的两大搜索引擎——Google 和百度进行比较分析。

2、Google 和百度的简介

2.1 Google 简介

Google 公司成立于 1998 年。Google 富于创新的搜索技术和典雅的用户界面设计使 Google 从当今的第一代搜索引擎中脱颖而出。2004 年 8 月在纳斯达克上市，开创了第二代搜索引擎技术。

Google 利用“蜘蛛程序”在互联网上抓取各个网站的网页，对网页内容进行分词处理，并对抓取到的网页进行超链接分析，然后做 Page Rank 方式排序（按照此网页被引用次数多少进行排序）。Google 还独创了多种语言搜索技术，支持世界上几十种主要语言。Google.com 是世界上最流行的网站之一，Google 每月独特访问用户数为 3 亿 8 千万 (Nielsen/NetRatings2005 年 8 月统计)，这一数字还在不断增长。Google 数据库存有 30 亿个 Web 文件。属于全文 (Full Text) 搜索引擎。

2.2 百度简介

百度公司(Baidu.com, Inc) 由李彦宏先生和徐勇先生于 1999 年底成立于美国硅谷。2000 年 1 月，百度公司在中国成立了他的全资子公司百度网络技术 (北京) 有限公司，随后于同年 10 月成立了深圳分公司，2001 年 6 月又在上海成立了上海办事处。百度公司立足国内、

专注中文搜索市场,研究汉语特点和国内网民的上网习惯,主要提供中文(简/繁体)网页搜索服务。如无限定,默认与关键词精确匹配方式搜索。支持“-”号、“.”号、“link”、“《》”号等特殊搜索命令。百度自己研发并较早的推出国人喜爱的MP3搜索,吸引了大量的年轻网民。2003年12月,百度推出了“贴吧”这样的个性搜索服务来聚集人气,因而逐步奠定了百度成功的基石。2003年、2004年,百度推出了“新闻搜索”、“下吧”、“IE搜索工具条”等一系列搜索产品。2004年8月,收购Hao123.com进入网站网址搜索市场,逐步成为了中国搜索引擎的领头羊。其功能完备,搜索精度高,除数据库的规模及部分特殊搜索功能外,其他方面可与当前的搜索引擎业界领军人物Google相媲美,在中文搜索支持方面有些地方甚至超过了Google,是目前国内技术水平最高的搜索引擎。为包括Lycos中国、Tom.com、21CN、广州视窗等搜索引擎,以及中央电视台、外经贸部等机构提供后台数据检索及技术支持。

3、Google与百度的对比分析

3.1 功能服务比较

Google与百度的基本服务情况如下表:

搜索引擎	语种	网页	新闻	图片	音频	视频	建议搜索词	系统更新	翻译
Google	30余种	有	有	有	无	无	无	月更新	有
百度	中文	有	有	有	有	无	有	周更新	无

在个性服务方面,Google中的“手气不错”功能为用户挑选最佳网页,提高用户的搜索效率;“论坛搜索”功能专门用来查询并观看各讨论区里的流言信息;“网页搜索”功能能够帮助了解各分类主题的相关程度,使用网络目录检索能将搜索范围局限于特定的分类内,略去类似但无关的其他网页。而百度最新推出的“百度指数”服务功能是以百度网页搜索和百度新闻搜索为基础的免费海量数据分析服务。它以曲线图、相关新闻、专业评论的方式为用户透彻分析,把握用户所关心的“关键词”趋势,让用户在第一时间获得互联网上最有价值的信息和资讯;百度“贴吧”类似BBS的功能,可以在里面随意发表信息,另外,贴吧搜索还可以在那些发表的信息中进行搜索;“音乐掌门人”建立了最流行的音乐互动共享平台。由于Google和百度各有专长,Google常常被用来搜索网页、企业产品、商情、交通旅游等内容。而百度的娱乐功能更加强大,因此用户用它来搜索音乐(mp3)等娱乐内容的使用率相对较高。

3.2 排序技术比较

Google应用了多项排序技术,其中最核心、最关键的是Page RankTM技术(网页级别)和“超文本匹配分析”技术(Hy-pertext Matching Analysis)。Page RankTM技术根据网页之间的链接结构对网页的重要性进行客观的评价,并将网页的Page Rank值应用于检索结果的排序。Google在排列其检索结果时,都会考虑每个网页的Page Rank值,将重要的、高质量的网页排在结果列表的前面。这样,Page Rank技术在很大程度上避免和减少了人为因素,客观地把最恰当的检索结果呈现给用户。即使网页的Page Rank值再高,如果不能满足用户的检索需求也是毫无意义。因此Google将Page Rank算法与“超文本匹配分析”技术结合在一起。一般的搜索引擎仅仅考虑检索词出现的次数,而Google还在此基础上分析关键词的字体、字号以及关键词在网页中出现的精确位置,并且对该网页的邻近网页(包括链入网页和链出网页)的内容加以分析。Google认为,关键词在网页中出现的位置以及字体等因素都体现了该关键词在该网页中的重要程度,比如说,如果检索词出现在网页A的标题中,而只是出现在网页B的正文中,那么在不考虑其它因素的情况下,说明网页A与用户

需求更匹配。百度搜索引擎应用的排序技术其核心技术是“超链分析”技术 (Link Analysis), 百度总裁李彦宏是“超链分析”专利的惟一持有人。它与 Page Rank 技术的内容相似: 统计每个网页被其它网页链接指向的情况, 次数越多则级别越高, 排名也就越靠前。因此, “超链分析”就是通过分析链接网站的多少来评价被链接的网站质量, 这保证了用户在百度搜索时越受用户欢迎的内容排名越靠前。此外, “超链分析”技术将情报学中的引文索引技术同 Web 中最基本的东西——“超级链接分析”的技术相结合, 在查找的准确性、查全率、更新时间、响应时间等方面与其它技术相比都有很大的优势; 同时, 百度应用“内容相关度评价”技术, 并且运用了中文智能语言的处理方法, 依靠字与词的不同切割方法, 弥补了单纯依靠字或词的引擎技术的缺陷, 并且能够在不同的编码之间转换, 这就使得简体字和繁体字的检索结果可以自然结合。

从综合实力来看, 两大搜索引擎可说是各有千秋, 旗鼓相当。Google 的优势在于信息量大, 资源丰富。百度的优势在于对中文的理解较好, 更贴近中文用户的搜索习惯。用户可根据各自的需要和使用习惯选择使用。

推荐阅读:

基于 PageRank 算法的一种搜索引擎优化方法及实现

<http://www.0931seo.cn/seo-book/pagerank-in-seo/>

网站搜索引擎优化 36 计

<http://www.0931seo.cn/seo-book/WangZhanSouSuoYinQingYouHua36Ji/>

PageRank 科学排名遏止关键字垃圾

<http://www.0931seo.cn/seo-book/PageRankKeXuePaiMingEZhiGuanJianZiLaJi/>