

对于搜索引擎优化(SEO)的研究

马晓玲 吴永和

(华东师范大学 上海 200062)

摘要 在分析 Google 的可能排名算法的基础上,研究了针对 Google 进行优化的若干重要策略,讨论了对网站优化设计具有普遍指导意义的原则。

关键词 搜索引擎优化 Google 算法 网站排名

1 搜索引擎优化(SEO, search engine optimization)

搜索引擎优化通过了解搜索引擎如何抓取互联网页面,进行索引以及确定其对某一特定关键词的搜索结果排名等技术,来对网页进行相关的优化,使其提高搜索引擎排名,从而提高网站访问量,最终提升网站的销售能力或宣传能力的技术。而从事这方面工作的就是搜索引擎优化师,他们利用各种工具或者技术手段使某网站符合搜索引擎的搜索与排名规则,从而获得较前的网站排名。

根据 Jupiter MediaMetrix 2002 年公布的统计数据,信息搜索成为仅次于电子邮件的互联网第二大应用。因此搜索引擎营销目前已经成为网络营销的热点,是企业网站推广的首选。国外这方面工作发展得比较早,早在 1997 年左右就有人从事相关方面的工作。中国目前 SEO 正处于初步发展阶段,但发展很快。

在众多搜索引擎当中,Google 以其极高的知名度成为搜索引擎的代名词,前述的搜索引擎优化在实际操作中往往就是针对 Google 的优化。一个网站针对 Google 进行优化后,其在百度、雅虎、搜狐等搜索引擎中通常也会排在前面。

2 Google 可能的算法猜想

搜索引擎的算法应该是保密的,它是 Google 的商业秘密,所以针对 Google 的算法只能是一种猜想。Google 以前采用的算法据说是 Pagerank 算法,该算法最早出现在其创始人的一篇论文中,是利用页面相关性得分(RelevancyScore) + 页面等级得分(PageRank)来决定页面的相关性与重要性。该算法首先找到所有与查询关键词相匹配的网页,然后根据页面因素等进行排名,最后通过 PageRank 得分调整网站排名结果。

假设网页 A 有网页 t_1, t_2, \dots, t_n 的链接指向它,则 PageRank 的原始公式为:

$$PR(A) = (1 - d) + d(PR(t_1)/C(t_1) + \dots + PR(t_n)/C(t_n))$$

其中 d 为阻尼因子,一般设为 0.85;公式可描述为:A 网页的其 PR 值等于链向该页面的所有链接页面的 PR 值分别除以各自的外部链接数量的总和,再加上 $(1 - d)$ 。在此 PR(A) 是网页 A 的网页级别(PageRank); $C(t_n)$ 是网页 t_n 的外向链接的数量。

在“Google Florida”更新中,Google 对其排名算法进行了大规模改动,对于 Google 新算法目前有两种主要理论的争论:一种是 Hilltop 算法;另外一种则是 TSPR(Topic - Sensitive - PageRank:主题性页面级别技术)理论。表 1 为两算法的比较表。

从表 1 可以看出,两种算法都是为了克服 Pagerank 页面等级系统中存在的一个基本缺陷,即它根据一个网页上被链接的站点数量和质量来给该网页分配一个绝对的 PR,但是 PR 并非针对查询词语,因而一个网页如果只是在内容中偶然提到了一个和查询主题偏离的关键词语,也会因其居高的 PR 值而获得一个比较高的排名。

3 搜索引擎优化的策略

Google 推广分为右侧竞价广告排名和左侧自然排名,一般不加特别说明均指左侧排名,也叫 Google 排名。如何使网站能在 Google 搜索引擎上排名更靠前?根据自己对搜索引擎优化的研究,有以下一些原则是必须考虑的。

3.1 网站内容与主题 虽然有许多所谓技巧来提升网页在 Google 搜索引擎上的排名,但网站的实际内容才是网络优化策略的本质因素,即网站有明确的主题,有关于某主题的丰

富内容,专注于某领域的变化,要有较多的与主题相关的页面,时常更新,这才是网站优化的一个开始。

表 1 两算法的比较表

	HillTop 算法(行业得分)	主题性页面级别算法(TSPR)
Google 新算法	页面相关性得分(RelevancyScore) + 页面等级得分(PageRank) + HillTop 得分	页面相关性得分(RelevancyScore) + 页面等级得分(PageRank) + TSPR 得分
提出者	Krishna Bharat, 现在 Google 工作	Taher H. Haveliwala, 现在 Google 工作
后台支持	HillTop 专利技术	CIRCA 专利技术
运作流程	基于查询关键词进行一次普通查询,找出所有匹配的“专家网页”,剔除其中的成员和镜像站点,然后根据目标网页获得的上述专家文档的链接的数量和质量分配一个“行业分”,再将此值送入 Google 其它两个排名因素中进行相应整合处理	将查询条件与 CIRCA 数据库中主题进行匹配,并根据查询词语与数据库中所包含的主题的远近关系分配一个 TSPR(主题性页面级别)得分,再将此值送入 Google 其它两个排名因素中进行相应整合处理
定义/内涵	满足查询条件(词语)的网页的所有外部“专家文档”链接所形成的“行业分”,即在考虑链接页面等级(重要性)的同时,还须考虑该网页与查询主题的相关度	可确定对特定词语或短语的相关概念,并能够计算“短语 A”对“概念 B”的相关程度,从而可根据查询条件与 CIRCA 数据库中主题的相关程度提供主题性查询结果
相同点	不但要考虑满足查询结果的网页上的外部链接页的重要性(页面等级),还要考虑这些链接页对查询主题的相关性;相同主题的网站链接较之非相关网站具有更高价值	不但要考虑满足查询结果的网页上的外部链接页的重要性(页面等级),还要考虑这些链接页对查询主题的相关性;页面等级相同条件下,与 CIRCA 数据库中主题越相近的网页价值越高
对商业站点排名受到冲击的解释	Google 设置了搜索次数阈值,凡搜索频率高于此阈值的查询关键词均被纳入 Hilltop 系统并定期进行批处理,“热门商业词”一般都会“人圈”,该算法从而受到较大冲击。 Google 设置了关键词搜索次数上限,凡搜索频率低于此阈值的查询词都无缘 Hilltop 系统,因而 Google 仍将使用原来的算法并显示原来的排名结果	缺乏对查询条件的相关主题性,若 CIRCA 数据库中无任何主题与用户的查询条件匹配,则 Google 仍使用原来的算法系统
优势	提高了搜索结果的相关性和质量;有效降低了人工操纵排名的可能性	解决了搜索结果的质量性方面的一系列关键性问题;无须精确的页面等级计算结果
不足	无法保证“专家文件”的公正性、质量性;为保证 Hilltop 算法所需的处理能力,该算法只能周期性运行,从而无法保证随时向用户提供“最新、最好”的查询结果;新站点对热门查询词的排名的难度加剧	查询条件的多意性可能会导致主题的错误匹配,从而会影响搜索结果的正确性
优化建议	建立链接的侧重点放在“专家文件”上,尽可能为更多“专家文件”如网上目录、贸易目、黄页、贸易协会和资源页等收录;避免低质量或带有 SPAM 性质的链接机制;对网站内容进行正确和适当的优化;获得相关网上目录及相关站点群体的链接	锁定所有可能相关的关键词,扩大对搜索结果的涵盖面;保证网站设计结构的有效性,必要时可创建网站地图;对网站内容进行正确和适当的优化;获得相关网上目录及相关站点群体

搜索引擎的蜘蛛基本上只能根据网页文本内容判断网站的质量,而不能从图片、flash 动画上判断。因此在所有的页面中有充足的内容给搜索引擎进行索引是很重要的,否则没

什么内容的网站即使排上去,也是没意义的,是对广大网络用户的欺骗。随着用户不断积累,流量越来越大,Google 排名自然会向前。

3.2 关键字选择 搜索就得用关键词。关键词分析和选择是 SEO 最重要的工作之一。某网站的排名是与某关键字息息相关的,因此一定要正确选择适当的关键词。选择关键词时,要注意以下几点:

3.2.1 选取那些常为人们人们在搜索时所用到的,而且是与该网站所要重点推广的产品、服务、信息相关的关键词。在选择时,要尽量从客户方面考虑,要为用户着想,用户可能寻找什么,可能使用哪些关键词,多征求其他人的意见。

3.2.2 给网站确定主关键词不要太多。要符合搜索引擎的要求,尽量避免采用热门关键词,一般在 5 个左右,然后针对这些关键词进行优化。但也不能只用单一词汇,多词短语比单一词汇更有用。因为搜索引擎很难就单一词汇搜索相关结果,这样的搜索条件会产生太多结果,而搜索者是不会去浏览数百个结果页面的。

3.2.3 可以利用关键词分析工具软件。如 Overture 搜索引擎免费提供关键词分析工具,或在网页 <http://www.trafficzap.com/> 上有测试关键字使用频率的工具 Keyword Generator,输入某关键字后,该工具会列出高频率的与该关键字相关的短语或者词组可以作为参考。WordTracker 网络营销软件的功能是查看你的关键词在其它网页中的使用频率,以及在过去 24 小时内各大搜索引擎上有多少人在搜索时使用过这些关键词。

3.2.4 避免以含义宽泛的一般性词语作为主要关键词。要根据业务或产品的种类,尽可能选取具体的词。

3.3 关键字密度 关键字的个数与该页面字数的比例称为关键字密度,这也是搜索引擎优化策略中一个重要的因素。一张网页上通常会有众多的词语,搜索引擎会统计该页面的字数,再利用自身的算法来统计页面中每个字的重要程度,当然那些重复出现的词或短语被认为比较重要些。因此关键字的密度最好是不要超过文本数的 3% 或者更少,可以在搜索引擎中搜索要优化的关键字,然后统计一下排在前面的几个网站中该关键字的密度,也可以利用很多专门统计关键字密度的工具软件。切记不能盲目堆砌,尽量做到自然,符合基本的文法规则。不要在同一行连续使用某个关键词 2 次以上,更不能用作弊技术如“关键词堆砌欺骗(Keyword Stuffing)”,它利用搜索引擎对网页正文和标题中及 HTML 元标识中出现的关键词的高度关注来对关键词进行不合理的(过度)重复。这种技术很容易被搜索引擎察觉并受到相应惩罚。

3.4 关键字位置 关键字要放在有价值的地方,吸引搜索引擎注意。搜索引擎专注于网页中某一部位的内容,处于这部位的词语显得比其他部位的词语要重要得多。但据说 Google 在逐步降低这些部位标签的作用,以避免被操纵,但是也不能说毫无作用。这些部位包括以下方面:

3.4.1 title 和 meta 标签。title 标签是网页中最重要的

标签,最好是把它放在前面,它是网页中最先看到的部分。所以在 title 中放置关键字显得非常重要。标题设计不要超过 20 个汉字。meta 标签包含着网页的一些隐藏信息,其中 Description 属性内容可以是一小段句子,来正确地反映网站的主题,尽量在描述中加入主要的关键字;在 keyword 属性中多个关键字用逗号分开,一般只需关注 1~5 个关键词。

3.4.2 标题。在“标题”标签中能出现关键字对于提高网站排名有很大的好处。

3.4.3 超链接文本。链接到的一个网页肯定与该网站内容相关,所以在链接文本中出现关键字是重要的。在另一个网站与你网站建立链接时,尽量使用关键字作为链接文字,这有利于提高你网站的重要性。

3.4.4 URL 文本。URL 是“统一资源定位符”,在该网页 URL 中直接出现关键字对于搜索引擎排名会产生很大的影响。这样的关键字被称为“URL 文本”。对于搜索引擎优化来说,域名的易记性不是最主要的,最主要的是域名里是否包含了所优化的关键字。

3.4.5 页首、页尾和段首。网页顶部的文本、底部的文本和每段开头的内容都是关键句或是导航条,所以尽量在这些地方把关键字包含进来。

3.4.6 仅通过一个首页来命中所有希望推广的关键词是不可能的。因此让网页尽可能多地进入搜索引擎的索引,把握好整个网站的主题风格是非常重要的,让网站的主题关键词能够比较均匀地按照金字塔模式分布到网站中是比较理想的。

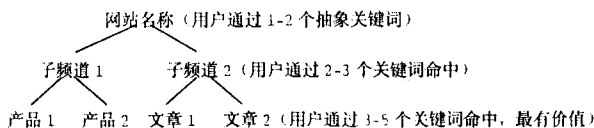


图 1 网站的主体关键词金字塔设计

3.5 网页设计 HTML 代码要精简,网页内容要充实,远离对搜索引擎来说是不友好的东西,如 FLASH、JAVA SCRIPT、图片等等,要记住浏览者是来看信息而不是来看动画的。要在页面颜色搭配方面多下功夫,本人认为这也是国内网站与国外网站差距较大的地方。网页占的空间尽量少,一个比较理想的页面不超过 15K 的大小。可以用加粗,也可以用相对大一点的字体突出关键词。

注意提高服务器响应速度,导航系统要清晰,栏目使用纯文本的网站导航系统要比图片格式的导航条在搜索引擎中反映出更多的信息,最好构建“站点地图”页面,拥有优化的结构,方便网页爬虫(spider)快速遍历网站所有需要发布的内容。

3.6 链接策略 增加链接数被认为是搜索引擎优化的一

个重要因素。根据 Pagerank 算法,搜索引擎会认为外部链接较多的网站重要性也相对较高。但链接数量只是关键因素之一,不是所有的链接权重都是一样的,从 PageRank 高的网站引用会给你网站更多的 Pagerank。外部链接不容易控制,因为是要其它网页链接到你这儿来,例如竞争对手就不可能相互链接,但可以按照以下的做法来提高链接数。

3.6.1 在网站外部链接上,链接要有质量,而不仅仅是有数量。做一个高质量的有价值的网站,其它网页自然会主动地进行链接,这是最重要的。不要随便与那些质量低且存在作弊的网站交换链接,最好与有相同网页级别的网页链接。同时交换外部链接时,尽量要求对方以你所优化的关键字作为链接的关键字。交换链接一旦完成,也要具有一定的相对稳定性,需要不定期回访友情链接伙伴的网站,看对方的网站是否正常运行,自己的网站是否被取消或出现错误链接。同时考虑在重要的网站中做广告或者向多个搜索引擎、收费目录提交你的网站。

3.6.2 在网站内部链接上,要使网站的结构合理化。星型结构与树型结构相结合,每个网页间是容易相通的,但一个页面中的链接应少于 100 个。保持网页内容链接的稳定性和持久性,如果自己网页中有链接更新时,最好能保留旧的页面并做好链接转向,以保持内容的连续性。

4 结束语

以上介绍了一些 SEO 技术,对网站优化设计具有普遍的指导意义。网站设计对搜索引擎要友好,但千万不要使用一些被搜索引擎拒绝的优化技术,Google 明令禁止的属 SPAM 性质的优化技术,以免被 Google 惩罚,将其 Pagerank 值设为 0。如隐藏文本/隐藏链接、误导性或重复性关键词、欺骗性重定向(Deceptive redirects)、加入链接工厂(link farms)等等。尽量遵从 Google 官方的站点设计指南,做一个网站或内容在搜索引擎中的排名,关键心态是要“长期培养”,而不是“拔苗助长”。

参考文献

- 1 L Page, S Brin, R Motwani et al. The PageRank Citation Ranking: Bring order to the Web. Stanford University, Tech Rep:1997-0072, 1997
- 2 Boutin P. Search Engine Optimization FREE. <http://webmonkey.wired.com/webmonkey/01/23/index1a.html>
- 3 Karen 编译.“HillTop”论——探索 Google 排名新算法. <http://www.google8.net/archives/000182.html>
- 4 李福良,张 辉. 企业网站 SEO 技术研究. 合肥工业大学学报(自然科学版),27;(1)
- 5 吴泽欣. 网站优化方法. <http://www.seochat.org>

(责编:亦思)