

# 搜索引擎优化 (SEO) 的探索与实现

易丹

(广州铁路职业技术学院 基础部 广东 广州 510665)



**摘要** 本文从分析搜索引擎 Google 的搜索排名原理入手,探索了如何针对 Google 进行搜索引擎优化,并从实践中总结出一些网站设计的优化策略。

**关键词** 搜索引擎优化(SEO);Google;页面级别;HillTop 算法

## 0 引言

搜索引擎优化(SEO, Search Engine Optimization),是通过分析搜索引擎的算法,优化网站结构、网页代码和内容,使网站对全文搜索引擎友好,从而帮助全文搜索引擎找到含有最好内容的网页,提高网站在搜索结果中的自然排名,从而提高网站的影响力。搜索引擎自 90 年代出现以来,深受网民喜爱,据中国互联网络信息中心发布的《第 15 次中国互联网络发展状况统计报告》指出,用户在互联网上获取信息最常用的方法是使用搜索引擎查找相关网站,比例高达 70.7%。因此针对搜索引擎的优化技术也随之出现,目前在中国还处于初步发展阶段。在众多搜索引擎当中,Google 知名度最高,占了全球 56% 的搜索引擎市场。如果网站在 Google 中能获得较好的排名,那么在 Yahoo、百度、新浪的排名自然不会靠后。

## 1 Google 排名原理分析

Google 的排名算法是公司机密,无法得到,任何对排名系统的一点泄漏都有可能被一些人钻了空子,并用来恶意操纵搜索结果,从而破坏搜索结果的公正性。这里只是针对 Google 的可能算法进行一些分析猜测。

最新的 Google 的排名运算法应该使用了三个算法的结合:

Google 排名公式 =  $\{(1-d) a (RS)\} * \{(1-e) b (PR * fb)\} * \{(1-f) c (LS)\}$

其中 RS = 页面相关性得分 (Page Relevance Score):基于网页标题 (Title)、元标识 (Meta tags)、正文标题 (Headlines)、正文 (Body text)、URL、图片 Alt 文字、锚文字 (Anchor Text) 等元素中出现的关键词得出的分数。

PR = 页面等级得分 (PageRank Score):基于链接至你网站的网页数量及其 PR 值得出的分数。

LS = 行业得分 (Local Score):根据专家文件计算得出的分数。

a, b, c = 调节控制比重:Google 用于精工调整查询结果。

d, e, f = 阻尼控制:Google 用于精工调整查询结果,目前“f”值可能为“0”。

fb = 因子基数:Google 工具栏所显示的页面等级得分范围从 1 到 10,但它实质上并不是一个线性函数,而是一个指数/对数函数。经过分析,认为它有一个趋近于 8 的基数。即 PR 为 5 的网页的值比 PR 为 4 要多 8 倍之多 ( $PR_5 = PR_4 \times 8$ ),同样,一个 PR 值为 8 的网站比一个 PR 值为 4 的网站的值要高出 4000 多倍 ( $PR_8 = PR_4 \times 8 \times 8 \times 8$ )。

页面相关性的算法就不在这里介绍了,此类文章很多,在这里重点介绍页面等级算法和计算行业得分的 HillTop 算法。

### 1.1 页面等级算法

PageRank 取自 Google 的创始人 LarryPage,用来标识网页的重要性,级别从 1 到 10 级,10 级为满分,PR 值越高说明该网页越受欢迎(越重要),例如:一个 PR 值为 1 的网站表明这个网站不太具有流行度,而 PR 值为 7 到 10 则表明这个网站非常受欢迎(或者说

极其重要)。假设网页 A 有网页 t1、t2、……tn 链向它,那么网页 A 的 PR 值算法为:

$$PR(A) = (1 - d) + d \{ PR(t1)/C(t1) \dots PR(tn)/C(tn) \}$$

其中 PR(A)表示网页 A 的 PR 值,PR(t1)表示外部链接网页 t1 的 PR 值,C(t1)则表示外部链接站点 t1 拥有的外部链接数量,d 为阻尼因子,一般设为 0.85。

如:广州铁路职业技术学院网站 www. gtxy. cn,被 PR 值为 7、外部链接数为 9 的新浪网站链接,则得到的 PR 得分为:

$$PR(gtxy) = (1 - 0.85) + 0.85 * (7/10) = 0.745$$

这说明,若你的网站被很多其他网站链接,且这些网站的 PR 值高而它们的链接数又少,则你的网站的 PR 值就高。

### 1.2 HillTop 算法

页面等级算法存在一个基本的缺陷: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“重要性值”,但是“页面等级值”并非针对查询词语,因而一个网页即使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,也会因其居高的页面等级值而获得一个比较高的排名。美国加州的克利须那·伯哈拉特提出“Hilltop”算法,Google 成为受让方。

Hilltop 算法以如下方式计算一个网页的“权威值”:以该关键词进行一次普通查询,找专家文档“文集”。“专家文档”的定义有严格标准,“文集”应该是数量上易于管理的一组网页。从返回的专家列表中把成员站点(来自同一个域或域相同而后缀不同的站点,如 ibm. com, ibm. co. uk 等等,或指来自相邻的 IP 地址的站点)和镜像站点去掉。根据所获得的上述专家文档对其投票的数量和质量,网页被分配以一个“局部分数”(LocalScore),然后按网页的“局部分数”进行排名。Hilltop 算法若没有找到搜索引擎认为数量足够的“专家文件”(要求至少需有两票),则该算法失效,即返回结果为零。换言之,对于高度明确化的查询条件(查询词语),Hilltop 算法的结果很可能为“0”,这是该算法的一个独特特性。简单地说,页面等级决定一个网页的“权威性”,而 Hilltop 算法则决定匹配一个查询条件的一个网页的“权威性”。

## 2 网站设计的优化策略

如何使网站能在 Google 搜索引擎上排名更靠前? 根据对搜索引擎优化的研究,可以总结出一些基本策略。

### 2.1 优化网站设计

优化网站主要包括:网站结构优化、网站标签优化、网站页面优化,为的是让 Google 更容易搜索你的网站并且关注你想排的关键词。

#### 2.1.1 结构优化

网站应具有清晰的层次结构和文本链接,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。

#### 2.1.2 标签优化

网页标签的优化主要是在网页标题、简介标签、关键词标签三个部分。标签中要适当的突出关键词。例如一个关于“电子商务”的网站的首页标签可以如下设计:

```
<title> 电子商务指南 </title>
```

```
<meta name = "description" content = "电子商务指南是一个专业的电子商务交流学习站点,主要有以下栏目:电子商务新闻、电子商务研究、电子商务论文、电子商务法律、电子商务论坛。另外提供电子商务与网络营销咨询策划服务。 >
```

```
<meta name = "keywords" content = "电子商务 >
```

这三段标签要放在 < head > 与 </head > 之间,标题标签长度不可超过 40 个字符(20 个汉字)。

#### 2.1.3 页面优化

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、JAVASCRIPT、图片等等。要在页面颜色搭配方面多下功夫,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。可以用加粗或相对大一点的字体突出关键词。

### 2.2 关键词设计的技巧

关键词的设计是整个网站登录过程中最基本也是最重要的一步,要考虑诸多因素。

#### 2.2.1 选择相关的关键词

选择和你的网站内容密切相关的关键词,不要听信那些靠毫不相干的热门关键词吸引更多访问量的宣传,比如你的网站是进行“在线团购”业务,若你在选关键词上为了吸引客户而选择了“在线电影”,那么在引擎上查找“在线电影”的网民就算进了你的网站也不会感兴趣,相反会产生厌恶,这样的做法毫无意义,甚至可能会受到搜索引擎的惩罚。

#### 2.2.2 选择具体的关键词

在挑选关键词时还有一点要注意,就是避免拿含义宽泛的一般性词语作为主打关键词,而是要根据你的业务或产品的种类,尽可能选取具体的词。含义宽

泛的词语在搜索引擎必然会有大量的搜索结果,必然影响到你的网站的排名。

### 2.2.3 关键词个数不宜太多

有些网站设计者恨不得把所有与网站有关联的关键词都加进去,其实这种做法完全错误。如果一个网站其主页的标题标签中包含 10 个以上的关键词,则没有一个关键词能够满足较高排名所要求的关键词密度,排名都无法上去。应该确定三个最重要的关键词,总关键词个数在 7 个左右就够了。

### 2.2.4 关键词金字塔设计

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(如图 1 所示)。

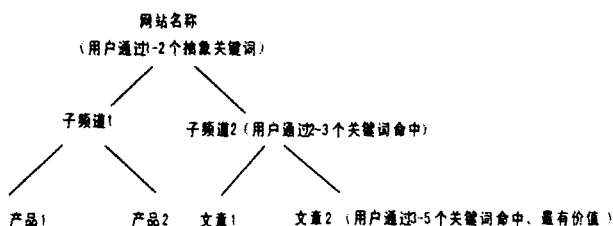


图 1 网站的主题关键词金字塔设计

## 2.3 做好网站的推广工作

网站做得再好,没人知道也是毫无意义的。做好网站的推广工作,对于提高网站流量可以起到立竿见影的效果。

### 2.3.1 登录搜索引擎

登录搜索引擎是指将网站信息,如关键词、网站描述、网站名称、URL、联系人等信息,提交到搜索引擎,目的是让这些信

息进入搜索引擎数据库。Google、百度、Yahoo、新浪等都提供免费的自动登录服务,可以每个月登录搜索引擎一次(频繁的登录被视为作弊),而且针对不同搜索引擎设计最合适的资料提交可以增加被搜索引擎收录的可能性。

### 2.3.2 提高链接广泛度(Link Popularity)

前面介绍了 Google 通过网站的外部链接数及链接网站的质量来评价你的网站的重要性(PR 值),即需要提高你的链接广泛度。可以通过交换链接的方法达到目的。交换链接是指网站之间相互加入对方的链接。要进行交换链接的网站内容要相近,且不是竞争对手。但是并不代表可以毫无策略地与任何网站交换链接,最好选择网页级别数大于 4 的网站交换链接,千万不要和低质量的甚至是受过搜索引擎惩罚的网站交换链接。

## 3 结束语

针对搜索引擎的优化策略见仁见智,但是最重要也是最有效的方法还是丰富你的网站内容,一个有内涵的网站自然会吸引网民的光顾,也会得到搜索引擎的青睐。另外,随着搜索引擎技术的变化发展,也要相应地调整 SEO 策略,去适应新的技术。

### 参考文献

- [1] 康桂英,新一代智能搜索引擎网典研究[J],情报理论与实践,2000,(3):218-220
- [2] 李福良,张辉,企业网站 SEO 技术研究[J],合肥工业大学学报(自然科学版),2005,10
- [3] 陈会安,JavaScript 网页制作彻底研究[M],北京:人民邮电出版社,2002,396-423

· 油海采珠 ·

## 国内首套天然气水合物二维开采模拟系统问世

在研制天然气水合物一维开采实验模拟系统的基础上,中国科学院广州能源研究所天然气水合物开采技术团队,成功研制出国内第一套天然气水合物二维开采实验模拟系统。

天然气水合物是一种优质洁净能源。中国南海已经发现天然气水合物存在的间接证据,如果开采成功,将对中国未来能源战略产生深远影响。初步测试结果表明,该系统能有效模拟海底天然气水合物的生成及分解过程,可以对现有的开采技术进行系统的模

拟评价。

较之一维开采模拟系统,该系统采用更加先进的手段测量多孔介质中气、液、固(水合物)的含量及分布,并能够更加真实地模拟实际水合物地层特征。

该系统能综合模拟注热、降压、注入化学试剂以及二氧化碳置换开采天然气水合物,以及开采过程井网布置,并对开采过程系统的动态特征进行实时监控。这是国内第一套二维水合物开采模拟系统,同类模拟系统在国外也鲜见报道。(于明)