

# 基于 Google 的网站优化技术

杨帆<sup>1</sup>, 王秀伟<sup>2</sup>, 白振兴<sup>1</sup>

(1. 空军工程大学 工程学院 陕西 西安 710038; 2. 葫芦岛有色金属集团 电力分厂 辽宁 葫芦岛 125003)

**摘要:**搜索引擎已经成为现在检索信息的重要手段,为了提高网站在搜索引擎中的排名,产生了网站优化的概念,针对目前流行的 Google 搜索引擎,通过对网站结构的分析,详细阐述了网页结构对基于 PageRank 算法的网站搜索引擎排名结果可能产生的影响,就如何提升网站在搜索引擎中的排名问题提出了几点优化策略,分析结果表明:对网站结构的优化可以显著提高网站在 Google 搜索引擎上的排名。

**关键词:**优化;Google;PageRank;链接

**中图分类号:**TP393

**文献标识码:**B

**文章编号:**1004-373X(2006)19-149-03

## Optimization Technology for Website Based on Google

YANG Fan<sup>1</sup>, WANG Xiuwei<sup>2</sup>, BAI Zhenxing<sup>1</sup>

(1. Engineering Institute, Air Force Engineering University, Xi'an, 710038, China;

2. Electric Power Branch Factory, Huludao Nonferrous Industry Co. Ltd., Huludao, 125003, China)

**Abstract:** Search engine become the import means for searching information. For the sake of improving the rank in search engine, bring forth the concept of search optimization. Aiming at the Google search engine, it analyzes the possible influence that various kinds of web structures exerted on the ranking results of search engines based on PageRank. Putting forward the optimize strategy for how to improve the rank of website. The result of analysis shows that optimizing the structure of the website can improve the rank for Google search engine.

**Keywords:** optimize; Google; PageRank; link

随着 Internet 的发展,人们已经进入信息量极为丰富的时代。Internet 已经成为当今和未来人们获取资源和信息交流的主要场所。为了帮助用户快速找到需要的信息,搜索引擎应运而生。目前,Google 已经成为使用最为广泛的搜索引擎之一。Google 等新一代搜索引擎的优势在于他不仅索引量大,而且还能将最好的结果排在搜索结果的最前面,有利于用户尽快找到所需的信息<sup>[1]</sup>。通过研究 Google 的排名规律,合理地优化网站链接结构可以提高搜索结果排名,而目前国内互联网站在处理网页之间的链接结构时,往往存在很大的随意性。殊不知这种随意性往往导致网站在搜索结果中不能取得理想的排名。

### 1 网站优化概述

网站优化就是通过对网站功能、网站结构、网页布局、网站内容等要素的合理设计,使得网站内容和功能表现形式达到对用户友好并易于宣传推广的最佳效果,充分发挥网站的网络营销价值,提高在搜索引擎上的排名。

网站优化设计具体表现在 3 个方面:用户优化、网络环境优化以及网站运营维护优化<sup>[2]</sup>。

#### 1.1 用户优化

经过网站的优化设计,用户可以方便地浏览网站的信息、使用网站的服务。具体表现是:以用户需求为导向,网站导航方便,网页下载速度尽可能快,网页布局合理并且适合保存、打印、转发,网站信息丰富、有效,有助于用户产生信任。

#### 1.2 网络环境优化

通过搜索引擎推广网站的角度来说,经过优化设计的网站使得搜索引擎顺利抓取网站的基本信息,当用户通过搜索引擎检索时,企业期望的网站摘要信息出现在理想的位置,用户能够发现有关信息并引起兴趣,从而点击搜索结果并到达网站获取进一步信息,直至成为真正的顾客。对网络环境优化的表现形式是:适合搜索引擎检索(搜索引擎优化),便于积累网络营销网站资源(如互换链接、互换广告等)。

#### 1.3 网站运营维护优化

网站运营人员方便进行网站管理维护(日常信息更新、维护、改版升级),有利于各种网络营销方法的应用,并且可以积累有价值的网络营销资源(获得和管理注册用户资源等)。

## 2 Google 采用的主要算法

Google 凭借其先进的排序机制已成为目前使用最为广泛的搜索引擎之一。Google 搜索引擎使用 PageRank 算法来进行网页的排名,该算法最早出现在其创始人的一篇论文中,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。

在计算某个页面 PageRank 值时,所有的入链接都要考虑在内。页面 A 的 PageRank 值计算公式如下:

$$PR(A) = (1 - d) + d(PR(T_1)/C(T_1) + \dots + PR(T_i)/C(T_i))$$

其中:PR(A) 表示网页 A 的 PageRank 值;PR( $T_i$ ) 表示链接到 A 的网页  $T_i$  的 PageRank 值; $C(T_i)$  表示网页  $T_i$  的出站链接数量; $d$  为阻尼系数, $0 < d < 1$  (Google 通常取值 0.85)。

由上面公式可知,PageRank 并不是将整个网站排等级,而是以单个页面计算的。其次,页面 A 的 PageRank 值取决于其他的相关页面,所以计算 PageRank 值实际上是一个迭代的过程,计算结果的精确程度取决于初值的选取和迭代的次数。对于初值一般取 1,而为了保证实际应用中的这个结果总是收敛的,则加入了阻尼系数  $d^{[3]}$ 。

## 3 面向 Google 的优化策略

面向 Google 的网站优化就是推测 Google 的排名规则,充分利用好各种有利于排名的因素,让网页的排名尽量靠前。对 Google 搜索引擎而言,链接就是一切,所以网站的链接结构一定要设计合理。由上文可知,网站的 PageRank 的最大值理论上等于网页的个数,但实际结果往往达不到<sup>[4]</sup>。下面重点研究网站的链接结构对 PageRank 值的影响,给出优化策略。

### 3.1 内部链接的影响

网站 PageRank 值就是网站内部所有页面 PageRank 值的和。很明显,一个网站的 PageRank 最大值就是他的页面数量。网站内部链接组织不好,网站可能会达不到最大的 PageRank 值。根据 PR 原理,网站上每一个已被 Google 收录的内部网页都是对该网站的一记投票。一个网站若能够拥有更多已被 Google 收录的内部网页,就可能获得更多的投票,因此可以通过创建大量的内部网页来提高网站整体的 PageRank 值,但也不是随便增加什么页面都能达到效果。那些完全相同或几乎相同的页面称为 cookie2cutter,Google 认为是垃圾信息并会引发相应的警报机制使得页面甚至是整个网站受到处罚,所以从根本上来说网页要有一定的质量。

以下分析一下网站内部链接如何影响 PageRank 值。在这里考虑的是一个相对独立的网站,外部链接的影响暂不考虑。

假设网站包含 3 个页面,没有外部链接,如图 1 所示。

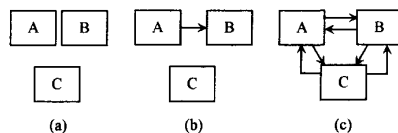


图 1 内部链接对 PageRank 的影响

在图 1(a~c)的情况下,为每个网页分配初值 1,阻尼系数保持与 Google 一致(0.85),经过迭代收敛后,3 种情况的 PageRank 值如下:

(a) PageRank A=0.15, PageRank B=0.15, PageRank C=0.15;

(b) PageRank A=0.15, PageRank B=0.277 5, PageRank C=0.15;

(c) PageRank A=1, PageRank B=1, PageRank C=1。

网站(a)的 PageRank 值为 0.45,(b)的情况稍好一些,总值 0.577 5,但仍然只是最大值的一小部分页情况。在(c)的链接结构下,网站达到了 PageRank 最大值。

由此可见,链接的不好,完全可能浪费潜在的 PageRank 值。根据分析结果可以得出:任意两个页面都有相互链接时能达到网站 PageRank 大值。

### 3.2 外部链接的质量与数量

在计算网站排名时,PageRank 会将网的外部链接数考虑进去。这并不是说一个网站的外部链接数越多其 PR 值就越高,Google 对一个网站上的外部链接数量的重视程度并不意味着可以不求策略的与任何网站建立链接,因为 Google 并不是简单地由计算网站的外部链接数来决定其等级<sup>[5]</sup>。

首先,外部链接的 PageRank 值很重要,但同时链接的数量也相当关键。常说的阻尼因数  $d$  就是投票或链接到另外一个站点时所获得的实际 PR 分值。阻尼因数一般是 0.85。当然比起网站的实际 PR 值,他就显得微不足道了。

前面已经介绍 PR 值的计算公式,这里要注意:一个网站的投票权值只有该网站 PR 分值的 0.85,而且这个 0.85 的权值平均分配给其链接的每个外部网站。

一个具体的例子:假设一个网站,被链接的 PR 值为 4,外部链接数为 10 的网站,则计算公式如下:

$$PR(A) = (1 - 0.85) + 0.85 * (4/10) \\ = 0.15 + 0.85 * (0.4) = 0.15 + 0.34 = 0.49$$

由此可以看出,如果网站获得一个 PR 值为 4,外部链接数为 10 的网站的链接,最后我的网站将获得的 PR 分值为 0.49。

如果网站获得的是一个 PR 分值为 8,外部链接数为

32 的网站链接,那么我将获得的 PR 分值将是:

$$PR(A) = (1 - 0.85) + 0.85 * (8/32)$$

$$= 0.15 + 0.85 * (0.25) = 0.3625$$

由此可见,外部链接站点的 PR 值固然重要,该站点的外部链接数也是一个需要考虑的重要因素。根据分析结果可以得出:在建设网站的外部链接时,应尽可能找那些 PR 值高且外部链接数又少的网站。网站上这样的外部链接站点越多,PR 值就会越高,排名将会得到显著提升。

### 3.3 页面关键字与关键字密度优化

很多搜索引擎都是通过页面的关键词密度(keyword density)来决定该页面对关键字的关联度,关联度越高,该页面的 PR 值就越高,所以应确保关键字在整个网页中的充分利用和合理分布<sup>[6]</sup>。即充分利用所有可以利用的因素,但不要过分重复或简单排列关键字。应遵循必要的语法规则,形成自然流畅的语句,使网页不失吸引力。从页面因素的优化角度出发,可考虑将关键字分布在网页标题元、网页描述/关键字元、正文标题、正文内容、文本链接、ALT 标识当中。

作者简介 杨帆 男,1983 年出生,河南潢川人,空军工程大学工程学院硕士研究生。主要研究方向为人工智能。  
白振兴 男,1953 年出生,山西太原人,空军工程大学工程学院教授、博士后。主要研究方向为人工智能。

(上接第 148 页)

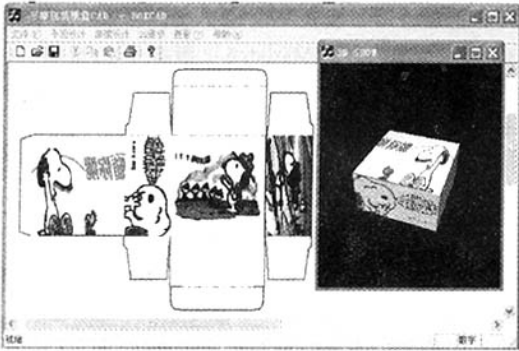


图 2 纸盒结构 2D 设计图及带贴图的 3D 显示

## 4 结 语

通过优化网站结构可以有效提高网站在基于 PageRank 算法的搜索引擎的搜索结果排名。这种效应也许短期内尚不明显,但随着页面的增加和网站间链接的增多,这种效果是十分明显的。在优化网站链接结构的同时,对网页内容的适当优化也可以提高网站的排名。

### 参 考 文 献

[1] Jeh G, Widom J. Scaling Personalized Web Search [R]. Stanford University, 2002.

[2] 吴泽欣. 网站优化方法[EB/OL]. <http://www.seochat.org>.

[3] <http://www.sowang.com/sousuo.20040115.htm> (Accessed sept. 2004)

[4] 赵荣,黄燕云,张露. 搜索引擎检索结果的组织技术[J]. 情报学报, 2004, 23(1): 69-72.

[5] 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, 2004, 27(1): 92-96.

[6] 车东. 如何提高网站在 Google 中的排名[EB/OL]. <http://www.Ch-edong.com/tech/google.html>, 2003-6-18.

## 4 结 语

本文对包装纸盒的三维仿真问题进行了研究,提出了基于 OpenGL 技术的一种由面模型构造三维盒体的方法。开发了一个三维包装纸盒设计系统,基本上可以实现从平面结构设计到三维效果观察,解决了在屏幕上对包装纸盒设计结果进行打样的问题,节约了成本,提高了效率。

### 参 考 文 献

[1] 张海东. 包装盒型的三维仿真研究[J]. 计算机辅助设计与图形学学报, 2005, 17(3): 617-622.

[2] Richard S. Wright, Jr, Micheael Sweet. OpenGL 超级宝典 [M]. 2 版. 北京:人民邮电出版社, 2001.

[3] 孙诚. 包装结构设计[M]. 北京:中国轻工业出版社, 1995.

## 欢迎订购《现代电子技术》杂志合订本

为满足广大读者要求,2003 年,2004 年,2005 年《现代电子技术》合订本正在销售。合订本包括单行本全部内容及广告页。需要者按以下方式订购。合订本单价 180 元(含邮费)。

银行汇款	邮局汇款
收款单位:陕西电子杂志社	地 址:西安市雁塔西路 158 号双鱼花园广场 B 座 1606 室
帐 号:611301074018000794620	邮 编:710061
开 户 行:交通银行西安含光路支行	电 话:(029)85393376
发行信箱:faxing@xddz.com.cn	传 真:(029)85397961
	联 系 人:薛进良