网站策划运营教材

网络营销-

Search Engine Optimisation

(SEO)

版本: 1.1 /2006-12-30

本文下载地址: http://www.digda.cn http://itcaogen.digda.cn http://www.itcaogen.cn

版权声明:本文档由 KEE 编写,本文档的版权归 KEE 所有,凡有采用,请标明出处和保证文档的完整性。阅读材料由网上汇集而成,标明收集的网址,阅读材料的版权由原创者所有,版权问题和本文档无关。

KEE 网站策划=策划+执行=市场策划+内容管理+技术+网络推广

MSN: qiuqiying_gzb@hotmail.com

QQ: 411422088

本文可以在网络上自由传播,但要保证本文的完整性,但不授权在杂志(包括电子杂志)。

目录

- ●前言
- SEO 工作原理
 - SE 工作原理
 - ■名词解释
- 实操 SEO
 - SEO 框架
 - 实操 SEO
 - ■案例
 - ■作弊



一、前言

1、什么是 SEO

SEO (Search Engine Optimisation),顾名思义就是搜索引擎优化。SEO 是通过研究各类搜索引擎如何抓取互联网页面和文件,及研究搜索引擎进行排序的规则,来对网页进行相关的优化,使其有更多的内容被搜索引擎收录,并针对不同的关键字获得搜索引擎的左边更高的排名,从而提高网站访问量,最终提升网站的销售能力及宣传效果。

SEO 不是技术手段,是网站推广的一部分,是主推被动式营销。 SEO 是针对网络的传播模式,把目标内容(广告、产品、品牌)传 递给目标授众的最有效的途径。

2、为什么要做 SEO

做 SEO,早期的目的让搜索引擎更容易对你的站点进行识别和 收录,同时相应的关键字排位靠前,让你的<u>客户更容易找到你的网站和你的产品</u>,增加你的产品和公司的曝光率,是适应搜索经济发展的 必经之路。完整的 SEO 还是对用户心理需求分析,适应浏览者需求的过程。

3、谁要做 SEO

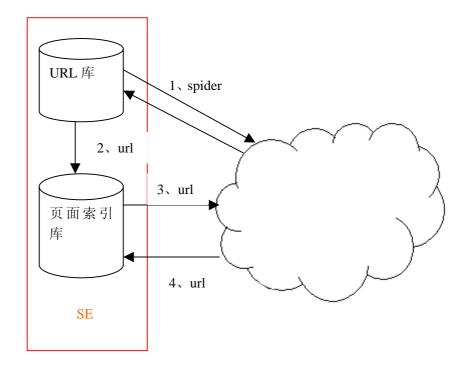
互联网站点,均需要做 SEO。SEO 即不是作弊,也不是一门学科,只是搜索引擎衍生出来的行业,是普通人可以掌握的简单的互联网站应用的技巧。做 SEO,不需要支付任何费用,不存在高深的知识,不存在垄断性,也不是不道德的行为,根本不需要昧着良心说话。

二、SEO 工作原理

1、搜索引擎工作原理

解释: spider、开放目录 DMOZ、登录、外连接、搜索频率、外连接、sandbox、内连接、pr 值、搜索深度、收录优先规则、爬虫 crawler、分词技术、关键字、spam kill、dance

1、搜索引擎(SE)工作原理



1、 搜索引擎释放网络蜘蛛 spider, spider 先检查开放目录 DMOZ

上**登录**网站网址进行访问,并根据收录的网站的外部链接找到更多的网站(**外连接**,对于二级域名 spider 等同于外部连接。)根据网站的的更新频率及权重级别,安排 spider 的**搜索频率**。对于新站,google 进入 sandbox 处理。

对于站内的连接 (内连接),(google)根据网站的 pr 值, 计算出 搜索深度, 检查收入的页面中获取的的网址是否更新, 有则收录新的网址, 无则关闭 spider。Baidu 及 yahoo 则根据自己的算法, 收录访问深度内的网址。

- 2、 搜索引擎读取 spider 收集的网址库,按照网址**收录优先规则**,调度**爬虫 crawler**。Google 对于提交了 **sitemap** 的网站,抓取深层次的页面。
 - 3、 搜索引擎释放 crawler 抓取网址页面
- 4、 收录网站为中文网站根据**分词技术**进行**关键字**索引,按照 页面的权重进行排位。
 - 5、 进行反作弊 spam kill 和 dance (google)

2、名词解释

spider: Spider 是搜索引擎用来访问 Internet 上网页的自动程序。 spiders 根据 html 的语法和格式,对读取的页面进行代码过滤,收入 相关的文字内容。目前搜索引擎无法象人那样去读相应的图片、

Flash、影片里面的内容。图片中的文字对 Spider 来说毫无意义。对于 javascript 里面的内容,现在已经有部分网站开始收录。

开放目录 DMOZ

开放目录 - Open Directory Project 是互联网上最大的,最广泛的人工目录。它是由来自世界各地的志愿者共同维护与建设的最大的全球目录社区。

pider

开放目录

开放目录专案效仿一些在 20 世纪最重要的编辑员/志愿者项目。 就像牛津英文字典通过志愿者的努力成为权威的工具一样,Open Directory 跟随着它的脚步也成为了权威的互联网分类目录。

开放目录专案是建立在开放资源共享的理念上的,是唯一 **100%** 免费的大型目录。提交一个网站或使用目录数据不需要支付任何费用。开放目录专案的数据在同意遵守免费使用条款的情况下,任何人都可以免费使用。

DMOZ - http://dmoz.org/ YAHOO - http://dir.yahoo.com/ GoGuides.org - http://www.goguides.org/ JoeAnt - http://www.joeant.com Zeal) - http://www.zeal.com/ Gimpsy - http://www.gimpsy.org HotRate) - http://www.hotrate.com Xoron - http://www.xoron.com/ Lycos Directory - http://www.lycos.co.uk/dir/ business.com - http://www.business.com/ Proudly Canadian - http://www.porthardybc.ca/ LocalSites.ca - http://www.localsites.ca/ CanadaOne.com - http://www.canadaone.com/ StartPage.ca - http://www.startpage.ca/ Limey Search - http://www.limeysearch.co.uk/ dotukdirectory.co.uk http://www.dotukdirectory.co.uk/ Splut.com - http://www.splut.com/ Ausie.com - http://www.aussie.com.au/ AAA Smart Search - http://www.aaa.com.au/ The Australian Internet Directory http://www.directory.com.au/ Sohu Directory - http://dir.sohu.com/

登录: 提交到搜索引擎的行为。

常用工具: 登陆奇兵

登陆类型: GOOGLE 和百度,只需要填写网址和说明(网站描述)。yahoo 目录式搜索引擎,采用人工方式,必须注册有 YAHOO

外部链接

Tips

搜索频率

的帐号才能提交网址。国外的部分搜索引擎,有部分会通过 EMAIL 确认你提交的网址。

外部连接: 指向友情链接和其他网站连接到您的某个页面外部的链接。外部链接越多,表示您的网站有更多的人关注。和大网站交换友情链接,可以使得网站的权重增加。

Tips:不是所有的外部网站的链接是有用的链接。PR 值越高的网站,搜索的权重越高。为了增加链接数量而引入垃圾链接网站,会遭到搜索引擎的惩罚。

查看搜索引擎收录的外部链接数量: link:网站域名

例如: link:www.baidu.com

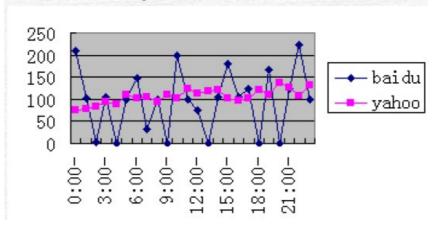
搜索频率:搜索引擎对网站的访问频率。网站内容更新越快,一定的周期内搜索引擎对网站的访问次数增多。

引用以前公布的《搜索引擎行为分析》为例,抽取某个网站的某天0:00-24:00的访问记录,**百度**的 spider 访问地址为:202.108.23.70,一天访问的次数为 20 次,crawler 的 IP 地址是 202.108.250.246,访问次数为 2312 次。访问时段如下:企业数据的页面收录时间为分插在各个时间段,对于文章的收入在 14:00-20:00,这个时间和网站更新新闻的时间比较一致。

yahoo 的 spider 访问 IP 为 202.160.178.114 和 202.160.179.127 和 202.160.180.172 每天的访问频率 25 次, crawler 访问次数为 2976 次。收录侧重于文章系统。

Google 访问 154 次,只访问有更新的页面。

Baidu与yahoo访问频率对比



用 site:http://www.???.com 在搜索引擎查搜索引擎收集数量为:

Google 收录数量	20,000
Baidu 收录数量	22,900
Yahoo 收录数量	6,170

可见,搜索引擎的访问频率不等于网站被收录的数量。Baidu 和yahoo 收集的方式需要耗费更多的服务器的 CPU 和硬盘访问的资源。

检查网站被收录网页的数量,可以在搜索引擎中输入 site:www.???.com 假设网站为 http://www.abc.com, 输入 为:site:www.abc.com

sandbox

沙盒效应指的是,新的网站在 Google 里面很难得到好的排名, 无论你怎么优化这个网站。换句话说,一个新的网站,可以有很丰富 的相关的内容, 可以有大量的高质量的链接,网站既搜索引擎友好,





Sandbox

也用户友好,所有一切都优化的很好。但是在一段时间之内,就是很难在 Google 里面得到好的排名。

通常沙盒效应会维持六个月,应该利用这段时间来增加网站的内容,寻求更多的导入链接。

(**内部连接**): 网站内部中某个网页的其他站内页面的链入的数量。内部链接越多,表示该页面在整个网站中的权重越重要。

内连接数量不适宜太多,否则搜索引擎可能认为存在桥页,对网 站进行惩罚。

pr 值: Google 使用 PageRank™ 技术检查整个网络链接结构,并确定哪些网页重要性最高。然后进行超文本匹配分析,以确定哪些网页与正在执行的特定搜索相关。在综合考虑整体重要性以及与特定查询的相关性之后,Google 可以将最相关最可靠的搜索结果放在首位。

PageRank 技术:通过对由超过 50,000 万个变量和 20 亿个词汇组成的方程进行计算,PageRank 能够对网页的重要性做出客观的评价。PageRank 并不计算直接链接的数量,而是将从网页 A 指向网页 B 的链接解释为由网页 A 对网页 B 所投的一票。这样,PageRank 会根据网页 B 所收到的投票数量来评估该页的重要性。

此外,PageRank 还会评估每个投票网页的重要性,因为某些网页的投票被认为具有较高的价值,这样,它所链接的网页就能获得较高的价值。重要网页获得的 PageRank(网页排名)较高,从而显示在搜索结果的顶部。Google 技术使用网上反馈的综合信息来确定某

内部连接



PR 值

个网页的重要性。搜索结果没有人工干预或操纵,这也是为什么 Google 会成为一个广受用户信赖、不受付费排名影响且公正客观的信息来源。

不同网页的地址,有不同的 pr 值。查询网页的 pr 值,可以通过安装 google toolbar 进行查询。也可以通过查询软件查询。Google toolbar 下载地址为:

http://toolbar.google.com/T4/intl/zh-CN/index.html?pro mo=abt-tb-cn

搜索深度

搜索深度

理论上,搜索引擎可以搜索到网站的每一个链接。搜索引擎每增加一个访问的层次,就会产生组合爆炸,对于高权重的网站,搜索引擎会增加网站的搜索深度,对于普通的网站,搜索的深度为一般为3层。对于访问深度4层以上的页面不再继续收集。

访问层次,假如搜索引擎访问网站的首页为访问入口,首页上所有列出的链接,为访问的第一层,第一层连接进去的页面上的链接为第二层、以此类推。

Tips

网站的 SEO 优化方向就是尽量降低页面的层次,有效的方法可以是: 1) 把重要的栏目和内容的链接提到首页, 2) 在页面尽可能多设置链接, 3) 建立网站地图 4) 优化分页的显示方法, 将传统的"首页、上一页、下一页、尾页", 可以采取如下的方法, 降低访问的层数:

url 收录优先

url 收录优先规则

搜索引擎优先收录网页权重或者 pr 值比较高的页面。象博客类、gov 类网站,由于其权重比较高,搜索引擎会优先收录。经常更新的网站,可以提高其网站的权重。

crawler

爬虫 crawler

搜索引擎根据 spider 收集回来的 URL 链接库去收集网站的程序。

sitemap

sitemap

GOOGLE 的定义为: Google Sitemaps 是您向 Google 索引提交所有网址并详细了解您网页在 Google 可见度的一条捷径。通过 Google Sitemaps, 我们始终可以自动得到您所有网页的信息及您更 改网页的时间,帮助您提高在 Google 抓取中的覆盖率。

Google Sitemap 最直接的作用体现在:

提供了一个向 Google 介绍自己网站信息、内容的迅捷途径。 了解 Google 对自己网站价值的判断及如何对待自己网站。

Google Sitemap 正是 Google 着力于改善这种误区的体现。通过google Sitemao,网站建设者可以告诉 Google,自己认为网站中哪些页面更为重要哪些页面则无关紧要。实现方法为给相应页面设定0.0 到 1.0 的优先级,比如说,可以将主页的优先级设为 1.0,栏目优先级为 0.8,内容页面优先级为 0.5,而其他网页如联系信息之类设为 0.2 等,这样,Google 便会对高优先级的网页更为关注。

更重要的是,网站建设者可以告诉 Google 网站更新的频率,以及每个网页的创建或最新修改时间,让 Google 在规定期限内检查相应网页的更新。比如说主页每天更新,而某些内容网页则可能每月更新甚至以年计。

Google Sitemap 另一个重要的功能是自 2006 年 2 月后提供的网页分析功能。这项功能分两部分表示: 网站内容中涉及的常用字词和主要的外部链接字词,这两项内容直观地反映 Google 对相应网站的看法,如果 Google 认为的您网站的主要字词与您希望强调的关键词不一致,或者,网站的反相链接中包含的字词与待强调的关键词有异,那么,意味着网站的优化工作还有很长的路要走。

Sitemap 申请入口有 Gmail 的帐号就可以申请 https://www.google.com/webmasters



Google 可接受多种格式的 Sitemap,但 Google 鼓励使用采用 Sitemap 协议的 Sitemap。因为这使网站拥有者可以提供除网址以外 的其他有关网页的信息,同时也因为 Google 是按创意公用授权-相



同方式共享 (Attribution-ShareAlike Creative Commons License) 的 条款提供此协议的,这样其他搜索引擎也可以使用此协议。我们希望 网站所有者只创建一个 Sitemap 文件供所有搜索引擎使用。

Sitemap 文件的可用格式为:

Sitemap 协议 - 这是 Google 创建的开放源码的 XML 格式的协议。使用 Sitemap 协议的 Sitemap 文件可包含您网站的网址列表以及有关这些网址的有用信息。您可以使用 Google Sitemap 生成器创建使用此协议的 Sitemap,还可以手动创建或者使用第三方生成工具。Google 建议使用此协议,因为它可扩展并且能让您向 Google 提供关于您每个网址的具体信息(如,每个网页上次修改的时间以及哪个网页拥有最高优先级)。

RSS2.0 和 Atom 0.3 供稿源。请注意可能仅提供有关最近网址的信息。

文本文件 - 可以向 Google 提供纯文本文件,其中每行包含一个网址。不过,Google 建议在拥有您网站的文本 Sitemap 文件后,即使用 Sitemap 生成器,根据 Sitemap 协议,通过这一文本文件来创建 Sitemap。然后您可以修改此文件,提供有关您网址的其他信息,如网址的上次修改时间以及更改频率。

Sitemap 可以是 xml 文件,可以是 ASP 程序,可以是 PHP程序,可以是文本文件...只要遵循 SITEMAP 的格式就可以,详细的格式可以参见下图。Sitemap 的文件不能超过 2M 或者 2万条记录。xml 文件必须是 utf-8 的编码格式。



我的 sitemap 文件

XML 标签

changefreq:页面内容更新频率可用参数:: "always", "hourly",

"daily", "weekly", "monthly", "yearly", o

lastmod:页面最后修改时间

loc:页面永久链接地址

priority:相对于其他页面的优先权,介于 0.1-1 之间 登陆 Google 提交你的 SiteMap 文件,给 Google 提交你的 Sitemap URL 后可以看见在列表里已存在,不过这时候还没有生效,必须过几个小时后 Status 栏变成 OK 表示正式生效。

我的 sitemap 程序,替换黑体字部分,可以成为你的 sitemap 程序。

```
<%
'sitemap_gen_db.asp
'A simple script (using database) to automatically produce sitemaps for a webserver, in the Google
Sitemap Protocol (GSP)
' by Francesco Passantino
'www.iteam5.net/francesco/sitemap gen
'v0.1b released 5 june 2005
'v0.2 released 17 june 2005 iso8601dates http://www.tumanov.com/projects/scriptlets/iso8601dates.asp
MAXURLS_PER_SITEMAP = 50000
'modify this to change website, baseurl and table
baseurl="http://www.gdgsyp.com/company/view.asp?companyid="
xDb_Conn_Str = "provider=sqloledb;server=localhost;database=数据库;uid=SQL 用户;pwd=密码"
'strsql = "SELECT ID, date from cor_data where id >= 1430000 AND id < 1460000 order by datex
desc"
'see http://www.time.gov/ for utcOffset
utcOffset=1
response.ContentType = "text/xml"
response.write "<?xml version='1.0' encoding='UTF-8'?>"
'response.write "<!-- generator='http://www.iteam5.net/francesco/sitemap_gen'-->"
response.write "<urlset xmlns='http://www.google.com/schemas/sitemap/0.84'>"
Set conn = Server.CreateObject("ADODB.Connection")
conn.Open xDb_Conn_Str
Set rs = Server.CreateObject("ADODB.Recordset")
rs.Open strsql, conn
Do while not rs.eof
    if URLS<MAXURLS_PER_SITEMAP then
         'modify this to change database field
         id_page=rs("id")
         fileImdate=rs("datex")
         'priority=rs("priority")
         if not isdate(fileImdate) then fileImdate=now()
         filedate=iso8601date(filelmdate,utcOffset)
         if priority="" or priority>1.0 then priority="1.0"
         response.write
"<url><loc>"&server.htmlencode(baseurl&id_page)&"</loc><lastmod>"&filedate&"</lastmod><priorit
y>"&priority&"</priority></url>"
         URLS=URLS+1
         Response.Flush
    rs.movenext
end if
Loop
response.write "</urlset>"
rs.Close
```

Function iso8601date(dLocal,utcOffset)

Dim d

'convert local time into UTC

d = DateAdd("H",-1 * utcOffset,dLocal)

'compose the date

iso8601date = Year(d) & "-" & Right("0" & Month(d),2) & "-" & Right("0" & Day(d),2) & "T" & _ Right("0" & Hour(d),2) & ":" & Right("0" & Minute(d),2) & ":" & Right("0" & Second(d),2) & "Z"

End Function

%>

关键字

关键字

当用户使用搜索引擎的时候,在搜索框输入的字是关键字。当 关键字查询次数名列前矛的时候,就是热门关键字。百度查阅热门关 键字的网址是: http://top.baidu.com/。在很多行业,所选的关键字是 查询次数非常少的,因此只要做关键字的堆砌,就可以轻易把关键字 排名做上去,这也是 SEO 行业的主要操作点。

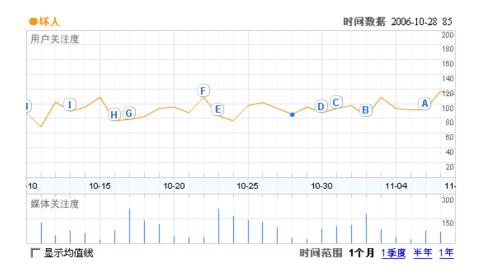
百度指数—SEO 的利器。查询你的关键字是否热门及每天/月/季的查询次数,可以使用百度指数。百度指数入口:

http://index.baidu.com/, 以"坏人"为关键字

坏人		
时间范围	用户关注度	媒体关注度
今日	112	87
1周	99 † +6%	92 ↓ -8%
1个月	92 † +10%	110 † +29%
1季度	87 ↓ -10%	105 ↓ -23%

注:1周、1个月、1季度 系均值





热门的关键字更容易引起客户的关注,带来的点击流量。热门的 关键字的排位也相对难做。

同一个关键字在同一个页面出现更多,会增加此页面在搜索引擎中此关键字的权重。关键字的挑选,关系到客户是否有更大的机率访问到你的网页。同义词由于语言的使用习惯,在搜索引擎中的搜索比率差别非常大,通过百度指数可以选择热门的关键字带来更多的潜在客户。大量的关键字堆砌能迅速提高 baidu 的排名,已经是不争的事实。尤其是 alt 注释,baidu 似乎不认为是作弊,alt 注释的堆砌成为当今最快捷安全的作弊方法了。控制关键词密度(Keyword Density),一般认为 2%-8%比较适合。文章的标题也是放置关键字的首选位置。一页中关键字的出现不是根据内容的需要而安排,而是为了讨好搜索引擎人为堆积关键字(Stuffing)。这已经被搜索引擎归入恶意行为(Spamming),有网站遭到惩罚的危险。可以使用关键字密度查询工具查询。

分词技术

分词技术





Alert

英文是以词为单位的,词和词之间是靠空格隔开,而中文是以字为单位,句子中所有的字连起来才能描述一个意思。例如,英文句子lamastudent,用中文则为:"我是一个学生"。计算机可以很简单通过空格知道student是一个单词,但是不能很容易明白"学"、"生"两个字合起来才表示一个词。把中文的汉字序列切分成有意义的词,就是中文分词,有些人也称为切词。我是一个学生,分词的结果是:我/是/一个/学生。不同的词组,是不同的关键字。

spam kill

搜索引擎会对恶意进行 SEO 的网站进行惩罚,如清除所有的链接。

百度对作弊的判定条件

- A 在网页源代码中任何位置,故意加入与网页内容不相关的关键词。
- B 在网页源代码中任何位置,故意大量重复某些关键词。即使与网页内容相关的关键词,故意重复也被视为作弊行为。
- C 在网页中加入搜索引擎可识别但用户看不见的隐藏文字。无论是使用同背景色文字、超小字号文字、文字隐藏层、还是滥用图片 ALT 等方法,都属于作弊行为。
- D故意制造大量链接指向某一网址的行为。
- E 对同一个网址, 让搜索引擎与用户访问到不同内容的网页(包括利用重定向等行为)。
- F 作弊行为的定义是针对网站而不是网页的。一个网站内即使只有一个网页作弊,该 网站也被认为是有作弊行为。
- G 有链接指向作弊网站的网站,负连带责任,也会被认为是作弊(但是,作弊网站上链接指向的网站,不算作弊)。

Google 对作弊的判断

- 请不要使用隐藏文本或隐藏链接。
- 请不要采用隐藏真实内容或欺骗性重定向手段。
- 请不要向 Google 发送自动查询。
- 请不要使用无关用语加载网页。
- 请不要创建包含大量重复内容的多个网页、子域或域。
- 请不要创建安装病毒(如,特洛伊木马)或其他有害软件的网页。
- 请不要采用专门针对搜索引擎制作的"桥页",也不要采用如联属 计划这类原创内容很少或几乎没有原创内容的"俗套"(cookie cutter)方式。



百度怎么老是只收我 20800 页

如果 A 级,百度会收 1000 页,B 级收 2000 页,D 级收 20800 页,E 级收 20900 页.如果百度判断你的网站等级为 A 级,那么即使你的页面数目远远大于 1000 页,site 出来也只会是 1000 页.相反,如果你的等级是 B 级,其实我只有 1000 多个页面,你 site 出来却是收录 2000 页.

Google dance

google dance

Google Dance 这个名词通常用于描述 Google 搜索引擎对其搜索结果进行更新、重组的过程。Google 一般每个月对其搜索数据库进行一次更新。在 Google Dance 时期(三到五天内),你可以发现Google 的搜索结果会有大幅度的波动。Google dance 是正常的行为。

Google dance 检测工具 http://www.google-dance-tool.com/

三、SEO 实操

- 1、SEO 框架
- 2、SEO 实操
 - a) 第一步: 优化网站(优化、网站地图)
 - b) 第二步:外连接(登录 SE、交换友情链接)
 - c) 第三步: 关键字
- 3、案例
 - 1) SEO 大师的网站是怎么做 SEO
 - 2) 网上商城怎么做 SEO
 - 3)一些 SEO 的例子
- 4、SEO常见作弊手法
- 5、SEO 不能做的事

1、SEO 框架



2、SEO 实操

第一步: 网站优化(结构和内连接)

- 1)、请按以下规则来策划或者规划你的网站:
- 尽可能多把重要的页面的链接和栏目放置到首页,尽可能把第二、三层的栏目标题抽取到首页,而不是没有意义堆砌。



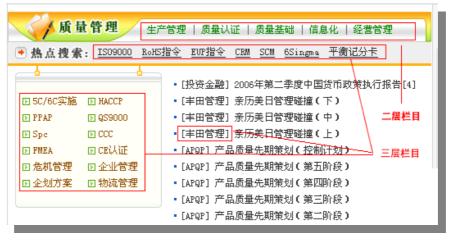
例 1: IT 信息门户网站

STEP ONE

示范



例 2 B2C 网上商城



例 3: 行业网站

公司库	加工市场	代理加盟	库存市场	二手转让
产品库	招投标	设备租赁	价格行情	商务服务
札械及行业 阀门及配件 泵机床 干燥设备	电焊、切割设备		金、工具 <u>[具 刀具/夹具 植</u> 量仪 <u>五金配件</u> 包	
▶ 电工电气、	照明	▶ 电·	子元器件 仪器仪	表
				
<u> 变频器</u> PLC 高	<u>压 低压</u>	正里	容器 电阻器 二根	<u> 20 显示器件</u>
	<u>压 低压</u> 电机 <u>灯具</u> 电光源		容器 电阻器 二极 则量仪器 示波器	

例 4: B2B 门户网站

		产品简介
◎ 油漆产品	◎ 油墨产品	□ 其它产品
PU高光金刚油系列 ->	pp油墨 ->	珠光粉 ->
电镀漆 ->	玻璃油墨 ->	金粉 ->
五金烤漆 ->	镜面油墨 ->	银浆 ->
塑胶机壳漆 ->	电镀油墨 ->	电镀油漆树脂 ->

例 5: 某企业网站

●尽可能采用静态页面,搜索引擎可以把您的页面全部收录,但使用 asp/jsp/php/cgi 程序您可以很方便调用您的数据库里面的数据,这不是搜索引擎友好的方式,解决此类问题的方法是您可以使用 sitemap 直接向搜索引擎提交所有的数据,也可以如下图示方式显示 更多的分页;



例: GOOGLE 的分页

●如果没有网站地图,请建立网站地图,并在首页下方加上 网站地图的链接入口;您还可以制作 sitemap 提交到 google 里面;

示范

设置网站地图,可以方便您的客户浏览,也可以让搜索引擎更清楚认识你的网站的结构,更容易抓取您网站的页面。

$\overline{}$	HOLD.
я	衻

一部由全体网民共同撰写的百科全书

传情

丝丝情意,字字传递

地区搜索

锁住特定区域,寻找身边信息

风云榜

一亿网民在找啥?网络热点大盘点

苗面

找电话最方便,为您网罗身边商家

TP9

歌曲大搜索,歌迷大聚会

手机娱乐

超级搜霸

在工具栏里 百度一下 立即下载

词典

最快速的英汉互译词典,支持中文成语翻译

地图

方便的出行指南,快速获取周边信息

国学

百度国学搜索,尽览文化精华

教育网站搜索

帮助您轻松查找各类教育信息

少儿捜索(侧试版)

面向少儿,绿色、健康的搜索工具

贴吧

例 1 某搜索引擎的网站地图

IT市场行情

下载

IT经销商

下载分类

软件下载

驱动下载 游戏下载

素材下载

资料下载

源码下载 下载专题

行业软件下载

PDA软件下载

软件自助上传

<u>笔记本</u>

頻道导航

导购指南 新品线报 产品评测 应用技巧 笔记本报

笔记本频道简介:提供最新笔记本电脑市场信息、业界动态的热点评论分析内容,为广大笔记本电脑爱好者、使用者打

手机

<u>手机产品库</u> PDA手机专区 智能手机 高像素手机 玄

<u> ชีโ</u>

手机频道简介:天极手机频道是集手机市场、手机产品导购设有手机游戏、手机八卦、智能手机专区等栏目满足不同用

数码

MP3随身听 数码相机DC 数码摄象机DV MP4 学习机

数码频道简介:关注数码类产品新闻报道、产品应用以及技产品、新技术和市场最新动态,即时报道全球最新数码产品

例 2 某 IT 网站地图

示湖

● 如果可能,请把友情链接放在首页,但不要包含在 javascript 程序里面。javascript 程序可以为页面带来很多动态的效果, 但不是所有的搜索引擎能访问 javascript 里面的数据。

什么是 javascript 效果-例如页面的文字上下滚动效果, 走马灯效果等:

● 如果您的网站有很多层目录,请把深层次的目录变成浅层次。



例 1 分页的深层次页面变为浅层次

把目录结构深达4层以上的网站,尽可能调整低。

● 对于内部的页面,您可以增加显示相同主题的文章和热 点文章的链接来增加页面的内连接;

◆ 上一篇文章: 禁止FSO

● 下一篇文章: 2003系统IIS设置404错误页无效解决方法

例 1 增加"上一篇 下一篇"的链接



例 2:增加同类文章列表

示范

示范

STEP TWO

🤼 Tags...... <u>Tags(标签云集)</u>

<u>观点(108) * web2.0(61) * google(50) * 人物</u>

(40) * 数据(39) * 视频(33) * 社区(33) * 互联网

(33) * UE(32) * seo(31) * 博客(29) * 搜索

(27) * 百度(27) * 创业(25) * 风险投资(25) * 游

<u>戏(24) * VC(23) * YouTube(22) * 案例</u>

(20) *设计(20) *微软(19) * sns(19) *广告

(19) * 网络招聘(18) * blog(18) * 译文

(17) * yahoo(16) * wap(16) * keso(16) * UI

(15) * 电子商务(14) * 忧化(13) * 谢文

(13) * MySpace(12) * 新浪(12) * 产品经理

(11)*开源(11)*3G(11)*阿里巴巴(11)*雅虎

<u>(11)</u> * <u>live(10)</u> * <u>网易(10)</u> * <u>方兴东(10)</u> * <u>Digg</u>

(10) *融资(10) * AdSense(10) * 电信(9) * QQ

(9) * blog network(9) * 营销(9) *

例 3: WEB2.0 的网站,使用 tag 链接更多的文章

第二步:外连接

1)、网站登录

A、必须登录的搜索引擎

Google http://www.google.com/intl/zh-CN/add_url.html http://www.baidu.com/search/url_submit.html http://cn.yahoo.com/docs/info/suggest.html

搜狐(搜狗) ttp://db.sohu.com/regurl/regselect.asp?step=REG

中搜

http://d.zhongsou.com/NetSearch/pageurlrecord/front pageurl.jsp

B、以下的如果有时间可以加入进去,可以提高外部连接的广度

搜豹免费登录入口

http://search.sobao.com/Computers_and_Internet/Personal/

协通免费登录入口

http://www.net2asp.com.cn/search/regform.htm 银河免费登录入口 http://search.inhe.net/navigate.asp 中国假日免费登录入口

http://www.china-holiday.com/newterms/hall/it/sort.as p?sortid=259

山西互连网免费登录入口

http://www.shanxi.net.cn/sxhulian/addcate.asp 网络奇兵登陆入口

http://www.net7b.com/net7b_site/denglu/index.asp 6128 搜索引擎免费登陆入口 http://www.6128.cn/add.htm 深圳网狐登陆入口 http://www.szfox.com/search/

亦凡信息娱乐网络一亦凡搜索

http://www.gotofind.com/opendir/ 法律网免费登录入口 http://www.law-lib.com/lawseek/wzdl.asp 温州信息港免费登录入口 http://www.wzonline.com.cn/

孙悟空免费登录入口 http://www.sunwukong.cn/add.php

猫头鹰 http://www.owlchina.com/publish.asp?l=pub

凯希 http://www2d.biglobe.ne.jp/~kinryou/cgi/url.html

zhao5.com 登陆入口 http://www.zhao5.com/deng_9.htm

三百六十行在线登陆入口 http://www.360hang.com/

好东西网址登陆

http://dir.haodx.com/computer/internet/search/search _egines/

知识引擎

http://www.kmcenter.org/web/User.asp 泰达搜索引擎 http://www.tedanet.com/search/index.asp Y28 Search http://www.y28.com/cgi-bin/link/join.pl

C、快速登陆-登陆

主要的软件有登陆奇兵、网站推广专家、名扬四海网站登陆系统。小的搜索引擎带来的流量并会如您不如预期的高。主要的流量集中在大的搜索引擎。

Visiting Spiders				
	Spider	Hits	% of Total Hits	Visits
1	Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slurp)	2,880	51.1%	1,550
2	Baiduspider (http:	2,312	41.02%	3
3	Openfind data gatherer, Openbot	19	0.33%	1
4	TurnitinBot	2	0.03%	
5	RSS-SPIDER (RSS Feed Seeker http:	1	0.01%	
6	SpiderMan	420	7.45%	
7	Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slurp))		0.01%	
8	SpiderMan)	1	0.01%	
Γota	For Spiders Above	5,636	100%	1,59

参考:搜索引擎某天访问数据分析

登录 DMOZ 等分类列表及搜索引擎网站,可以为您带来额外的访问者的流量和提高你的网站的权重。

D、sitemap 提交到 google

用 sitemap 提交到 google,并不等于 google 一定会收录 sitemap 里面的文件。Google 收录的文件数量,取决于您网站的 pr 值和 google 收录规则。

2) 友情链接

友情链接,**建议不要采取 flash 内建链接**和 **javascript 引发的链接**,搜索引擎不能访问到 flash 里面的内容。

添加友情链接,尽量添加 pr 值高的同类型的网站。文字链接和图片链接影响不会太大。

第三步: 关键字

1、title 处理

Tips



STEP THREE





title 是网站的描述,清楚表达网站的定位。一般为了让搜索引擎 更清晰知道内页的信息,**可以要求网站建设人员把文章的标题放在网 站的描述语之后。**

2、META 标签

META 标签是 HTML 语言 HEAD 区的一个辅助性标签。meta 标签通常用来为搜索引擎 robots 定义页面主题,或者是定义用户浏览器上的 cookie;它可以用于鉴别作者,设定页面格式,标注内容提要和关键字;还可以设置页面使其可以根据你定义的时间间隔刷新自己,以及设置 RASC 内容等级。

META 标签分两大部分: HTTP 标题信息(HTTP-EQUIV)和页面描述信息(NAME)。

★HTTP-EQUIV:回应给浏览器一些有用的信息,以帮助正确和精确地显示网页内容。常用的 HTTP-EQUIV 类型有: 1)、Content-Type 和 Content-Language (显示字符集的设定) 2)、Refresh (刷新): 让网页多长时间(秒)刷新自己,或在多长时间后让网页自动链接到其它网页。3)、Expires (期限):指定网页在缓存中的过期时间,一旦网页过期,必须到服务器上重新调阅。4)、Pragma (cach 模式)禁止浏览器从本地机的缓存中调阅页面内容。

★NAME 变量是描述网页的,对应于 Content(网页内容), 以便于搜索引擎机器人查找、分类(目前几乎所有的搜索引擎都 使用网上机器人自动查找 meta 值来给网页分类)。

name 的 value 值(name="")指定所提供信息的类型。有些值是已经定义好的。例如 description(说明)、keyword(关键字)、refresh(刷新)等。还可以指定其他任意值,如:creationdate(创建日期)、document ID(文档编号)和 level(等级)等。

name 的 content 指定实际内容。如:如果指定 level(等级)为 value(值),则 Content 可能是 beginner(初级)、intermediate(中级)、advanced(高级)。

1)、Keywords (关键字)

说明: 为搜索引擎提供的关键字列表

用法: <Meta name="Keywords" Content="关键词 1, 关键词 2, 关键词 3,关键词 4,·····">

注意:各关键词间用英文逗号","隔开。**META 的通常用处是指定搜索引擎用来提高搜索质量的关键词**。当数个 META 元素提供文档语言从属信息时,搜索引擎会使用 lang 特性来过滤并通过用户的语言优先参照来显示搜索结果。例如:

<Meta name="Kyewords" Lang="EN"
Content="vacation,greece,sunshine">
 <Meta name="Kyewords" Lang="FR"</pre>

Content="vacances,grè:ce,soleil">



Tips

2)、Description:告诉搜索引擎你的网站主要内容。

用法: <Meta name="Description" Content="你网页的简述">

3)、Robots (机器人向导)

说明: Robots 用来告诉搜索机器人哪些页面需要索引,哪些页面不需要索引。默认是 all。

用法: <Meta name="Robots"

Content="All|None|Index|Noindex|Follow|Nofollow">

all: 文件将被检索,且页面上的链接可以被查询;

none: 文件将不被检索,且页面上的链接不可以被

查询; (和 "noindex, no follow" 起相同作用)

index: 文件将被检索; (让 robot/spider 登录)

follow:页面上的链接可以被查询;

noindex: 文件将不被检索,但页面上的链接可以被

查询; (不让 robot/spider 登录)

nofollow: 文件将不被检索,页面上的链接可以被查询。(不让 robot/spider 顺着此页的连接往下探找)

4)、Author (作者)

说明: 标注网页的作者或制作组

查看 title 和,可以如下操作:



示范

Tips

<meta http-equiv="Content-Type" content="text/html;
charset=gb2312" />
<meta name="description" content="首节战报: 姚明 6 分
麦迪爆发 14 分 火箭 27-23 马刺">
<meta name="keywords" content="首节战报: 姚明 6 分
麦迪爆发 14 分 火箭 27-23 马刺">
<meta name="robots" content="all">
<meta name="robots" content="all">
<script src="http://www.sohu.com/sohuflash_1.js"
type="text/javascript"></script>
<title>首节战报: 姚明 6 分麦迪爆发 14 分 火箭 27-23 马刺-搜狐体育</title>

进行 title 和 meta 标注,可以方便搜索引擎对网页进行分析和收录。

3) 关键字选定

关键字是 SEO 的重心。

● 选择关键字的依据

关键字的选择不只是选择反映行业的一些特别词汇,选择的关键字要符合用户的搜索习惯。



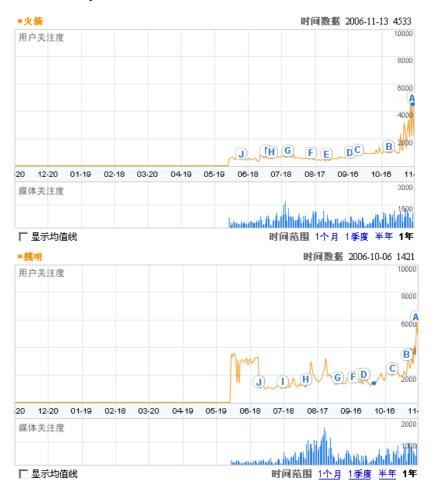
检查是否热门关键字,可以用如下的方法:

示范

到 google 和百度用该关键字去搜索下,检查收录的网页的数量。 也可以用百度指数去分析该关键字的情况。Google 可以用 google 趋势。

例如:要做一个"姚明"的专题,选择什么样的关键字是最合适的?初步选择为:火箭、姚明、NBA和篮球

进入百度指数: http://index.baidu.com





平均用户的关注度为: NBA》姚明>火箭》篮球

篮球		
时间范围	用户关注度	媒体关注度
今日	2752	570
1周	3199 † +5%	486 ↓ -17%
1个月	2881 † +27%	566 † + 5 %
1季度	2707 † +21%	722 † +20%

注:1周、1个月、1季度 系均值

姚明		
时间范围	用户关注度	媒体关注度
今日	4878	1109
1周	3843 † +62%	829 † +14%
1个月	2545 † +65%	655 † +89%
1季度	1931 † +4%	583 † +69%

注:1周、1个月、1季度 系均值

nba		
时间范围	用户关注度	媒体关注度
今日	79580	865
1周	92976 † +35%	933 ↓ -43%
1个月	57162 † +154%	944 † +138%
1季度	31237 † +28%	636 † +26%

注:1周、1个月、1季度 系均值

火箭		
5407## ED		444 AL 34 344 PT
时间范围	用户关注度	媒体关注度
今日	2172	1160
1周	3156 † +76%	976 ↓ -3%
1个月	1694 † +93%	915 † +64%
1季度	1023 † +82%	695 † +11%

注:1周、1个月、1季度 系均值

当日用户关注度: NBA》姚明>篮球>火箭



网页收录数量为:篮球》NBA>火箭>姚明

google 趋势 http://www.google.com/trends



结论:由上可见,搜索引擎上用户的访问习惯是:NBA》姚明>火箭>篮球,而收录的网页的数量,并不一定带来更多的用户的关注。 所以选择以NBA为主关键字,姚明和火箭为副关键字。搜索引擎收录的网页数量多少,并不是反映访问者的真实的情况,所谓收录数量,

Tips

Tips

只是搜索引擎本身对用户习惯的研究结果所制定的规则,然后按照该规则去网上采集网页。关键字的选定,请用好百度指数和 google 的趋势。

● 关键字密度

关键越多,带来的流量会越大,但对于某一个页面而言,**某个关 键字越多,被搜索引擎收录后,该关键字的排名会越高。但故意堆砌** 关键字,可以被认为是作弊行为,有可能会导致站点就搜索引擎清除。

关键字的放置,可以在如下地方放置: 网页的 title, 网页的 meta, 图片的 alert, 及文章的标题和文章的内容。所以网站的文章中,也要修改为包含适当的关键字。这是网络媒体和传统的印刷媒体的一个本质的区别。网络的文章标题,不单止要吸引眼球,引发浏览者点击的欲望,还要考虑进行设置关键字更符合网络传播的规律。

关键字的密度有软件可以检查。

建议:1个网页设置1个主关键字,关键字不超过2-3个。不同的关键字,可以分散设置在不同的页面作为主关键字,然后通过内连接的手段把页面联合起来,这样可以充分发挥选择的不同的关键字的作用。

● 关键字的更新

由于网络热点不断改变,由"木子美""芙蓉姐姐""天仙 MM", 热点不断产生,过去的热门关键字,也有可能成为冷门的关键字。搜索引擎会根据用户的查询关键字不断调整收录的网页,所以不断进行 关键字的追踪和更新,保持热点,可以更好适应网络营销的发展。









对于热门的关键字,可以采取专题和不断更新包含热门关键字的内容的文章来充实站点。热门的关键字的文章,通常也是浏览者最关心的文章。搜索引擎会根据您的网站的表现,不断提高网站的权重,收录更多的网页。

对于热门的关键字, baidu 会人工干预排位。(阅读材料)

很多从事 SEO 的人会选择行业的一些专业的词汇做优化,因为 这些关键字没有人做优化所以很容易排名在前面的第一第二位,但因 为冷门,被查询的机会会非常少。关键字应该选择行业内常用的热门 关键字,这样才能带来有用的浏览者。

- 3、SEO 实施案例分析
- 1) SEO 从业人员是怎么做自己的网站的

从 baidu 和 google 中输入 "SEO",以各自收录前 5 位的网站来综合分析各 SEO 大师是怎么进行 SEO 的。

以 2006 年 11 月 15 日星期三下午 17: 10 分 KEE 查询结果为:

Google.com 前五位收录的站点是

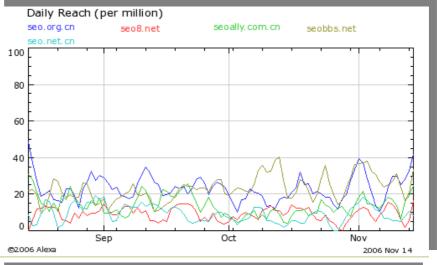
www.seo.org.cn www.seo8.net seoally.com.cn www.seobbs.net www.seo.net.cn

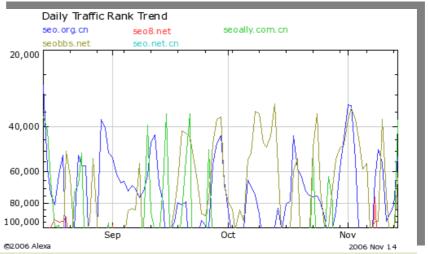
baidu.com 收录前 5 位置的站点是

www.seo.org.cn

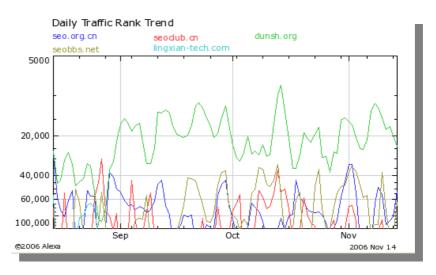
www.seoclub.cn www.dunsh.org www.seobbs.net www.lingxian-tech.com

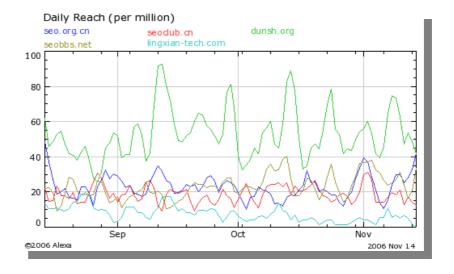
以以上10个站点进行分析。





上: google 前 5 名对比,下: baidu 前 5 对比





网址		www.seo .org.cn		seoally. com.cn	www.seo bbs.net	l	www.se oclub. cn	www. dun sh. org	www.ling xian- tech.com
google表现	排名 (SEO)	1	2	3	4	5			
	收录数量	5,330	2,790	1,990	46,500	72	3,540	927	316
	收录数量 (SEO)	3,420	1,380	1,690	29, 300	72	1,640	487	281
	链接数量	133	135	156	20	315	140	306	78
	PR	4	4	4	5	4	4	4	4
baidu表现	排名 (SEO)	1			4		2	3	5
	收录数量	1	0	0	38,600	0	38,600	34,800	10,100
	收录数量(SEO)	1	0	0	38,600	0	10,700	28,000	4,420
	链接数量	145	64	8	1,640	23	158	2,050	9
ALEXA表现	排名 (3mos)	53,411	117,447	69,937	50,823	181,762	61,115	17,404	139,378
	排名 (1wk)	49,881	141,629	59,961	55,762	255,938	84,125	13,847	306,537
	PV (3mos)	4.5	4.7	7.3	5.1	1.7	4.2	8.4	4.6
	PV (1wk)	3.1	3.9	7.5	4	1.6	3.2	10.2	1.6
	REACH(3mos)	23	9.2	13.5	23	9.25	18.5	53	7.5
	REACH(1WK)	30.5	9	18	23.5	7	17.5	60.5	5.5
	linkin	184	14000	100	1500	124	59	10	10
	网站系统组成							wordpre ss+discu	
		joke+bbs	joke+bbs		phpwin			z	
	首页关键字-seo	137	142	80	287	41	327	109	74
	网址含SEO	102	97	64	255	23		73	
	页面关键字	35	45	16	32	18	34	36	26

页面表述:

www.seo.org.cn

- 其他关键字数量: google15 搜索引擎 15 百度 11ALEXA13
- KEYWORDS: SEO,搜索引擎排名,搜索引擎优化

● CONTENT:中国最大的 SEO 研究社区,提供 SEO 动态、SEO 技术资料、SEO 工具软件、SEO 交流论坛

www.seo8.net

其他关键字数量: GOOGLE:47 搜索引擎 47 网络营销 5 由上可见, SEO 是综合应用的结果,并不是外连接、内连接或者是关键字堆砌数量的单一描述,每个搜索引擎有自己的排序规则。

2) 网上商城是怎么做 SEO

以"鲜花"为关键字,到 baidu 和 google 搜索,取第一个网站(非广告投放)为例

A. 关键字

<title>XXX 鲜花 - 专业的鲜花速递服务商|北京鲜花|商务鲜花|鲜花、花礼品</title>

<meta name="Description" content="中国最专业的鲜花速递服务商,xxx 鲜花,鲜花,鲜花特卖上海鲜花北京鲜花广州鲜花深圳鲜花武汉鲜花西安鲜花预订鲜花,礼品,鲜花商城,北京鲜花,国际鲜花,鲜花礼品,商务鲜花,爱情鲜花,友情鲜花,祝福鲜花,探望鲜花,生日鲜花,鲜花时尚,迷你花卉,flower,gift,interflower,flowershop,sendflower" />
<meta name="Keywords" content="xxx 鲜花鲜花礼品 商务鲜花 爱情鲜花 友情鲜花 祝福鲜花 探望鲜花 生日鲜花</p>



示范







Tips

鲜花时尚 鲜花生活 国际送花 天使花房 迷你花卉 flower gift interflower flowershop sendflower" />

以上已经包含 41 个"鲜花"页面有 13 个"鲜花",

- B、页面优化
 - 首页增加了二级页面的链接



■ 二级页面增加了内连接,使得各二级页面互相并联,增加了内连接。



■ 网站最下方增加了日常更新的栏目,让搜索引擎知道网站不断在更新和完善,不断调整访问频率。



- 建设了完善的网站地图。
- ◆ 文章的显示标题反映了当前页面的内容。

Tips



google 收录网页总数量为

site:www.salaXX.com.cn

9370

google 收录页面包含"鲜花"网页数量为

鲜花 site:www.salaXX.com.cn

5080

baidu 收录网页总数量为

site:www.salaXX.com.cn

38700

baidu 收录页面包含"鲜花"网页数量为

鲜花 site:www.salaXX.com.cn 2630

C、交换外部链接

google 外部链接

link:www.salaXX.com.cn

2200

3)一些极端的 SEO 例子

极端的 SEO 页面,会破坏用户体验,有可能被搜索引擎认为是作弊。



例:某网站的 SEO 页面见左图,将数据库中的文章按照不同的关键字,抽出整理成该关键字的文章列表。如以IIS为列表的网页,里面包含了650多篇文章的链接。这类页面,是人无法进行浏览和阅读的。

4、SEO 作弊

针对搜索引擎和 ALEXA 排名是 SEO 走进极端的例子,搜索引擎根据作弊的严重行会给网站不同的惩罚。

常见的作弊手法有:

1) 内部连接

- 泛域名解释:利用 spider 的特性,故意把内页变成泛域名, 然后提交大批量的泛域名给搜索引擎。
- LINK FARM 作弊者通过建立若干不同的域名但是是同一 IP 地址来构建 LINK FARM 提高目标页面排名,并欺骗搜索 引擎这些链接页面分属于不同拥有者.对于这种情况可以进

行 DNS解析分析,大量不同域名映射到同一IP地址很有可能是作弊网站.

- 桥页,桥页就是在第三方网站上做一个页,堆砌了大量的 关键词,让该页在搜索引擎收录里排在前边,如第一页, 然后,在桥页中注入转向代码,用户来到该页,实际上却 去了另外一页。这里,第三方网站一般是一些 BLOG,这 些 BLOG 允许一些转向 HTML 代码及 Javascript 代码,包 括 blog.donews.com。搜索引擎的对策是:如果 A 页是来 自于搜索引擎,下边的算法启动:如果 A 页被 A 用户浏览 且不超过 N 毫秒,然后马上导向到 B 页。则标识 A-->B 为 桥页 spam。如果被标识次数大于 N,N 可以是 3,且来自 于不同的 IP,就可以降低 A 页的权重了,但不应该去动 B 页。
- 对同一个网址,让搜索引擎与用户访问到不同内容的网页(包括利用重定向等行为);用脚本或者 frame 形式强行把用户从一个页面跳转到另一个与前页内容毫不相关的页面,被视为重定向作弊。储存重定向作弊页面的网站以及被导向的目的网站将做屏蔽处理。

2) 关键字

 大量的关键字堆砌能迅速提高 baidu 的排名,具体表现如 在标题域、Meta 的 Keywords、Description 域、图片 Alt 域、正文内等地方,大量重复堆砌相关或者不相关关键词。 无论用户是否能看见,一律视为内容作弊。若关键词与网 页内容相关,则做降权处理;如果不相关,一律屏蔽。

- 页面多使用重复的关键词,甚至可以使用一些隐藏层
- 目录作弊,把目录名字取成关键字
- 堆砌和文章和网站根本无关的热门关键字

3) 外部连接

● 链接诱饵(link baiting)。发布一些独特性的资料,诱使人主动增加你的网站的链接。

BLOG SPAM

- A、 作弊者建立 blog,拷贝排名比较高的 blog 内容,这样能够尽快提高自己 blog 排名,然后在评论里面增加指向目标页面的链接或者做广告.
- B、 自动建立若干 blog 站点,但是用户点击会自动转向目标站点.类似于 URL 跳转方式.
- Honey pot 作弊方式是作弊网页作者发布有价值信息比如介绍某项技术的技术文章,但是在发布的信息里面隐含指向作弊页面的隐藏链接,这样通过网页本身内容的价值来诱导其它网站增加指向该页面的链接从而间接增加作弊页面的排名.比如往 blog 站点,BBS 站点,留言簿或者 wiki 等可以任意发言的站点增加评论.在发表的评论里面增加指向目标页面的链接,这样会间接增加目标页面的 pagerank 排名.根据发表评论类型可以分为以下两种:一种是直接发广告性质的与主题无关评论,这个我们经常在 BBS 或者留言版上看到.

一种是评论与主题相关,但是在评论中隐藏无法看到的链接信息.

5、SEO 不能做的事情

- SEO 带来的流量,能不能把潜在的客户变成自己的客户,需要企业和网站去努力;
- SEO 会破坏用户体验,如何把用两者协调好,需要网站策划的时候就要考虑。
- SEO 不能代替其他的营销模式,只是主动去实施,但结果不由自己控制的的营销手段,需要持之以恒。

阅读材料

《21 世纪》采访百度 CTO 刘建国

对百度,CTO 刘建国的采访过程就是一个挤牙膏的过程,他永远不会借题发挥,对采访者的问题,回复得简练而专注。

这位与李彦宏一起创立百度的"技术新官僚",对搜索引擎技术运用的思考,在一定程度上体现了这个新上市公司的商业方向及这个团队的技术想象力。

采访中,刘建国不断接到要求解封的"作弊网站"的电话,大多是熟人才会直接将电话打到他这里,或是打给李彦宏。采访由此而展开。

反作弊是我们的技术

《21世纪》:李彦宏选择搜索引擎作为百度的主业,当时是怎么确定的?

刘建国:2000 年之前,中国有许多搜索引擎公司,Goyoyo 是香港一个公司做的;科学院有一个网站,还有北大天网。当时创立公司时搜索引擎已经很多了。我们机会在哪里?在当时我和李彦宏的通信中,他说要做就做最好的。当时已有的搜索引擎速度慢,索引小,用户体验很差。我们都认为做一个中文搜索引擎有很大前景,所以我就从北大出来,跟李彦宏一起创业了。

《21世纪》:现在来看,当时搜索的核心技术是什么?

刘建国:分成好几块,一块是网页抓取,由蜘蛛程序选择、抓取互联网中的网页,还要持续不断地更新网页,将过期的网页剔除掉;二是超链分析和内容分析;三是针对用户的搜索服务:如何处理海量数据和大规模的用户访问?如何提供稳定的服务?有很多技术问题需要解决。

《21世纪》:你们是一帮人在做?

刘建国:好多个组,重点是相关性的评价。哪些网页是用户想要的?后面就有好多技术;比如超链分析;这里有李彦宏的专利。超链在 Infoseek 之前就有成熟的技术,李彦宏在那儿工作时就研究得很细,但我们没有仅用这套方法,而是根据当时搜索引擎全球的发展做了改进。

《21世纪》:百度在相关性方面有什么特色?

刘建国:有两个基本方面:超链分析和内容分析。我们这帮技术人员,包括李彦宏,根据搜索引擎当时的技术发展,做出了一套百度自己的超链分析系统。另外还得分析网页内容,

比如中文处理就很重要,要理解中文的语法和特定文字意义,因为中文与英文还是不一样的,只有判断文字是什么意思,才能给用户提供搜索结果。

《21世纪》:"欺骗搜索"是什么意思?你们有什么独特的技术来对付它?

刘建国:就是所谓的网络优化,比如 SEO 之类,链接一些域名,好象**链接很多**,但实际上是人为做出来的。还有一些链接,就是专门帮别人做链接,这也是超链分析。你不是链接分析吗?专门针对你。还有一些网站,专门给别人做,相关度也高,我们必须发现哪些是LINKFORM,然后分析出来。

《21世纪》:这如何看得出来?有简单工具吗?

刘建国:我们已经可以通过技术手段及时发现这种欺骗或者叫作弊。是否作弊,有一个很简单的标准:用户看到的内容和搜索引擎看到的内容是否一致?前一段搜索"芙蓉姐姐",进去之后很多是买手机的,买其他信息,这就是作弊。我们就分析,他们是通过什么方法来"欺骗搜索"。比如 LINKFARM,我们靠技术发现出来,它已经"堆砌关键词"了,我们看出它是作弊的,我就不要它。这就是为何后来反弹这么大,有了"反百度联盟"的原因。有"反百度联盟"成员说,你为什么屏蔽我?前一段搜狐网的一个峰会,一个医药网的说我们很好啊,你为什么把我给封了?它的"堆砌"是作弊的,如果我们把它放开,对用户是一种欺骗。

《21世纪》:比如在百度搜"芙蓉姐姐",你们把有些"芙蓉姐姐"链接给屏蔽了,那就不能给用户提供更多的结果了?

刘建国:是更准确,比如用户到这里来搜,得到的是真正的"芙蓉姐姐"而不是手机铃声。 这样做的目的是给用户很好的用户体验。

《21世纪》:在反作弊技术里,有什么可以明确识别作弊?

刘建国:比如,来回互相链接、堆砌关键词等。还有前景色和背景色完全一致,在 HTML 里面,用户看不见,可以骗搜索引擎,就这样骗用户。

《21世纪》:仅仅背景色也可以用来作弊?

刘建国:背景是黑色的,搜索看见的仅仅是内容,是看不到背景的。前景与后景一样,机器看不着,人能看出来,景色也是识别的因素。其中的文字就是有些网站想欺骗的东西。 有很多方法作弊,层出不穷的。搜索与作弊网站是永远不会结束的战争。

《21世纪》:有没有可以"招安"的作弊网站?

刘建国:不会的。他们的利益很大,骗一点是一点,不会到这里来做。

《21世纪》:有多少这样的网站?

刘建国:每天碰到成千这样的网站。

《21世纪》:他们到哪里拉客户?

刘建国:不一定拉客户,他们可以自己做铃声、短信,SEO会拉企业,给钱,我给你排位。

另一个"敌人"是 Google 吗?

《21 世纪》:与 Google 的反作弊比, 你们有什么优势?

刘建国:说 Google 的技术比我们好,我不认同,他们主要是不够集中。比如反作弊的,Google 中文就做不过我们,这有技术问题,还有人力,要知道作弊惯用的手法,而 Google 不知道。这个模式我们比 Google 发展得更快,我们更接近用户。我们中文的理解做得比他们好,比如像分词、切词,中文是一个一个字,放一块儿,那些字和词是不一样的。我做得比 Google 好。他们没有专门研究中文。

《21世纪》:你们是用自己的技术?

刘建国:是的。我们中文分词,这个技术里面没有英文,中文本身的处理我们已经到位,还有语法的应用,我们做得很好。语法是主谓宾,如何用,老外不理解。

《21世纪》:你们用以前的文字识别技术吗?

刘建国:对中文我们有一套独特的做法,现在成熟的技术都不太适合我们。

《21世纪》:有没有用哪个平台?比如大的软件?

刘建国:这个软件是我们自己做的。

《21世纪》:是应用软件?

刘建国:不是一套。我们直接用自己的东西。比如说词典、分词、词汇。这不是释义的词典,可能有很多资源。比如我们用人民日报的语料库,我们购买这样的资料。更多的平台是我们自己的搜索引擎。我们有日志,总结新的词汇。我们对 10 亿计的网页也能分析。这是 Google 中文做不到的。至少没有聚焦。另外从中文用户体验来说,我们也做得比他们好,比如用户体验是什么、习惯是什么。比如相关搜索,在中文里面,用户心里想的一个事情,比如"软件"这个词,用户心里想的是软件下载,但搜索的是"软件"这个词,他输入"软件",与搜索不符。百度会提示得更加精细化,到底是软件下载还是加密,全自动服务,这样用户会觉得很好。类似的工作我们下了很多功夫。其他像拼音检索,错别字纠正,我们也能正确提示。现在有些搜索正在抄我们。我认为不能崇洋。

从竞价排名平衡用户体验

《21 世纪》:从用户体验来看,曾经一度 Google 要比你们好些,比如竞价排名不是他们主

要标准,信息的直接有效很重要。

刘建国:这是另外一个误区。搜索结果的首页我们有多少竞价排名呢?没多少,最多不超过20%,我们的原则是用户体验是第一位的,而且是真正相关的。现在20%是被人定制的,其他是免费体现的,是有用的。实际上,20%的竞价排名也是在严格的人工审核下去通过的。

《21世纪》:人工审核能不能有一个例子?

刘建国:我们有一个30人的组,比如用户提一个词,看是不是真正跟商品有关的。第二,链接的这个网站,是不是做商品的。即使是客户定制的竞价,竞的是某个关键词,我们也要看是不是商品、是不是服务。

《21世纪》:如何决定?

刘建国:人工能查出来,当然也有很多调整。比如 IBM,我们考虑把 IBM 放在最前,后来还是把 IBM 中国放在前面。虽然阿里商城(IT 销售商)交钱了,但第一个仍是 IBM 中国,我们也是考虑到用户的体验。

《21世纪》:就是说要在用户与客户之间平衡?在网页上还有哪些直接体现?

刘建国:收费的内容均有标志,比如"百度快照"不是收费的。

《21世纪》:你们卖得最好的关键词是哪些?

刘建国:比如"鲜花",就有很多人买。第一条是买的,是实时变化的,谁的价高就买到最前面去了。

百度不再用"插件"营销

《21世纪》:你们的营销方式曾经是以"一个插件"(业界通称为"流氓软件")来解决?

刘建国:这不是最关键的。当时主要是与3721打,竞争太激烈了,这是环境所致的。

《21世纪》:现在"插件"还用吗?

刘建国:我们插件装得挺少的。与有的搜索比,百度这种"流氓软件"最少,插件不是主流的,现在都已经走向正规了。插件只是推广手段,今后变成用户自己主动。百度的插件是可以卸掉的,比如搜霸。但有些搜索,没有点就生成了,这种推广手段我们以前做过,但后来停了。如果用户要觉得好,可以自主选择,我们可以保证,用户不会被动选择、不遵从他的意愿——被动强制性的。

《21世纪》:你们去掉了哪些"插件"呢?

刘建国:我们的不会自动弹出来,也不会误装。

《21世纪》:我们可不可以说,百度以前用过"插件",现在不用了?

刘建国:现在都不用了。我们就是一个搜霸,百度联盟,而且推广方式不是强制性的。搜霸确实也有好处,节省了用户时间,一步到位。我们不会用以前的推广方式。但有些网站点进去就是网络实名,就是靠弹出,非意愿的。我们原来也用得很少,也是百度联盟成员自己用的因为装上后有收益。现在谁利用这种方式装,就不让加入联盟了。

《21世纪》:有多少个联盟成员?

刘建国:现在有好几万。如何管理这些联盟也是一个大问题。

技术与人工、谁排第一?

《21世纪》:刚才提到相关性、超链分析、反作弊,结合起来形成了你们的核心价值吗?

刘建国:是。但如何调节这些东西,各家都不一样。超链分析是共用的。大家都在用。没有一个公司去公布搜索技术,因为这是一个商业秘密。我们是经过 5 年多积累才出来这样好的结果。这些都是核心技术,一定得积累。

《21世纪》:共享的经验有哪些?

刘建国:各个公司不一样,不能公布细节。

《21世纪》:你们重要在分析用户还是客户?

刘建国:重要的是分析网页,网页与网页之间的分析。WEB 的方式,是由 HTML 组成的。网页与网页之间的是 URL,最大好处是网页与网页之间可以用超链来指,叫 HYPERLINK,指向另外一个网页是做什么的。比如"中国人民银行"会指向很多东西,每个都是链。他们指向他,LINK 有一个叫评价,U1 对 U2 的评价。第一是指向,第二是评价。这是一个网页,"中国人民银行"就是超链,他指向另外一个——有说明中国银行是做什么的,指向后,另一个是指向的 U2,另外还有 U3。很相关的排前面,找到最权威的。最后综合加权,中国人民银行大,其他商业银行小,那么就加权,算出来结果谁排前面。

《21世纪》:你们做这些,有没有编辑?

刘建国:有的网站有,我们没有,但我们有人工。网站不需要编辑,我们是靠技术,要把这些东西都拿过来。刚才说的抓取、分析是否相关,生成抓取的排名。抓取要分析,之后建立索引,把内容到 URL 的一个检索,之后提供服务。抓取,分析、检索、服务四个环节。

《21世纪》:你们的技术部门是按四个环节来分的?

刘建国:我们有很多组,做不同工作。

《21世纪》:你们的生产环节与通常的产品定义不一样。

刘建国:是有些怪。比如说我们是如何攻关:要分析、提出问题,每个部门都提,李彦宏也会参与。

《21世纪》:你们最近提过什么问题?

刘建国:我还不想透露太多的细节,比如抓取。我最近要提出一个问题:网站的覆盖率。要研究中国所有的网站。我们平常与 Google 比较,我们来看交集,相同的内容是多少?各自抓多少?我们发现交集不多,我们独自的较多,这是所有搜索的共性。我提出说,要去调研,为什么会有这种差距,这种差别是如何产生的,没抓的是什么特性?Google 抓的是什么特性?Google 是海外的吗?这都是一种猜想。我们要去调查,然后去做方案,方案再讨论是不是可行。这样持续不断地做,技术的含量很高。效果都是很客观的,都是真刀真枪的。

《21世纪》:技术队伍如何分工?

刘建国:除了研发还有测试的,另外还有运营的,给我们提供服务是在中国电信或网通,另外要管机器,硬盘坏了,负责换,系统要上线,数据有哪些问题,都得分析。研发、运营、测试还有客服(有面向用户及企业的)。有问题可以直接发到电子邮件,我们暂时没有呼叫中心。为用户解决问题,先整理一些问题,再传到其他部门。

《21世纪》:你们的产品,是以什么形态呈现出来的?

刘建国:网页、MP3、"知道"等等,每个频道都是一个产品。

《21世纪》:光靠技术行不行?还是需要编辑?

刘建国:主要靠技术,这么大量信息处理,靠人工不行。有人工,也是IT辅助下的人工。

《21世纪》:专业性的频道,比如百度传情、影视频道,百度"知道",也是技术生产出来的?

刘建国:这是搜索"人脑",技术生成就行了。

《21 世纪》:你们作为纯技术派的人,认为百度在技术上与微软、Google 竞争,真得那么轻松吗?

刘建国:我认为百度这么多年为何能成功一个主要的原因,就是我们的专注,只做搜索,中文搜索。对于要赶超的人,第一是专注,要超越自我。我们在不同的场合说过,我们的对手是我们自己,不是其他的人。我们能否保持旺盛的斗志,我们在艰难岁月里保持的那些东西能否保留下来,我们每个人是不是都很进取、协作。这是决定我们最关键的东西。此外还要继续招优秀的人。

"知道"、贴吧核心解密

《21世纪》:搜索技术会有大的变化吗?

刘建国:有可能要变化,一句说不清楚。我觉得要保持创新,要监控业界的变化,要跟得上变化。以后的方向很多,比如说个性化需求,不同的人查到不同的结果,也许真有不同的发展。

《21世纪》:解决个性化需求依靠什么技术?

刘建国:了解这个人的个性,知道这个人的背景、身份、年龄和习惯,这样来调整搜索;另外无线搜索也是新市场。通过手机搜索及其他智能终端,针对不同的需求,专门建立无线搜索频道;其他的应用,比如地图搜索,现在 Google 正在做,百度是否要做,现在还没有完成研发。

《21世纪》:让我们来说说"知道",它的哪一点技术比较关键?

刘建国:"百度知道"是用来找知识的。没有所谓核心技术,它是一个平台,让问的人和回答的人都能在这里自动滚动,提问和回答都很方便。平台的界面和流程都要建得好,另外评价的系统很好,加权平均的系统,没有一个核心的技术。

《21世纪》:你们对客户进行管理用什么系统?

刘建国:竞价的用户,一个是客户(企业),一个是用户(网民),我们不需要管理,我们只需要服务就行了。比如对客户,主要是提升服务效果,对生意有帮助。

《21世纪》:那你们不要考虑客户的忠诚度,也就是贴性?

刘建国:我们是看效果,不怕他流失,他能找到知识就来,要靠产品的价值来吸引,而不是管理。如何建立很有粘性的平台,比如"知道"有评分,我们还有百度币、贴吧。

《21世纪》:贴吧的关键技术是什么?

刘建国:用户量大、发言多,如何把平台做到能够支持,网页检索定期更新,这都很难。 现在暂时不是完全实时的。我们专门有 TEAM,叫搜索应用组,实时性、数据量大,这两 个很难。

《21世纪》:贴吧可以无尽的扩展?

刘建国:我们现在有82万个吧,理论上可以无尽地扩展。一个吧里面的话题帖子有的话就会保留着。

《21世纪》:有人管贴吧吗?

刘建国:有,比如作广告的。BBS 只有几类,不能无尽的扩展,我们可无尽地扩,用户可以自己申请成为吧主。

《21世纪》:贴吧能产生哪些作用?

刘建国:网民以兴趣为单位组成社区,这样就很有粘性,可以倾诉、发表看法、言论,志趣相投的人讨论东西,这是比较独特的。

《21世纪》:贴吧之外,还有没有体现技术含量的东西?

刘建国:有许多,比如 MP3,要有用户习惯的积累;新闻频道,我们的新闻是靠技术分析,没有人工、零成本。哪套新闻出现得多,转载得多,它就是权威。而且是实时的。新闻是分类,国内、国际、财经没有分类,暂时没有更细的了。

《21世纪》:比如你们转载新浪新闻,需他们给授权吗?

刘建国:不需要。我们只是直接链接。我们也能带来广告的,如有编辑内容的,可以定制新闻。

阅读材料

搜索引擎使用诀窍

- Intitle,表示后接的词限制在网页标题范围内。
 - 例:要在标题是物流的网页中找论文

关键词 摘要 intitle:物流

市场 消费 需求 intitle:调查报告

- site 语法把搜索范围局限在这些网站内,以提高搜索效率。
 - 例:在 skycn.com 里面找有关网络快车的网页

网际快车 site:skycn.com

例:则表示询问搜索引擎收集 skycn.com 的所有的页面,

site:skycn.com

- "inurl"语法返回的网页链接中包含第一个关键字,后面的关键字则出现 在链接中或者网页文档中
 - 例: 查找微软网站上关于 windows 2000 的安全课题资料

inurl:security windows2000 site:microsoft.com

- "filetype:"这个语法来对搜索对象做限制,冒号后是文档格式,如 PDF、DOC、XLS 等
- 例: 查找网络营销的 PPT 文档,可以用:

网络营销 filetype:ppt

网站 SEO 对照表

网站 SEO 工具一览表