

目 录

关于这本书.....	1
第一章 搜索引擎基础.....	2
第一节 什么是搜索引擎.....	2
一、搜索引擎的作用.....	2
二、搜索引擎的定义.....	4
第二节 搜索引擎的基本工作原理.....	4
一、抓取.....	5
二、索引.....	5
三、排序.....	5
第三节 搜索引擎的分类.....	6
一、网页级搜索.....	6
二、垂直搜索.....	6
三、元搜索引擎.....	6
四、目录搜索.....	6
五、集成搜索.....	7
第四节 搜索引擎的未来.....	7
一、快速化.....	7
二、多样化.....	7
三、智能化.....	7
四、社会化.....	7
五、个性化.....	8
第五节 主要搜索引擎介绍.....	8
一、Google.....	8
二、Yahoo.....	12
三、百度.....	13
四、搜狗.....	15
五、中文搜索引擎列表.....	16
第二章 搜索引擎营销基础.....	17
第一节 什么是搜索引擎营销.....	17
一、搜索引擎营销的定义.....	17
二、搜索引擎营销的价值.....	17
三、搜索引擎营销原理.....	18
第二节 搜索引擎营销的特点.....	19
一、广泛使用.....	19
二、用户主动查询，针对性强.....	19
三、获取新客户.....	19
四、竞争性强.....	20

五、动态更新，随时调整.....	20
六、门槛低，投资回报率高.....	20
第三节 搜索引擎营销的目标.....	21
一、被收录.....	21
二、排名靠前.....	21
三、被点击.....	21
四、客户转化.....	21
五、提高品牌知名度.....	21
第四节 搜索引擎营销的形式.....	22
一、搜索引擎登录.....	22
二、固定排名和广告.....	23
三、按效果付费排名.....	24
四、自然排名.....	24
第三章 搜索引擎优化基础.....	25
第一节 为什么要搜索引擎优化.....	25
一、搜索引擎优化的定义.....	25
二、搜索引擎优化的价值.....	25
三、SEO和SEM的关系.....	26
四、SEO和付费排名的关系.....	26
第二节 搜索引擎对SEO的态度.....	27
一、Google.....	27
二、Yahoo.....	29
三、百度.....	30
四、搜狗.....	34
第三节 白帽SEO和黑帽SEO.....	37
第四节 搜索引擎优化的发展.....	39
一、SEO的发展历史.....	39
二、SEO在中国.....	40
三、SEO的发展方向.....	42
第五节 关于搜索引擎优化的一些观点.....	43
一、搜索引擎优化合法吗？.....	43
二、会不会和搜索引擎对着干？.....	43
三、搜索引擎是否会人工干预？.....	44
四、我的网站有大量的优质原创内容，所以我不需要SEO了。.....	44
五、SEO没有技术含量或SEO需要很高深的技术.....	44
六、大网站那样做了，我也向他们学习.....	44
七、SEO企业或SEO工具的保证排名宣传.....	44
八、稳定的排名.....	45
九、SEO的目标就是获得排名第一的位置或尽可能多的访问量.....	45
十、竞争对手无法破坏我的排名.....	45
第四章 SEO相关的搜索引擎技术.....	46

第一节 搜索引擎机器人	46
一、什么是搜索引擎机器人?	46
二、如何辨别搜索引擎机器人身份	47
第二节 超链分析	47
一、基本原理	48
二、工作步骤	48
三、搜索效果	49
四、超链分析技术的应用	49
五、存在的缺陷	50
第三节 中文分词	50
一、什么是中文分词?	50
二、中文分词的意义和作用	50
三、中文分词技术	51
四、分词中的难题	53
五、中文分词的应用	53
六、中文搜索引擎技术	54
第四节 基于词意的文本分析	55
一、文本分析技术特点	55
二、文本分析应用	55
第五章 了解网站的搜索引擎表现	57
第一节 基础查询	57
一、域名信息查询	57
二、域名历史查询	57
三、主机连接速度和IP地址查询	57
四、同IP网站查询	57
第二节 收录查询	58
一、是否被收录	58
二、网页收录数量	58
第三节 反向链接查询	59
第四节 网页RANK查询	59
一、Google PageRank查询	59
二、Sogou Rank查询	59
第五节 关键词排名查询	59
第六节 ALEXA查询	60
一、什么是Alexa?	60
二、Alexa主要数据	60
三、用Alexa查什么?	60
第六章 基于搜索引擎友好的网站设计	62
第一节 域名策略	62
一、不同后缀的域名在搜索引擎中的权重	62
二、域名基本常识	62

三、域名的关键词策略.....	69
四、中文站点的域名的拼音策略.....	70
五、谨防注册一个遭受搜索引擎惩罚过的域名.....	70
第二节 空间策略.....	71
一、安全元素一：稳定性.....	71
二、安全元素二：良好的共处环境.....	71
三、本节附录.....	72
第三节 网站结构规划.....	72
第四节 站点导航设计.....	73
第五节 IFRAME、JAVASCRIPT和AJAX.....	73
一、IFRAME.....	74
二、JS.....	74
三、AJAX.....	74
第六节 URL设计.....	75
一、便于用户记忆.....	75
二、URL静态化.....	75
三、URL的关键词部署.....	78
第七节 网页代码的编写.....	80
一、遵循WEB标准.....	80
二、代码的逻辑性.....	80
第八节 面向搜索引擎的文案写作指导.....	81
一、网页标题Title.....	81
二、Meta元标签.....	84
三、图片ALT替换文本.....	85
四、Robots.txt.....	85
五、链接锚文本.....	89
六、Sitemap.xml.....	89
七、版权与隐私.....	97
第九节 用户友好性检测.....	97
一、链接的可用性.....	97
二、访问速度体验.....	97
三、查找信息的便捷度.....	98
第七章 创建搜索引擎喜欢的内容.....	99
第一节 搜索引擎喜欢什么内容.....	99
一、Google对网站内容的说明：.....	99
二、百度对网站内容的说明：.....	99
第二节 内容的四大来源.....	99
一、自己撰写原创文章.....	99
二、使用他人的内容.....	100
第三节 常见问题.....	102
一、网站内容的更新.....	102
二、内容匮乏怎么办.....	103

三、采集系统.....	104
四、商城程序的内容问题.....	105
第八章 提交你的网站.....	106
第一节 向搜索引擎提交网站.....	106
一、常用搜索引擎登录入口.....	106
二、提交注意事项.....	106
三、登录搜索引擎的最佳办法.....	106
第二节 登录分类目录站.....	107
第三节 登录网址导航站.....	108
第四节 登录垂直行业门户.....	109
第九章 链接策略（一）.....	111
第一节 链接的基本知识.....	111
第二节 为什么链接很重要.....	111
一、超链分析技术.....	111
二、网页级别.....	111
三、网站的更新频率.....	112
四、内容的相关性.....	113
第三节 网站内部页面的链接策略.....	113
一、给重要的网页更多的关注.....	113
二、相关链接（人工和自动）和频道关联.....	114
三、链接的导出.....	114
第十章 链接策略（二）.....	115
第一节 登录开放目录DMOZ.....	115
一、为什么登录DMOZ如此重要.....	115
二、DMOZ开放目录登录实战攻略.....	115
三、一些对你有帮助的建议.....	116
第二节 交换链接.....	116
第三节 购买链接.....	117
第四节 链接诱饵.....	118
一、如何制造链接诱饵.....	118
二、对链接诱饵应如何看待.....	118
第五节 外链建设中的几个误区.....	119
一、外部链接数目增长的自然化.....	119
二、不要忘记你最能做好的事情.....	119
三、不宜使用软件群发.....	120
第十一章 数据监测和分析.....	121
第一节 网站流量数据统计与分析.....	121
一、什么叫网站流量及网站流量统计分析.....	121
二、网站流量指标.....	121

三、用户行为指标.....	123
四、用户浏览网站的方式.....	126
第二节 中文常见流量统计系统介绍.....	126
一、51.la统计服务.....	126
二、网站管理员统计.....	126
三、ItSun统计.....	126
四、51Yes网站流量统计.....	126
五、武林榜、太极链.....	126
六、一统天下统计服务.....	127
七、趋势流量分析系统.....	127
八、Google Analytics.....	127
九、Measure Map.....	127
十、Statcounter.com.....	127
第三节 流量统计分析实例.....	128
一、典型性数据采集抓取.....	128
二、IP与PV.....	128
三、网页访问入口分析.....	130
四、搜索引擎流量导入.....	131
五、搜索引擎关键词分析.....	132
第四节 行业热点跟踪.....	133
一、Yahoo风向标:.....	133
二、百度风云榜:.....	133
三、百度指数.....	134
第十二章 高端技巧.....	137
第一节 搜索行为分析.....	137
一、什么是搜索行为分析.....	137
二、为什么要进行用户搜索行为分析.....	137
三、用户搜索行为的几个特点.....	137
四、搜索行为分析的重点对象.....	138
五、搜索行为变化情况.....	138
六、用户的关键词搜索行为.....	140
第二节 关键词分析.....	143
一、什么是关键词.....	143
二、如何挑选正确的关键词.....	143
三、竞价关键词的选择.....	147
四、关键词竞争程度分析.....	148
五、关键词查询工具.....	153
第三节 关键词密度.....	153
一、什么是关键词密度.....	153
二、正确理解关键词密度.....	153
三、适当的关键词密度.....	153
四、关键词位置和形式.....	154

五、Google 和Yahoo的区别	156
六、关键词密度的基本原则	156
七、如何增加关键词密度	157
八、关键词密度工具	158
第四节 信息干扰现象	158
一、信噪比与干扰信息	158
二、信息干扰的副作用	158
三、如何减小信息干扰	158
第五节 重定向	159
一、什么是重定向	159
二、301 重定向	159
三、302 重定向	161
四、404 重定向	162
五、其他重定向	164
第六节 垂直搜索引擎优化概览（图片搜索、移动搜索等）	165
一、图片搜索	165
二、Flash	167
三、视频搜索	167
四、避免搜索中的搜索	168
第十三章 几个重要的问题	169
第一节 搜索引擎排名作弊	169
一、什么是搜索引擎作弊？	169
二、搜索引擎排名作弊的常见形式	169
三、常见的作弊载体	171
第二节 搜索引擎的反作弊	172
一、为什么搜索引擎要反作弊	172
二、搜索引擎对待SEO与作弊的态度	172
三、主要搜索引擎反作弊的一些规定	173
四、搜索引擎惩罚站点事件举例	175
第三节 搜索引擎惩罚和解除	176
一、搜索引擎惩罚的方式	176
二、站点被惩罚后的影响	176
三、如何投诉作弊	177
四、你的站点是否被惩罚？	177
五、搜索引擎作弊惩罚解除流程	178
第四节 搜索引擎反作弊任重道远	178
一、搜索垃圾永不会消失	178
二、搜索引擎反作弊的历程	179
三、石头记录的 2005 年SEO作弊状况	180
四、曾经的作弊有多疯狂	182
五、SEO的职业道德	186
六、搜索引擎反作弊的人性化	189

七、搜索引擎反作弊的法律支持.....	190
第十四章 附录.....	191
第一节 搜索引擎发展史.....	191
一、搜索引擎编年史：全球.....	191
二、搜索引擎编年史：中国.....	197
第二节 搜索引擎行业八卦.....	200
一、中国SEO市场概况.....	200
二、杨致远为Yahoo做优化.....	201
三、SEO大赛.....	201
第三节 SEO工具介绍.....	202
一、SEO工具软件种类.....	202
二、Google Sitemaps.....	206
三、通过sitemaps分析死链.....	206
第四节 GOOGLE的几个著名现象.....	207
一、Google Trust Rank.....	207
二、Google Sandbox.....	208
三、Google Dance.....	210
四、Sitelinks.....	211
五、补充材料.....	213

关于这本书

2007年初，浩维互动的一次内部讨论中，产生了一个想法：写一本关于搜索引擎优化的书。这个想法很快被转化成实际的行动，柳焕斌、陈小华和我开始了白天工作晚上码字的生活。

写作过程中，因为个人的原因，柳焕斌、陈小华先后离开了公司，开始了他们自己的事业。虽然留下了一部分书稿，但看着还有些残缺和混乱的文档，我有些茫然。

在很多朋友的支持下，我们还是决定把这本书写完。于是，胡利和方琳加入了写作的队伍，经过一个多月的后续工作，书稿完成了。

已经不止一家出版社和我们联系这本书的出版事宜，但最终我们还是放弃了出版的机会，决定以免费的电子书的形式推出。

中文搜索引擎优化发展过程中，胡宝介女士免费的电子书和王通先生收费的电子书都产生了重要的影响，希望我们的这本电子书也能发挥些作用。

书名是受了《哈利·波特》的影响，但不是想表现搜索引擎优化的神秘，而是想说明，通过实施搜索引擎优化，能化腐朽为神奇，为网站带来更强的生命力。

感谢所有写这本书的人，大家都很辛苦，却没有稿费了；也感谢所有看这本书的人，时间仓促，有很多疏漏，大家看在免费的份上，多多包涵吧。

有任何问题和建议，请联系我们：nanjing@timev-inc.com

浩维互动 姚志国

2007年8月28日

第一章 搜索引擎基础

第一节 什么是搜索引擎

一、搜索引擎的作用

如果非要列举出搜索引擎能做什么，恐怕没有人可以给出完整的答案。那么搜索引擎到底在我们的生活中扮演着怎样的角色呢？以下是一些普通网民使用搜索引擎的过程，从中我们不难看出搜索引擎正在发挥着怎样的做用。

小张喜欢和朋友聊天，却经常被朋友口中冒出的一些新名词搞得头晕，又不太好意思经常问。昨天几个朋友闲聊，有人提到“导弹打卫星”，却是一带而过，小张有些犯迷糊了。晚上一回到家中，小张就打开电脑，上百度，搜索一下“导弹打卫星”，来龙去脉就一清二楚了。

重装了电脑，常用的 MP3 播放软件没有了。搜索一下“MP3 播放软件”，找到了一个叫“千千静听”的，好像和以前用的不一样啊。再搜索一下“千千静听”，发现大家对这个软件的评价挺高的，而且知道了这个 MP3 播放软件还能自动下载歌词、显示歌词。搜索“千千静听官方网站”，去下载最新版本的软件了。

周五照例要和朋友小聚一下，可是每周聚会，吃来吃去就是那几个餐馆，都腻味了。打开 Google，试试看本地搜索。搜索“南京川菜”，一下子出来几百个结果，好多都是没有听说过的餐馆。本地搜索还有地图，把查找范围缩小到家附近，这样吃完饭回家就方便了。

小李是个旅游爱好者，每年五一长假都要和朋友们一起出游。可是随着黄金周旅游越来越火爆，一般景点五一期间简直是人头碰人头，根本不能去。春节一过，小李就开始琢磨今年的五一计划了，一定要着一个风景不错，又不是那么商业化的地方。往边远地区找吧。搜索“贵州户外旅游”，找到当地一家不错的户外俱乐部网站，上他们的论坛泡泡，再咨询咨询，很快，小李就确定了一条位于黔东南的背包游线路。在制定具体计划的过程中，一些交通衔接的问题有点麻烦。小李又在百度知道和新浪爱问提了一些问题，很快就得到了满意的回答。

老吴是个摄影爱好者，最近一直想买一个新的数码单反相机。那天咖啡馆取阅的杂志上看到一款佳能新型号相机的广告，漂亮的外观和看上去不错的功能参数让老吴有点心动了。老吴用记下了这款产品的型号，回家直接用百度搜索一下，再阅读了几百篇关于这款相机的使用心得和测评文章后，老吴终于决定出手了。不过老吴可没有直奔商店，而是通过比价搜索引擎找到了最便宜的网上商店购买，比商场卖了便宜了将近 1000 元。

小王今年高考，填报志愿的事情让父母犯了愁。小王平时成绩算是中等，而且他一定要去北京读大学。虽然清华北大大家都很熟悉，却不是小王的成绩能达到的，而其他学校大家就都不熟悉了。父母想知道北京其它高校的教学水平、学习风气以及生活条件，说要专门去趟北京考察考察。小王劝住了父母，打开电脑，搜索“北

京大学”，一下子北京各大高校的网站出现在面前。小王和父母把这些网站的内容仔仔细细的看了一遍，还分析了一些感兴趣的学校往年的考分情况，基本上心中有数了。

小赵所在的公司原来提供某个配件的供应商出了一些问题，不能继续提供配件了，老板急得团团转。小赵通过搜索引擎找到了十多家具有生产这种配件能力厂家，很快就落实了新的供应商。

搜索引擎已经成为广大网民日常生活中不可缺少的重要部分。中国互联网信息中心（CNNIC）2007年第二十次调查报告显示：搜索引擎以74.8%的比率已经成为中国网民最常使用的网络服务之一。详细数据见下表：

表 1-1-1 网民使用网络的功能及其所占比例

互联网功能	使用者占网民的比例
浏览新闻	77.3%
获取信息	76.3%
搜索引擎	74.8%
即时通信	69.8%
网络音乐	68.5%
网络影视	61.1%
收发邮件	55.4%
网络游戏	47.0%
网络购物	25.5%
网络教育	24.0%
网上银行	20.9%
博客	19.1%
网络求职	15.2%
网上炒股	14.1%
网络销售	4.3%
网上旅行预订	3.9%

二、搜索引擎的定义

给搜索引擎下一个定义是非常困难的。一千个人就有一千个搜索引擎：有人认为搜索引擎就是百度，有人认为搜索引擎就是 Google，有人认为搜索引擎就是 Yahoo。

如果非要给个定义，就让我们来看看微软大百科全书（MSN Encarta）是怎么定义搜索引擎的吧：

Search Engine, software program that helps users find information stored on a personal computer, or a network of computers, such as the Internet. A user enters search terms, typically by typing a keyword or phrase, and the search engine retrieves a list of World Wide Web (WWW) sites, personal computer files, or documents, either by scanning the content stored on the computers or computer networks being searched or by parsing (analyzing) an index of their stored data.

搜索引擎是帮助用户查找存储在个人电脑、计算机网络如互联网上的信息的软件程序。用户输入搜索项目，通常是通过录入一个关键词或短语，搜索引擎通过扫描被搜索的计算机和计算机网络，或者分解（分析）它们数据的索引，返回万维网站点、个人电脑文件或文档的列表。

这个定义可能越看越模糊，那么以下章节的内容将会让你清晰地认识搜索引擎。

第二节 搜索引擎的基本工作原理

现在互联网上存在的网页数量是数以百亿千亿计的，这些网站存储在不同的服务器上，分布在世界各地的数据中心和机房。

对于搜索引擎来说，要抓取互联网上所有的网页几乎是不可能的，从目前公布的数据来看，容量最大的搜索引擎也不过是抓取了整个网页数量的百分之四十左右。一方面原因是抓取技术存在瓶颈，无法遍历所有网页，有许多网页无法从其它网页的链接中找到；另一个原因是存储技术和处理技术的问题，如果按照每个页面平均大小 20K 计算（包含图片），100 亿网页的容量是 $100 \times 2000G$ 字节，即使能够存储，下载也存在问题（按照一台机器每秒下载 20K 计算，需要 340 台机器不停的下载一年时间，才能把所有网页下载完毕）。同时，由于数据量太大，在提供搜索时也会有效率方面的影响。因此，许多搜索引擎的网络蜘蛛只是抓取那些重要的网页，而评价重要性的主要依据是某个网页的链接深度。

有人会认为搜索引擎在接收到搜索请求时，会实时地从全球所有的服务器上查询信息，并把查询结果展示在用户面前，这其实是一种误解。如果搜索引擎是这样工作的，那么查询一条信息可能要等上好几年才能得到搜索结果，这还不包括期间网页发生的变化。

实际上，搜索引擎会预先去拜访大量的网站，并把这些网页的部分信息预先存储在自己的服务器上，这样，当用户搜索的时候，其实是在搜索引擎自己的服务器中进行查询，就像我们在自己的电脑中查询文件一样。

搜索引擎是非常复杂的技术，但是其基本原理并不复杂，其基本技术包括抓取、索引、排序。

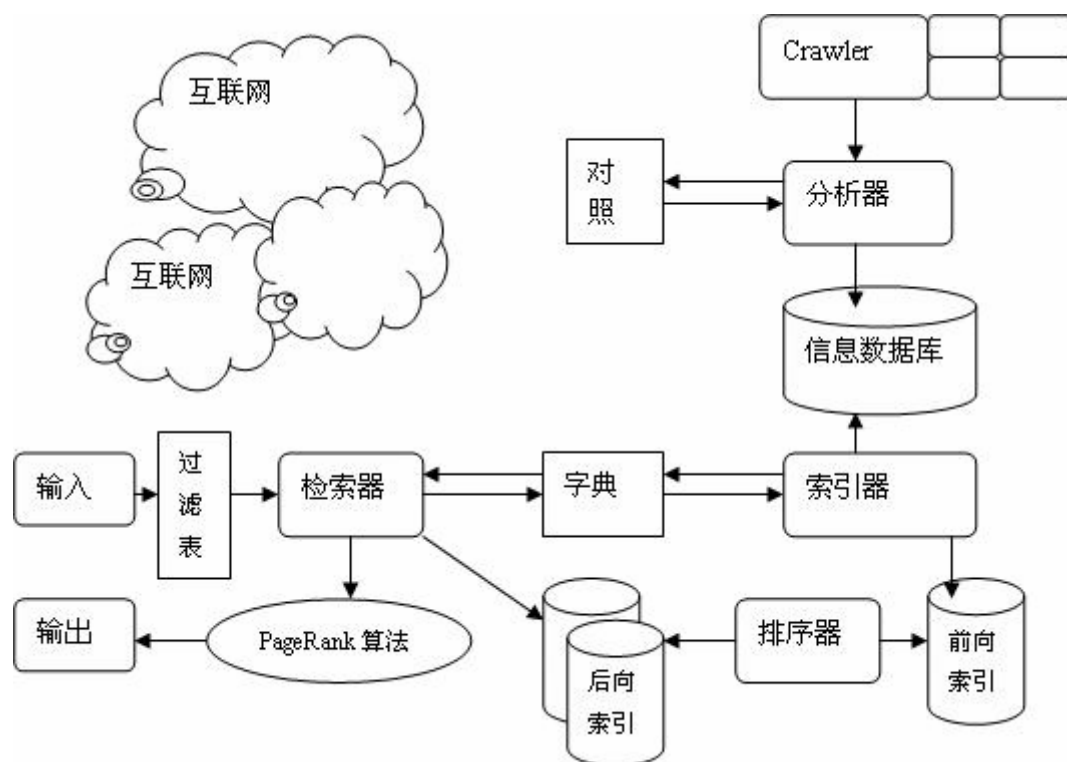


图 1-2-1 搜索引擎工作原理

一、抓取

搜索引擎首先会派出一种被称作“蜘蛛”或者是“机器人”的软件，根据一定规则扫描存在于互联网上的网站，并沿着网页上的链接从一个网页到另一个网页，从一个网站到另一个网站。为保证采集的资料最新，它还会回访已抓取过的网页。

二、索引

由分析索引系统程序对收集回来的网页进行分析，提取相关网页信息（包括网页所在 URL、编码类型、页面内容包含的所有关键词、关键词位置、生成时间、大小、与其它网页的链接关系等），根据一定的相关度算法进行大量复杂计算，得到每一个网页针对页面文字中及超链中每一个关键词的相关度（或重要性），然后用这些相关信息建立网页索引数据库。

三、排序

当用户输入关键词搜索后，由搜索系统程序从网页索引数据库中找到符合该关键词的所有相关网页。因为所有相关网页针对该关键词的相关度早已算好，所以只需按照现成的相关度数值排序，相关度越高，排名越靠前。

最后，由检索器将搜索结果的链接地址和页面内容摘要等内容组织起来，返回给用户。

第三节 搜索引擎的分类

按照不同的技术特点，可以把搜索引擎分为网页级搜索、垂直搜索、元搜索引擎、目录搜索和集成搜索等五类。

一、网页级搜索

网页级搜索引擎是名副其实的搜索引擎，国外具代表性的有 Google、Yahoo、MSN Live Search、Ask，国内著名的有百度、搜狗等。它们都是通过从互联网上提取的各个网站的信息（以网页文字为主）而建立的数据库，检索与用户查询条件匹配的相关记录，然后按一定的排列顺序将结果返回给用户。网页级搜索引擎也是目前常规意义上的搜索引擎。

二、垂直搜索

垂直搜索是针对某一个行业的专业搜索引擎，是搜索引擎的细分和延伸。它将网页库中某类专门的信息进行一次整合，定向分字段抽取出需要的数据进行处理，然后再以某种形式返回给用户。

垂直搜索引擎与普通网页搜索引擎的最大区别是对网页信息进行了结构化抽取，也就是将网页的非结构化数据抽取成特定的结构化信息数据，然后将这些数据存储到数据库，进行进一步的加工处理，如：去重、分类等，最后分词、索引，再以搜索的方式满足用户的需求。可以说，网页搜索是以网页为最小单位，基于视觉的网页块分析是以网页块为最小单位，而垂直搜索是以结构化数据为最小单位。

各大搜索引擎都有自身的垂直搜索功能，比如 MP3 搜索、图片搜索、新闻搜索、Blog 搜索等。独立的垂直搜索网站也受到有相关兴趣的用户的青睐，比如旅游搜索引擎、商业搜索引擎、比价搜索引擎等。

三、元搜索引擎

元搜索引擎在接受用户查询请求时，同时在其它多个引擎上进行搜索，并将结果返回给用户。元搜索引擎并不直接抓取网页，而是抓取多个搜索引擎的索引数据库，并根据自己的算法对抓取结果重新筛选排序。

虽然说元搜索引擎综合了多个搜索引擎的结果，但由于是建立在其他搜索引擎基础之上，自身的排序算法也难以得到认可，所以一直没有特别的影响力。

比较知名的元搜索引擎有 Dogpile (www.dogpile.com)、Vivisimo (www.vivisimo.com)、比比猫 (www.bbmao.com) 等。

四、目录搜索

目录索引虽然有搜索功能，但在严格意义上算不上是真正的搜索引擎，仅仅是按目录分类的网站链接列表，并提供站内搜索而已。用户完全可以不用进行关键词查询，仅靠分类目录也可找到需要的信息。

Yahoo 和搜狐都提供分类目录功能，其它知名的目录还有 DMOZ (www.dmoz.org)、

Looksmart (www.looksmart.com) 等。

五、集成搜索

集成搜索算不上是搜索引擎，只是运用了一定的网页技术，把多个搜索引擎的搜索功能整合在一个搜索页面，并对各个搜索引擎的功能进行分类，方便用户在多个搜索引擎的搜索结果中进行比较和选择。部分集成搜索甚至没有对用户的搜索请求做出任何记录。所以，集成搜索更像一个在线的、集合了多个搜索引擎入口的软件，而不是搜索引擎。

国内有很多集成搜索的网站，如优客搜索 (www.yok.com)。

第四节 搜索引擎的未来

随着计算机和互联网应用的普及，大众对搜索的要求也越来越高，人们期望搜索引擎的速度更快、返回结果更准确、可以搜索的内容更丰富、更加智能化和个性化。

一、快速化

速度的提高是现在各个搜索引擎都在努力达到的目标。快不仅仅是指搜索引擎返回结果的速度，更是指搜索引擎获取新信息的速度。现在博客和论坛已经成为信息最重要的来源，一条爆炸性的新闻很可能来自某个人博客，而不是知名媒体。

二、多样化

多样化主要包括两个方面：内容多样化和搜索途径多样化。

内容多样化 除文字之外还包括图片、音频、视频内容的搜索。

搜索途径多样化 可以通过语音指令、动作指令、甚至是思维指令进行搜索。

三、智能化

语义输入 可以通过输入自然语言来查找结果，就像在问朋友问题。比方说，“第一个登上太空的中国宇航员是谁？”这是很多搜索引擎研究的防线。虽然如 ask.com 一直宣称自己是提问式搜索，但实际上并不能完全做到和用户的自然对话。

互动功能 逐层缩小搜索范围，使用户更快更准地找到所需信息。目前有些搜索引擎已实现了部分互动性。

相关搜索 在用户搜索过程中结合大量其他用户的搜索习惯统计，给出可能用户要查找的内容，引导和启发搜索者找到想要找的内容。

四、社会化

有人把社会化搜索称作“搜索人脑”。搜索引擎借助对大量用户的粘性，提供用户提问的功能，而这一提问的回答者不是搜索引擎本身，而是其他所有看到这一问题的用户。通过大量用户的群策群力，以及对应的专业人士的帮助，这些问题往往能得到比较好的解答。比

如新浪的爱问、百度知道等都属于社会化搜索的范畴。

目前，社会化搜索的不足是还不能对回答内容的准确性进行评价，只能依靠提问者自己的判断。另外问题不能及时得到解答也会影响用户的耐心。

五、个性化

因为每个用户的喜好、目的不同，对搜索结果的期待就有很大差异。比如同样是搜索“绿茶”，有人希望查找“绿茶”的商品信息，有人则是要找一部叫《绿茶》的电影。

所以现在很多搜索引擎开始建议用户注册登录，比如 Google 和 Yahoo，这样用户的搜索习惯就会被搜索引擎记录和分析。如果你经常查询电影的相关资料，那你下次输入“绿茶”的时候，返回结果中《绿茶》电影的信息就会出现在靠前的位置了。

个性化搜索对用户搜索习惯的数据收集有很高的要求，不过在某种程度上，又涉及到用户的隐私问题。

第五节 主要搜索引擎介绍

一、Google

1、关于 Google

“Google”来源于“Googol”，这是一个数学名词，表示一个“1”后面跟着100个“0”。这个词是由美国数学家 Edward Kasner 的外甥 Milton Sirotta 创造的，随后通过 Kasner 和 James Newman 合著的“Mathematics and the Imagination”（《数学与想象力》）一书广为流传。Google 使用这一术语体现了公司整合网上海量信息的远大目标。

Google 的使命是整合全球信息，使人人皆可访问并从中受益。完成该使命的第一步始于 Google 创始人 Larry Page 和 Sergey Brin，他们在斯坦福大学的学生宿舍内共同开发了全新的在线搜索引擎，然后迅速传播给全球的信息搜索者。目前 Google 被公认为全球规模最大的搜索引擎，它提供了简单易用的免费服务，用户可以在瞬间得到相关的搜索结果。

访问 www.google.com 或众多 Google 域之一时，用户可以使用多种语言查找信息，查看股价、地图和要闻，查找美国境内所有城市的电话簿名单，搜索数十亿计的图片并详读全球最大的 Usenet 信息存档——超过十亿条帖子，发布日期可以追溯到1981年。

用户甚至不必特意访问 Google 主页，也可以访问所有这些信息。使用 Google 工具栏，用户可以从网上的任何位置执行 Google 搜索，而 Google 桌面栏（测试版）将 Google 搜索框放在 Windows 任务栏中，这样用户可以从任何正在使用的应用程序中执行搜索，而不必打开浏览器。即使身边没有 PC 机时，用户也可以通过 WAP 和 i-mode 手机等无线平台使用 Google。Google 的实用性及便利性赢得了众多用户的青睐，它几乎完全是在用户的交口称颂下成为全球最知名的品牌之一的。

作为一个企业，Google 通过提供广告服务来获取收入，使广告客户能够刊登与特定网页内容相关、重要而又经济实效的在线广告。这不仅为用户提供了实用的广告信息，同时也给刊登广告的客户带来了好处。Google 认为，搜索用户应该知晓在他们面前展示的讯息是自然排名还是付费信息，因此始终在搜索结果或网页上将广告与其他内容区分开来。Google 不出售搜索结果中的排名，也不允许有人通过付费来提高在搜索结果中的排名。

成千上万的广告客户加入了 Google AdWords 计划，利用定位广告在网上推广他们的产品和服务，可以相信 AdWords 是同类计划中规模最大的一个。此外，成千上万的网站管理人员利用 Google AdSense 计划刊登与其网站内容相关的广告，增加收入并改善用户体验。

Google 是全球最大的搜索引擎，借助和 America Online、Netscape 及其他公司的合作伙伴关系，它所回应的查询远远多于其他在线服务商。

Google.com 是互联网上 5 大最受欢迎的网站之一，在全球范围内拥有无数的用户。

Google 是阿根廷、澳大利亚、比利时、巴西、加拿大、丹麦、法国、德国、印度、意大利、墨西哥、西班牙、瑞典、瑞士、英国和美国的头号搜索引擎（根据 Nielsen//NetRatings 05 年 6 月调查报告唯一身份访问者总数结果）。

每月全球唯一身份用户：3 亿 8000 万（Nielsen/NetRatings，05 年 8 月调查报告）。

全球受众：超过百分之五十的点击量来自美国境外。

Google 在硅谷和全球各地拥有最顶尖的技术人员和业务专家。截止到 2005 年 12 月 31 日，共有 5,680 名全职员工，其中许多是技术人员或工程师。

2、Google 技术

Google 秉持着开发“完美的搜索引擎”的信念，在业界独树一帜。所谓完美的搜索引擎，就如公司创始人之一 Larry Page 所定义的那样，可以“确解用户之意，切返用户之需”。为了实现这一目标，Google 坚持不懈地追求创新，而不受现有模型的限制。为此，Google 开发了自己的服务基础结构和具有突破性的 PageRank™ 技术，使得搜索方式发生了根本性变化。

Google 的开发人员从一开始就意识到，要以最快的速度提供最精确的搜索结果，就需要一种全新的服务器设置。大多数的搜索引擎依靠少量大型服务器，这样，在访问高峰期速度就会减慢，而 Google 却利用相互链接的 PC 来快速查找每个搜索的答案。这一创新技术成功地缩短了响应时间，提高了可扩展性，并降低了成本。这也是其他公司一直在效仿的技术。与此同时，Google 从未停止过对其后端技术的改进，以使其技术效率更高。

Google 搜索技术所依托的软件可以同时一系列运算，且只需片刻即可完成所有运算，而传统的搜索引擎在很大程度上取决于文字在网页上出现的频率。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 5 亿个变量和 20 亿个词汇组成的方程进行计算，PageRank 能够对网页的重要性作出客观的评价。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，某些网页的投票被认为具有较高的价值，它所链接的网页就能获得较高的价值。重要网页获得的 PageRank 值较高，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个网页的重要性，搜索结果没有人工干预或操纵，不受付费排名影响。这也是 Google 成为一个公正客观、并且广受用户信赖的信息来源的重要原因。

超文本匹配分析 Google 的搜索引擎同时也分析网页内容。然而，Google 的技术并不采用单纯扫描基于网页的文本这一方式（网站发布商可以通过元标记控制这类文本），而是分析网页的全部内容以及字体、分区及每个文字精确位置等因素。Google 同时还会分析相

邻网页的内容，以确保返回与用户查询最相关的结果。

Google 的创新并不限于台式机 为了确保通过便携式设备访问网络的用户能够快速获得精确的搜索结果，Google 还率先推出了业界第一款无线搜索技术，将 HTML 即时转换为针对 WAP、i-mode、J-SKY 和 EZWeb 优化的格式。

3、Google 的搜索过程

Google 查询的全过程通常不超过半秒时间，但在这短短的时间内需要完成多个步骤，然后才能将搜索结果交付给搜索信息的用户。

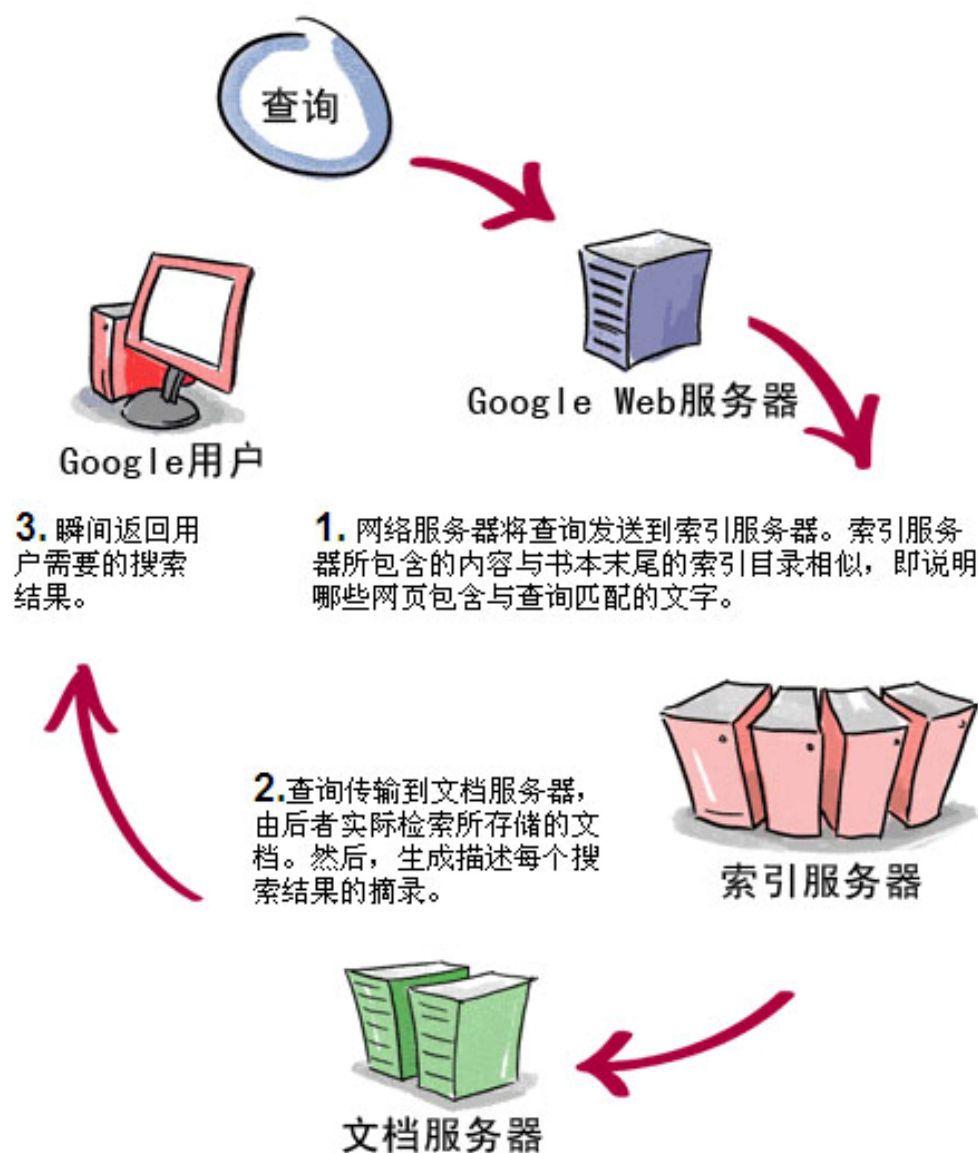


图 1-5-1 Google 的搜索过程

4、Google AdWords

Google AdWords 是 Google 提供的搜索引擎关键词广告，采用按点击付费方式的网络广告模式，只有当用户点击广告时广告主才支付广告费用。

Google AdWords 的表现形式是这样的：当用户用某个关键词在 Google 进行检索时（例如“鲜花”），在搜索结果页面右侧会出现与鲜花这一关键词相关的广告（如果有公司购买这一关键词广告的话）。广告的显示顺序取决于广告主愿意为每次点击支付的价格，比如广告主 A 愿意为每次点击支付 1 元，而广告主 B 愿意支付 1.2 元，那么广告主 B 就会排在广告主 A 的前面。

当然，出价并不是影响 Google AdWords 排名的唯一因素，广告点击率等也会对广告排名产生影响，但是总体还是以单次点击的出价为主。

Google 关键词广告出现在搜索结果的右侧，并明显地标注有“赞助商链接”字样，在每页搜索结果中最多显示 8 条 Google AdWords 关键词广告信息。

Google 严格区分自然排名和广告，一般用户能一目了然地分辨，这也是 Google 能吸引如此众多用户的原因之一。

Google AdWords 是 Google 最主要的收入来源，虽然每次点击带来的收入可能只有几个美分，每个客户带来的价值可能只是几百美金，但是不容忽视的是，Google AdWords 的全球客户数量是相当可观的。

5、Google AdSense

Google AdSense 是一种获取收入的快速简便的方法，适合于各种规模的网站发布商。它可以在网站的内容网页上展示相关性较高的 Google 广告，并且这些广告不会过分夸张醒目。由于所展示的广告同用户在你的网站上查找的内容相关，因此，最终你的内容网页不仅会为你带来经济效益，还能够得以充实。

另外，网站发布商还可以利用 AdSense 向自己的网站用户提供 Google 搜索功能，并通过在搜索结果页展示 Google 广告获得收入。

这一免费计划将按点击次数付费的广告和按展示次数付费的广告相结合——也就是说，网站上或搜索结果页中广告的有效点击次数，以及内容网页上的展示次数都会为你带来收入。你只需填写在线申请，并在 AdSense for content 和 AdSense for search 中选择一种，或两种都选。申请一个产品即可获得 AdSense 和 AdSense for search 这两者的批准，可在自己的网页上使用这些产品的任意组合。

6、谷歌：Google 中国

Google 全球中文名称为“谷歌”，2006 年 4 月 12 日由 Google 公司 CEO Eric Schmidt 在中国北京发布。根据新闻发布，“谷歌”的意思解作“山谷之歌”（亦解作“丰盛之歌”），另外，该公司也希望能够获得五谷丰登般的成就。

易观国际对中国搜索引擎市场 2007 年第 2 季度的最近监测研究表明，不计渠道收入，2007 年第 2 季度，搜索引擎厂商收入为 6.575 亿元，其中谷歌的市场份额从上季度的 18.7% 上升到 22.8%。总结 Google 2006 年下半年至今的举措可以看到 Google 本土化的努力。

技术产品改进 Google 陆续推出包括图书搜索、谷歌地图、热榜、谷歌金融、导航等产品，推出搜索提示、翻译功能等中文网页。这些都属于 Google “整合搜索”战略的一部分（谷歌的“整合搜索”战略是力图利用自己的技术，通过智能分析和导航系统，把信息、新闻、地图等不同搜索领域结合起来，呈现给用户）。

渠道布局 谷歌在中国市场，不再坚持全国总代模式，开始发展区域代理，经过一年多的时间，谷歌在全国近 20 个城市发展了 20 多家授权经销商，销售能力提高。

多方面的本地合作 谷歌正通过资本并购、合作捆绑、结盟门户网站等形式来提升在华

综合实力，包括与中国移动、中国电信、腾讯、新浪、Flashget、暴风影音合作，以及对迅雷、天涯等战略投资。通过合作，谷歌吸纳了异业资源，为开展新业务打基础。

通过上述本地化措施，谷歌在中国搜索引擎市场的份额逐渐提高，根据易观国际的研究，谷歌的市场份额从 06 年 1 季度的 13.2% 提高到目前的 22.8%，增加 9.6 个百分点。

二、Yahoo

1、关于 Yahoo

90 年代初，搜索引擎的应用起源于少数高校和科研机构中对研究论文的查找。1994 年 4 月，美国斯坦福大学电机工程系的博士生大卫·费罗 (David Filo) 和杨致远 (Jerry Yang) 建立了自己的网络指南信息库，将其作为记录他们个人对互联网的兴趣的一种方式。这就是 Yahoo 的前身。

但是不久，他们就发现他们自己编写的列表变得很长，不便于处理。于是他们开始在 Yahoo 身上花费越来越多的时间。

随后，他们将 Yahoo 变成了一个可定制的数据库，旨在满足成千上万的、刚刚开始通过互联网社区使用网络服务的用户的需要。他们开发了可定制的软件，帮助他们有效查找、识别和编辑互联网上存储的资料。

不久后，David Filo 和杨致远共同创办了雅虎，通过著名的雅虎目录为用户提供导航服务。雅虎目录有近 100 万个分类页面，14 个国家和地区当地语言的专门目录，包括英语、汉语、丹麦语、法语、德语、日语、韩文、西班牙语等。雅虎目录已成为最常用的在线检索工具之一，并成功地使搜索引擎的概念深入人心。

随着网页数量的增长和用户关键词查询的需要，从 1996 年到 2004 年，雅虎先后选用 AltaVista、Inktomi、Google 等第三方的搜索引擎作为自己网页搜索的后台服务提供商。

YST 是雅虎搜索技术 (Yahoo! Search Technology) 的简称。2003 年，雅虎公司斥资近 20 亿美金收购 Inktomi、Overture、Alta Vista、Fast 四家国际知名搜索服务商，经过数百名工程师一年的开发打磨，于 2004 年 2 月正式推出了国际领先的搜索引擎。雅虎 2004 年 2 月宣布中断与 Google 的合作，在全球范围内转而采用 YST。

进入 2005 年后，雅虎搜索逐步确立了社区化搜索 (Social Search) 的策略，将积极发挥全球庞大的注册用户群来积累大批高质量内容和元数据 (Meta Data)，从而改善用户的搜索体验。在这种策略下，雅虎不断推出新的社区化搜索服务，例如：“知识堂”、“收藏+”等，并收购了著名的照片共享网站 Flickr 和社会书签网站 Del.icio.us，进行产品上的优势互补。雅虎搜索以人为本的理念 (Better Search Through People) 开始引领新的潮流。

2、YST 技术

YST 是一个涵盖全球 100 多亿网页的强大资料库，是拥有数十项技术专利和近万台服务器的分布式集群架构，支持 38 种语言，服务全球 50% 以上网友的搜索需求。

雅虎网页搜索引擎的基本技术原理如下：

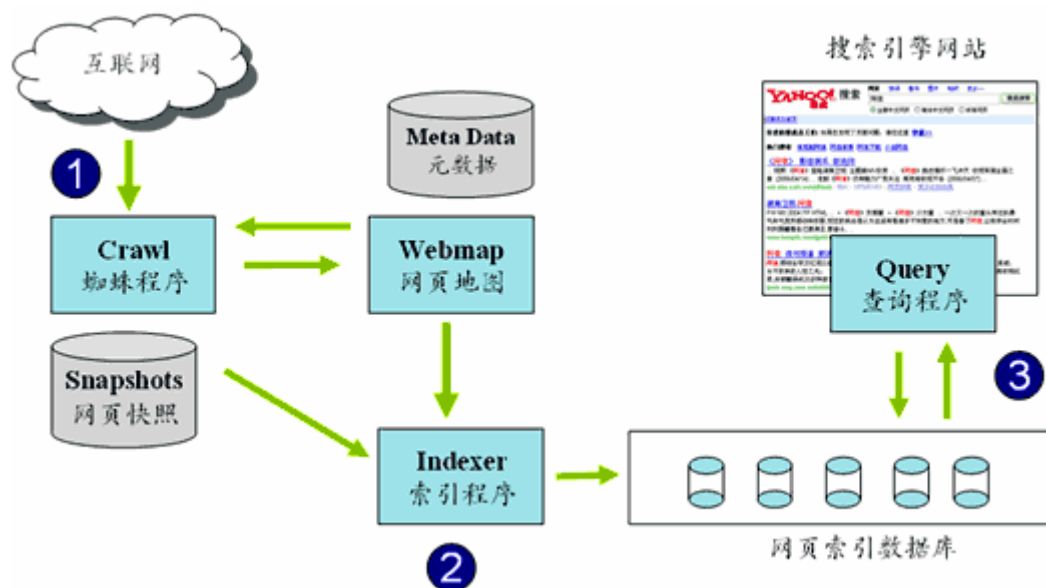


图 1-5-2 YST 基本原理

首先，搜索引擎会建立一个网页地图（Webmap），记录互联网的链接结构，再使用网页抓取程序（Robot）根据网页地图来抓取（Crawl）质量好的网页，存储到网页快照数据库中。接下来，索引程序（Indexer）会将快照数据库中的网页编号存储到网页索引数据库中，在这个过程中，会利用相关技术去掉作弊网页（Spam）。

当用户输入一个查询词搜索时，查询程序会使用这个查询词到索引数据库中比较，经过相关性计算后，按照相关程度对网页进行排序，相关性越高的排得越靠前。相关性的计算是包括各种因素的综合结果，例如：网站标题或网页内容对查询词的匹配、网页被链接的次数等。

3、雅虎中国

雅虎在全球共有 24 个网站，12 种语言版本，其中雅虎中国网站（www.yahoo.com.cn）于 1999 年 9 月正式开通，它是雅虎在全球的第 20 个网站。

2003 年 11 月 21 日，雅虎中国收购 3721 公司，获得了网络实名的相关技术和巨大的分销渠道。随后，雅虎中国推出了独立域名的搜索引擎“一搜”（www.yisou.com）。

2005 年 8 月 11 日，阿里巴巴收购雅虎中国，并整合原先的产品，推出了“雅虎百业窗”。

2007 年 3 月底，雅虎中国重新推出了 Yisou.com 并改名“易搜”。

三、百度

1、关于百度

百度，2000 年 1 月创立于北京中关村，是全球最大的中文搜索引擎。

2000 年 1 月 1 日，公司创始人李彦宏、徐勇携 120 万美元风险投资，从美国硅谷回国，创建了百度公司。创立之初，百度就将自己的目标定位于打造中国人自己的中文搜索引擎，并愿为此目标不懈地努力奋斗。

2000 年 5 月，百度首次为门户网站——硅谷动力提供搜索技术服务，之后迅速占领中

国搜索引擎市场，成为最主要的搜索技术提供商。2001年8月，发布 Baidu.com 搜索引擎 Beta 版，从后台服务转向独立提供搜索服务，并且在中国首创了竞价排名商业模式，2001年10月22日正式发布 Baidu 搜索引擎。

2005年8月5日，百度在美国纳斯达克上市，成为2005年全球资本市场上最为引人注目的上市公司，百度由此进入一个崭新的发展阶段。

七年来，百度一直孜孜不倦地追求技术创新，依托于博大精深的中文智慧，致力于为用户提供“简单，可依赖”的互联网搜索服务。

百度每天响应来自138个国家超过数亿次的搜索请求。用户可以通过百度主页，在瞬间找到相关的搜索结果，这些结果来自于百度超过10亿的中文网页数据库，并且，这些网页的数量每天正以千万级的速度在增长。

同时，用户不必访问百度主页，也可以搜索信息。超过3万个搜索联盟会员，通过各种方式将百度搜索结合到自己的网站，使用户在上网的任何时候都能进行百度搜索。百度还提供 WAP 与 PDA 搜索服务，即使身边没有 PC 机，用户也可以通过手机或掌上电脑等无线平台进行百度搜索。

百度一直致力于倾听、挖掘与满足中国网民的需求，秉承“用户体验至上”的理念，除网页搜索外，还提供 MP3、文档、地图、传情、影视等多样化的搜索服务，率先创造了以贴吧、知道、百科、空间为代表的搜索社区，将无数网民头脑中的智慧融入了搜索。“百度一下”已经成为了人们进行搜索的新动词。

百度还为各类企业提供竞价排名以及关联广告服务。每个月，有超过5千家的企业通过百度获得商机，5万家企业使用百度竞价排名服务，超过300家大型企业使用百度搜索广告服务。

2、关于李彦宏

1991年毕业于北京大学信息管理专业，随后赴美国布法罗纽约州立大学完成计算机科学硕士学位。在美国的8年间，李彦宏先生先后担任了道·琼斯公司高级顾问，《华尔街日报》网络版实时金融信息系统设计者，以及国际知名互联网企业 INFOSEEK 资深工程师，是新一代互联网技术领域的权威专家。他为道·琼斯公司设计的实时金融系统，迄今仍被广泛地应用于华尔街各大公司的网站，其中包括《华尔街日报》的网络版。

李彦宏最先创建了 ESP 技术，并将它成功的应用于 INFOSEEK/GO.COM 的搜索引擎中。GO.COM 的图像搜索引擎是他的另一项极具具有应用价值的技术创新。

1996年，他首先解决了如何将基于网页质量的排序与基于相关性排序完美结合的问题，并因此获得了美国专利。

1998年，根据在硅谷工作以及生活的经验，在大陆出版了《硅谷商战》一书，获得了各界的好评。

1999年底，携风险投资回国与好友徐勇先生共同创建百度。

2001年被评选为“中国十大创业新锐”之一。

2002年、2003年荣获首届、第二届“IT十大风云人物”称号。

2004年1月15日，当选第二届“京城十三新锐”。

2004年4月，当选第二届“中国软件十大杰出青年”。

2005年1月，当选全国青联委员。

2005年8月23日，荣获第十二届“东盟青年奖”。

2005年12月28日，荣获“CCTV2005中国经济年度人物”。

2006年12月10日，当选美国《商业周刊》2006年全球“最佳商业领袖”。

四、搜狗

搜狗是搜狐公司于2004年8月3日推出的全球首个第三代互动式中文搜索引擎，域名为www.sogou.com。搜狗以搜索技术为核心，致力于中文互联网信息的深度挖掘，帮助中国上亿网民加快信息获取速度，为用户创造价值。

搜狗的产品线包括了网页应用和桌面应用两大部分。网页应用以网页搜索为核心，在音乐、图片、新闻、地图领域提供垂直搜索服务，通过“说吧”建立用户间的搜索型社区；桌面应用则旨在提升用户的使用体验：搜狗工具条帮助用户快速启动搜索，拼音输入法帮助用户更快速地输入，PXP加速引擎帮助用户更流畅地享受在线音视频直播、点播服务。

搜狗网页搜索作为搜狗最核心的产品，经过两年半持续不断的优化改进，于2007年1月1日正式推出3.0版本。全面升级的搜狗网页搜索3.0凭借自主研发的服务器集群并行抓取技术，成为全球首个中文网页收录量达到100亿的搜索引擎；加上每天5亿网页的更新速度、独一无二的搜狗网页评级体系，这些确保了搜狗网页搜索在海量、及时、精准三大基本指标上的全面领先。

搜狗的其他搜索产品也各有特色。音乐搜索小于2%的死链率，图片搜索独特的组图浏览功能，新闻搜索及时反映互联网热点事件的看热闹首页，地图搜索的全国无缝漫游功能，使得搜狗的搜索产品线极大地满足了用户的需求，体现了搜狗强大的研发能力。

搜狗在产品研发的过程中追求技术创新。尤其值得一提的是，搜狗以一种人工智能的新算法，分析和理解用户可能的查询意图，对不同的搜索结果进行分类，对相同的搜索结果进行聚类，在用户查询和搜索引擎返回结果的人机交互过程中，引导用户更快速准确定位自己所关注的内容。该技术全面应用到了搜狗网页搜索、音乐搜索、图片搜索、新闻搜索、地图搜索等服务中，帮助用户快速找到所需的搜索结果。这一技术也使得搜狗的问世标志了全球首个第三代互动式中文搜索引擎诞生，是搜索技术发展史上的重要里程碑。

基于搜索技术，搜狗还推出了若干桌面应用产品。搜狗工具条是用户快速执行搜索的入口，同时集成了RSS订阅、文件下载加速、广告拦截、网页评级显示等丰富的功能。拼音输入法利用先进的搜索引擎技术，通过对海量互联网页面的统计和对互联网上新词热词的分析，使得首选词准确率领先于其他输入法。PXP加速引擎是一套基于PXP技术的互联网音视频直播和点播解决方案，能够支持内容和服务提供方以很少的带宽同时为上万用户提供流畅的视频服务。

设想一个用户在搜狗的使用体验：

通过拼音输入法，大幅提升了输入速度；

通过网址导航和工具条，直达目标网站或快速启动搜索；

通过网页搜索，从百亿中文网页中找到最相关的信息；

通过音乐搜索、图片搜索、新闻搜索、地图搜索等专项搜索，满足特定的查找需求；

通过说吧平台，和天南海北的用户交流信息；

通过PXP加速引擎，更快地下载文件，更流畅地享受在线音视频直播、点播服务。

可见，搜狗的全线产品，针对用户访问互联网过程中信息获取不畅的情况，在多个环节提供服务，把新、快、全、准的优质信息提供给用户，提升了用户访问互联网的体验。可以说，搜狗——更懂网络。

五、中文搜索引擎列表

表 1-5-1 中文搜索引擎列表

名 称	网 址	所用技术
百 度	www.baidu.com	百 度
谷 歌	www.google.cn	Google
雅虎中文	www.yahoo.cn	Yahoo
易 搜	www.yisou.com	Yahoo
有 道	www.youdao.com	网 易
SOSO	www.soso.com	Google
爱 问	www.iask.com	Google
搜 狗	www.sogou.com	搜 狗

第二章 搜索引擎营销基础

第一节 什么是搜索引擎营销

一、搜索引擎营销的定义

搜索引擎营销，英文是 Search Engine Marketing，缩写为 SEM。

搜索引擎营销是一种营销方法，它根据用户使用搜索引擎的习惯，采用付费形式或者技术手段，使网页在关键词搜索结果中排名靠前，引导用户点击，从而达到品牌展示和促进销售的目的。

二、搜索引擎营销的价值

美国知名搜索引擎营销专业服务商 iProspect 和市场研究公司 Jupiter Research 联合调查显示：互联网用户使用搜索引擎越来越没有耐心，越来越多的互联网用户仅关注搜索引擎结果第一页的内容，如果第一页面没有满意的结果，立刻变换关键词或者更换搜索引擎重新进行检索（数据来源：iProspect & Jupiter Research 2006 年 1 月）。

iProspect 调查结果如下：62% 的搜索用户只点击搜索结果第一页链接，2004 年的这个比例是 60%，2002 年是 48%。此外，高达 90% 的搜索者只查看搜索结果前三页，2004 年和 2002 年分别是 87% 和 81%。

该调查询问了 2369 个搜索引擎用户。当搜索用户被问及什么情况下会更换关键词或者更换搜索引擎重新搜索，他们的回复比例如下：

表 2-1-1 用户在什么情况下更换关键词或者搜索引擎

浏览搜索结果的方式	比例
看第一页前几个搜索结果	16%
看第一页搜索结果	25%
看前两页搜索结果	27%
看前三页搜索结果	20%
看三页以上的结果	12%

需要注意的是，以上的比例只是表明搜索用户可能会查看的搜索结果页面数量，并不意味着会点击查看这些页面上的所有网站和网页内容。如果网站的信息出现在搜索结果第三页之后，只有 12% 的用户可能会查看这些页面，而查看到某个网站或者网页的几率更是微乎其微了。

因此，让自己的网站在某个关键词的搜索结果中排到靠前的位置，而且是越靠前越好，这样，搜索引擎才可能为你带来更多的关注和点击，同时也带来更多的商业机会。

三、搜索引擎营销原理

那么，是不是只要网站在搜索结果排名第一，就一定会财源滚滚呢？答案是否定的。下面，我们来分析一个搜索用户的行为过程，从而明确搜索引擎营销怎样才能为企业带来收益。

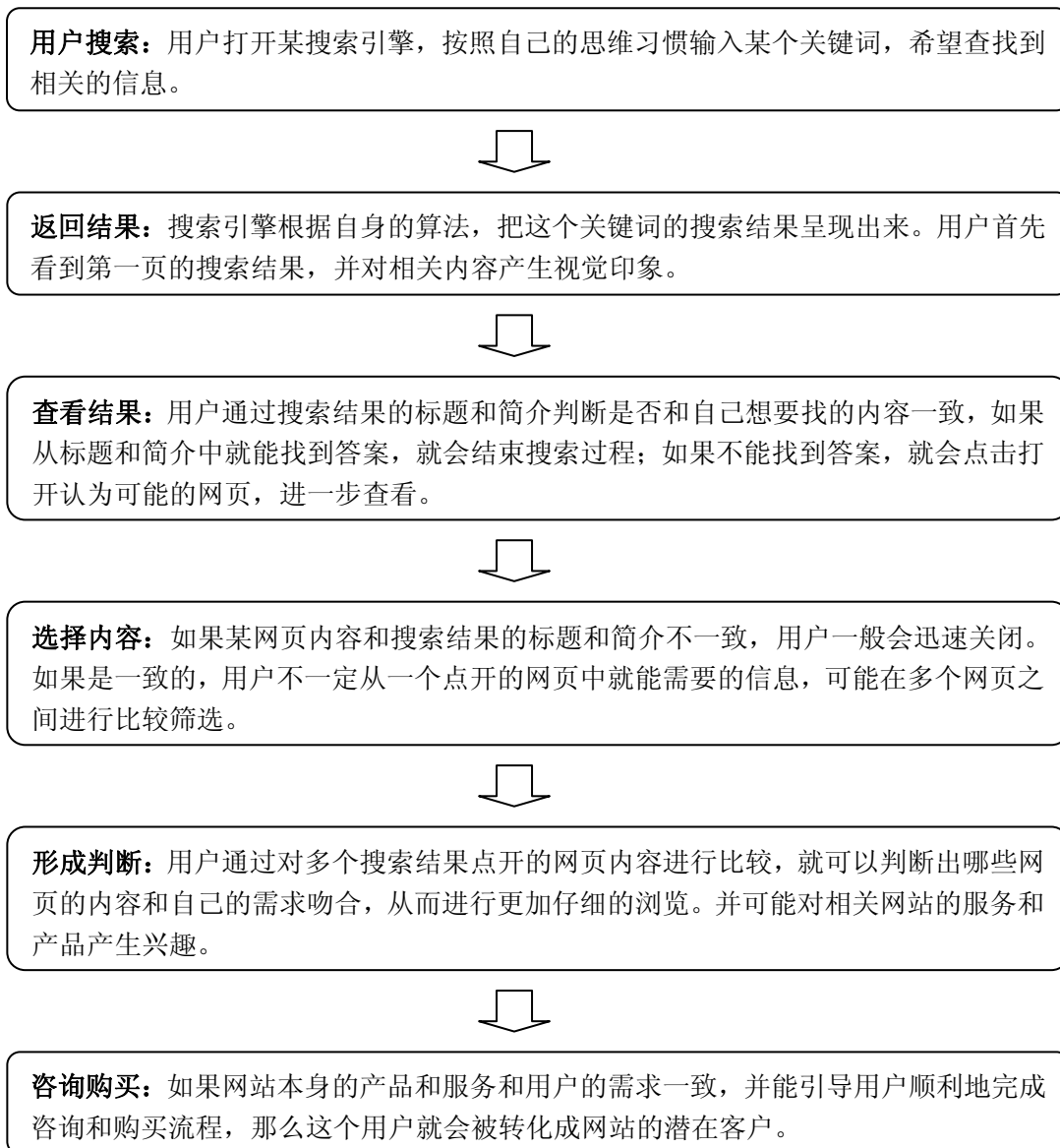


图 2-1-1 搜索用户行为过程图

从上面流程图可以看出，搜索引擎营销并不是简单地让网站排在搜索结果靠前的位置就行了，而是和目标设定、用户搜索习惯分析、网站设计制作以及业务流程相关的。

正确的搜索引擎营销理念应该是超出搜索引擎本身的，出售搜索引擎关键词广告或者是搜索结果排名的服务，仅仅是搜索引擎营销过程的一部分。

第二节 搜索引擎营销的特点

一、广泛使用

世界已经进入信息大爆炸的时代，互联网的网页数量已经达到数百亿级别，中文网页数也已经达到 44.7 亿个（数据来源：CNNIC，中国互联网络发展状况统计报告，2007 年 1 月）。可以说，对于广大网民来说，不是信息不够，而是信息太多了。如何让企业的商品和服务信息，在众多的信息中脱颖而出，呈现在用户面前，并引起用户的关注，成了广告主非常重视的问题。

搜索引擎在某种程度上解决了信息获取和信息筛选的问题。一方面，搜索引擎帮助用户找到想要找到资料；另一方面，搜索引擎又通过自身的算法，努力使与用户搜索请求更相关的内容出现在搜索结果靠前的位置。

搜索引擎是网民经常使用的第二大网络服务，根据市场研究公司 comScore Networks 公布的数据，2007 年 1 月，整个美国网络搜索行业的搜索总量为 69 亿次，比去年 12 月增长 2%，比去年同期增长 26%。这个数字让所有的广告主都无法忽略搜索引擎带来的商业价值。让自己的网站被用户搜索到，已经成为各类企业新的营销方向。

二、用户主动查询，针对性强

你在找客户，客户也在找你。网络用户是自己主动查询，针对性强。在搜索的时候，客户的需求已经通过关键词表现出来，搜索引擎根据客户需求，给出相应结果。根据 Jupiter Media Metrix 在 2001 年 7 月底的统计，47%的消费者通过搜索引擎来寻找和购买在线产品。这些消费者在线购物时寻找产品的方法和比例是：在搜索引擎中输入产品（28%）；在搜索引擎中输入商标名（9%）；在搜索引擎中输入商店名（5%）；访问搜索引擎的购物频道（5%）。

“你上网为什么使用搜索引擎？”“因为我想查找我需要的信息。”这是每一个使用搜索引擎查找信息的网民的必然回答。当一个上网的人，需要某个方面的信息的时候，他最先想到的就是去搜索引擎搜索。比如说，一个自来水公司的采购经理想要采购一批水表，但是手头又没有现成的水表生产厂家的资料，他自然想到的就是去搜索引擎搜索水表生产厂家的信息。他会打开一个搜索引擎，搜索水表、水表企业、水表公司等系列可能的关键词。这样，排在搜索结果前面的公司，被选中的几率就会更大，更有可能得到这笔生意。

从上面的两点可以看出，搜索引擎是绝大多数网民经常使用的网络服务，是大多数网民查找自己所需要信息的重要工具，从注意力经济的原则看，肯定会有很好的营销效果。并且，使自己出现在搜索结果的前面，是取得好的搜索引擎营销效果的前提。

搜索引擎是建立在互联网基础上的，它继承了互联网没有时间、空间、地域限制的优点，同时又具有传统媒体广告所无法比拟的低成本、高针对性的优点，因此，可以说是迄今为止最优秀的推广方式。

三、获取新客户

2002 年 7 月 22 日，中国互联网络信息中心发布了《CNNIC 第十次中国互联网络发展状况统计调查报告》，用户得知新网站的主要途径有：

表 2-2-1 用户得知新网站的主要途径

得知新网站的途径	占网民的比例
搜索引擎	82.2%
其它网站上的链接	68.9%
电子邮件	30.5%
朋友、同学、同事的介绍	54.0%
网友介绍	28.0%
网址大全之类的书籍	17.8%
报刊杂志	43.0%
广播电视	12.3%
黄页	3.2%
户外广告	11.6%
其它	0.4%

根据这些数据，在得知新网站的途径中，搜索引擎的重要性已经不言而喻。也就是说，在获取新客户方面，搜索引擎发挥了重要作用。特别对于已经拥有网站的企业来说，登录搜索引擎已是网站宣传推广最直接、最有效的途径，也是最重要的网络营销方法，如果企业的站点不在搜索引擎中，潜在的客户将不能找到。如何充分利用搜索引擎为企业和公司搞好网络营销并发挥最大限度的作用，应当是目前每一个上网企业所应该高度重视的问题。

四、竞争性强

每个企业或网站都希望自己的信息出现在搜索结构中靠前的位置，否则，就意味着被用户发现的概率比较低，因此，对搜索引擎检索结果排名位置的争夺成为许多企业网络营销的重要任务之一。

如果客户通过搜索引擎找到你的网站，而非你竞争对手的网站，你就已经在互联网的竞争中战胜了竞争对手。

五、动态更新，随时调整

根据企业产品和服务内容的变化，可以随时调整与更新广告发布的内容和网站内容，为客户提供及时的资讯，吸引更多新客户，留住更多老客户。

六、门槛低，投资回报率高

搜索引擎是开放性的平台，门槛比较低。任何企业，不论企业规模大小，也不论品牌知

名度高低，都可以在搜索引擎上推广宣传，并且机会均等，发布的信息都有可能排在前面。另外，与传统广告和其他的网络推广方式相比，搜索引擎网站推广更便宜，更有效。

第三节 搜索引擎营销的目标

在不同的发展阶段，搜索引擎营销具有不同的目标。最终的目标在于将浏览者转化为真正的顾客，从而增加销售收入。

一、被收录

是指在主要搜索引擎或分类目录中获得被收录的机会，这是搜索引擎营销的基础。搜索引擎登录包括免费登录、付费登录、搜索引擎关键词广告等形式。被收录的含义就是让网站中尽可能多的网页（而不仅仅是网站首页）被搜索引擎收录，增加网页的搜索引擎可见性。

二、排名靠前

是指在被搜索引擎收录的基础上尽可能获得好的排名，在搜索结果中有良好的表现。因为用户关心的只是搜索结果中靠前的少量内容，如果利用主要的关键词检索时，网站在搜索结果中的排名靠后，那么还有必要利用关键词广告、竞价广告等形式作为补充手段来实现这一目标。同样，如果在分类目录中的位置不理想，则需要同时考虑在分类目录中使用付费等方式使得排名靠前。

三、被点击

指通过搜索结果点击率的增加来达到提高网站访问量的目的。仅仅做到被搜索引擎收录并且在搜索结果中排名靠前是不够的，这样并不一定能增加用户的点击率，更不能保证将访问者转化为顾客。要通过搜索引擎营销实现访问量增加的目标，则需要从整体上进行网站优化设计，并充分利用关键词广告等有价值的搜索引擎营销专业服务。

四、客户转化

指将访问者转化为客户，将网站访问量的增加转化为企业实现收益的提高。从各种搜索引擎策略到产生收益，期间的中间效果表现为网站访问量的增加。网站的收益是由访问量转化所形成的，从访问量转化为收益则是由网站的功能、服务、产品等多种因素共同作用决定的。这个目标在搜索引擎营销中属于战略层次的目标。其他三个层次的目标则属于策略范畴，具有可操作性和可控制性的特征，实现这些基本目标是搜索引擎营销的主要任务。

五、提高品牌知名度

搜索引擎营销专业组织（SEMPO）首期 SEM 全面调研活动报告结果显示，61% 的搜索引擎营销商利用付费排名模式提高品牌知名度，同时更多的营销商对品牌影响力的重视程度远远高于产品销售、产生引导或者增加访问流量，虽然后者也被认为是重要目标。

这份名为《2004 搜索引擎营销状况》的报告由 Executive Summary Consulting 公司从 2004 年 10 月份开始执行，共计 288 家搜索营销商参与此次调查。报告旨在对北美广告商在搜索引擎营销商的投资规模进行评估，并对成长中的搜索引擎市场营销进行前瞻性预测。报告由该公司资深市场研究分析师、现加入 DoubleClick 调查开发部门的 Rick Brunner 负责完成。

一直以来搜索引擎营销被认为是一种直销媒介，大多数人认为它的主要作用是实现销售及产生引导。但是据 Brunner 介绍，更多被调查者看重付费搜索产生的品牌价值。数据显示，61% 的被调查者主要利用 SEM 提高品牌知名度；58% 希望直接实现在线销售，54% 的被调查者回答利用 SEM 的目的是为了产生引导，他们可以自己开拓其他销售渠道。

“我并不是刻意宣传这一点，但是事实确实如此，产生引导和在线销售的作用确实排在品牌宣传之后。从调查结果的图示上可以更加清楚的表现出来，”他说，“许多大型代理公司及搜索引擎已经耳闻到某些关于该特点的信息。人们在使用搜索时表现得非常复杂，各不相同。他们通常进行若干次不同搜索后才决定购买某种东西。或许正是因为这样，才给品牌宣传制造了机遇。”

与其他调查不同的是，SEMPO 的报告结果发现大多数广告客户由内部人员运作搜索营销的大部分工作，而不是外包给代理机构。52% 的广告客户承认将由内部人员自己操作 2005 年的全部营销花费。较大型的广告客户则更加愿意采用外包方式。最近的 JupiterResearch 研究报告显示只有三分之一的搜索营销商选择使用代理，但是由于大型营销商倾向于外包，代理机构控制了一半以上的搜索营销成本。

两家公司得出的调查结论出现如此差异，Brunner 认为是由于双方调查问题不同所致。“他们提问的问题可能是‘你是否计划外包？’……，而我们提问的问题则是‘总体投入的百分之多少将用于外包’。”

另外，该调查发现付费列表成本的急剧上升成为人们关注的一个问题。广告客户反映，在过去一年内点击付费的单位价格平均上涨了 26%，但是营销商同时承认仍然可以承受 33% 及以上的上漲价格，并且仍然能够获得较好的投资回报。

调查结果还显示，广告客户对欺诈点击的关心仅对次于蹩脚的 SEO 技术的担忧，大部分认为后者对行业造成了很大的危害。

该调查还发现，许多公司采取吸收原本投入到传统或网络营销方面的资金用于搜索营销的做法，而不是为搜索营销设立新的预算条目。预算分配比例大致呈现以下规律：获得最大份额的是购物搜索列表（13%），电子邮件营销次之（9%），网上展示广告（9%），印刷杂志广告（9%）和报纸广告（9%）。

据对 2005 年广告客户的营销投资计划的调查，搜索引擎营销全部模式的总体投入比例比 2004 年平均上涨 39 个百分点。

第四节 搜索引擎营销的形式

一、搜索引擎登录

实施搜索引擎营销，首先就是要使得网站被搜索引擎收录，有被搜索到的可能。要被搜索引擎收入，除了等待搜索引擎的蜘蛛找到你的网站后进行索引之外，还可以主动向搜索引擎提交网站。

一般的搜索引擎都支持网站主提交，但是免费的登录在时间上没有保障。有些搜索引擎

已经不提供免费在网站登录服务，比如搜狐只对非商业型网站，如政府机构、慈善机构、学校等，提供免费登录服务。搜狐、Yahoo 等有收费型登录，只要缴纳一定的费用，就可以加速收录的过程。

几年前，由于大大小小的搜索引擎有几百种之多，所以就出现了一些向搜索引擎提交网站信息的工具。但是由于现在几大搜索引擎已经占据了绝大部分的市场，而且搜索引擎也比较反感这种群发性质的网站提交请求，建议不要使用类似的工具和方法了。

这里需要指出的是，最好的被搜索引擎的收录的形式，还是让搜索引擎主动找到你。至于怎样让搜索引擎找到你的网站以及如何加速这个过程，这是后面章节重点阐述的内容。

二、固定排名和广告

1、固定排名

固定排名一般是一种收费服务，搜索引擎根据某个关键词的搜索量和所排列的位置，制定一个按年或者按月收取的价格，保证网站的排序位置。可以是前几名，或者是第一页，一般会用特殊的形式来显示，以和自然排名结果区别。固定排名有很多种说法，比如搜狐的“固定排序”、“推广型登录”等。

对于广告主而言，固定排名比较容易理解，而且因为费用固定也比较容易做预算，所以在搜索引擎营销发展的初期很受欢迎，一度是最主要的搜索引擎营销形式。

固定排名的广告价格往往取决于关键词每天被搜索的次数。比如“电影”这样的关键词，每日的搜索量是巨大的，所以固定排名的价格高达每月数万元。几年前很多电影下载网站，就是通过购买相关关键词的固定排名，来大赚短信分成的。

固定排名服务也不限于搜索引擎，很多分类目录网站也提供此项服务，让客户网站出现在某个分类收录结果靠前的固定位置。但并不是所有的搜索引擎都提供固定排名服务。有些公司提供所谓的 Google、百度的固定排名服务，其实是一种欺骗客户的行为。

随着搜索引擎营销的发展，广告主发觉固定排名无法体现搜索引擎营销的特点，所以这种广告形式渐渐被更加灵活的按效果付费的广告所取代。

2、地址栏广告

地址栏广告是一种特殊的固定排名形式，比较知名的有网络实名和通用网址。

网络实名 主要由 3721 公司提供的地址栏搜索服务——3721“网络实名”，属于第三代中文上网方式，用户无需记忆复杂的域名，直接在浏览器地址栏中输入中文名字，就能直达企业网站或者找到企业、产品信息。地址栏搜索基本上覆盖了 80% 以上的中国互联网用户，每日使用量超过 3000 万人次，是用户常用网络搜索服务之一。广告客户将自己的公司名、产品名注册为 3721 网络实名，用户输入自己的名字就可以直达自己的网站，为企业带来更多的商业机会。

通用网址 国家信息产业部信息中心的通用网址是一种新兴的网络名称访问技术，通过建立通用网址与网站地址 URL 的对应关系，实现便捷的浏览器访问。用户只需要使用自己所熟悉的语言告诉浏览器要去的通用网址即可。使用通用网址，企业可以最终对品牌进行保护；在地址栏直接输入中文，可以实现直达网站，便捷快速；可以在各大门户网站页面的显著位置，有文字或图片说明，实现免费宣传。企业还可以通过搜狐、新浪的直通车用一点通直接打开网站，保护流量。

3、弹出式广告

弹出式广告指当用户打开某特定网页时，该网站广告客户的宣传广告就会同时弹出，这种广告独立于用户浏览需要浏览的页面。弹出式广告出现于 2001 年，很快就盛行一时。

使用弹出式广告的商家认为，在这场弹出式广告之战中，对于弹出式广告或者其他网络广告形式谁优谁劣下定论还为时过早。现在还不是很清楚哪种互联网广告是比较合适的，而哪种又比较奏效，因为这种媒体宣传方式还太新了。在这场战争尘埃落定之前，可以做几件事来避免公司陷入尴尬，以及毫无必要地激怒用户。如果必须使用弹出式广告，至少可以控制它们不要那么无礼地打搅消费者。

而互联网用户则认为，弹出式广告是心头大患。当人们访问那些受人欢迎的网址时，弹出式广告总是一个一个地出现在用户面前，使得电脑桌面混乱不堪。广告商希望人们点击窗口的任何一个地方来访问网页以提高点击率，但人们大多只是选中右角顶端的关闭按钮了事。

三、按效果付费排名

赛迪评测报告显示，在网络营销服务的付费方式中，按效果付费的方式得到广大企业用户的青睐（60.51%），其次是收取年费的方式（26.25%），这表明多数企业很希望通过网络营销为企业带来新的推广效果，并很乐意为此效果付费。

竞价排名是按效果付费的主要推广方式，主要是企业在搜索引擎购买关键词，当用户搜索这些关键词信息时，企业的推广内容就会出现在用户面前。这种方式一般按照用户点击的效果付费，企业可以竞价搜索引擎里的排位。赛迪评测报告显示：百度在巨大的个人用户群基础上，由于其竞价排名服务按效果付费的特性受到较多企业认可，目前已成为国内最大的网络营销平台。

四、自然排名

自然排名是通过网站优化（SEO）使关键词获得较高的自然排位。相对于竞价排名来说，自然排名稳定持久，无需为每一次点击付费，长远看费用低于竞价排名费用。更重要的是，自然排名比竞价排名更令搜索用户信任，获得的点击是竞价排名的 2 倍以上。

但是自然排名同样会有不足：

- 影响自然排名的因素多，无法精确控制，排名位置无法保证；
- 专业网站优化公司收费高，一次性投入成本比较高；
- 自然排名效果显现需要等待较长时间（通常需要 2-6 个月），相较于竞价排名来说需要更多的耐心；
- 搜索引擎排序规则时常变化，排名维护是个长期系统的工作，网站需要有独特有价值的内容并需要经常更新；
- 关键词优化与网页阅读和视觉效果经常产生冲突；
- 如果优化过度很可能造成网站受到惩罚；
- 热门关键词的高排名越来越困难，排名竞争总是越来越激烈。

第三章 搜索引擎优化基础

第一节 为什么要搜索引擎优化

一、搜索引擎优化的定义

1、WIKI 定义

SEO 是 Search Engine Optimization 的缩写,用英文描述就是 to use some technics to make your website in the top places in Search Engine when somebody is using Search Engine to find something, 翻译成中文就是“搜索引擎优化”,一般可简称为搜索优化。

SEO 的主要工作是通过了解各类搜索引擎如何抓取互联网页面、如何进行索引以及如何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。

2、国内的一些解释

搜索引擎优化是搜索引擎营销的常见形式之一。根据《网络营销基础与实践》(第2版)第4章“搜索引擎营销基础”中的相关内容,所谓搜索引擎优化,也就是针对各种搜索引擎的检索特点,让网页设计适合搜索引擎的检索原则(即搜索引擎友好),从而获得搜索引擎收录并在排名中靠前的各种行为。如对于基于 META 标签检索的搜索引擎,在 META 标签中设置有效的关键词和网站描述,对于以网页内容相关性为主的蜘蛛型搜索引擎,则通过在网页中增加关键词的密度,或者专门为搜索引擎设计一个便于检索的页面(如 sitemap.htm、robots.txt)。

Robin 根据自己对搜索引擎优化两年多的研究而总结出的搜索引擎优化定义:搜索引擎优化是这么一种技术,即是遵循搜索引擎科学而全面的理论机制,对网站结构、网页文字语言和站点间的互动外交策略等进行合理规划部署来发掘网站的最大潜力而使其在搜索引擎中具有较强的自然排名竞争优势,从而对促进企业在线销售和强化网络品牌起到作用。

目前为止,比较权威的定义是:搜索引擎优化,指遵循搜索引擎的搜索原理,对网站结构、网页文字语言和站点间互动外交策略等进行合理规划部署,以改善网站在搜索引擎的搜索表现,进而增加客户发现并访问网站的可能性的这样一个过程。搜索引擎优化也是一种科学的发展观和方法论,它随着搜索引擎的发展而发展,同时也促进了搜索引擎的发展。

二、搜索引擎优化的价值

1、自然结果商机更大

从搜索习惯来看,用户更倾向于点击查看自然排名的结果。

各大搜索引擎都很好的区分了自然排名结果和广告,比如 Google 仅仅把广告显示在右侧,左侧全是自然排名结果,百度也在广告结果中加入了“推广”的字样。一般用户对广告都有一定的排斥,一旦他们分辨出哪些是广告,就会尽量不去点击。

所以,如果企业的网站出现在在自然排名中,就会加大被用户点击的可能性,从而带来

更多的商业机会。

2、节省费用

网站通过实施搜索引擎优化,可以使一些关键词的排名提高。而此前如果希望通过这些关键词带来潜在客户,则需要不断地投放广告。这无疑为企业节省了一定的广告开支。即便企业不去削减广告预算,也大可以把这些费用投放在更多的关键词上。

3、全面覆盖

随着搜索技术的发展,越来越多的搜索引擎出现了。但是企业不可能在所有的搜索引擎针对所有的相关内容都投放广告,况且还有一些搜索引擎暂时不提供收费的广告业务。因此,只有根据搜索引擎的普遍规律去优化网站,才有可能做到对搜索引擎更多的覆盖。

4、更好的用户体验

正确的搜索引擎优化并不会为了迎合搜索引擎而把网站改得面目全非,相反,优化会遵循“用户喜欢的才是搜索引擎喜欢的”这一基本观点,去充分考虑用户的体验。

三、SEO 和 SEM 的关系

国外有一些相关的书籍和文章把 SEM 和 SEO 放在并列的位置看待,认为 SEM 就是付费排名,SEO 就是自然排名。这样的提法也无不可,但本书的观点还是把 SEO 看作是 SEM 的一个部分。

作为 SEM 的重要形式,SEO 和 SEM 有着共同的目标,就是使网站出现在搜索结果更靠前的位置,从而带来更多的访问量和潜在客户。

SEO 工作涉及的领域有不仅仅局限于 SEM 的范畴。一个好的 SEO 方案及其实施,将不可避免地涉及到网站策划、网页设计、程序编写,甚至要考虑用户体验、购物流程等。

四、SEO 和付费排名的关系

有些 SEO 服务商提出要 SEO 不要付费排名的观点,认为网站只要通过搜索引擎优化,就可以达到从搜索引擎带来访问量和潜在客户的目的,不需要另外花钱购买任何形式的关键词广告了。这种观点很容易让人认为 SEO 和搜索引擎是对立的,好像 SEO 抢了搜索引擎的客户。

搜索引擎的付费广告有一个前提,就是必须事先规划好关键词,并针对这些关键词投放广告。如果对一个网站的访问情况进行分析,我们不难看出,有很多来自搜索引擎的访问是在我们预期之外的。

那么如何有效抓住这些预期之外的关键词,让它们也能为网站带来访问量和潜在客户,并不断提升呢?有效的搜索引擎优化将能很好地解决这一问题。

搜索引擎优化不仅能有目的地针对一些关键词进行优化,使其在搜索结果的排名上升,同时还通过调整网站结构、代码书写规范、文本写作等一系列工作,使得网站在搜索引擎的表现得到改善。网站绝大多数页面都能被搜索引擎收录,网页中绝大多数文字内容都能被搜索引擎索引到,这就意味着出现在网站的任何文字都有可能成为目标关键词。也许一个非预期的关键词带来的访问量是很小的,可能一天一次甚至一个月一次,但是这样的关键词数量

巨大，综合起来的整体流量依然是不容小视的。

按照以上的说法，是不是根本不用考虑付费排名了呢？答案也是否定的。上面提到的仅仅是一种可能性。虽然用户搜索某个关键词，网站可能在搜索结果中出现，但是可能排在数十页之后，那网站被用户点击的可能性也就几乎为零了。如果有一些关键词通过分析和实践，确实能带来有效的访问者和潜在客户，而网站在这个关键词的搜索结果排名又不是非常理想，同行的竞争也比较激烈的话，就很有必要购买相关关键词的付费广告了。

网站实施 SEO 确实会减少某些关键词的广告投放量，但是因为通过 SEO，网站本身各方面都有了改善，客户转化率提高了，就可能加大企业在其他关键词上的投放量。而且通过 SEO 工作，能够帮助分析出更多的相关关键词，从而开始更多关键词的广告投放。

综上所述，SEO 和付费排名并不是矛和盾的关系，从事搜索引擎营销，必须把二者有机结合起来，以期从搜索引擎带来尽可能多的目标客户，让搜索引擎带来的价值最大化。

第二节 搜索引擎对 SEO 的态度

对 SEO 有明确的态度是搜索引擎自身成熟的表现。主流搜索引擎都对 SEO 有明确的态度，一般来说，搜索引擎都有关于网站如何在搜索引擎取得更好排名的建议，同时也对作弊的行为作出了明确的界定。

从根本上说，SEO 与搜索引擎有一个共同的目标，那便是如何有效地组织内容，帮助用户以最有效的方式找到所需的内容。

一、Google

以下内容来源于 Google 推广 (<http://googlejscom.bokee.com/>)。

1、SEO 与 Google

■ 搜索引擎优化服务商。

SEO 也是“Search Engine Optimizer”的缩写。许多 SEO 为网站所有者提供了诸多有用的服务，包括撰写稿件、为网站结构出谋划策以及帮助查找网站可以提交到的相关目录等。但是，也有少数不道德的 SEO 采用极为过分的市场营销手段，并试图暗箱操纵搜索引擎结果，从而损害了这个行业的声誉。

Google 与任何 SEO 都没有关联，也不会提供这方面的建议，但 Google 提供了一些提示，帮助你区分哪些 SEO 可以改善你的网站效果，哪些却只能徒使搜索引擎将你从搜索结果中删除。

■ 对突然向你发送电子邮件的 SEO 公司要提高警惕。

令人吃惊的是，Google 也收到过这类垃圾邮件：

“尊敬的 Google.com，

我访问了你的网站，发现大多数主要的搜索引擎和目录都没有将你列入其中……”

不要轻易相信那些涉及搜索引擎的垃圾邮件，它们就像“让脂肪在睡眠中燃烧”的减肥药丸一样可疑。

■ 没有人可以保证在 Google 排名第一。

如果 SEO 宣称可以确保你名列前茅，或声称与 Google 有特殊关系，可以优先向 Google 提交你的网站，千万不要相信。Google 从来都没有优先提交一说。实际上，向 Google 直接提交网站的唯一方法是通过 Google 的添加网址页或通过 Google Sitemaps（测试版）程序提交。你可自己完成这一工作，无需支付任何费用。

■ **如果一个公司遮遮掩掩，或不明确说明自己的意图，请多加小心。**

如果有任何不明白之处，都应当要求对方解释。如果 SEO 为你制作了一些欺骗性或误导性内容，如桥页或“一次性”域，则你的网站可能会彻底从 Google 索引中删除。归根结底，你要对所雇佣公司的行为负责，因此，你最好弄清楚他们打算采用何种方法来“帮助你”。

■ **切勿链接到 SEO。**

请远离那些高谈“免费链接站点”的好处，宣传链接流行度方案，或将你的网站提交给成千上万个搜索引擎的 SEO。因为这些做法是徒劳无益的，并不会影响你在各大搜索引擎结果中的排名——至少不会产生你期望的正面影响。

■ **一些 SEO 可能会尝试向你出售直接在浏览器地址栏中输入关键词的功能。**

这些方案大都要求用户安装额外的软件，而很少有用户这样做。在评估此类方案时，务必极其小心，不要轻信对方提供的下载所需应用程序的用户人数。

■ **做出明智的选择。**

在考虑是否与 SEO 合作时，你应当对该行业做一些调查。Google 自然可以提供此方面的信息。你可能还需要查找报刊上所登载的一些提醒性质的报道，譬如以下这篇有关某个做法激进的 SEO 的文章：

http://seattletimes.nwsourc.com/html/business/technology/2002002970_nwbizbriefs12.html

尽管 Google 不会对具体的公司加以评论，但确实遇到了一些自称是 SEO 而其行为却违背了合理的商业行为准则的公司。请务必小心。

■ **了解资金去向。**

虽然 Google 从未将搜索结果中的较好排名进行出售，但其他一些搜索引擎已在其常规的网络搜索结果中实施了按点击次数或列入结果的次数进行收费。一些 SEO 会承诺使你在搜索引擎中居于靠前的位置，但是却将你放在广告部分而非搜索结果中。还有少数 SEO 甚至会实时更改他们的报价，制造出他们可以“控制”其他搜索引擎并能够手眼通天、为所欲为的假象。这些伎俩在 Google 是行不通的，因为 Google 的广告服务已明确标示，并与搜索结果相互独立，但是对于任何 SEO，你都要问清楚哪些费用用于将你永久性地列入搜索结果，哪些费用用于短时间的广告刊登。

■ **与多个 SEO 洽谈，询问其他 SEO 对你正在考虑公司的看法。**

他人的推荐是一个良好的开端，但是推荐人的介绍往往并不全面。你应当了解公司的成立时间，雇佣的全职员工数量等信息。如果你感到有压力或怀疑，应顺从你内心的感觉，不要贸然行事，在找到可以信赖的公司之前不要做出任何决定。询问你的 SEO 公司是否使用

位于以下网址的垃圾邮件投诉表，向 Google 报告所发现的每封垃圾邮件：

<http://www.google.com/contact/spamreport.html>

有道德的 SEO 公司会报告违反 Google 垃圾邮件指南的所有欺骗性网站。

■ 确保你受法律保护。

为了你自身的安全，你应当坚持要求无条件全额退款保证。无论什么原因，只要有所不满，或如果 SEO 的行为导致你的域名从搜索引擎索引中删除，不要惧怕要求退款。务必签订包括定价在内的书面合同。合同中还应要求 SEO 遵循每个搜索引擎针对要列入的网站所做出的规定。

2、网站所有者可能会遇到哪些最常见的欺诈行为？

一个常见的欺骗手段是创建“影子域名”，使用欺骗性的重定向功能，将用户链接到某个网站。这些影子域名通常为那些声称为客户工作的 SEO 所有。但是，如果双方关系恶化，SEO 可能会将该域指向另外一个网站，甚至指向竞争对手的域。此时，客户等于付费开发了一个完全由 SEO 所有的竞争网站。

另外一个不轨的做法是在客户网站的某个位置放置充斥着关键词的“桥页”。SEO 承诺这种方法会增强网页的相关性，从而可以匹配更多的查询。这种观点根本就是错误的，因为单个网页几乎不可能与大量的关键词都相关。但更大的隐患是这些桥页还常常包含指向 SEO 其他客户的隐藏链接。这些桥页会将网站的链接流行度分流到 SEO 其客户那里，其中可能包含一些提供令人生厌或非法内容的网站。

3、还需要注意哪些事项？

你可以通过某些迹象来判断是否正在与不道德的 SEO 打交道。下面所列出的只是其中很少的一部分，因此，只要有任何疑虑，你都应当相信你的直觉。如果 SEO 具有以下行为，一定要想尽办法果断摆脱它：

- 拥有影子域名；
- 在桥页上放置指向其他客户的链接；
- 出售在地址栏中输入关键词的功能；
- 对实际搜索结果和搜索结果中的广告不加区分；
- 保证排名，但只依靠随意可得的晦涩冗长的关键词词组；
- 使用多个别名或伪造的 WHOIS 信息；
- 从“虚假的”搜索引擎、间谍软件（spyware）或变脸软件（scumware）获取点击量；
- 自身的域名或者所拥有的域已从 Google 索引中删除，或者在 Google 的索引中没有被列出。

如果认为某 SEO 采用其中某种欺骗方式，请最好向 Google 报告。

二、Yahoo

以下内容来自雅虎帮助中心（<http://help.cn.yahoo.com/>）。

1、如何让雅虎更多更快的收录我的网页？

你需要注意以下几点：

- 正确配置网站的 robots.txt 文件，错误的配置会导致雅虎无法抓取；
- 雅虎搜索喜欢独特原创内容的网站；
- 请维护并且经常更新你的网站，避免被人恶意利用群发垃圾讯息；
- 与和你相关类型的网站交换链接（请注意不要在 BBS，blog，留言本等增加链接，这样会被雅虎搜索引擎认为是作弊行为）；
- 欺骗用户以及采用作弊手段欺骗搜索引擎的行为，都会在雅虎搜索中删除。

2、网站排列位置由什么决定？

网站排列位置主要由网站自身质量以及与查询词的相关性强弱决定的。你可以采取以下方法来提高网站在搜索结果中的排名：

- 方法一：高质量的网站内容，雅虎搜索引擎喜欢抓取原创的、专业的、特殊的、有意义的网页内容；
- 方法二：尽量使用搜索引擎能够识别的文本信息，避免过多的 Javascript、Cookie、框架、DHTML 或 Flash 等复杂技术；
- 方法三：不要在留言本或论坛中大量添加你网站的链接；
- 方法四：专业、认真、用心的去对待访问你网站的用户；
- 方法五：对于采用不正当手段骗取排名的网站，搜索引擎会自动降低排名。例如：在网页标题中堆积大量无关查询词，在网页的 meta keywords 中堆积大量无关查询词等，而主动链接这些作弊网站的网站，也会被搜索引擎降低其排名或者删除。

3、哪些类型的网站是作弊网站？

如果你的网站存在以下行为，就有可能被雅虎搜索引擎判断为作弊网站，而不予收录：

- 在网页中存在隐藏文字或隐藏链接；
- 在网页中存在大量与网页内容不相关的查询词；
- 在网页中使用不必要的跳转或者重定向技术；
- 在页面中存在大量链向其他网站的链接；
- 在整个网站中存在大量重复内容的页面；
- 一个域名下存在大量不必要的子域名；
- 在页面中存在大量恶意广告、恶意代码或有害插件的网站；
- 在页面中存在主动链向已经被雅虎搜索引擎删除的网站；
- 在留言本、论坛等地方群发链接。

备注：如果你的网站确实存在以上现象，请清除这些作弊行为，然后重新向雅虎搜索引擎提交你的网站。搜索引擎会在下一个更新周期，对已经没有作弊行为的网站进行重新收录。

三、百度

以下内容来自百度网页搜索帮助 - 网站管理员 FAQ (<http://www.baidu.com/search/guide.html>)。

1、网页收录问题

■ 如何让我的网站（独立网站或者 blog）被百度收录？

百度会收录符合用户搜索体验的网站和网页。为促使百度 Spider 更快发现你的站点，你也可以向 Baidu 提交一下你的网站的入口网址。提交地址是：

http://www.baidu.com/search/url_submit.html

你只需提交首页即可，无需提交详细的内容页面。百度的网页收录机制，只和网页价值有关，与竞价排名等商业因素没有任何关联。

■ 如何让我的网页不被百度收录？

百度严格遵循搜索引擎 Robots 协议（详细内容参见 <http://www.robotstxt.org/>）。

你可以写一个 Robots 文件以限制你的网站全部网页或者部分目录下网页不被百度收录。

如果你的网站是在被百度收录之后再设置 Robots 文件，则 Robots 文件通常在两周内生效，被文件限制的内容，将从百度搜索结果中移除。

如果你的拒绝被收录需求非常急迫，也可以发邮件给 webmaster@baidu.com 请求处理。

■ 为什么我的网站内一些不设链接的私密性网页，甚至是需要访问权限的网页，也会被百度收录？

百度 Spider 对网页的抓取，是通过网页与网页之间的链接实现的。

网页之间的链接类型，除了站点内部的页面链接之外，还有不同网站之间的互相链接。因此，某些网页即便通过你的网站内部链接无法访问到，但是，如果别人的网站上有指向这些页面的链接，那么这些页面还是会被搜索引擎所收录。

百度 Spider 的访问权限，和普通用户是一样的。因此，普通用户没有权限访问的内容，Spider 也没有权限访问。之所以看上去某些访问权限限制内容被百度收录，原因有两点：

- 该内容在 Spider 访问时是没有权限限制的，但抓取之后，内容的权限发生了变化；
- 该内容有权限限制，但是由于网站安全漏洞问题，导致用户可以通过某些特殊路径直接访问。而一旦这样的路径被公布在互联网上，则 Spider 就会循着这条路径抓出受限内容。

如果你不希望这些私密性内容被百度收录，一方面可以通过 Robots 协议加以限制；另一方面，也可以联系 webmaster@baidu.com 进行解决。

■ 我的网页为什么会从百度搜索结果中消失？

百度并不允诺所有网页都可从百度搜索到。

如果你的网页长时间无法从百度搜索到，或者突然从百度的搜索结果中消失，可能的原因有：

- 你的网站所在服务器不稳定，被百度暂时性去除。稳定之后，问题会得到解决；
- 你的网页内容有不符合国家法律和法规规定的地方；
- 你的网页不符合用户的搜索体验；
- 其他技术性问题。

以下的说法是错误的和毫无根据的:

- 参与了百度竞价排名但未续费, 会从百度搜索结果中消失;
- 参与了其他搜索引擎的广告项目, 会从百度搜索结果中消失;
- 与百度旗下网站产生了竞争, 会从百度搜索结果中消失;
- 从百度获得的流量太大, 会从百度搜索结果中消失。

■ 什么样的网页会被百度认为是没有价值而不被百度收录或者从现有搜索结果中消失?

百度只收录百度认为有价值的网页。任何网页在搜索结果中的去留变化, 都是机器算法计算和调整的结果。下述类型的网页, 百度明确不会欢迎:

- 网页做了很多针对搜索引擎而非用户的处理, 使得用户从搜索结果中看到的内容与页面实际内容完全不同, 或者使得网页在搜索结果中获得了不恰当的排名, 从而导致用户产生受欺骗感觉。如果你的网站中有较多这种页面, 那么这可能会使你的整个网站的页面收录和排序受到影响;
- 网页是复制自互联网上的高度重复性的内容;
- 网页中有不符合中国法律和法规的内容。

■ 如果我的网站因为作弊行为而从百度搜索结果中消失, 是否还有被重新收录可能?

凡是作出完全修正的网站, 是有机会被百度重新收录的。百度会定期对被处理站点进行自动评估, 并对符合条件者重新收录。

需要说明的是, 百度技术和产品部门只对用户搜索体验负责。以下的说法都是错误的和毫无根据的:

- 我成为百度的广告客户或者联盟网站, 就可以重新被收录;
- 我给百度若干钞票, 就可以重新被收录;
- 我认识百度的某某人, 就可以重新被收录。

■ 我的网站更新了, 可是百度收录的内容还没更新怎么办?

百度会定期自动更新所有网页 (包括去掉死链接, 更新域名变化, 更新内容变化)。因此请耐心等待一段时间, 你的网站上的变化就会被百度察觉并修正。

2、网页排序问题

■ 我的网站首页被收录了, 但搜索网站名称却排不到第一个, 怎么办?

排序算法非常复杂。百度的目标, 即在于通过算法改进, 让用户以最小的成本, 搜索到所需要的信息。这个过程中还是会有各种各样不尽如人意的地方。百度会非常欢迎你把你遇到的困惑和问题, 反馈给百度。百度的工程师, 对每一个问题都会有细致的跟踪和分析, 以期将之最终解决。百度搜索结果页下方的搜索框右侧, 有“与百度对话”链接, 你可以在那里提交你的问题, 以协助百度改进。

百度一直在改进搜索算法, 以使得搜索结果更加符合用户的搜索需求。

■ 搜索某关键词, 我的网页在百度搜索结果的排序短期内变化剧烈, 这正常吗?

通常情况下, 这是正常的变化。一般来说, 有三类原因导致排序发生变化:

- 特定关键词所涉及的你的网页发生了变化;
- 特定关键词所涉及的其他网页发生了变化;
- 百度的排序算法发生了变化。

■ **搜索某关键词，我的网页在百度的排序位置，和在其他搜索引擎的排序位置，差异非常大，这正常吗？**

通常情况下，这是正常的现象。因为不同搜索引擎的算法，都是不同的。

■ **我请一些“SEO”来为我的网站或者网页做优化，会有什么后果？**

合理的搜索引擎优化，参见百度的“给网站管理员的建站指南”。

外界很多打着 SEO 旗号的公司或者个人，也许能为你的网站带来短期的排序收益，但是，这会使你将面临更大损失的风险。在你把网站资源交托给别人之后，很多 SEO 甚至会利用你的资源进行他们个人的运营项目，最终导致你的利益受损。

不要因为 SEO 们以下的说法，而冒险将自己的网站托付给他们随意处置：

- 我和百度的人很熟，想怎么干就怎么干，没风险；
- 我是搜索引擎专家，对百度的算法一清二楚，玩玩火也不要紧；
- 我把 xxx、yyy、zzz 这些关键词都搞到第一了，所以我是牛人啊。

你也可以向百度投诉搜索中遇到的垃圾网站或者网页，帮助百度保持搜索结果的质量。

3、商业客户相关的问题

■ **我是百度的竞价排名客户，如果我不续费，百度是否会因此对我进行惩罚？**

这是绝对不可能的。

百度的网页搜索策略的唯一标准，在于用户的搜索体验。竞价排名和网页搜索自然排名，是完全独立的两个技术服务系统。一个网站是否是百度竞价排名客户，对于网页搜索自然排序无任何影响。

如果你收到任何类似威胁的说辞，请你直接发邮件至 ask@baidu.com 举报。

■ **我的网站因为作弊而从百度消失了，是否可以通过成为百度竞价排名客户、广告客户或者联盟站点的方式重新被百度收录？**

不可以。百度对网站的收录，唯一标准是用户搜索体验。被惩罚网站重新被百度收录的说明，见网页收录问题中的叙述。

■ **我的网站加入百度竞价排名、百度联盟，或者成为百度的广告客户，是否能在网页的收录和排序上获得特别的照顾？**

不可能。

4、给网站管理员的建站建议

只有当搜索引擎、网站管理员、互联网用户之间，能有一种默契的利益均衡，这个行业才会顺畅发展。竭泽而渔式的网站建设，只会使你与用户、与搜索引擎越来越远。搜索引擎与网站管理员之间，宜和谐发展，共同拥抱美好的愿景。

以下是百度给出的一些建站建议：

■ 站点结构宜简洁明晰

不要让你的用户一进你的站点就因为纷繁芜杂而不知所措。从某种意义上来说，百度的 Spider 也是一个相对特殊的访客而已。每一个子域名，每一个目录，都最好有明确的内容区隔，避免不同子域名或者目录对相同内容的互相串用。

■ 创造属于你自己的独特内容

百度更喜欢独特的原创内容。所以，如果你的站点内容只是从各处采集复制而成，很可能不会被百度收录。

■ 保持经常的更新

经常的更新，蜘蛛程序就会经常的光顾；而长期不更新的网站，蜘蛛到访会日趋减少。

■ 谨慎设置你的友情链接

如果网站上的友情链接，多是指向一些垃圾站点，那么你的站点可能会受到一些负面影响。参与各类以 SEO 为目的的自助链接活动，很可能“过犹不及”。

■ 把自己的网站做成常青树

如果没有搜索引擎，你的网站仍然访客盈门，那么你的网站就属于“常青树”了。面向用户做网站，而不要面向搜索引擎做网站，这是成为常青树网站的真谛。

四、搜狗

以下内容来自搜狗帮助中心网站管理员指南(<http://www.sogou.com/docs/help/help.htm>)。

1、如何使你的站点被搜狗收录

- 请给每个网页加上与正文相关的标题。如果是网站首页，则标题建议使用站点名称或者站点代表的公司/机构名称；其余的内容页面，标题建议做成与正文内容的提炼和概括。这可以让你的潜在用户通过搜索引擎结果中的标题快速访问到你的页面；
- 请不要在标题中堆积与正文无关的关键词；
- 请确保每个页面都可以通过至少一个文本链接到达。搜狗无法识别 Flash 中的链接，这些 Flash 中的链接所指向的网页，搜狗就无法收录了；
- 页面间的链接，尽量使用平实的超链，而不是重定向跳转。使用自动跳转的页面，可能会被搜狗丢弃；
- 尽量少使用 Frame 和 Iframe 框架结构，搜狗有可能会丢弃掉 Frame 的内容页面；
- 如果是动态网页，请控制一下参数的数量和 URL 的长度。搜狗更偏好收录静态网页；
- 在同一个页面上，不要有过多的链接。在那些站点地图类型的页面上，请把重要的内容给出链接，而不是所有细枝末节。链接太多，也可能导致无法被搜狗收录。

对于非商业性网站，搜狐/搜狗（Sogou）搜索引擎提供免费登录服务。只需提交网站的网址，搜狐/搜狗（Sogou）搜索引擎会在一个工作日内自动收录网页。

2、什么样的站点会受到搜狗欢迎

- 站点应该是面向用户的，而不是面向搜索引擎的。一个受到用户欢迎的站点，最终也会受到搜索引擎欢迎；反过来，如果你的站点做了很多针对搜狗的优化，但却给用户带来大量负面体验，那么，你的站点最终可能还是会受到搜狗的冷落；
- 搜狗更喜欢有独特内容的网页，而不是简单抄袭和重复互联网上已有内容的网页。对于已经被重复了千百遍的内容，搜狗可能会不予收录；
- 请谨慎使用你的站点链接。与一些垃圾站点做友情链接，很可能会使你的网站受到负面影响。因此，当有人很热情的要求你为他的站点提供友情链接时，请审视以下两点：
 - A. 对方的站点在他那个领域是否是高质量的？网站管理员间所谓的流量以及排名，很多都是用欺骗手段获取的，无法保持长久。
 - B. 对方要求的链接名称是否和对方网站地位相称？用一个涵盖范围广泛的关键词来做一个内容非常局限的网站的链接名称，很可能会使你的网站受到负面影响。
- 经常保持内容更新。经常有新内容产生的站点，搜狗是会注意到，并且大力欢迎，而且会频繁造访。

3、搜狗对作弊网站的定义

作弊网站是指在网页设计中，为了提升网页在搜索引擎中的排序，设计者采用的一系列欺骗搜索引擎的做法，主要表现是普通用户看到的页面与搜索引擎抓取到的内容不一致。对搜狗来说，以下行为都可能被认为是作弊：

- 隐藏文本或隐藏链接。在网页中加入搜索引擎可识别但用户看不见的文本内容或链接。包括使用与背景色相同或十分接近的文本或链接文字、超小号文字、文字隐藏层、页面底部下出现的多余内容、滥用图片 ALT 等；
- 滥用关键词。在网页源代码中大量堆积、重复某些关键词，或者加入与网站内容毫不相关的热门关键词；
- 恶意互换链接。在网页中互相建立大量指向对方网站的链接；
- 欺骗性重定向、欺骗性更换网页内容。故意制造大量链接指向某一网址或者使用跳转域名欺骗用户进入与搜狗描述不相符的网站；
- 建立大量镜像网站。复制网站或网页的内容并分配以不同域名和服务器，欺骗搜索引擎对同一页面内容进行多次索引；
- 页面包含恶意广告、恶意代码或恶意插件；
- 页面有链接指向作弊网站的，可能会被视为作弊网站。

4、搜狗对作弊网站的处罚

搜狗反作弊系统会实时监控作弊网站，并根据作弊等级对作弊网站进行相应的处罚。

- **常见处罚** 降低网站在搜狗的整体排名；
- **一般处罚** 仅索引正常网页，其他网页不予收录；
- **严厉处罚** 包含非法内容及恶意程序、用户集中投诉的网站，搜狗将拒绝收录整个网站。

5、如何举报作弊网站

除了搜狗反作弊系统, 搜狗也非常欢迎广大用户积极举报作弊网站或网页, 举报信箱是: anti_spam@sohu.com。如果你发现有以上作弊行为或者包含非法内容的网站, 请及时与搜狗联系。

6、如何申请解除处罚

被搜狗视为有作弊行为的网站, 在清除作弊内容后, 可以写信申请解除处罚: anti_spam@sohu.com, 经过搜狗监控人员的审查, 如果确认网站已经不存在作弊行为, 搜狗将恢复对该网站的正常收录。

解除处罚的网站一旦被发现有再次作弊, 将会受到更为严厉的处罚。

7、关于搜狗 spider 的 FAQ

■ 什么是 Sogou spider?

Sogou spider 是搜狗搜索引擎的一个自动程序。它的作用是访问互联网上的网页, 存储到本地数据库中, 并从中发现新的链接, 走访互联网, 使用户能在搜狗搜索引擎中搜索到你网站的网页。

■ Sogou spider 访问我的网站过快怎么办?

Sogou spider 对于同一个 IP 地址的服务器主机, 只建立一个连接, 抓取间隔速度控制在几秒一次。一个网页被收录后, 最快也要过几天以后才会去更新。如果持续不断地抓取你的网站, 请注意你的网站上的网页是否每次访问都产生新的链接。如果你认为 Sogou spider 对于你的网站抓取过快, 请与搜狗联系, 最好能提供访问日志中 Sogou spider 访问的部分。联系邮箱是 sogou_spider@sohu-rd.com。

■ Sogou spider 喜欢收录什么样的页面?

- 内容优良而独特的页面, 如果你的页面内容和互联网上已存在的其他页面有高度的相似性, 可能不会被 Sogou spider 收录;
- 链接层次较浅的页面, 过深的链接层次, 尤其是动态网页的链接, 会被丢弃而不收录;
- 如果是动态网页, 请控制一下参数的数量和 URL 的长度。搜狗更偏好收录静态网页;
- 重定向次数越多的页面, 越有可能被 Sogou spider 丢弃。

■ Sogou spider 如何反应我网站上页面的更新?

Sogou spider 根据网页的重要性和历史变化情况来动态调整更新时间, 更新已经抓取过的页面。

■ 如何让 Sogou spider 不抓我的网站?

Sogou spider 支持 robots 协议, 你可以在网站的根目录放置 robots.txt。robots.txt 的规则请参阅 <http://www.robotstxt.org/> 但是新更新的 robots.txt 可能得过几个星期才能体现出效果来。

需要注意的是，被你禁止 Sogou spider 收录的网页将不能在搜狗搜索引擎上检索到。

■ Sogou spider 会封锁那些网站？

- 作弊网站；
- 内容低劣的网站；
- 违反国家政策和法律的网站。

■ 如何让 Sogou spider 解除对我网站的封锁？

在你的网站的问题改正之后，请与搜狗联系，联系邮箱是：sogou_spider@sohu-rd.com。

■ 为何在搜狗上检索不到我的网站

首先请确认你的网站没有因作弊等问题而被搜狗封锁，确认无误可以到：
<http://db.sohu.com/regurl/regform.asp?step=regform&class=>
免费登录。

第三节 白帽 SEO 和黑帽 SEO

在 2004 年 12 月 13 日美国芝加哥举办的搜索引擎战略大会上，搜索引擎优化专家 Andrew Goodman 发表了题目为“Search Engine Showdown: Black Hats vs. White Hats at SES”（搜索引擎摊牌：黑帽 VS 白帽）的演讲，第一次正式提出了“黑帽 SEO”和“白帽 SEO”的说法。

在搜索引擎优化业界，人们把使用作弊手段的称为黑帽（blackhat），使用正当手段优化网站的称为白帽（whitehat）。笼统地说，所有使用作弊手段或可疑手段的，都可以称为黑帽 SEO。比如说垃圾链接、隐藏网页、桥页、关键词堆砌等等。

不过，正如人不能简单地以非好即坏评判一样，这样的非黑即白、非白即黑事实上也不能准确概括 SEO 市场的实际情况，毕竟中间存在着很多过渡色。

近一两年，最典型的 blackhat SEO 是，用程序从其他分类目录或搜索引擎抓取大量搜索结果做成网页，然后在这些网页上放上 Google AdSense。这些网页的数目不是几百几千，而是几万几十万。所以即使大部分网页排名都不高，但是因为网页数目巨大，还是会有用户进入网站，并点击 Google AdSense 广告。

姑且不谈黑帽 SEO 是否道德，两者更多的是一种企业模式和生活方式的选择。

白帽 SEO 可能会觉得用黑帽技术虽然能得到排名和利润，但是过不了多久，网站就会被封掉，你又得重新去做另外一个网站。为什么不用白帽手段踏踏实实地做一个健康的、对人们真正有用的网站呢？白帽网站不用担心会被搜索引擎封掉，你也可以骄傲地说：这个网站是我的。十年二十年以后，这个网站还在给你带来利润，何乐而不为呢？

但站在黑帽 SEO 的角度，他们也有他们的道理。因为大部分黑帽 SEO 都是使用程序，他们建立一个几万几十万页的网页，不费吹灰之力，只要放他们的蜘蛛出去抓取就可以了。就算过几个月他们的网站被惩罚，他可能已经赚了几千几万美金了。对他来说，投资报酬率还是相当高的。

一个白帽 SEO 所关注的更多的是长远利益，也就是两年三年，甚至十年八年以后的利益。应该说如果你坚持不使用作弊手段，坚持十年，不出意外的话，你的网站应该可以得到好的流量。有了流量，就有了盈利（如果你有了流量，却不知道怎么把它转化成盈利，那么

你应该检讨当初建站时候的计划)。

当十年以后，你的网站有稳定的盈利的时候，你对搜索引擎的依靠度就变得小多了。你也不必每天花那么长时间在你的网站上，你的网站会自动带来源源不断的流量。

站在黑帽 SEO 的立场上，这种放长线掉大鱼的策略，即使很正确，有的人也不愿意这么做。认真建设一个网站，有的时候是一件很无聊的事。你要写内容，要做调查，要做分析流量，要分析用户浏览路径，要和用户交流沟通。

黑帽 SEO 要做的就简单多了。买个域名，甚至可以就使用免费虚拟主机，连域名都省了。程序一打开，放上 AdSense 编码，到其他留言板或博客留一些言（这些留言也有可能是程序自动生成的），然后就等着收支票了。

而且黑帽 SEO 有一个无法否认的论据是：你不能保证完全遵守搜索引擎的规则，就能在十年以后得到一个受搜索引擎重视的网站。谁知道搜索引擎在什么时候会对它的算法做一个大的改变，让成千上万白帽网站从搜索引擎里消失呢？

黑帽 SEO 赚钱的短平快方法，也有它的优势。所以说，抛开道德观念不谈，黑帽和白帽 SEO 的选择，更多的是自己生活方式和企业模式的选择。你是要花时间和精力建立一个长久健康的企业？还是轻松简单的赚一笔快钱，但是要冒随时被封杀的危险？

这里说的对生活方式的选择，是在黑帽们完全清楚自己在干什么的前提下，自主做出的选择。但令人忧心的是，很多人并不清楚黑帽技术所含有的高风险，却出于某种原因，在不停的使用这些技术。在清楚知道风险的情况下，选择黑帽 SEO 是自己的选择。但稀里糊涂就用了黑帽技术，却是梦魇的开始。

与国内 SEO 行业情况相比，国外 SEO 行业的话语主动权基本上是掌握在白帽 SEO 手里。无论是搜索引擎本身，还是网站设计公司或者客户，对 SEO 都有比较正确的认识。所以国内还需提倡白帽 SEO 技术，把整个行业的氛围健康化、成熟化。

近日，任职于 Yahoo 的 Tim Converse 给出了更详细的分类方法，是否合理或是否仍有遗漏值得商榷，不过，挺有意思。在他名为 Search engine optimization (SEO) from black to white 文中，给出了如下九种 SEO：

- **Dark Inky Black Hat SEO** 排名第一的是最不道德的一种情况，其目标不仅针对搜索引擎，也针对访问用户，比如说使用 spyware；
- **Charcoal Hat SEO** 在法律的界限以内，通过欺骗搜索引擎，尽一切可能让与搜索项无关的网页出现在搜索结果前列；
- **Dark Gray Hat SEO** 通过窃取其他网站内容来创建网页的 SEO；
- **Slate Gray Hat SEO** 通过 link farms、linkage 或网页跳转等方式实施的 SEO；
- **Gray Hat SEO** 研究搜索引擎的规范、要求、指南，在其允许的范围内尽可能地打擦边球；
- **Light Gray Hat SEO** 这类 SEO 创建大量的原创内容，但出发点首先考虑搜索引擎的排名机制、如何逃搜索引擎的 duplicate content 检测，最后才是该内容是否真正对用户有价值，也即是针对搜索引擎而不是用户；
- **Off-White Hat SEO** 这类 SEO 在保证网站具有良好架构内部页面可爬行可索引的基础上，针对重要的关键词优化页面。在链接方面，这类不通过 linkfarms 之类的方式获得链接，但使用大量来自朋友、联盟网站的链接；
- **White Hat SEO** 在网站结构合理的基础上，通过有价值的原创内容和优化手段，让网页无论对搜索引擎还是对访问用户而言都是有价值的，其优化目标为让网页出现在最相关的关键词搜索中；

- **Luminescent Pearly White Hat SEO** 在 White Hat SEO 的基础上，避免网页出现在不相关的关键词搜索中（Tim Converse 给出一个这类 SEO 的例子）。

你属于哪个类型的 SEO 呢？

第四节 搜索引擎优化的发展

SEO 是随着搜索引擎的普及而出现，并伴随搜索引擎的发展而发展的。关于搜索引擎和搜索引擎优化 SEO 最初是怎么诞生的有很多种说法。有一个非常有趣的事实是，最早的一批 SEO 甚至可以追溯到 Yahoo 出现之前，我们可以把 Yahoo 的传世人 David Filo 和杨致远看作是第一批 SEO。

一、SEO 的发展历史

不知在什么时候，人们开始习惯在互联网上使用搜索引擎找所需的东西。可能是 Yahoo 出现时，或者更早些，SEO 也因此得了飞速的发展。

1、20 世纪九十年代

Yahoo 是最早使用 SEO 技术的公司之一，那个时候，他们还没有估计到搜索产业的巨大增长潜力。世纪之交，Yahoo 的联合创始人大卫·费罗和杨致远为了增加站点的曝光度，想办法让他们的网站能被互联网上其他搜索引擎发现。通过对站点结构的优化和一些“狡猾”的代码编写技巧，他们的网站变得很容易被那些新用户检索到。他们当时也没有考虑将这种服务的商业化操作，因为当时没有人清楚这样做到底是否符合商业道德，也没有可以参考的简单标准。

随着搜索引擎开始对网上的信息进行归类，企业主开始意识到站点被显示在搜索引擎上的价值了，最显而易见的是可以增加站点的访问量。他们开始不断向搜索引擎递交他们站点的网址，为了适应搜索引擎蜘蛛（Robots）而对他们的站点做出改变。前期的努力就集中在站点的递交，不久以后，自动向搜索引擎递交站点的程序诞生了，同样，臭名昭著的搜索引擎垃圾（Spam）也开始存在了。

- 1994 年，Yahoo、Lycos 等分类目录型搜索引擎相继诞生，搜索引擎表现出网络营销价值，搜索引擎营销思想开始出现；
- 1995 年，自动提交到搜索引擎的软件诞生，搜索引擎营销“智能化”，此后不久许多搜索引擎开始拒绝自动登录软件提交的信息；
- 1995-1996 年，基于网页 HTML 代码中 META 标签检索的搜索引擎技术诞生。利用 META 标签改善在搜索引擎中排名的技术很快成为搜索引擎营销的重要内容，这就是搜索引擎优化方法的萌芽；
- 1997 年，搜索引擎优化与排名自动检测软件问世，网络营销人员可以据此制定针对性的搜索引擎营销策略；
- 1998 年，“搜索引擎算法”开始关注网站外部链接，“网站链接广度”（link popularity）概念诞生。

2、二十一世纪

SEO 专家开始出现还是近 5 年的事情。因为在较早的 2001 年，精力充沛的网站管理员们迅速认识到他们可以通过反复向搜索引擎递交来征服搜索结果。然后，不幸的是，随着互联网行业的发展，对于那些不管公正不公正，道德不道德，也不管任何代价，试图将用户引向他们客户站点的 SEO 公司来说，搜索引擎已经变得很危险。一些常用伎俩比如关键词堆积、门页、隐形页面或文字（比如将网页字体和背景都设置为白色），已经超出了搜索引擎可以容忍的范围。最终导致的后果是，搜索引擎将这些乱七八糟的手法列为作弊。这是一个好消息，他迫使那些有道德的 SEO 公司开始使用正确的手法去帮助他们客户的站点在搜索引擎上获得好的排名。

Google、Msn、Yahoo 三大搜索引擎，也开始意识到搜索引擎优化这个行业的存在，可以维持和改善搜索结果的有效性，他们需要敞开怀抱接受这个行业。搜索引擎公司也开始跟一些成功的、有道德的 SEO 公司合作，建立了一系列公正的合法的优化标准。确保信息对于使用者来说是相关的和有效的非常重要，这样做对于那些在自己站点提供相关内容的人也不存在任何偏见。

- 2000 年，出现按点击付费（Pay-per-click）的搜索引擎关键词广告模式，搜索引擎广告诞生；
- 2001 年，搜狐等部分中文分类目录开始收费登录，网站登录每年要交纳数百元到数千元不等的费用，付费搜索引擎营销开始走向主流；
- 2002 年后半年，在网络广告市场最低潮中，搜索引擎关键词广告市场增长强劲，占 2002 年网络广告市场的 15%，搜索引擎带动整个网络经济复苏；
- 2003 年，出现基于内容定位的搜索引擎广告（Google AdSense）；
- 2004 年，搜索引擎全面引领互联网经济潮流，搜索引擎营销的价值被企业普遍接受。

3、目前 SEO 现状

今天，搜索引擎的工作方式和排名规则已经有了很大的变化。在正确的搜索优化服务的帮助下，网站管理员们可以有一个公平的机会去获得较高的搜索排名。SEO 是一个非常专业的服务，它不仅需要一定的技术技巧，还要有一定的商业营销知识。只有很好把这二种能力结合的人才可以正确地运用 SEO 技术来获得较高的搜索引擎排名。

二、SEO 在中国

总体来说，SEO 在中国还属于新生事物。

1、中文网站 SEO 变迁

- 2001 年之前：免费搜索引擎营销阶段，以免费分类目录登录为主要方式；
- 2001 年中期-2003 年中期：因网络经济环境、搜索技术、收费等原因带来搜索引擎营销市场进入调整期；
- 2003 年后期开始：各种形式的搜索引擎广告快速发展，同时基于自然检索结果的搜索引擎优化开始受到重视；
- 2004 年-2005 年：搜索引擎广告进入快速增长期，并且营销效果逐步为企业所认可；

新的搜索引擎不断出现；

- 2005 年开始：垃圾 SEO 泛滥，严重破坏搜索引擎营销市场秩序，引起主流搜索引擎大规模清理。

2、中文网站 SEO 现状

A. SEO 在国外已经是一门相当正规、成熟的行业，有大批的专业公司、专业技术人员在为广大客户网站提供全方位的搜索引擎优化服务。同时，搜索引擎优化技术也得到了广大网站管理者甚至搜索引擎在内多方面的认同。但据说在国内的发展情况并不理想，那么 SEO 在国内目前的状况如何？

相对于 03、04 年的混乱来说，经过了优胜劣汰的洗礼，国内 SEO 市场正逐渐向正规化、专业化发展，但是仍然存在着规模小、实际操作性不强等情况。

规模方面：从事 SEO 行业的仍然以个人工作室和小公司为主体，同时，部分国外知名企业也开始进驻中国，竞争更为激烈。

技术方面：由于国内缺乏学术交流，从业人员缺乏对技术的探索与追求等因素，SEO 技术在国内来说仍然处于较低水平。不过随着国外企业的到来，必将推动技术方面的升级。

市场方面：由于算法的改进，原有的所谓优化方法已经行不通，国外正规企业的加入迫使那部分对 SEO 理解比较粗浅者放弃或转向正当方式。行业得到了一定程度的净化。

B. Google 曾对搜索引擎规则进行调整，据说很多 SEO 公司做过的网站从原先较好的排名消失得无影无踪，那么，这次调整对中国的 SEO 行业有没有什么特殊的意义，都有哪些改观（主要是由于中国一些 SEO 的作弊行为，对留言版进行海量信息和非法链接）？

05 年以来，Google 在算法方面有两个比较大的改进。

第一，调整反向链接的权重，加强了反向链接相关度的权重。

第二，“沙盒效应”。“沙盒效应”主要是针对新网站，以域名正常运作时间，网站内容丰富与否，更新频率等作为一个决定最终排名的权重，而且占着极大的份量。

这使得原来所谓的优化方法已经不能奏效，也加长了作业的时间。之前那种靠垃圾链接及堆叠关键词方式作业的 SEO 难以为继。

C. 搜索引擎优化的方法有许多种。国内 SEO 公司或者团队主要采用哪些方法，这些方法和国外的有什么不同吗？

搜索引擎优化并没有什么像外界说的那样有各种各样的方法，百变不离其宗，都是在围绕搜索引擎排名算法而进行。但是国外对于搜索引擎算法变动比较敏感，看法也比较深刻，从而能及时做出相应对策。而国内只能跟着国外的步伐艰难前进。这取决于学术环境、学术气氛及业内人士对知识的追求程度等因素。

D. 一提及 SEO，在部分人印象当中都说是一种作弊行为或者是违规行为，是什么造成大家对 SEO 有这样的印象呢？

不可否认这种现象是存在的，而且可能还要存在很长一段时间。一方面，企业意识到搜索引擎优化的重要性，但又片面追求低投入、短时间、高回报；另一方面，部分从事 SEO 的企业或个人为了迎合客户需求，不惜作出种种承诺、降低投入来控制成本。这很大程度上取决于从业者的立场、原则，更重要的是他们对 SEO 的片面理解及对搜索引擎优化认识的严重缺乏。许多所谓的 SEO 认为，做优化不外乎关键词、外部链接，从而陷入作弊误区不

能自拔。

E. 我们前面提到的 Google 搜索规则改变，是不是暗示着国内 SEO 发展畸形呢？

搜索引擎算法完善的过程是一个不断改进的过程。当搜索引擎开发者发现自身算法缺陷时就会进行改进，发现者可能是搜索引擎的开发者本身，也可能是使用者。两者是一种相互监督、相互促进的关系。

Google 排名规则的改变并不可以说是针对中国 SEO 业界，更准确地说是国外从事 SEO 的部分人员利用了算法的缺陷使用了某种非法手段获得更好的排名，这种手法到了中国就变得更加普遍，从而迫使 Google 进行算法改进。

目前国内存在的一些问题与矛盾并不能说国内 SEO 发展出现畸形了，而应该说是不断进化。如果哪一天，中国从事 SEO 的人士能让 Google 因其行为而改进算法，那说明我们已经在一定程度上超越国外了。

F. 从各方面看来，SEO 在中国的发展并不顺利，并且造成了一定的负面影响，那接下来我们的路又该怎么走？SEO 如何朝一个健康的方向发展？

大概是从 02 年开始，国内陆陆续续有人涉足 SEO 这一领域。经历过市场洗礼、优胜劣汰，国内 SEO 行业正慢慢朝着正规化、规模化方向前进。

SEO 行业能否健康发展，在很大程度上取决于从业者自身的技术水平与职业素质，也与广大关心及参与搜索引擎营销者息息相关。首先，从业者应该端正作风、坚持立场、坚持原则、坚持以用户为中心。这也是一个 SEO 从业者所必需的素质；其次，在学术交流方面，我们应该更积极地参与各种形式的交流与研讨，慢慢形成一种相互交流、相互促进的学术环境；最后，网络营销者应加深对搜索引擎优化行业的认知与理解，搜索引擎优化是一个长期的过程也是一个循序渐进的过程。

三、SEO 的发展方向

1、白帽 SEO 占据话语主流

与国内 SEO 行业情况相比，国外 SEO 行业的话语主动权基本上掌握在白帽 SEO 手里。无论是搜索引擎本身，还是网站设计公司或者客户，对 SEO 都有比较正确的认识。

白帽 SEO 们在私底下并不一定就那么排斥黑帽 SEO 和作弊，甚至可能也在偷偷使用黑帽手段。但基本上，在台面上，大家还是都在提倡白帽 SEO 技术，所以整个行业的氛围比较健康。

2、国外 SEO 大多关注高端市场

主要 SEO 人物和公司很少做低端项目，比较有名的 SEO 公司都会提供以咨询为主的服务，在需要的时候，也会提供整套服务，包括咨询、写作、关键词研究、页面的修改、建立链接等。

3、SEO 服务价格

国外和国内一样，有很多网站都声称自己提供 SEO 服务，价格相差也很大，不过有号召力、可信的 SEO 公司服务价格相差不是很多。一些小的不太正规的公司也在宣传几百块钱做一个关键词之类的服务，但真正可信赖的 SEO 公司都不会做几百几千块的项目。

4、业务多样化

几年前 Google 大规模数据更新使不少 SEO 服务公司陷入困境,这些公司都在分散风险,建立多种多样的收入来源。大部分 SEO 公司都在积极开拓不同的业务,建自己的电子商务网站,这是一个很重要的战略方向的改变,在这方面投入的精力往往在百分之三四十以上。

5、SEO 服务专门化

有一部分 SEO 专家和公司专注于做 SEO 的某一方面,比如专做关键词研究、专门做链接建立、专门写文案等。

6、建立信誉和地位

做得有声有色的 SEO 公司,绝大部分都有这样一个人:他(她)通过论坛或博客等方式,在业界建立一个标志性地位。客户在寻找 SEO 服务公司时,提到这个人,行业内都知道有这么一号人物,这对于 SEO 服务公司来说是非常有好处的,虽然这位人物并不一定直接做具体的项目。

第五节 关于搜索引擎优化的一些观点

一、搜索引擎优化合法吗?

SEO 并不神秘,写好页面标题、把动态网址改成静态的,这些就已经是在使用一些 SEO 原理了。在 SEO 这个概念出来以前,大家其实早就已经使用了其中的理念,只是大家没有意识到它而已。

事实上,SEO 跟网站早就已经是你中有我,我中有你了。那些称 SEO 非法的公司,自己都已经使用了部分 SEO 技术了。而且,几乎所有的搜索引擎都在其自身的网站上公布了什么样的网站是它们容易理解和受欢迎的,Google 左侧排名还针对这个行业特别推出了相关建议。

在互联网最为发达的美国,搜索引擎优化已经成为企业网络营销战略的基本组成部分,很多企业使用了搜索引擎优化服务。在日本,像索尼和松下这样的大型企业也使用搜索引擎优化服务来提升企业网站在搜索引擎的表现,从而推广其线上业务。中国目前也有非常多的企业实施了 SEO,而且越来越多的网络公司开始提供商业化的服务。各种主流的 IT 站点也开始纷纷开有专栏,介绍搜索引擎优化的相关技术。

在这样的情况下,说 SEO 违法,只是一种基于不理解的可笑偏执而已。

二、会不会和搜索引擎对着干?

不少人认为搜索引擎与 SEO 之间是相互对立的关系,认为 SEO 的目的便是意图破坏搜索结果的公正与有效,这种说法在一定范围内很有市场。

不过,对此要客观地认识。首先,不要将 SEO 与 Spam 混为一谈,对于那些通过抄袭别人内容、关键词填充甚至隐形文本的 Spammer,搜索引擎肯定是要打击的;其次,从根本上说,SEO 与搜索引擎有一个共同的目标,那便是如何有效地组织内容,帮助用户以最有效的方式找到所需的内容,从这个角度,搜索引擎与 SEO 更像是一种相互储存相互补充

的共生关系。

三、搜索引擎是否会人工干预？

几乎可以肯定的是，搜索引擎会采取人工的方式，降低某些网站的权重甚至删除某些网站。但多数的人工干预，是基于对象网站有明显的 Spam 行为，而这些行为暂时还没有归入搜索引擎的判断序列。

四、我的网站有大量的优质原创内容，所以我不需要 SEO 了。

任何时候，内容都是 SEO 的基础，但是，内容只是 SEO 成功与否的必要条件而非充分条件，也即是说，内容本身并不能保证网站能够在搜索引擎占据有利的地位。信息爆炸的时代，酒好也怕巷子深的。

五、SEO 没有技术含量或 SEO 需要很高深的技术

SEO 从业者往往是“半路出家”：很多从广告或策划公司转向 SEO 行业的人，往往笔头功夫尚可，做方案能够自圆其说，但技术上的缺陷是其力图避开的；技术人员出身的 SEO 业者则往往沉溺于技术而在整体把握上有所匮乏，但他也要强调自己的优势。因此，无论是说“SEO 没有技术含量”或是说“SEO 需要很高深的技术”，究其根源，都是在寻找己之长以及人之短。

六、大网站那样做了，我也向他们学习

很多人喜欢研究某些大网站如 sina（或网易、sohu、阿里巴巴）在优化方面采取的“技术”，比如说关键词填充、隐形文本等，然后照模学样，应用到自己的网站上。其理由往往很充分：既然那些权威网站可以这么做，就说明那样没问题，因此，我也可以。

但客观说来，这是很危险的。首先，对那些网站来说，由于自身多年积累的高知名度，来自搜索引擎的访问量在其总访问量中的比例要远远低于一般的网站，因此，即便 SEO 效果受损对其伤害也要远远小于一般的网站；其次，世上并没有绝对的公平，不管承认与否，很多事确属他们可以做而一般的网站不能做的；第三，他们有庞大的信息量为基础，那么做即便存在负面影响，也是足以抵销甚至忽略的。因此，最起码的，等做到像他们那么庞大的信息量、每天那么多的更新量时，再考虑与他们攀比吧。

七、SEO 企业或 SEO 工具的保证排名宣传

我们都喜欢购买放心的产品或服务，“保证”往往是让我们放松警惕的有效手段，很多情况下，敢于“保证”往往蕴含着背后强大的技术能力或丰富的经验，让人刮目相看。但是，也必须清醒地看到，所谓的“保证”也许只是一种销售策略。

在很多时候，以一定的技术为保障，经验丰富的 SEO 人员能够让相应的网站或网页获得预期的排名，但是，没人能真正保证在不可知的未来也能做到这一点。因为，谁也无法确知搜索引擎的算法在下一次更新时将做如何调整。SEO 在某种程度上与股票市场类似，正

如没人能准确掌握股票未来的走向那样，也没人确知搜索引擎下一步将如何变化。从这个角度看，只有搜索引擎自身才具有决定网站排名的权力。

如果你要的是“保证排名”的效果，也许你应考虑竞价排名，而不是这些 SEO 企业或 SEO 工具。

另一方面，对于这些信誓旦旦做保证的企业，应更谨慎一些：他们是否会采取某些搜索引擎深恶痛绝的 spam 手段？从长期来看是否会对你的网站造成很难挽回的伤害？他们保证排名位置的关键词是什么？事实上，如果选择一个没人使用的关键词，通过添加相应的内容页面、进行简单的优化、建立必要的链接等，也许即便傻瓜也能做到排名第一的位置，但那样有意义么？

同样的保证，也许有的可以达到预期的效果，有的则让你的网站彻底从搜索引擎中消失。

八、稳定的排名

时常有人困惑于“为什么我的排名掉下来了”之类的问题，必须明确的是，没有永远不变的排名。Internet 上一直都有新的网页在产生，一直都有旧网页被删除或更新，搜索引擎索引和算法也在不断更新，排名变化是绝对的。

九、SEO 的目标就是获得排名第一的位置或尽可能多的访问量

这就有点买椟还珠的意思了。从根本上说，无论 SEO 还是网站优化，都只是手段，是为了网站的 ROI（Return On Investment），无论让网站获得高排名还是高访问量都是为了一个简单的目的：为网站带来更多的潜在消费者。

十、竞争对手无法破坏我的排名

人人都希望这是真的，尤其在道德感缺失的 SEO 行业，但很多时候，事实并不那么乐观。很多例子表明，恶意的竞争对手还是能够通过某些方式伤害你的网站的。

第四章 SEO 相关的搜索引擎技术

第一节 搜索引擎机器人

一、什么是搜索引擎机器人？

搜索引擎机器人也称网络蜘蛛 (Web Spider)，这是一个很形象的名字。把互联网比喻成一个蜘蛛网，那么 Spider 就是在网上爬来爬去的蜘蛛。网络蜘蛛通过网页的链接地址来寻找网页，从网站某一个页面（通常是首页）开始，读取网页的内容，找到在网页中的其它链接地址，然后通过这些链接地址寻找下一个网页，这样一直循环下去，直到把这个网站所有的网页都抓取完为止。如果把整个互联网当成一个网站，那么网络蜘蛛就可以用这个原理把互联网上所有的网页都抓取下来。

在抓取网页的时候，网络蜘蛛一般有两种策略：广度优先和深度优先（如下图）。

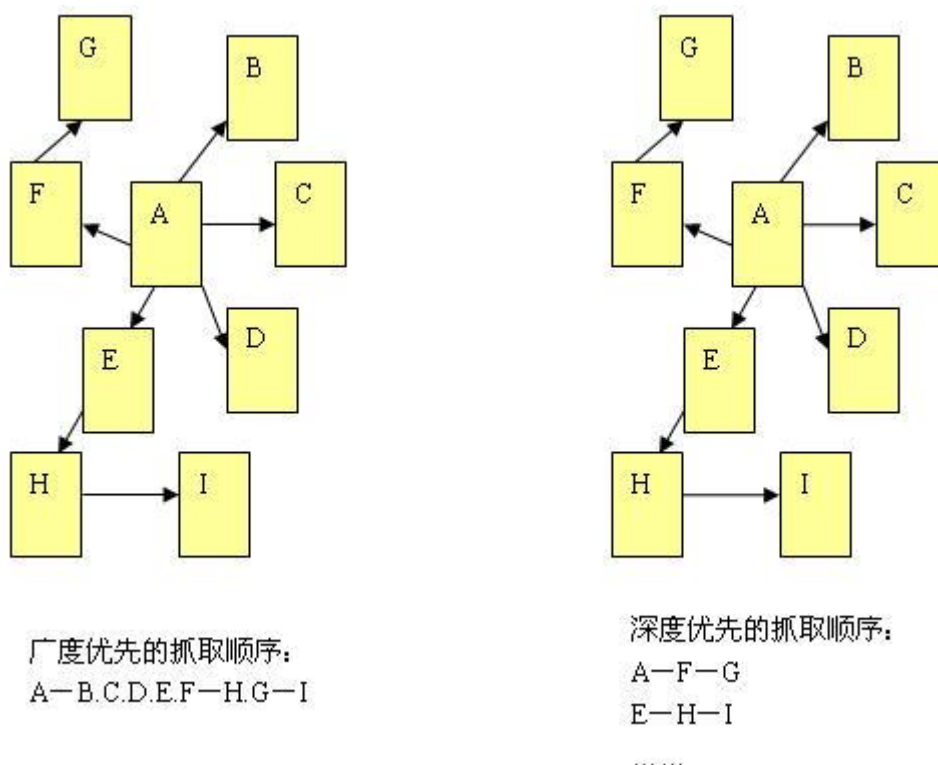


图 4-1-1 广度优先与深度优先

广度优先 是指网络蜘蛛会先抓取起始网页中链接的所有网页，然后再选择其中的一个链接网页，继续抓取在此网页中链接的所有网页。这是最常用的方式，因为这个方法可以让网络蜘蛛并行处理，提高其抓取速度。

深度优先 是指网络蜘蛛会从起始页开始，一个链接一个链接跟踪下去，处理完这条线路之后再转入下一个起始页，继续跟踪链接。这个方法有个优点是网络蜘蛛在设计的时候比较容易。

由于不可能抓取所有的网页,有些网络蜘蛛对一些不太重要的网站,设置了访问的层数。例如上图 A 为起始网页,属于 0 层,B、C、D、E、F 属于第 1 层,G、H 属于第 2 层,I 属于第 3 层。如果网络蜘蛛设置的访问层数为 2 的话,网页 I 是不会被访问到的。这也让有些网站上一部分网页能够在搜索引擎上搜索到,另外一部分不能被搜索到。对于网站设计者来说,扁平化的网站结构设计有助于搜索引擎抓取其更多的网页。

网络蜘蛛在访问网站网页的时候,经常会遇到加密数据和网页权限的问题,有些网页是需要会员权限才能访问。当然,网站的所有者可以通过协议让网络蜘蛛不去抓取,但对于一些出售报告的网站,他们希望搜索引擎能搜索到他们的报告,但又不能完全免费的让搜索者查看,这样就需要给网络蜘蛛提供相应的用户名和密码。网络蜘蛛可以通过所给的权限对这些网页进行网页抓取,从而提供搜索。而当搜索者点击查看该网页的时候,同样需要搜索者提供相应的权限验证。

二、如何辨别搜索引擎机器人身份

搜索引擎派出他们的搜索机器人去访问、索引网站内容,网站管理员也默认、欢迎它们来访。但是由于搜索引擎派机器人来访会在一定程度上影响网站性能,因此并非所有的机器人都是无害的,有一些非法机器人伪装成主流搜索引擎漫游器对网站大量遍历,并且不遵循 robots.txt 规范,会严重拖垮网站性能而又无其它益处。因此,网站管理员需要验证每个机器人身份是否合法。

在你的服务器日志文件中,可见每次访问的路径和相应的 IP 地址,如果是机器人来访,则 user-agent 会显示 Googlebot 或 MSNBot 等搜索引擎漫游器名称,每个搜索引擎都有自己的 user-agent,但仅有这个还不足以证明这个机器人的合法性,因为很多垃圾制造者可能将他们的机器人也命名为 Googlebot,以伪装蒙混进入网站,大肆采掘内容。

目前,主流搜索引擎都建议网站管理员通过这种方式来辨别真实的机器人身份:通过 DNS 反向查询找出搜索引擎机器人 IP 地址对应的主机名称;用主机名查找 IP 地址以确认该主机名与 IP 地址匹配。

首先,使用 DNS 反向查询 (Reverse DNS Lookup),找出机器人 IP 地址对应的主机名称。主流搜索引擎的主机名称通常情况下应是这样的:

- Google : 主机名称应包含于 googlebot.com 域名中,如: crawl-66-249-66-1.googlebot.com;
- MSN : 主机名称应包含于 search.live.com 域名中,如: livebot-207-46-98-149.search.live.com;
- Yahoo: 主机名称应包含于 inktomisearch.com 域名中,如: ab1164.inktomisearch.com。

最后,做一次 DNS 查询,用主机名查找 IP 地址 (Forward DNS Lookup),以确认该主机名与 IP 地址匹配。由此证明该机器人是合法的。

现在,如果发现一个机器人将自己伪装成合法搜索引擎漫游器,你就可以通过服务器上的设置来阻止这个非法机器人。

第二节 超链分析

最近几年,许多研究者发现,WWW 上超链结构是个非常丰富和重要的资源,如果能

够充分利用的话，可以极大地提高检索结果的质量。超链分析技术，是新一代搜索引擎的关键技术。

一、基本原理

超链分析的基本原理是：在某次搜索的所有结果中，被其他网页用超链指向得越多的网页，其价值就越高，就越应该在结果排序中排到前面。

超链分析是一种引用投票机制，对于静态网页或者网站主页，它具有一定的合理性，因为这样的网页容易根据其在互联网上受到的评价产生不同的超链指向量，超链分析的结果可以反映网页的重要程度，从而给用户提供更重要、更有价值的搜索结果。

搜索引擎，并不能真正理解网页上的内容，它只能机械地匹配网页上的文字。它收集了互联网上几千万到几十亿个网页并对网页中的每一个文字（即关键词）进行索引，建立索引数据库的全文搜索引擎。当用户查找某个关键词的时候，所有在页面内容中包含了该关键词的网页都将作为搜索结果被搜出来。在经过复杂的算法进行排序后，这些结果将按照与搜索关键词的相关度高低，依次排列。

搜索引擎在查询时主要根据一个站点的内容与查询词的关联程度进行排序。对于一个站点的内容搜索引擎则是根据标题、关键词、描述、页面开始部分的内容以及这些内容本身之间的关联程度以及一个站点在整个网络上的关联程度来确定的。

使用超链分析技术，除要分析索引网页本身的文字，还要分析索引所有指向该网页的链接的 URL、AnchorText，甚至链接周围的文字。所以，有时候，即使某个网页 A 中并没有某个词，比如“软件”，但如果有别的网页 B 用链接“软件”指向这个网页 A，那么用户搜索“软件”时也能找到网页 A。而且，如果有越多网页（C、D、E、F……）用名为“软件”的链接指向这个网页 A，或者给出这个链接的源网页（B、C、D、E、F……）越优秀，那么网页 A 在用户搜索“超链分析”时也会被认为相关度越高，排序也会越靠前。

二、工作步骤

1、从互联网上抓取网页

专门用于检索信息的 Robot 程序像蜘蛛一样在网络间爬来爬去，利用能够从互联网上自动收集网页的 Spider 系统程序，自动访问互联网，并沿着任何网页中的所有 URL 爬到其它网页，重复这过程，并把爬过的所有网页收集回来。随着互联网的迅速发展，检索所有新出现的网页变得越来越困难。因此，在 Wanderer 基础上，一些编程者将传统的 Spider 程序工作原理作了些改进。其设想是，既然所有网页都可能连向其他网站的链接，那么从一个网站开始，跟踪所有网页上的所有链接，就有可能检索整个互联网。

2、建立索引数据库

由分析索引系统程序对收集回来的网页进行分析，提取相关网页信息（包括网页所在 URL、编码类型、页面内容包含的所有关键词、关键词位置、生成时间、大小、与其它网页的链接关系等），根据一定的相关度算法进行大量复杂计算，得到每一个网页针对页面文字中及超链中每一个关键词的相关度（或重要性），然后用这些相关信息建立网页索引数据库。

3、索引数据库中搜索排序

当用户输入关键词搜索后，由搜索系统程序从网页索引数据库中找到符合该关键词的所有相关网页。因为所有相关网页针对该关键词的相关度早已算好，所以只需按照现成的相关度数值排序，相关度越高，排名越靠前。最后，由页面生成系统将搜索结果的链接地址和页面内容摘要等内容组织起来返回给用户。

三、搜索效果

搜索引擎 Spider 一般要定期重新访问所有网页（各搜索引擎的周期不同，可能是几天、几周或几月，也可能对不同重要性的网页有不同的更新频率），更新网页索引数据库，以反映出网页文字的更新情况，增加新的网页信息，去除死链接，并根据网页文字和链接关系的变化重新排序。这样，网页的具体文字变化情况就会反映到用户查询的结果中。

互联网虽然只有一个，但各搜索引擎的能力和偏好不同，所以抓取的网页各不相同，排序算法也各不相同。大型搜索引擎的数据库储存了互联网上几千万至几十亿的网页索引，数据量达到几千 G 甚至几万 G。但即使最大的搜索引擎建立超过 20 亿网页的索引数据库，也占不到互联网上普通网页的 30%，不同搜索引擎之间的网页数据重叠率一般在 70% 以下。人们使用不同搜索引擎的重要原因，就是因为它们能分别搜索到不同的网页。而互联网上有更大量的网页，是搜索引擎无法抓取索引的，也是无法用搜索引擎搜索到的。

使用超链分析的搜索引擎未能收录的网页有：Spider 未能正确处理的网页性质及文件类型（如 Flash、script、JS，某些动态网页及 Frame、数据库）；没有主动登录搜索引擎而且没有指向链接的孤岛网页；Spider 访问时因为某些原因正好是死链接的网页；被认为是劣质网页而不抓；因为色情、反动、spam 等问题而不抓的非法网页；需要输入用户名、密码方可打开的网页；网站用 robots 协议拒绝搜索引擎抓取的网页；搜索引擎还未来得及抓取的新网页；go-pher、newsgroups、Telnet、np、wais 等非 http 信息的网页。

任何地址中带“?”和“&”号（及其他类似符号）的网页都会被“蜘蛛”程序挡在门外。这些网页通常由 CGL、PHP、ASP 等程序产生，技术上较先进，但不适合搜索引擎的“蜘蛛”程序。虽然目前有的大型搜索引擎（如 Google）已具备检索动态网页的能力，但相当一部分引擎还是不支持它的。而且即使是能够索引动态网页的 Google，也在多个场合中明确表示不保证检索全部的动态网页。

四、超链分析技术的应用

超链分析技术已为世界各大搜索引擎普遍采用，在我国使用该项技术的搜索引擎有：

百度（<http://www.baidu.com>）搜索引擎使用了高性能的“网络蜘蛛”程序自动在互联网中搜索信息，可定制、高扩展性的调度算法使得搜索器能在极短的时间内收集到最大数量的互联网信息。百度在中文互联网拥有天然优势，支持搜索 1.3 亿个中文网页，是现在最大的中文搜索引擎。并且，百度每天都在增加几十万新网页，对重要中文网页实现每天更新。百度除了用超链分析排名外还开展竞价排名。具有网页快照，相关搜索、中文人名识别、简体中文自动转换、网页预览等功能，还可以进行专业的 MP3 搜索、Flash 搜索、新闻搜索、图片搜索、信息快递搜索。百度总裁李彦宏就是超链分析专利的唯一持有人。

引入人工智能的慧聪（<http://www.huicong.com>）行业搜索引擎对于商务人士而言尤为有用。如果在 Google 上输入“化工”，搜索出相关网页有 138 万个，里面包罗万象，有用的、

无用的混杂在一起，如果用户想找的网站正好排名在上千、上万个以后，无疑面对的是一个茫茫网海，找到这个信息十分困难。而在慧聪行业搜索引擎，弹出的首先是一个分类页面，列有有机化工、石油化工、精细化工、印刷化工等 43 个行业分类的条目，虽然只有 1 万多条查询结果，针对性却非常强，用户可以快速找到自己需要的信息，网站与有效用户的见面机会大大增强。模拟人类智慧，更好地使用技术来完成更具智慧的搜索，必然是下一代搜索引擎技术的发展趋势。

五、存在的缺陷

用户在搜索关于某些内容的有效信息时，最大的特点是各异性。利用超链分析技术，用户将接受一种根据某种标准进行网页排名的信息服务，从而演变成为各网站想尽办法追求网页排名的商业活动。

海量的网页被收集回来，用姓名、电话、单位名称或网名都可以搜索到许多含有此关键词的信息，这些信息有不少侵权、侵犯隐私、泄露机密的信息，尤其是大量论坛的帖子被收录，不少帖子言论含有攻击的成分。所以如何及时处理好这些链接又是搜索引擎急需解决的问题。

第三节 中文分词

一、什么是中文分词？

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“I am a student.”，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”。

二、中文分词的意义和作用

要想说清楚中文分词的意义和作用，就要提到智能计算技术。智能计算技术涉及的学科包括物理学、数学、计算机科学、电子机械、通讯、生理学、进化理论和心理学等等。简单的说，智能计算就是让机器“能看会想，能听会讲”。要想实现这样的一个目标，首先就要让机器理解人类的语言，只有机器理解了人类的语言文字，才使得人与机器的交流成为可能。再反观我们人类的语言中，“词是最小的能够独立活动的有意义的语言成分”，所以对于中文来讲，将词确定下来是理解自然语言的第一步，只有跨越了这一步，中文才能像英文那样过渡到短语划分、概念抽取以及主题分析，以至自然语言理解，最终达到智能计算的最高境界，实现人类的梦想。

从现阶段的情况来看，英文已经跨越了分词这一步，也就是说在词的利用上已经先一步，并且已经展现了良好的应用前景，无论是信息检索还是主题分析的研究都要强于中文，究其根本原因就是中文要通过分词这道难关，只有攻克了这道难关，我们才有希望赶上并超过英文在信息领域的发展，所以中文分词对我们来说意义重大，可以说直接影响到使用中文的每一个人的方方面面。

中文分词到底对搜索引擎有多大影响？对于搜索引擎来说，最重要的并不是找到所有结果，因为在上百亿的网页中找到所有结果没有太多的意义，没有人能看得完，最重要的是把最相关的结果排在最前面，这也称为相关度排序。中文分词的准确与否，常常直接影响到对搜索结果的相关度排序。在现有三个中文搜索引擎上做测试，测试方法是直接在 Google (<http://www.google.com>)、百度 (<http://www.baidu.com>)、中搜 (<http://www.zhongsou.com>) 上以“和服”为关键词进行搜索：

在 Google 上输入“和服”搜索所有中文简体网页，总共结果 507,000 条，前 20 条结果中有 14 条与和服一点关系都没有。在第一页就有以下错误：

- “通信信息报：瑞星以技术和服务开拓网络安全市场”
- “使用纯 HTML 的通用数据管理和服务- 开发者- ZDNet ...”
- “陈慧琳《心口不一》化妆和服装自己包办”
- “.: 外交部：中国境外领事保护和服务指南（2003 年版） ...”
- “产品和服务”

等等。第一页只有三篇是真正在讲“和服”的结果。

在百度上输入“和服”搜索网页，总共结果为 287,000 条，前 20 条结果中有 6 条与和服一点关系都没有。在第一页有以下错误：

- “福建省晋江市恒和服装有限公司系独资企业”
- “关于商品和服务实行明码标价的规定”
- “青岛东和服装设备”

在中搜上输入“和服”搜索网页，总共结果为 26,917 条，前 20 条结果都是与和服相关的网页。

这次搜索引擎结果中的错误，就是由于分词的不准确所造成的。Google 的中文分词技术采用的是美国 Basis Technology (<http://www.basistech.com>) 公司提供的中文分词技术，百度使用的是自己公司开发的分词技术，中搜使用的是国内海量科技 (<http://www.hylanda.com>) 提供的分词技术。由此可见，中文分词的准确度，对搜索引擎结果相关性和准确性有相当大的关系。

三、中文分词技术

中文分词技术属于自然语言处理技术范畴，对于一句话，人可以通过自己的知识来明白哪些是词，哪些不是词，但如何让计算机也能理解？其处理过程就是分词算法。

现有的分词算法可分为三大类：基于字符串匹配的分词方法、基于理解的分词方法和基于统计的分词方法。

1、基于字符串匹配的分词方法

这种方法又叫做机械分词方法，它是按照一定的策略将待分析的汉字串与一个“充分大的”机器词典中的词条进行匹配，若在词典中找到某个字符串，则匹配成功（识别出一个词）。按照扫描方向的不同，字符串匹配分词方法可以分为正向匹配和逆向匹配；按照不同长度优

先匹配的情况，可以分为最大（最长）匹配和最小（最短）匹配；按照是否与词性标注过程相结合，又可以分为单纯分词方法和分词与标注相结合的一体化方法。常用的几种机械分词方法如下：

- 正向最大匹配法（由左到右的方向）；
- 逆向最大匹配法（由右到左的方向）；
- 最少切分（使每一句中切出的词数最小）。

还可以将上述各种方法相互组合，例如，可以将正向最大匹配方法和逆向最大匹配方法结合起来构成双向匹配法。由于汉语单字成词的特点，正向最小匹配和逆向最小匹配一般很少使用。一般说来，逆向匹配的切分精度略高于正向匹配，遇到的歧义现象也较少。统计结果表明，单纯使用正向最大匹配的错误率为 1/169，单纯使用逆向最大匹配的错误率为 1/245。但这种精度还远远不能满足实际的需要。实际使用的分词系统，都是把机械分词作为一种初分手段，还需通过利用各种其它的语言信息来进一步提高切分的准确率。

一种方法是改进扫描方式，称为特征扫描或标志切分，优先在待分析字符串中识别和切分出一些带有明显特征的词，以这些词作为断点，可将原字符串分为较小的串再来进行机械分词，从而减少匹配的错误率。另一种方法是将分词和词类标注结合起来，利用丰富的词类信息对分词决策提供帮助，并且在标注过程中又反过来对分词结果进行检验、调整，从而极大地提高切分的准确率。

对于机械分词方法，可以建立一个一般的模型，在这方面有专业的学术论文，这里不做详细论述。

2、基于理解的分词方法

这种分词方法是通过让计算机模拟人对句子的理解，达到识别词的效果。其基本思想就是在分词的同时进行句法、语义分析，利用句法信息和语义信息来处理歧义现象。它通常包括三个部分：分词子系统、句法语义子系统、总控部分。在总控部分的协调下，分词子系统可以获得有关词、句子等的句法和语义信息来对分词歧义进行判断，即它模拟了人对句子的理解过程。这种分词方法需要使用大量的语言知识和信息。由于汉语语言知识的笼统、复杂性，难以将各种语言信息组织成机器可直接读取的形式，因此目前基于理解的分词系统还处在试验阶段。

3、基于统计的分词方法

从形式上看，词是稳定的字的组合，因此在上下文中，相邻的字同时出现的次数越多，就越有可能构成一个词。因此字与字相邻共现的频率或概率能够较好的反映成词的可信度。可以对语料中相邻共现的各个字的组合的频度进行统计，计算它们的互现信息。定义两个字的互现信息，计算两个汉字 X、Y 的相邻共现概率。互现信息体现了汉字之间结合关系的紧密程度。当紧密程度高于某一个阈值时，便可认为此字组可能构成了一个词。这种方法只需对语料中的字组频度进行统计，不需要切分词典，因而又叫做无词典分词法或统计取词方法。但这种方法也有一定的局限性，会经常抽出一些共现频度高、但并不是词的常用字组，例如“这一”、“之一”、“有的”、“我的”、“许多的”等，并且对常用词的识别精度差，时空开销大。实际应用的统计分词系统都要使用一部基本的分词词典（常用词词典）进行串匹配分词，同时使用统计方法识别一些新的词，即将串频统计和串匹配结合起来，既发挥匹配分词切分速度快、效率高的特点，又利用了无词典分词结合上下文识别生词、自动消除歧义的优点。

到底哪种分词算法的准确度更高，目前并无定论。对于任何一个成熟的分词系统来说，不可能单独依靠某一种算法来实现，都需要综合不同的算法。比如，海量科技的分词算法就采用“复方分词法”。所谓复方，相当于用中药中的复方概念，即用不同的药才综合起来去医治疾病，同样，对于中文词的识别，需要多种算法来处理不同的问题。

四、分词中的难题

有了成熟的分词算法，是否就能容易地解决中文分词问题呢？事实远非如此。中文是一种十分复杂的语言，让计算机理解中文语言更是困难。在中文分词过程中，有两大难题一直没有完全突破。

1、歧义识别

歧义是指同样的一句话，可能有两种或者更多的切分方法。例如：表面的，因为“表面”和“面的”都是词，那么这个短语就可以分成“表面”“的”和“表”“面的”。这种称为交叉歧义。像这种交叉歧义十分常见，前面举的“和服”的例子，其实就是因为交叉歧义引起的错误。“化妆和服装”可以分成“化妆”“和”“服装”或者“化妆”“和服”“装”。由于没有人的知识去理解，计算机很难知道到底哪个方案正确。

交叉歧义相对组合歧义来说是还算比较容易处理，组合歧义就必需根据整个句子来判断了。例如，在句子“这个门把手坏了”中，“把手”是个词，但在句子“请把手拿开”中，“把手”就不是一个词；在句子“将军任命了一名中将”中，“中将”是个词，但在句子“产量三年中将增长两倍”中，“中将”就不再是词。这些词计算机又如何去识别？

如果交叉歧义和组合歧义计算机都能解决的话，在歧义中还有一个难题，是真歧义。真歧义意思是给出一句话，由人去判断也不知道哪个应该是词，哪个应该不是词。例如：“乒乓球拍卖完了”，可以切分成“乒乓”“球拍”“卖”“完”“了”、也可切分成“乒乓球”“拍卖”“完”“了”，如果没有上下文其他的句子，恐怕谁也不知道“拍卖”在这里算不算一个词。

2、新词识别

新词，专业术语称为未登录词，也就是那些在字典中都没有收录过，但又确实能称为词的那些词。最典型的是人名，人可以很容易理解句子“王军虎去广州了”中，“王军虎”是个词，因为是一个人的名字，但要是让计算机去识别就困难了。如果把“王军虎”做为一个词收录到字典中去，全世界有那么多名字，而且每时每刻都有新增的人名，收录这些人本本身就是一项巨大的工程。即使这项工作可以完成，还是会存在问题，例如：在句子“王军虎头头虎脑的”中，“王军虎”还能不能算词？

新词中除了人名以外，还有机构名、地名、产品名、商标名、简称、省略语等都是很难处理的问题，而且这些又正好是人们经常使用的词，因此对于搜索引擎来说，分词系统中的新词识别十分重要。目前新词识别准确率已经成为评价一个分词系统好坏的重要标志之一。

五、中文分词的应用

中文分词主要应用于信息检索、汉字的智能输入、中外文对译、中文校对、自动摘要、自动分类等很多方面。下面就以信息检索为例来说明中文分词的应用。

通过近几年的发展，互联网已经离我们不再遥远。互联网上的信息也在急剧膨胀，在这海量的信息中，各类信息混杂在一起，要想充分利用这些信息资源就要对它们进行整理，如果由人来做这项工作，已经是不可能的，而如果面对中文信息不采用分词技术，那么整理的结果就过于粗糙，而导致资源的不可用，例如：“制造业和服务业是两个不同的行业”和“我们出口日本的和服比去年有所增长”中都有“和服”，而被当作同一类来处理，结果是检索“和服”的相关信息，会将他们都检索到，在信息量少的情况下，似乎还能够忍受，如果是海量信息，这样的结果就会令人讨厌了。通过引入分词技术，就可以使机器对海量信息的整理更准确更合理，在“制造业和服务业是两个不同的行业”中“和服”不会被当做一个词来处理，那么检索“和服”当然不会将它检索到，使得检索结果更准确，效率也会大幅度的提高。

所以中文分词的应用会改善我们的生活，使人们真正体会到科技为我所用。

六、中文搜索引擎技术

信息的飞速增长，使搜索引擎成为人们查找信息的首选工具，Google、百度、中国搜索等大型搜索引擎一直是人们讨论的话题。随着搜索市场价值的不断增加，越来越多的公司开发出自己的搜索引擎，阿里巴巴的商机搜索、8848 的购物搜索等也陆续面世，自然，搜索引擎技术也成为技术人员关注的热点。

搜索引擎技术的研究，国外比中国要早近十年，从最早的 Archie，到后来的 Excite，以及 altvista、overture、Google 等搜索引擎面世，搜索引擎发展至今，已经有十几年的历史，而国内开始研究搜索引擎是在上世纪末本世纪初。在许多领域，都是国外的产品和技术一统天下，特别是当某种技术在国外研究多年而国内才开始的情况下。例如操作系统、字处理软件、浏览器等等，但搜索引擎却是个例外。虽然在国外搜索引擎技术早就开始研究，但在国内还是陆续涌现出优秀的搜索引擎，像百度（<http://www.baidu.com>）、中搜（<http://www.zhongsou.com>）等。目前在中文搜索引擎领域，国内的搜索引擎已经和国外的搜索引擎效果上相差不远。之所以能形成这样的局面，有一个重要的原因就在于中文和英文两种语言自身的书写方式不同，这其中对于计算机涉及的技术就是中文分词。

目前在自然语言处理技术中，中文处理技术比西文处理技术要落后很大一段距离，许多西文的处理方法中文不能直接采用，就是因为中文必需有分词这道工序。中文分词是其他中文信息处理的基础，搜索引擎只是中文分词的一个应用。其他的比如机器翻译（MT）、语音合成、自动分类、自动摘要、自动校对等等，都需要用到分词。因为中文需要分词，可能会影响一些研究，但同时也为一些企业带来机会，因为国外的计算机处理技术要想进入中国市场，首先也是要解决中文分词问题。在中文研究方面，相比外国人来说，中国人有十分明显的优势。

分词准确性对搜索引擎来说十分重要，但如果分词速度太慢，即使准确性再高，对于搜索引擎来说也是不可用的，因为搜索引擎需要处理数以亿计的网页，如果分词耗用的时间过长，会严重影响搜索引擎内容更新的速度。因此对于搜索引擎来说，分词的准确性和速度，二者都需要达到很高的要求。目前研究中文分词的大多是科研院校，清华、北大、中科院、北京语言学院、东北大学、IBM 研究院、微软中国研究院等都有自己的研究队伍，而真正专业研究中文分词的商业公司除了海量科技以外，几乎没有了。科研院校研究的技术，大部分不能很快产品化，而一个专业公司的力量毕竟有限，看来中文分词技术要想更好的服务于更多的产品，还有很长一段路。

第四节 基于词意的文本分析

基于词意的文本分析是指运用词意对文本内容进行分析，得到文本内容的重点要素。文本分析是一种词意的自学习技术，是与整个技术相结合的独特分词技术。

一、文本分析技术特点

由于词意库是自学习形成，因此不需要预先设置庞大的词典库，而且最重要的是，学习是持续性的，系统对词意的理解会随着外界的变化而变化，并能不断增添新创词汇。

由于同一个词，对于不同的人，词意可能是不同的，因此可以根据每个用户的习惯形成自己的子词意库，可使用子词意库对文本进行二次分析，产生个性化结果。

系统的学习分为两种方式：

知识学习 系统自动在互联网上进行，无特定目的地吸收网上各种信息，并对收集到的信息进行分析后作为知识保留，整个过程无需人工干预，可一天 24 小时不间断地进行。

经验学习 每一次用户的具体使用，其结果也将作为经验保留下来，并对知识学习的结果进行修正。

分词技术不追求 100% 的准确，而讲究实用、快速，不依赖于庞大的词汇库或知识库，因此可以做到不针对特定领域，可解决人名、地名、新出现的词汇等的分词，这些问题是传统分词方法难以解决的，尤其是新词汇的分词，几乎是一个世界性的难题。

由于整个核心算法并非基于字、词典及语法，而是从模仿人类对语言文字的理解入手，比如一个儿童并不懂得查字典和语法，但能够听懂别人说的话，因此对核心只要稍作修改便能够用于英语及其它文字语言，就好象一个婴儿，你把他放到哪个国家，他就能学会当地的语言。

二、文本分析应用

1、相关性搜索

相关性搜索是根据关键词的词意，与文章提取的重点要素进行比对，以此生成搜索结果，这将是真正基于内容分析的搜索结果，而并非仅仅是简单的全文检索。

比如在一篇计算机相关的文章中，可能计算机这个词出现得很少、甚至根本没有，而大量出现的可能是软件、硬件、程序、内存等词汇，也有可能文章中不叫“计算机”而叫“电脑”，当使用“计算机”这个词进行检索时，系统可分析出从文章中提取的重点要素与“计算机”这个词是密切相关的，因此也会将其放入检索结果。

另外可以避免将错误结果放入，比如用“苹果”搜索，却把有关“苹果色素”的文章放在了搜索结果中，而“苹果”和“苹果色素”完全是两回事。

2、个性化搜索

根据用户爱好和习惯，自动搜索其所需要的内容。通过某种技术，可以从用户所浏览页面中分析出用户的爱好和习惯，进而在用户进入网站时自动将他最喜欢的内容呈现在他的眼前，而且这种分析也是持续性的，可以即时掌握用户习惯和爱好的变化。

个性化搜索最大的问题是用户嫌麻烦，不愿意用。而且习惯和爱好是发展变化的，这种

变化一般是潜移默化的，用户自己也难以察觉而去修改所设置的关键词，即使察觉，也往往会忘记和懒得去修改。

3、其他应用

A. 自动摘要 由于使用了词意的理解，使得摘要的准确性大大提高，而且因为核心算法并非针对某个特定领域，所以应用的范围也是全领域的。

B. 自动分类 比如输入一个产品，电脑系统能够自动将其归到一个产品类录下面。这对搜索引擎、电子商务、供求信息等网站也是非常有用的，然而这又是一个世界性的难题。目前一般都是采用人工来进行分类，比如搜狐就曾在网上发动大量的志愿者对其搜索引擎进行分类整理，但可想而知这样做在成本、效率、准确度等方面都难以满足要求。

C. 应用核心技术开发出其它特定功能的软件。

第五章 了解网站的搜索引擎表现

第一节 基础查询

一、域名信息查询

域名归属：查看域名所有人、注册时间、到期时间等。

英文域名信息查询：www.whois.com

中文域名信息查询：www.cnnic.cn

一般域名注册商网站也提供此类查询功能。

二、域名历史查询

有些没在使用也没有被注册的域名，可能是因为作弊被 Google 或者百度处罚过的。在注册新域名的时候，有必要留个心眼，以免捡了别人的垃圾。

<http://domain-history.domaintools.com>

三、主机连接速度和 IP 地址查询

开始-运行-输入“cmd”进入命令行窗口。

输入 ping www.timev.net，返回以下内容：

```
Ping www.timev.net [61.141.5.50] with 32 bytes of Data:  
Reply from 61.141.5.50: bytes=32 time=282ms TTL=54
```

其中 61.141.5.50 就是网站的 IP 地址，time=282ms 这个数值越小表明速度越快。

需要注意的是，这个数值仅仅反应相对的速度，比如今天是不是比昨天慢了好多。收到查询者和网站主机所处区域和运营商的限制，会有很大的差异。比如你是电信用户，查询网通的主机可能显示速度比较慢，但是网通用户查询起来就很正常。

可以让不同地区的朋友，特别是网站主要目标用户群体所在地区的朋友帮助你做一些测试，如果普遍反应很慢，就要考虑换主机服务商了。

四、同 IP 网站查询

用来查询在和你的网站同一个 IP 地址上，还有哪些网站。这往往对于虚拟主机比较有用，你可以看看和你在同一个服务器上有哪些邻居，也可以看看这些邻居是不是都是些安分守己网站，如果发现其中有明显搜索引擎作弊的网站，就要引起注意了，因为你的网站有可能收到牵连。

查询地址：www.114best.com

第二节 收录查询

一、是否被收录

直接在搜索引擎搜索域名。比如在百度搜索 www.timev.net

浩维互动
首页 关于我们 服务内容 客户案例 知识库 浩维博客 联系方式 互联网战略规划 网站诊断和网站策划 搜索引擎营销和优化 互联网开源计划 行业解决方案 超越客户期望 互联网是增强企业竞争优势的资源，互联网应用应该具有整体性和长期性，并应该融入...
www.timev.net/ 6K 2007-3-6 - 百度快照

这就说明网站已经被收录了。

如果显示下面的结果，说明网站还没有被收录。

你可以直接访问：www.timev.net/
百度建议你：
如果网址正确，请点击上面的链接
如果你在寻找包含“www.timev.net/”的所有网页，请点击[这里查询](#)

需要注意的是，有时候你解析的域名是 timev.net 而不是 www.timev.net，那要查询解析的那个。

二、网页收录数量

一般可以通过在搜索引擎搜索框输入“site:”命令来查看某网站被搜索引擎收录的网页数量。

site:163.com

找到相关网页约 40, 400, 000 篇。

查看网址中包含 163.com 的网页收录数，例如 news.163.com、tech.163.com、bbs.163.com、blog.163.com 开头的网页都会被统计在内。

site:www.163.com

找到相关网页 189 篇。

site:tech.163.com

找到相关网页约 1, 480, 000 篇。

（以上数据来自 2007 年 3 月 26 日百度查询结果）

一般来说，网页收录数量反应了网站内容的丰富程度。对于一个网站来说，如果内容不断更新，网页收录数也会随之增加。

目前搜索引擎已经能够收录动态的网页。由于各个搜索引擎由于算法不一样，对同一个网站的网页收录数量会有所区别。如果某个网站的网页收录数量忽然大幅下跌，可能说明网

站因为存在一些不符合搜索引擎要求的地方，被搜索引擎处罚了。

可以说，网页收录数是网站搜索引擎表现的一个晴雨表。

第三节 反向链接查询

前面的章节提到，搜索引擎会根据一个网站被其他网站链接的数量和质量来决定网站在搜索结果中的排名。

有的网站链接是网站主管理员主动寻求、添加和交换的，而有的时候，别的网站会主动链接过来，一些包含网站链接的文章被转载或者发布在论坛、博客，都有可能带来意料之外的链接。

看看哪些网站链接了你，可以使用 link 命令。

link:www.timev.net

Google 和百度的外部链接查询查询的结果比较少，推荐使用 Yahoo Site Explore 查询外链。

第四节 网页 Rank 查询

一、Google PageRank 查询

查询 Google PageRank 可以通过 Google 免费提供的浏览器工具条来实现。

在 <http://toolbar.google.cn> 可以下载到中文版本的 Google 工具条，并在设置选项中把“PageRank 和网页资料”勾选，当你访问一个网站的时候，如果改网页已经被 Google 收录并评级别，工具条上 PageRank 就会显示当前网页的 PageRank 值了。

目前 Google 的工具条不仅支持 Internet Explorer 浏览器，也支持 Firefox 浏览器了。

二、Sogou Rank 查询

网页评级（Sogou Rank）是搜狗衡量网页重要性的指标，不仅考察了网页之间链接关系，同时考察了链接质量、链接之间的相关性等特性，是机器根据 Sogou Rank 算法自动计算出来的，值从 0 至 100 不等。网页评级越高，该网页在搜索中越容易被检索到。

Sogou Rank 评级范围从 0 到 100，同时显示链接情况，提供比 Google Rank 更精细的优化参考。并且无需安装任何工具，即可查询到每个页面的评级。

第五节 关键词排名查询

如果你了解某个关键词的搜索结果中自己网站的排名，最简单的办法就是直接到搜索引擎输入关键词。

如果你的网站表现不错，它可能排在前几位，第一页或者是前几页。

如果你连续翻了十多页还没有找到你的网站，而你仍然想看看你的网站在这个关键词搜索结果中到底落后到什么程度，就可以借助一些在线工具：

www.timev.net/tools/keyword-rank.html

第六节 ALEXA 查询

一、什么是 Alexa?

1996 年 4 月, Alexa 在美国创立, 那时它只是一个小网站, 主要做分类导航。在那时候, 很多网站都是雅虎的 Fans, Alexa 只是众多跟风者中的一员罢了。但 Alexa 自己也有一些技术, 就是对访问某个 Web 站点的流量进行统计和分析。后来, Alexa 还向网民提供自己开发的搜索引擎服务。

1997 年 7 月, Alexa 发布了一款软件, 就是现在著名的 Alexa Toolbar (Alexa 工具条), 这是一个嵌入到微软 IE 浏览器中的工具, 它在用户访问每个 Web 页面时都向 Alexa 发回一串代码, 将该次浏览的相关信息告诉 Alexa。而用户会在该工具条上看到其浏览的网站在全球所有网站中的排名信息, 当然, 这只是 Alexa 给出的网站排名, 就是我们通常所说的 alexa 排名。

Alexa 在自己的网站上说, 从 1997 年 9 月发布 Alexa Toolbar 的第一个版本起, Alexa Toolbar 已累计被下载了 1000 多万次, 现在全世界使用 Alexa Toolbar 的用户已有上百万人。通过对搜集到的 Alexa Toolbar 用户的浏览信息进行分析, 并综合页读数 and 到达用户数的数据, Alexa 顺势推出了自己的网站流量排名。不过这个排名很少被正规网站引用, 与全世界近 6 亿互联网用户相比, 即使真有上百万人使用 Alexa Toolbar, 其样本的有效性仍然值得怀疑。Alexa 自己也承认, “流量数据是基于 Alexa 用户的设置, 可能不是全球互联网人口的一个有代表性的采样”, 最多可以说, 这个排名大致反映了 Alexa Toolbar 用户的喜好。

二、Alexa 主要数据

某个特定网站被排名时, 依据的浏览率数据是基于该网站 3 个月访问量记录的累积。也就是说 Alexa 每三个月发布一次排名结果, 即通常说的名次。它的计算主要取决于访问用户数 (Users Reach) 和页面浏览数 (Page Views)。Alexa 系统每天对每个网站的访问用户数和页面浏览数进行统计, 通过这两个量的三个月累积值的几何平均得出当前名次。

访问用户数 (Users Reach) 指通过 Internet 访问某个特定网站的人数。用访问某个特定网站的人数占有 Internet 用户数的比例来表示。即: $\text{访问用户数} = (\text{访问人数} / \text{全部 Alexa 用户数}) * 100\%$, Alexa 以每百万人作为计数单位。以世界网络 (linkwan.com) 为例, 如果它的访问用户数为 2% 的话, 就是说, 随意抽取一百万的 Internet 用户, 其中有 20, 000 人访问世界网络。

页面浏览数 (Page Views) 是指用户访问了某个特定网站的多少个页面。是所有访问该网站的用户浏览的页面数之和。每个用户浏览的页面数取平均值, 是所有访问该网站的用户每天每人浏览的独立页面数的平均。同一人、同一天、对同一页面的多次浏览只记一次。

三、用 Alexa 查什么?

Alexa 工具条的使用率在全球各地有所差异, 受用户的语言、地域、文化等各方面的影响。比如英文网站相对于其它语言的网站, 访问量数据更容易被充分地统计, 而同样语种的网站中, IT 类网站由于用户群中 Alexa 工具条的用户比较多, 所以排名也容易比较高。

所以，不同类别的网站有时没有可比性，所以不能一味地比较综合排名。如专业性的网站在同类别网站中排名非常靠前，但和门户类网站相比，浏览率可能差别很大。

浏览率太小的网站统计数字可能不准确，总体上排名越靠前（浏览率越大）的网站统计数字就越可靠。一般来说，月访问量 1000 以下或排名 100,000 以后的网站统计数字是不准确的。

可以肯定，任何的计算方法都不完美，所以我们也无法评述 Alexa 排名的公正性和科学性。但既然 Alexa 将相关统计信息通过 Internet 全球公开发布，大家又都用 Alexa 的统计作为标杆，从中国到世界各国权威的新闻媒体在讨论一个著名网站的规模时，莫不以 Alexa 为标准。世界上也没有第二个像 Alexa 那样的网站不需要你在网页中插入他的代码等等来计算排名，他对全世界所有网站一视同仁，我们就姑且相信其专业性和权威性。

所以，使用 Alexa，应该关注网站的流量趋势，关注各个子域名的访问比例，以及和同类网站的对比情况，从中找出对网站运营以及搜索引擎优化有用的信息，而不是整天盯着排名。

第六章 基于搜索引擎友好的网站设计

网页设计师一般在设计网页的时候，考虑更多的是如何让网页具有视觉冲击效果，从而让用户对网站有深刻的访问印象，这并没错。但是，如果互联网用户不能很容易的找到这个站点，就算网页的视觉效果再好，他们的工作也是徒劳的。所以，我们在规划设计一个网页的时候，也应该为网站的营销拓展做好工作。在前面章节，大家应该已经了解到搜索引擎优化的营销推广优势，所以设计一个既受访问用户喜欢又能迎合搜索引擎喜好的网站就十分重要了。

第一节 域名策略

域名的选取在整个搜索引擎优化工作过程中是一个很重要的环节。选择一个好的域名，并期望它为 SEO 效果提供尽可能最大的辅助作用，这无疑是个搜索引擎优化工作人员都想看到的。

小提示： 在开始本节之前，大家可以事先阅读本章附录部分的域名基本知识。

一、不同后缀的域名在搜索引擎中的权重

一般来说，权威专业的站点更能获得搜索引擎的喜爱。由于.edu 和.gov 后缀的域名并不是任何人都能注册，需要注册单位出示相关机构证明材料才可注册，这两类的站点上发布的信息具有权威和原创的特性，所以我们会经常发现教育和政府的站点更能在搜索引擎中获得好的关键词排名。

像 Google 那样的对全球网页进行索引的搜索引擎，当用户查询的时候，它会把与用户所处地域和使用语言等信息匹配的搜索结果展示给用户，而不会出现 Google 向一个中文本土搜索用户返回其他非中文国家的网页信息。基于以上分析，有很多的观点认为，带有.cn 后缀的域名在中文搜索引擎里更加具有搜索结果展示优势。

二、域名基本常识

互联网上的域名就相当于我们现实生活中的门牌号码一样，可以在纷繁芜杂的网络世界里准确无误地把我们指引到我们要访问的站点。在互联网发展之初并没有域名，有的只是 IP 地址。IP 地址就是一组类似这样的数字，如：162.105.203.245。由于当时互联网主要应用在科研领域，使用者非常少，所以记忆这样的数字并不是非常困难。但是随着时间的推移，连入互联网的电脑越来越多，需要记忆的 IP 地址也越来越多，记忆这些数字串变得越来越困难，于是域名应运而生。域名就是对应于 IP 地址的用于在互联网上标识机器的有意义的字符串。例如 CNNIC 的域名 www.cnnic.net.cn，比起 IP 地址而言就更形象也更容易记忆。

为了便于大家进一步了解域名的实质，有必要在这里谈谈域名的体系结构。从 www.cnnic.net.cn 这个域名来看，它是由几个不同的部分组成的，这几个部分彼此之间具有层次关系。其中最后的.cn 是域名的第一层，.net 是第二层，.cnnic 是真正的域名，处在第三层，当然还可以有第四层，如：inner.cnnic.net.cn，至此我们可以看出域名从后到前的层次结构类似于一个倒立的树型结构。其中第一层的.cn 叫做地理顶级域名。

目前互联网上的域名体系中共有三类顶级域名：一是地理顶级域名，共有 243 个国家和地区的代码。例如.CN 代表中国，.JP 代表日本，.UK 代表英国等等，另一类是类别顶级域名，共有 7 个：.COM(公司)，.NET(网络机构)，.ORG(组织机构)，.EDU(美国教育)，.GOV(美国政府部门)，.ARPA(美国军方)，.INT(国际组织)。由于互联网最初是在美国发展起来的，所以最初的域名体系也主要供美国使用，所以.GOV，.EDU，.ARPA 虽然都是顶级域名，但却是美国使用的。只有.COM，.NET，.ORG 成了供全球使用的顶级域名。相对于地理顶级域名来说，这些顶级域名都是根据不同的类别来区分的，所以称之为类别顶级域名。随着互联网的不断发展，新的顶级域名也根据实际需要不断被扩充到现有的域名体系中来。新增加的顶级域名是.BIZ(商业)，.COOP(合作公司)，.INFO(信息行业)，.AERO(航空业)，.PRO(专业人士)，.MUSEUM(博物馆行业)，.NAME(个人)。

在这些顶级域名下，还可以再根据需要定义次一级的域名，如在我国的顶级域名.CN 下又设立了.COM，.NET，.ORG，.GOV，.EDU 以及我国各个行政区划的字母代表如.BJ 代表北京，.SH 代表上海等等。

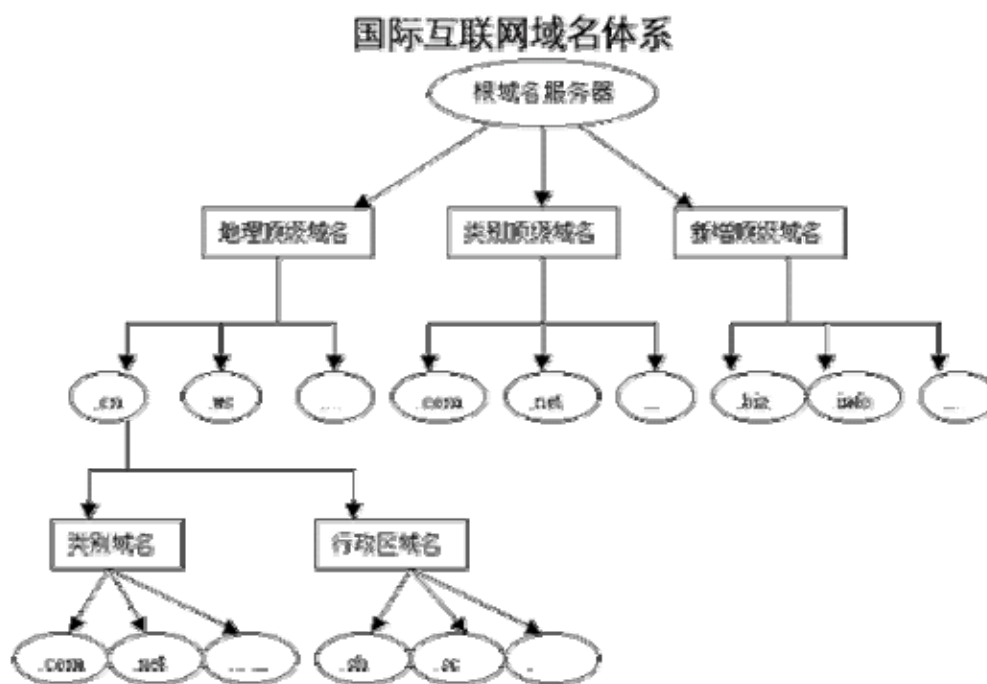


图 6-1-1 国际互联网域名体系

1、组织域名

- com: Commercial organizations, 商业组织、公司
- edu: Educational institutions, 教研机构
- gov: Governmental entities, 政府部门
- int: International organizations, 国际组织
- mil: Military (U. S), 美国军部
- net: Network operations and service centers, 网络服务商
- org: Other organizations, 非盈利组织
- biz: business, 企业站点
- name: 个人网站

info: information, 信息类网站

2、国家和地区域名

ad: Andorra, 安道尔
ae: United Arab Emirates, 阿联酋
af: Afghanistan, 阿富汗
ag: Antigua and Barbuda, 安提瓜和巴布达
ai: Anguilla, 安格拉
al: Albania, 阿尔巴尼亚
am: Armenia, 亚美尼亚
an: Netherlands Antilles, 荷兰属地
ao: Angola, 安哥拉
aq: Antarctica, 南极洲
ar: Argentina, 阿根廷
as: American Samoa, 东萨摩亚
at: Austria, 奥地利
au: Australia, 澳大利亚
aw: Aruba, 阿鲁巴
az: Azerbaijan, 阿塞拜疆
ba: Bosnia Hercegovina, 波黑
bb: Barbados, 巴巴多斯
bd: Bangladesh, 孟加拉国
be: Belgium, 比利时
bf: Burkina Faso, 布基纳法索
bg: Bulgaria, 保加利亚
bh: Bahrain, 巴林
bi: Burundi, 布隆迪
bj: Benin, 贝宁
bm: Bermuda, 百慕大
bn: Brunei Darussalam, 文莱达鲁萨兰国
bo: Bolivia, 玻利维亚
br: Brazil, 巴西
bs: Bahamas, 巴哈马
bt: Bhutan, 不丹
bv: Bouvet Island, 布韦群岛
bw: Botswana, 伯兹瓦纳
by: Belarus, 白俄罗斯
bz: Belize, 伯利兹
ca: Canada, 加拿大
cc: Cocos Islands, 科科斯群岛
cf: Central African Republic, 中非共和国
cg: Congo, 刚果

ch: Switzerland, 瑞士
ci: Ivory Coast, 象牙海岸
ck: Cook Islands, 库克群岛
cl: Chile, 智利
cm: Cameroon, 喀麦隆
cn: China, 中国
co: Colombia, 哥伦比亚
cq: Equatorial Guinea, 赤道几内亚
cr: Costa Rica, 哥斯达黎加
cu: Cuba, 古巴
cv: Cape Verde, 佛得角
cx: Christmas Island, 圣诞岛 (英属)
cy: Cyprus, 塞浦路斯
cz: Czech Republic, 捷克共和国
de: Germany, 德国
dj: Djibouti, 吉布提
dk: Denmark, 丹麦
dm: Dominica, 多米尼加联邦
do: Dominican Republic, 多米尼加共和国
dz: Algeria, 阿尔及利亚
ec: Ecuador, 厄瓜多尔
ee: Estonia, 爱沙尼亚
eg: Egypt, 埃及
eh: Western Sahara, 西萨摩亚
es: Spain, 西班牙
et: Ethiopia, 埃塞俄比亚
ev: El Salvador, 萨尔瓦多
fi: Finland, 芬兰
fj: Fiji, 斐济
fk: Falkland Islands, 福克兰群岛
fm: Micronesia, 密克罗尼西亚
fo: Faroe Islands, 法罗群岛
fr: France, 法国
ga: Gobon, 加蓬
gb: Great Britain (UK), 大不列颠联合王国
gd: Grenada, 格林纳达
ge: Georgia, 格鲁吉亚
gf: French Guiana, 法属圭亚那
gh: Ghana, 加纳
gi: Gibraltar, 直布罗陀
gl: Greenland, 格陵兰群岛
gm: Gambia, 冈比亚

gn: Guynea, 几内亚
gp: Guadeloupe, 瓜德罗普岛 (法属)
gr: Greece, 希腊
gt: Guatemala, 危地马拉
gu: Guam, 关岛
gw: Guinea-Bissau, 几内亚比绍
gy: Guyana, 圭亚那
hk: Hong Kong, 香港
hm: Heard McDonald Isl.
hn: Honduras, 宏都拉斯
hr: Croatia, 克罗蒂亚
ht: Haiti, 海地
hu: Hungary, 匈牙利
id: Indonesia, 印度尼西亚
ie: Ireland, 爱尔兰共和国
il: Israel, 以色列
in: India, 印度
io: British Indian Ocean Territory, 英属印度洋领地
iq: Iraq, 伊拉克
ir: Iran, 伊朗
is: Iceland, 冰岛
it: Italy, 意大利
jm: Jamaica, 牙买加
jo: Jordan, 约旦
jp: Japan, 日本
ke: Kenya, 肯尼亚
kg: Kyrgyzstan, 吉尔吉斯斯坦
kh: Cambodia, 柬埔寨
ki: Kiribati, 基里巴斯
km: Comoros, 科摩罗
kn: St. Kitts & Nevis, 圣茨和尼维斯
kp: Korea-North, 北朝鲜
kr: Korea-South, 南朝鲜
kw: Kuwait, 科威特
ky: Cayman Islands, 开曼群岛 (英属)
kz: Kazakhstan, 哈萨克斯坦
la: Lao People's Republic, 老挝人民共和国
lb: Lebanon, 黎巴嫩
lc: St. Lucia, 圣露西亚岛
li: Liechtenstein, 列支敦士登
lk: Sri Lanka, 斯里兰卡
lr: Liberia, 利比里亚

ls: Lesotho, 莱索托
lt: Lithuania, 立陶宛
lu: Luxembourg, 卢森堡
lv: Latvia, 拉脱维亚
ly: Libya, 利比亚
ma: Morocco, 摩洛哥
mc: Monaco, 摩纳哥
md: Moldova, 摩尔多瓦
mg: Madagascar, 马达加斯加
mh: Marshall Islands, 马绍尔群岛
ml: Mali, 马里
mm: Myanmar, 缅甸
mn: Mongolia, 蒙古
mo: Macau, 澳门
mp: Northern Mariana Islands, 北马里亚纳群岛
mq: Martinique, 马提尼克岛 (法属)
mr: Mauritania, 毛里塔尼亚
ms: Montserrat, 蒙塞拉特岛
mt: Malta, 马尔他
mv: Maldives, 马尔代夫
mw: Malawi, 马拉维
mx: Mexico, 墨西哥
my: Malaysia, 马来西亚
mz: Mozambique, 莫桑比克
na: Namibia, 纳米比亚
nc: New Caledonia, 新喀里多尼亚
ne: Niger, 尼日尔
nf: Norfolk Island, 诺福克岛
ng: Nigeria, 尼日利亚
ni: Nicaragua, 尼加拉瓜
nl: Netherlands, 荷兰
no: Norway, 挪威
np: Nepal, 尼泊尔
nr: Nauru, 瑙鲁
nt: Neutral Zone, 中立区
nu: Niue, 纽埃
nz: New Zealand, 新西兰
om: Oman, 阿曼
pa: Panama, 巴拿马
pe: Peru, 秘鲁
pf: French Polynesia, 法属玻利尼西亚
pg: Papua New Guinea, 巴布亚新几内亚

ph: Philippines, 菲律宾
pk: Pakistan, 巴基斯坦
pl: Poland, 波兰
pm: St. Pierre
pn: Pitcairn Island, 皮特克恩岛
pr: Puerto Rico, 波多黎各
pt: Portugal, 葡萄牙
pw: Palau, 帕劳
py: Paraguay, 巴拉圭
qa: Qatar, 卡塔尔
re: Reunion Island, 留尼汪岛 (法属)
ro: Romania, 罗马尼亚
ru: Russian Federation, 俄罗斯联邦
rw: Rwanda, 卢旺达
sa: Saudi Arabia, 沙特阿拉伯
sb: Solomon Islands, 所罗门群岛
sc: Seychelles, 塞舌尔
sd: Sudan, 苏丹
se: Sweden, 瑞典
sg: Singapore, 新加坡
sh: St. Helena, 海伦娜
si: Slovenia, 斯洛文尼亚
sj: Svalbard & Jan Mayen, 斯瓦尔巴特和扬马延岛
sk: Slovakia, 斯洛伐克
sl: Sierra Leone, 塞拉利昂
sm: San Marino, 圣马力诺
sn: Senegal, 塞内加尔
so: Somalia, 索马里
sr: Suriname, 苏里南
st: Sao Tome & Principe, 圣多美和普林西比
su: USSR, 苏联
sy: Syrian Arab Republic, 叙利亚
sz: Swaziland, 斯威士兰
tc: Turks & Caicos Islands, 特克斯群岛与凯科斯群岛
td: Chad, 乍得
tf: French Southern Territories, 法属南半球领地
tg: Togo, 多哥
th: Thailand, 泰国
tj: Tajikistan, 塔吉克斯坦
tk: Tokelau, 托克劳群岛
tm: Turkmenistan, 土库曼斯坦
tn: Tunisia, 突尼斯

to: Tonga, 汤加
tp: East Timor, 东帝汶
tr: Turkey, 土耳其
tt: Trinidad Tobago, 特立尼和多巴哥
tv: Tuvalu, 图瓦鲁
tw: Taiwan, 台湾
tz: Tanzania, 坦桑尼亚
ua: Ukrainian SSR, 乌克兰
ug: Uganda, 乌干达
uk: United Kingdom, 英国
us: United States, 美国
uy: Uruguay, 乌拉圭
va: Vatican City State, 梵蒂冈
vc: St. Vincent, 圣文森特和格林纳丁斯
ve: Venezuela, 委内瑞拉
vg: Virgin Islands, 维京群岛
vn: Vietnam, 越南
vu: Vanuatu, 瓦努阿图
wf: Wallis & Fortuna Is. , 瓦利斯和富图纳群岛
ws: Samoa, 东萨摩亚
ye: Yemen, 也门
yu: Yugoslavia, 南斯拉夫
za: South Africa, 南非
zm: Zambia, 赞比亚
zr: Zaire, 扎伊尔
zw: Zimbabwe, 津巴布韦

三、域名的关键词策略

域名的词干可由英文字母、阿拉伯数字和中横线组合构成，搜索引擎会因域名词干字符组合方式的不同而对其有不同的理解。接下来，我们将会通过一个具体实例来说明这个问题。

这里有两个域名：www.madeinchina.com 和 www.made-in-china.com，它们有什么不同？在我们看来这两个域名的词干都可以理解为“made in china”英文词组的意思。但是，搜索引擎不比我们人脑，它只会根据已经设置好的程式来对以上域名进行识别。通常情况下，搜索引擎会把一个“-”当然一个空格来看待，所以我们也就明白了它会如何去理解上面的两个域名的词干了。搜索引擎会认为 [Madeinchina](http://www.madeinchina.com) 是一个单词，而 [made-in-china](http://www.made-in-china.com) 则会被其认为是“made in china”的词组。

基于以上分析，如果域名中包含关键词，或者以复合关键词作为域名的时候，我们最好用“-”来使其分隔开来。

特别说明，以上的域名的关键词策略主要应用于英文语种的 SEO。

四、中文站点的域名的拼音策略

由于中文语言的特点，拼音字母仅是为汉字的发音提供辅助作用，所以它们并不经常出现在我们的常用书写中。有人曾经提出过这么一种观点，即是域名词干由中文关键词的字母全拼构成会在搜索引擎排名中具有竞争优势。我们并不认同这种观点，通过对用户的搜索行为分析，可以发现如下事实：用户一般在中英文输入法切换错误的情况下，才会出现使用关键词的拼音形式在搜索引擎中进行查询，被搜索网页的域名的词干与用户查询拼音关键词极其匹配，这些网站能排在前面也就是理所当然了。

而有些站点，在其规划构建之初，网站所有者为了考虑访问用户对域名的记忆，会启用由中文目标关键词的字母全拼构成的域名。通常情况下，域名所有者一旦注册此类型的域名，就说明其对网站的内容定位相当明确，随着时间的推移，其网站能在搜索引擎中获得好的排名表现也并不是很难理解的。

至于域名中应该如何部署关键词，这个问题将会在后面的章节中有所阐述。

五、谨防注册一个遭受搜索引擎惩罚过的域名

域名的注册费用不再像前几年那么高得让人犹豫了，所以很多 SEO 从业人员或网站管理员往往会注册大量的域名来做网站群。一旦这些网站因违反了搜索引擎的相关规则而被搜索引擎惩罚了，域名所有人将会立即抛弃这些被惩罚了的域名。

如果一不小心注册了一个被搜索引擎惩罚过的域名，这会为你的网站登录搜索引擎付出不少时间成本。一般来说，等搜索引擎刷新被惩罚域名所有人的信息后，它把该域名视为一个新域名，并对其站点网页进行重新索引，但是谁也不知道这个过程需要多少时间。

那么，我们怎样才能知道一个域名是否已经遭受搜索引擎惩罚了呢？

首先，我们可以使用 `site` 命令在主流搜索引擎中查询，来获取该域名在搜索引擎中的相关数据。比如，你想注册域名 `www.abc.com`，那么针对该域名的 `site` 命令查询形式就为：`site:abc.com`。

如果该域名在搜索引擎尚未存在收录数据，则有两种对该域名进行判断的可能：

- 该域名为一个新域名，未曾被人注册过。
- 该域名已经被人注册过，并被搜索引擎惩罚而删除了该域名所对应的网站收录数据。

到这一步，我们还不能最终决定该域名是不是要注册，因为其是否搜索引擎惩罚过暂时还不得而知。

通过<http://www.archive.org/>或者<http://www.infomall.cn/>查询域名历史。



Search Results for Jan 01, 1996 - Aug 29, 2007									
1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	0 pages	3 pages	5 pages	0 pages
							Apr 07, 2005 * Aug 28, 2005 * Dec 31, 2005 *	Jan 09, 2006 Jan 12, 2006 Feb 15, 2006 Feb 21, 2006 May 19, 2006	

上图是查询域名www.uf168.com的历史网页的结果，我们看到，这个域名 2005 年 4 月已经被使用，到 2006 年 5 月停止使用了。浩维互动的一个客户正是在 2007 年又注册并使用了这个域名，结果百度始终没有收录其网页，最后不得不更换域名。

通过浩维互动的检查发现，www.uf168.com这个域名曾经用来做过垃圾站点，被百度惩罚后还在网上公开销售过，显然已经被百度列入了黑名单了。而客户也因为这个域名的原因，白白浪费了将近两个月的时间。

当然，以上的判断仅仅是从实际操作经验中总结而来，对于大部分的域名判断是准确的，在这里提出来仅供参考。

第二节 空间策略

如何确保你的网页存放空间被搜索引擎认为是“安全”的，而不会被其视为有悖相关规则？这里所指出的安全包括两个方面，即是空间运行的稳定性和其所处的环境。

一、安全元素一：稳定性

你必须确保你所购买的主机运行稳定，这样才会确保搜索引擎机器人对你网站进行索引的过程十分顺利。如果，搜索引擎机器人在对一个网站进行爬行的时候，正值该网站空间所在的服务器宕机，那么搜索引擎机器人将不会带回任何关于该网站的最新信息，如果搜索引擎蜘蛛连续好几次都不能获取该网站的信息的话，搜索引擎就视为该网站不可用，就会清除其索引数据。

所以在购买空间的时候一定要选择一个口碑好的服务商，这会为你的网站的稳定运行提供强大的保障。许多企业为了贪图一时的便宜而去使用免费空间，这种做法是不可取的。免费空间提供商不会承诺任何有保证的服务，诸如确保空间的稳定运行、网站数据备份和相关应急响应等，一旦出现事故，将会对网站造成致命的打击。

二、安全元素二：良好的共处环境

一般来说，大部分企业所使用的空间应该称之为虚拟主机，也叫共享主机（关于虚拟主机的知识附录中有说明）。根据共享主机的特点，我们知道，大部分企业的空间是和其他很多网站的空间公用同一个独立 IP，那么我们可以形象的称这个 IP 下所有的网站都处在一个共处环境。

当你的网站准备入驻这个共处环境的时候，你应该有事先对这个环境有一个充分的考虑分析，从而来决定你的网站跟同一个 IP 下的网站呆在一起是否有风险。

那我们该如何来评估你将购买的空间是否有风险呢？

首先，请确保你的网站不要和哪些发布有违反基本社会伦理道德信息（诸如反动和色情等）的网站共处。因为搜索引擎已经在其相关条款中明确指出，它们不会对这些类型的网站进行收录，如果同一个 IP 下存在有大量的这样不受欢迎的站点的话，搜索引擎可能要求搜索引擎机器人不会对该 IP 下的所有网站进行爬行索引，如果你的网站身处其中的话，结果可想而知。查询同一个 IP 下的站点工具：

<http://whois.webhosting.info/hotmail.com>

其次，你的网站也不应该选择那些有搜索引擎排名作弊前科或嫌疑的网站为伍。由于一些网站的所有者怀有急功近利的心态，他们往往会采用一些搜索引擎排名作弊手法以期能快速获得排名效果。搜索引擎对这种行为是十分痛恨的，因为这些作弊网站一般都不会向用户提供有价值的信息同时也干扰搜索引擎排名的公平性。搜索引擎一旦发现了这些网站的作弊行为，就会对其处以相应的惩罚，它们或被搜索引擎降权或被删除数据。如果同一个 IP 下存在大量的有作弊行为的站点的话，该 IP 可能会被搜索引擎视为不友好，搜索引擎机器人也将不会对该 IP 下的所有站点进行索引，所以应该注意所启用的主机所属 IP 下的其他站点是不是有大范围作弊嫌疑。

至于如何判断一个网站是否被搜索引擎惩罚，在域名策略部分有相关讲解内容。

三、本节附录

■ 什么叫虚拟主机？

虚拟主机（Virtual Host/Virtual Server）是使用特殊的软硬件技术，把一台计算机主机分成一台台“虚拟”的主机，每一台虚拟主机都具有独立的域名和 IP 地址（或共享的 IP 地址），具有完整的 Internet 服务器功能。在同一台硬件、同一个操作系统上，运行着为多个用户打开的不同的服务器程序，互不干扰；而各个用户拥有自己的一部分系统资源（IP 地址、文件存储空间、内存、CPU 时间等）。虚拟主机之间完全独立， 在外界看来， 每一台虚拟主机和一台独立的主机的表现完全一样。 虚拟主机技术的出现，是对 Internet 技术的重大贡献，是广大 Internet 用户的福音。由于多台虚拟主机共享一台真实主机的资源，每个用户承受的硬件费用、网络维护费用、通信线路的费用均大幅度降低，Internet 真正成为人人用得起的网络。

第三节 网站结构规划

具有扁平化结构的网站更适于搜索引擎机器人爬行，从而能使网站在搜索引擎里有一个好的收录表现。如何来理解网站结构的扁平化？这主要取决于网站的物理结构和逻辑结构规划。一般来说访问用户通过少于 4 次的点击数到达最终内容页面的网站结构是符合扁平化的要求的。

物理结构 主要指用于存放网站文件的目录设置形式。网站的目录结构有不同层次之分，以下是目录的实例表现形式：

根目录：<http://www.example.com/>

二级目录：<http://www.example.com/test/>

三级目录：<http://www.example.com/test/test/>

.....

目录的级数由主域名后面斜杠的个数决定，一般来说目录结构级数控制在 3 或者 3 以内

为最佳。而对于那些信息容量大的网站来说，如果不能把目录的级数控制在上面的最佳范围的话，那则需要通过网站的逻辑结构来调整。

逻辑结构 根据内容的关联性，由超级链接构织成的网站结构。一个网站的逻辑结构如果构建合理，它就会为网站的访问用户查找信息提供便利，当然也会为用户提供更多可以参考的相关信息。由于逻辑结构最直接的表现形式为超级链接，势必可以使得搜索引擎机器人在网站上索引到更多的网页信息。

在网站扁平化原则下，逻辑结构部署所起到的作用就是通过内容的关联性把物理结构目录级数高的页面内容提升到物理结构目录级数低上来，具体的表现形式就是在目录级数低的页面上部署指向与其有内容关联性的目录级数高页面的超级链接。

第四节 站点导航设计

网站有一个清晰明了的导航结构，无论是对于用户还是搜索引擎都十分重要。用户可以通过优良的站点导航系统快速便捷地找到他们所需要的信息，而搜索引擎则可以通过合理的导航对网站的页面进行高效索引。

导航的设计原则是以纯文本链接形式为最佳，同时也应在导航链接锚文本描述中部署核心关键词。以下，我们将介绍大多数网站中不可或缺的导航种类。

栏目导航 展示网站的栏目设置情况，点击相应的链接可以抵达项目的栏目专题页面。该导航为网站中最重要的导航设置，一般位于网页头部或者头部下方的左右侧位置。

辅助导航 处于每个页面的页眉或者页脚的位置，用来链接网站中一些重要的信息。例如：网站首页、关于我们、联系信息、隐私条款等。

历史导航 也被称为面包屑导航，它准确地告诉访问者目前在网站的哪一个目录下，并说明网站的目录分类结构。通常是由一个首页链接开始，到浏览者当前所在部分结束。例如：“当前位置：首页 > 网络营销 > 搜索引擎营销”。

相关导航 当访问者在网站上阅读某一篇文章时，网页的下方常常会提供“相关文章”的链接，通常是与该文处在相同栏目里的一些文章。也有的网站还会提供与其相同或相似主题的网站链接。

网站地图 它是为了方便浏览者对网站所有的资源和服务有一个整体的了解而设置的。当然，从 SEO 的角度来看，一个站点建立网站地图是为了方便搜索引擎对站点重要网页进行索引。

网页设计师一般会采用 HTML 文本、JPEG 或 GIF 静态图片、FLASH 动画等格式来设计网站导航，我们建议采用 HTML 文本来制作网站地图为最佳，因为搜索引擎对纯文本链接的识别性最高。

第五节 IFRAME、JavaScript 和 AJAX

搜索引擎都会在其帮助指南中明示，它们对 IFRAME、JavaScript 和 AJAX 中的文字信息不能进行有效索引，我们也称这几种网页组成部分为搜索引擎优化不友好元素。

一般网页的核心信息最好不要放置在这些不友好元素中，以防搜索引擎不能成功进行索引。这些核心信息包括带有该页面核心关键词的文本区域以及相关辅助信息，诸如重要导航，相关导航和文章正文。

而那些内容描述与核心信息的相关性相去甚远的信息就建议放置在 IFRAME、

JavaScript、AJAX 中，因为它们会对核心信息起到干扰。可以放置于搜索引擎不友好因素中的信息包括网站尾部的辅助导航、版权信息和页面的广告信息等。

一、IFRAME

框架结构或称帧结构 (Frame) 网页表现为一个网页内的某一块保持固定，其它部分信息可以通过滚动条上下或左右移动显示。框架型网页的另一个表现是，深层页面的域名通常不会在 URL 中体现出来，即使进入深层子页面，浏览器显示出来的 URL 仍是主页的 URL。它的好处是：页面整体一致，更新方便。但大多搜索引擎无法识别框架，也没兴趣抓取其中内容。某些浏览器不支持框架页面。

如使用框架，必须在代码中使用“Noframes”标签进行优化，把 Noframe 标签看作一个普通文本内容的主页。在<Noframe></Noframe>区域中包含指向 frame 页的链接以及带有关键词的描述文本。同时在框架以外的区域也出现关键词文本。这样搜索引擎才能正确索引到框架内信息。

还有一个办法是采用 iframe 即内联框架 (Inner Frame) 技术来避免 Frame 带来的不便。所谓 iframe 也是框架的一种形式，它是相当于在主浏览器窗口内嵌一个子窗口，内容自动打开。Iframe 可以嵌在网页中的任意部分，也可以随意定义其大小，其代码显示为：

```
<iframe src=xx width=x height=x scrolling=xx frameborder=x></iframe>
```

对搜索引擎来说，iframe 中的文字是可见的，也可以跟踪到其中链接指向的页面，不过与用户所见不同的是，搜索引擎将 iframe 内容看成单独的一个网面内容，与被内嵌的页面无关。

二、JS

由于 JS 的内容不能被大多数搜索引擎检索，所以做网站的 SEO 时一般强调导航菜单等重要代码部分尽量不使用 JS。

部分网页左侧、右侧的相关文章、热门文章链接等，喜欢使用 JS 来调用。这又是一个 JS 产生的问题。如果大面积的链接都使用 JS，无疑将大大降低搜索引擎抓取收录的机会。

正常情况下，网页尽量以<html>作为代码开始端。但采用 JavaScript 技术的网页往往在页面一开始就堆积大量 java 代码，以致 meta 及关键词迟迟不能出现，被推至页面底部，对搜索引擎很不友好。

两种方法改善：

将脚本移至页面底部 大部分 java 代码可移到页面结束标签之上，从而不影响网站功能。

将 java 脚本置入一个 .js 扩展名的文件 包含你的 javascript 的 .js 文件通常在网站访问者的浏览器中被缓存下来，使得下次访问速度加快，也使网站修改和维护更加方便。

三、AJAX

AJAX 是异步 JavaScript 和 XML 的缩写形式，是一种网站互动应用开发技术，其直接效果是用户无加载一次，网页也可以实现页面内容刷新，由此极大提升速度和用户体验。

不过 AJAX 技术在搜索引擎优化方面被完全归入“糟糕”一类效果，因为搜索引擎不支

持 JavaScript，而 AJAX 是使用 JavaScript 发挥作用，因此搜索引擎机器人看不到 AJAX 发送的内容。比如，如果一个网站导航采用 AJAX 技术制作，那么搜索引擎无法通过导航系统深入抓取该网站首页以外的其它链接页面，同样，如果网页内容由 AJAX 生成，也无法为搜索引擎识别。

使用 AJAX 技术制作网站该如何兼顾用户体验的高效与搜索引擎优化呢？最简单的答案是确保导航与正文主体内容这两部分最核心的内容是 html 文件，这不仅有助于提升网站的搜索引擎可见度，还确保那些没有启用 JavaScript 的用户也能看到导航条内容和链接。以国际品牌 Gucci.com 为例，这个外观时尚新潮的网站制作几乎破坏了搜索引擎优化的所有规范，并且如果浏览器不支持 JavaScript，则所见完全一片空白。

第六节 URL 设计

URL 其英文完整拼写是 Uniform Resource Locator，中文翻译意思为“统一资源定位系统”，也就是我们经常所说的网站地址。面向访问用户和搜索引擎友好的 URL 设计必须要考虑三个方面的要求：便于用户记忆、URL 静态化和部署关键词。

一、便于用户记忆

URL 的设计是否对于用户友好，这也是我们必须考虑的。我们会经常看到很多站点内容页面的 URL 是一长串令人捉摸不透的数字和字母组合，很不便于用户记忆。

我们给出的设计便于记忆的 URL 的建议是：根据访问用户的语言环境和文化素质情况而灵活采用英文或者汉语拼音，然后再加上有规律的数字组合来设计 URL。

二、URL 静态化

搜索引擎对静态的 URL 索引效率更高，这一点是在其对网站管理员的帮助指南里都有明确指出过。所以，在架构网站的时候采取静态页面架构或者实现动态 URL 伪静态化无疑是对搜索引擎友好的方法。

一般来说，中小型网站由于其信息量不多，维护起来也比较方便，大多都采用静态 HTML 页面来构建。而对于大型网站来，由于信息量大，它们大都采用了 CMS (Content Management System) 来搭建网站，因为其架构安装和管理都比较方便。

但是，大部分的 CMS 系统开发商并没有考虑到 URL 静态化对于搜索引擎友好性的重要性，这些系统生成的页面 URL 都是带有参数变量的动态形式。

1、动态 URL 的伪静态化处理

已经是动态 URL 的站点如何实现 URL 静态化呢？

实现方式主要有 2 种：基于 `url rewrite` 和基于 `path_info`。

■ 把 URI 地址用作参数传递：URL REWRITE

最简单的是基于各种 WEB 服务器中的 URL 重写转向 (Rewrite) 模块的 URL 转换，这样几乎可以不修改程序，就能够实现将 `news.asp?id=234` 这样的链接映射成 `news/234.html`，从外面看上去和静态链接一样。Apache 服务器上有一个模块 (非缺省)：`mod_rewrite`：URL

REWRITE。

当需要将 news.asp?id=234 的映射成 news/234.html 时，只需设置：

```
RewriteRule /news/ (\d+) \.html /news\.asp?id=$1 [N, I]
```

这样就把 /news/234.html 这样的请求映射成了 /news.asp?id=234

当有对/news/234.html 的请求时：web 服务器会把实际请求转发给/news.asp?id=234

而在 IIS 也有相应的 REWRITE 模块：比如 ISAPI REWRITE 和 IIS REWRITE，语法都是基于正则表达式，因此配置几乎和 apache 的 mod_rewrite 是相同的。

比对于某一个简单应用可以是：

```
RewriteRule /news/ (\d+) \.html /news/news\.php?id=$1 [N, I]
这样就把 http://www.chedong.com/news/234.html 映射到了：
http://www.chedong.com/news/news.php?id=234
```

一个更通用的能够将所有的动态页面进行参数映射的表达式是：

```
把 http://www.myhost.com/foo.php?a=A&b=B&c=C
表现成 http://www.myhost.com/foo.php/a/A/b/B/c/C。
RewriteRule (.*\.\php) (\?[^/]*) ?/ ([^/]*) / ([^/]*) (.+)? ?$1 (?2$2&:\?) $3=$4?5$5:
[N, I]
```

以下是针对 phpBB 的一个 Apache mod_rewrite 配置样例：

```
RewriteEngine On
RewriteRule /forum/topic_ (.+) \.html$ /forum/viewtopic.php?t=$1 [L]
RewriteRule /forum/forum_ (.+) \.html$ /forum/viewforum.php?f=$1 [L]
RewriteRule /forum/user_ (.+) \.html$ /forum/profile.php?mode=viewprofile&u=$1 [L]
```

这样设置后就可以通过 topic_1234.html forum_2.html user_34.html 这样的链接访问原来的动态页面了。

通过 URL REWRITE 还有一些好处：

- **隐藏后台实现** 这在后台应用平台的迁移时非常有用：当从 asp 迁移到 java 平台时，对于前台用户来说，根本感受不到后台应用的变化；
- **简化数据校验** 因为像 (\d+) 这样的参数，可以有效的控制数字的格式甚至位数；

比如我们需要将应用从 news.asp?id=234 迁移成 news.php?query=234 时，前台的表现可以一直保持为 news/234.html。从实现应用和前台表现的分离：保持了 URL 的稳定性，而使用 mod_rewrite 甚至可以把请求转发到其他后台服务器上。

■ 基于 PATH_INFO 的 URL 美化

URL 美化的另外一个方式就是基于 PATH_INFO：PATH_INFO 是一个 CGI 1.1 的标准，经常发现很多跟在 CGI 后面的 "/value_1/value_2" 就是 PATH_INFO 参数，比如：
http://phpunixman.sourceforge.net/index.php/man/ls 中：\$PATH_INFO = "/man/ls"

PATH_INFO 是 CGI 标准，因此 PHP Servlet 等都有的支持。比如 Servlet 中就有 request.getPathInfo () 方法。

注意：/myapp/servlet/Hello/foo 的 getPathInfo () 返回的是/foo，而/myapp/dir/hello.jsp/foo 的 getPathInfo () 将返回的 /hello.jsp，从这里你也可以知道 jsp 其实就是一个 Servlet 的 PATH_INFO 参数。ASP 不支持 PATH_INFO。

PHP 中基于 PATH_INFO 的参数解析的例子如下：

注意：参数按"/"分割，第一个参数是空的：从/param1/param2 中解析出\$param1 \$param2 这 2 个参数。

```
if ( isset ( $_SERVER["PATH_INFO"] ) ) {  
    list ( $nothing, $param1, $param2 ) = explode ( '/', $_SERVER["PATH_INFO"] ) ;  
}
```

如何隐蔽应用：例如 .php 的扩展名：

```
在 APACHE 中这样配置：  
<FilesMatch "^app_name$">  
    ForceType application/x-httpd-php  
</FilesMatch>
```

如何更像静态页面：app_name/my/app.html？

解析的 PATH_INFO 参数的时候，把最后一个参数的最后 5 个字符 “.html” 截断即可。

注意：APACHE2 中缺省是不允许 PATH_INFO 的，需要设置 AcceptPathInfo on

特别是针对使用虚拟主机用户，无权安装和配置 mod_rewrite 的时候，PATH_INFO 往往就成了唯一的选择。

OK，这样以后看见类似于 <http://www.example.com/article/234> 这样的网页你就知道可能是 article/show.php?id=234 这个 php 程序生成的动态网页，很多站点表面看上去可能有很多静态目录，其实很有可能都是使用 1, 2 个程序实现的内容发布。比如很多 WIKIWIKI 系统都使用了这个机制：整个系统就一个简单的 wiki 程序，而看上去的目录其实都是这个应用拿后面的地址作为参数的查询结果。

利用基于 MOD_REWRITE/PATH_INFO + CACHE 服务器的解决方案对原有的动态发布系统进行改造，也可以大大降低旧有系统升级到新的内容管理系统的成本。并且方便了搜索引擎收录索引。

附：如何在 IIS 上利用 PHP 支持 PATH_INFO

PHP 的 ISAPI 模式安装备忘：只试成 php-4.2.3-Win32

解包目录

=====

php-4.2.3-Win32.zip c:\php

PHP.INI 初始化文件

```
=====
复制: c:\php\php.ini-dist 到 c:\winnt\php.ini
```

配置文件关联

```
=====
按照 install.txt 中的说明配置文件关联
```

运行库文件

```
=====
复制 c:\php\php4ts.dll 到 c:\winnt\system32\php4ts.dll
```

这样运行后: 会发现 php 把 PATH_INFO 映射到了物理路径上

```
Warning: Unknown(C:\CheDong\Downloads\ariadne\www\test.php\path): failed to create stream:
No such file or directory in Unknown on line 0
```

```
Warning: Unknown () : Failed opening 'C:\CheDong\Downloads\ariadne\www\test.php\path' for
inclusion (include_path='.;c:\php4\pear') in Unknown on line 0
```

安装 ariadne 的 PATCH

```
=====
停止 IIS 服务
```

```
net stop iisadmin
```

```
ftp://ftp.muze.nl/pub/ariadne/win/iis/php-4.2.3/php4isapi.dll
```

```
覆盖原有的 c:\php\sapi\php4isapi.dll
```

注:

ariadne 是一个基于 PATH_INFO 的内容发布系统,
PHP 4.3.2 RC2 中 CGI 模式的 PATH_INFO 已经修正, 照常安装即可。

2、不同具体情况下的 URL 静态化抉择建议

并不是所有的动态 URL 站点都需要进行 URL 静态化处理, 如果决策不当, 会对站点某一个阶段带来流量的重创。一般来说, 如果站点信息规模已经很庞大, 并且搜索引擎对其收录效果也不是很差的情况下, 就不需要 URL 伪静态化处理了。而那些站点信息量本来就少, 搜索引擎收录效果也很差的站点, 就必须要对其动态 URL 进行伪静态化处理。诸如国内的动网论坛, 虽然没有做 URL 的静态化处理, 但是搜索引擎对其索引效率十分好。

三、URL 的关键词部署

在 URL 里部署关键词的做法在英文语种的 SEO 工作中比较重视, 这里简单谈一下如何来操作。

比如在 <http://www.timev.com> 网站的 SEO 专栏下有一个网页, 这一网页主要是介绍 SEO 书籍的, 很多人可能会给它设定这么一个 URL: <http://www.timev.com/01/1.html>, 或许也可

能是这么一个 URL: <http://www.timev.com/a/a.html>。

如果这个 URL 被 SEO 部署了关键词后, 会是什么样的呢? 下面就是你想要的答案:

<http://www.timev.com/SEO/SEO-books.html>。

如果有人搜索引擎中搜索“SEO book”的时候, 以上网站地址中的关键词“SEO”和“SEO books”也被作为文本纳入了搜索引擎的文本分析中, 增加了用户搜索的“SEO book”的关键词密度, 这种做法有助增强关键词的搜索引擎排名竞争优势。

- 最好建议使用关键词或关键词组合构成 URL;
- URL 部分的各个关键词应使用“-”分隔;
- 如果有名词的关键词出现, 最好使用其的复数形式。

■ 为什么在关键词部署中, 名词要尽量使用复数形式?

如果你的关键词使用“books”, 当用户搜索“book”时, 关键词“books”中已经包含了它; 当用户搜索“books”, 正好是它。采用名词的复数形式能尽可能照顾用户的多样搜索习惯, 当然也可能会为你的网页赢得了更多被搜索用户找到的机会。

在这里值得指出的, 也有很多中文 SEO 人员也采取了网站 URL 的中文形式的关键词部署。譬如, 我们会经常看到的像 <http://www.xxx.com/数码相机.html> 类型的网站地址。一般来说, 服务器并不直接支持中文 URL, 所以 URL 中的汉字必须转换成对应的 UTF-8 编码, Google 能正常识别 UTF-8 编码的汉字, 而像百度这样的本土化中文搜索引擎则不行。

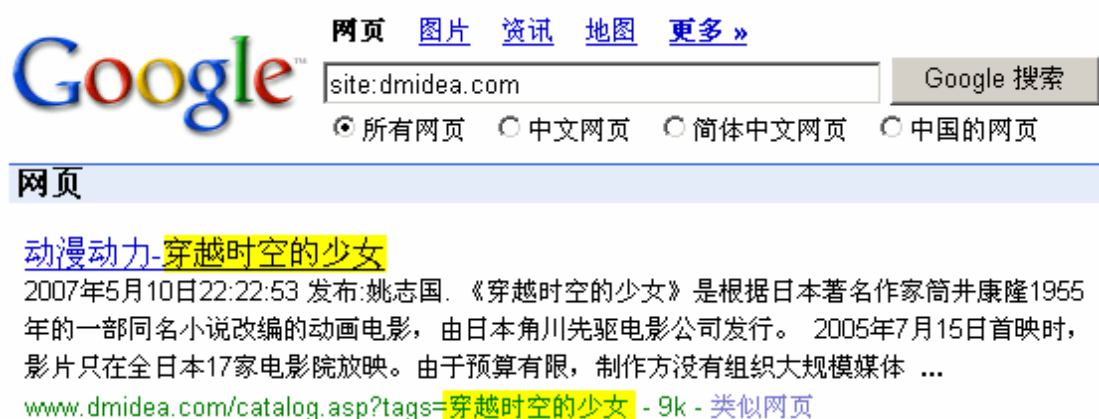


图 6-6-1 Google 识别 URL 中汉字的情况



图 6-6-2 百度识别 URL 中汉字的情况

从上面两个截图中，我们可以看出 Google 能识别 URL 中的汉字 UTF-8 编码。而百度则不会识别，其显示出了乱码。

在前些年，这种做法确实能对网页的搜索引擎排名起到作用，像阿里巴巴的某些专题站点就这样操作过。如今这样的方式似乎并不奏效，因为我们再也看不到那种带有汉字关键词的 URL 高居搜索结果榜首的局面了。

第七节 网页代码的编写

网页代码的编写是否简洁和具有逻辑性也是评估搜索引擎优化工作的一个重要指标。

一、遵循 WEB 标准

建议广大网页设计师遵循国际互联网标准组织（W3C）所推荐的 WEB 标准来编写网页源码，而不是继续沿用传统的 TABLE 表格布局方式来制作网页。

Web 标准是一些规范的集合，是由 W3C 和其他的标准化组织共同制定的，用它来创建和解释网页的基本内容。这些规范是专门为了那些在网上发布的可向后兼容的文档所设计的，使其能够被大多数人所访问。

遵循 WEB 标准来编写网页，可以从以下方面获益：

更简易的开发与维护 使用更具有语义和结构化的 HTML，将使得理解他人编写的代码更为容易和快速。

与未来浏览器的兼容 使用已定义的标准和规范的代码，向后兼容的文本就消除了不能被未来的浏览器识别的后患。

更快的网页下载、读取速度 更少的 HTML 代码带来的将是更小的文件和更快的下载速度。如今的浏览器处于标准模式下比它在以前的兼容模式下拥有更快的网页读取速度。

更好的可访问性 语义化的 HTML（结构和表现相分离）让使用浏览器以及不同浏览设备的读者都能很容易地看到内容。

更高的搜索引擎排名 内容和表现的分离使内容成为一个文本的主体。与语义化的标记结合会提高你在搜索引擎中的排名。

更好的适应性 一个用语义化标记的文档可以很好地适应于打印和其他的显示设备（掌上电脑和智能电话），这一切仅仅是通过链接不同的 CSS 文件就可以完成。同样也可以仅仅通过编辑单独的一个文件就完成跨站点般的表现上的转换。

Web 标准可以为网站的创建者节省时间与金钱，还可以为网站的浏览者提供一个更好的经历。如果想了解更多关于 WEB 标准的知识，推荐访问站点：www.w3cn.org。

二、代码的逻辑性

搜索引擎是通过分析网页源代码来分析页面文本信息的逻辑性，所以在编写网页代码的时候一定要尽可能使用合适的标签来体现文本表达的层次感，也即是让搜索引擎“看懂”哪些文本信息要重于其他部分。

1、标题标签

在 Html 代码中，标题的标签一共 6 个，它们的具体代码表现形式分别是：“<h1>标题 1</h1>”、“<h2>标题 2</h2>”、“<h3>标题 3</h3>”、“<h4>标题 4</h4>”、“<h5>标题 5</h5>”和“<h6>标题 6</h6>”，其中标题 1 的逻辑权重最大，其他的标题重要性随着 h 后的数字增大而减小。在搜索引擎的文本分析中，标题的信息权重要比文章正文的大，所以我们尤其要注意内容标题的规划和书写。

一篇文章中可能会出现很多标题，诸如正文大标题、副标题、段落标题等。我们应该根据标题的逻辑重要性分配合适的标题标签，正文大标题应该是<h1>，副标题可是是<h2>，而段落标题则可分配<h3>，其他的就可以以此类推了。我们不建议所有的标题都采用<h1>标签，这样不能向搜索引擎阐述哪些标题应该比其他的标题更加重要，因而显得标题没有层次逻辑性。

2、强调标签

为了说明文章正文部分的某些字词或者句子的重要性，我们应该要通过 html 强调代码来标识它们，从而使其与其他部分的字词或者句子区分开来。

一般强调标签：，Html 代码表现形式：“强调的字词或句子”，被标签包括了的字词或句子在网页中表现为斜体。

重点强调标签：和，具体 Html 代码表现形式：“强调的字词或句子”和“强调的字词或句子”，被重点强调标签包括了字词或句子在网页中表现为粗体。主要侧重于表现视觉上的强调，而则是指语意上的强调。

以上所谈的两类标签表达的文本逻辑性权重很大，所以其配合表现的字词句段在搜索引擎的文本分析中占有极大的评估优势，这也就不难理解为什么很多的 SEO 对它们十分重视了。网页设计师在编写代码的时候应当尽量使用丰富多样的标签，让网页的文本信息具有逻辑性，从而让搜索引擎能更好的读“懂”网页中的信息。对于搜索引擎优化人员来说，在这些标签中部署关键词是一个不错的选择。

第八节 面向搜索引擎的文案写作指导

文案写作在搜索引擎优化技能中占有很大的份量，专业的搜索引擎优化人员必须具有很娴熟的文字写作能力。在本节，我们将会向你介绍如何来撰写针对搜索引擎的文案。

一、网页标题 Title

通过下面的这张图片，可以看出 Title 在网页浏览器和 Html 源代码中的显示位置。



图 6-8-1 Title 在网络浏览器和 Html 源代码中的显示位置

由于网页的 Title 部分内容在搜索引擎的文本分析中占有很高的权重，所以网页 Title 写作质量的好坏对页面的优化效果有很大的影响。

1、Title 的写作原则

描述要简练有力，总字符数建议不要超过 60，而且要有核心关键词出现。切记勿出现堆积关键词的现象，Title 部分的行文描述一定要流畅。

■ 为什么要建议 Title 部分的字符数不要超过 60？



图 6-8-2 百度搜索网页 Title 显示实例

以上是一个网页在百度中的搜索结果展示效果图，蓝色部分为该网页的 Title 展示效果。细心的你或许已经发现，该网页的 Title 在搜索结果中展示并不完全，后面有两个小点来代替省略的部分。通过搜索结果的链接，可以看到实际网页的 Title 为“为什么会 IIS 连接数和论坛在线不符合的情况？-赢时代网络营销门户”，在搜索结果中被省略了 2 个汉字。60 个字符也相当与 30 个汉字，这个建议数字也只是根据经验总结出来的，其实只要把 Title 的字数控制在搜索引擎结果能完全展示出来的范围即可。

2、首页 Title 的写作原则

首页为网站最重要的入口，其重要性不言而喻。首页 Title 的构成形式为：“网站名称-网站内容简要描述或口号”。如果中横线分隔开的两个部分的总字数小于 30 的话，两个部分

的位置可以随意调换；如果两个部分的总字数超过 30 的话，可以把网站简要描述部分的内容放置于网站名称的前面。首页的 Title 部分要部署核心关键词，但不可泛滥堆积关键词。

- 值得参考的网站首页 Title 写法举例：



图 6-8-3 <http://www.timev.com/> 首页 Title 截图

- 不建议的首页 Title 写法举例：

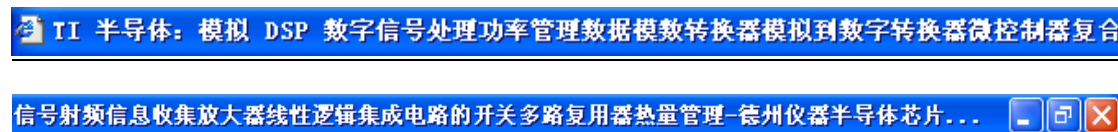


图 6-8-4 <http://www.ti.com.cn> 首页 Title 截图

3、网站栏目页的 Title 习作原则

站点栏目页的 Title 的构成形式为：“栏目名称-网站名称”或“栏目名称-栏目简要描述”。同样也应该根据栏目页的内容提炼出 1~3 个关键词布置于 Title 中。

- 建议的栏目页 Title 写法举例：



图 6-8-5 <http://www.timev.com/news/> 栏目页 Title 截图

- 不建议的栏目页 Title 写法举例：



图 6-8-6 不同栏目页使用相同 Title

4、最终内容呈现页面 title 写作

内容页面的 Title 构成形式为：“文章标题-网站名称”或者“文章标题”。

- 建议的最终页面 Title 写法举例：

图 6-8-7 <http://www.timev.com/news/china/20070815/6441.html> 网页 Title 截图

- 不建议的最终页面 Title 写法举例：



图 6-8-8 不同最终页面使用相同 Title

二、Meta 元标签

Meta 元标签位于网页源代码<head></head>之间，它是用来在 HTML 文档中模拟 HTTP 协议的响应头报文。Meta 的属性有两种：name 和 http-equiv。对于搜索引擎优化人员来说，我们只重点关注 Meta 的 name 属性。

Meta 的 name 属性主要用于描述网页，对应于网页内容，以便于搜索引擎机器人查找和分类。其中，对于 SEO 来说最重要的是 Description（站点在搜索引擎上的描述）、Keywords（分类关键词）和 Robots（机器人对具体页面进行索引的情况），所以应该给每页完善 Meta 值。

2005 年之前，Meta 部分的内容对搜索引擎优化的最终效果能起到很大作用，很多 SEO 人员在 Meta 的 Keywords 和 Description 部分泛滥堆积关键词来获得好的关键词排名，造成很多并不能给用户有价值信息的网页在搜索结果中排名很靠前，此举已经给用户的搜索体验造成了极大的伤害。后来，这种排名作弊行为被搜索引擎察觉，在其新一轮的算法更新中，搜索引擎淡化了 Meta 的 Keywords 和 Description 部分信息在排名因素中的权重。

虽然 Meta 部分的信息没有以前那么重要，其对排名效果影响并不那么大，但是搜索引擎仍然会把其作为评定网页质量的一个因素，所以我们应该尽可能完善每个页面的 Meta 值。

接下来我们就介绍以上三个 meta 值的书写方法：

1、Keywords

Keywords 的具体网页源码表现形式为：<meta name="keywords" content=" ">。

Keywords 向搜索引擎说明你的网页的关键词。每个页面的源代码 Keywords 的 content 部分填写 1-3 个关键词，关键词之间请以半角逗号分来。Keywords 部分的关键词应从当前页面的主要展示内容信息中提炼而来。

2、Description

Description 具体网页源码表现形式为：<meta name="description" content=" ">。

Description 告诉搜索引擎你的站点上的页面的主要内容。Description 的 content 部分文字内容最好控制在 100 个字符左右，行文流畅。Description 部分的内容应该是具体页面主要展示信息的简洁而强有力的概述。

3、Robots

Robots 的具体网页源码表现形式为：<meta name="robots" content=" ">。

Robots 的 content 部分文字可以填写的内容为：all、none、index、noindex、follow 和 nofollow。对于 Robots 的 content 部分填充内容有如下说明解释：

- 设定为 all：文件将被检索，且页面上的链接可以被查询；
- 设定为 none：文件将不被检索，且页面上的链接不可以被查询；
- 设定为 index：文件将被检索；
- 设定为 follow：页面上的链接可以被查询；
- 设定为 noindex：文件将不被检索，但页面上的链接可以被查询；
- 设定为 nofollow：文件将不被检索，页面上的链接可以被查询。

一般来说，大部分网站并没有太多的内容禁止搜索引擎机器人索引，Robots 的通用写法都是：<meta name="robots" content="all">。如果想对 Robots 做一些个性化的设置，可以参照上面的相关说明来完善 Robots 的 content 部分内容。

三、图片 ALT 替换文本

图片有一个 ALT 替换文本属性，在代码中填充某一个图片的 ALT 替换文本信息，如果这个图片在网页上显示不出来，在该图片的位置会出现替换文本的信息。

图片的 ALT 替换文本信息属性原本是考虑到残障人士（如盲人）的访问体验而准备的，他们在浏览网页的时候很难获取图片中的文字信息，只能通过 ALT 替换文本的描述来了解图片影像内容。

搜索引擎会把 ALT 部分的内容纳入整个页面的文本分析，所以建议在填写图片 ALT 部分的内容时要灵活地在描述中布置合适的关键词。

四、Robots.txt

搜索引擎都有自己的“搜索机器人”（Robots），并通过这些 Robots 在网络上沿着网页上的链接（一般是 http 和 src 链接）不断抓取资料建立自己的数据库。

对于网站管理者和内容提供者来说，有时候会有一些站点内容，不希望被 Robots 抓取公开。为了解决这个问题，Robots 开发界提供了两个办法：一个是 robots.txt，另一个是 The Robots META 标签。

Robots META 标签已经在前面有所叙述，这个部分我们将不再说明。

1、什么是 robots.txt?

robots.txt 是一个纯文本文件，通过在这个文件中声明该网站中不想被 robots 访问的部

分，这样，该网站的部分或全部内容就可以不被搜索引擎收录了，或者指定搜索引擎只收录指定的内容。

当一个搜索机器人访问一个站点时，它会首先检查该站点根目录下是否存在 robots.txt，如果找到，搜索机器人就会按照该文件中的内容来确定访问的范围，如果该文件不存在，那么搜索机器人就沿着链接抓取。

robots.txt 必须放置在一个站点的根目录下，而且文件名必须全部小写。

一般来说相应的网站就会拥有一个对应的 robots.txt 文件，如：

http://www.w3.org/

http://www.w3.org/robots.txt

2、robots.txt 的语法

“robots.txt”文件包含一条或更多的记录，这些记录通过空行分开（以 CR，CR/NL， or NL 作为结束符），每一条记录的格式如下所示：

“<field>:<optionalspace><value><optionalspace>”。

在该文件中可以使用#进行注解，具体使用方法和 UNIX 中的惯例一样。该文件中的记录通常以一行或多行 User-agent 开始，后面加上若干 Disallow 行，详细情况如下：

■ User-agent:

该项的值用于描述搜索引擎 robot 的名字，在"robots.txt"文件中，如果有多条 User-agent 记录说明有多个 robot 会受到该协议的限制，对该文件来说，至少要有一条 User-agent 记录。如果该项的值设为*，则该协议对任何机器人均有效，在"robots.txt"文件中， "User-agent: *" 这样的记录只能有一条。

■ Disallow:

该项的值用于描述不希望被访问到的一个 URL，这个 URL 可以是一条完整的路径，也可以是部分的，任何以 Disallow 开头的 URL 均不会被 robot 访问到。例如"Disallow: /help"对/help.html 和/help/index.html 都不允许搜索引擎访问，而"Disallow: /help/"则允许 robot 访问/help.html，而不能访问/help/index.html。

任何一条 Disallow 记录为空，说明该网站的所有部分都允许被访问，在"/robots.txt"文件中，至少要有一条 Disallow 记录。如果 "/robots.txt"是一个空文件，则对于所有的搜索引擎 robot，该网站都是开放的。

3、一些 robots.txt 基本的用法：

A. 禁止所有搜索引擎访问网站的任何部分：

User-agent: *

Disallow: /

B. 允许所有的 robot 访问

User-agent: *

Disallow:

或者也可以建一个空文件 "/robots.txt" file

C. 禁止所有搜索引擎访问网站的几个部分（下例中的 cgi-bin、tmp、private 目录）

User-agent: *

Disallow: /cgi-bin/

Disallow: /tmp/

Disallow: /private/

D. 禁止某个搜索引擎的访问（下例中的 BadBot）

User-agent: BadBot

Disallow: /

E. 只允许某个搜索引擎的访问（下例中的 WebCrawler）

User-agent: WebCrawler

Disallow:

User-agent: *

Disallow: /

F. 常见搜索引擎机器人 Robots 名字

表 6-8-1 常见搜索引擎机器人 Robots 名字

名 称	搜 索 引 擎
Baiduspider	http://www.baidu.com
Scooter	http://www.altavista.com
ia_archiver	http://www.alexa.com
Googlebot	http://www.google.com
FAST-WebCrawler	http://www.alltheweb.com
Slurp	http://www.inktomi.com
MSNBOT	http://search.msn.com

4、robots.txt 举例

下面是一些著名站点的 robots.txt:

http://www.cnn.com/robots.txt

http://www.google.com/robots.txt

http://www.ibm.com/robots.txt

http://www.sun.com/robots.txt

http://www.eachnet.com/robots.txt

5、常见 robots.txt 错误

A. 颠倒了顺序

错误写成:

User-agent: *

Disallow: GoogleBot

正确的应该是：

```
User-agent: GoogleBot
```

```
Disallow: *
```

B. 把多个禁止命令放在一行中

例如错误写成：

```
Disallow: /css/ /cgi-bin/ /images/
```

正确的应该是：

```
Disallow: /css/
```

```
Disallow: /cgi-bin/
```

```
Disallow: /images/
```

C. 行前有大量空格

例如写成：

```
Disallow: /cgi-bin/
```

尽管在标准没有谈到这个，但是这种方式很容易出问题。

D. 404 重定向到另外一个页面

当 Robot 访问很多没有设置 robots.txt 文件的站点时，会被自动 404 重定向到另外一个 Html 页面。这时 Robot 常常会以处理 robots.txt 文件的方式处理这个 Html 页面文件。虽然一般这样没有什么问题，但是最好能放一个空白的 robots.txt 文件在站点根目录下。

E. 采用大写。例如

```
USER-AGENT: EXCITE
```

```
DISALLOW:
```

虽然标准是没有大小写的，但是目录和文件名应该小写：

```
user-agent:GoogleBot
```

```
disallow:
```

F. 语法中只有 Disallow，没有 Allow

错误的写法是：

```
User-agent: Baiduspider
```

```
Disallow: /john/
```

```
allow: /jane/
```

G. 忘记了斜杠/

错误写作：

```
User-agent: Baiduspider
```

```
Disallow: css
```

正确的应该是：

```
User-agent: Baiduspider
```

```
Disallow: /css/
```

下面一个小工具专门检查 robots.txt 文件的有效性：

<http://www.searchengineworld.com/cgi-bin/robotcheck.cgi>

五、链接锚文本

什么是链接锚文本？举个例子解释说明一下：诸如，有一个网站做了个点石的友情链接，其具体代码表现为 `点石互动`，其中“点石互动”即为该链接的锚文本。

1、链接锚文本（Anchor Text）描述在 SEO 中有什么作用？

搜索引擎可以根据指向某一个网页链接的锚文本描述来判断该网页的内容属性，点石是一个搜索引擎优化研究团队，所以有很多指向点石的链接锚文本描述为“搜索引擎优化”，这样会让搜索引擎根据属性投票来给点石在搜索引擎优化的自然排名方面有一个考虑，尤其是这样的反向链接的质量和数量上升到一定层次时，这种排名优势也就越明显。

很多网站的锚文本的形式诸如以下形式：`more`、更多、阅读详细信息、点击这里等，这样的做法没办法让搜索引擎对某一个被链接的网页内容属性有一个好的判断，从而影响网页排名。

2、链接锚文本（Anchor Text）的优化

每个页面的内容都不同，因此我们在进行关键词策略部署的时候，就应该针对某一个目标页面有一个关键词考虑。一般来说一个页面的核心关键词在 1-3 个范围（范围太大，关键词的权重就有可能被稀释），所以我们一般在其他页面做指向该页面的链接的锚文本描述中就应该布置预先规划的核心关键词，而且锚文本描述的形式应尽量多样化。

六、Sitemap.xml

Sitemaps 协议使你能够告知搜索引擎网站中可供抓取的网址。最简便的方式就是，使用 Sitemaps 协议的 Sitemaps 就是列有某个网站所有网址的 XML 文件。此协议可高度扩展，因此可适用于各种大小的网站。它还能够使网站管理员提供有关每个网址的其他信息（上次更新的时间、更改的频率、与网站中其他网址相比它的重要性等），以便搜索引擎可以更智能地抓取该网站。

Sitemaps 协议补充而不是取代搜索引擎已用来发现网址的基于抓取的机制。通过向搜索引擎提交一个 Sitemaps（或多个 Sitemaps），可帮助搜索引擎更好地抓取你的网站。

1、XML Sitemaps 语法格式

Sitemaps 协议格式由 XML 标记组成，Sitemaps 的所有数据数值应为实体转义过的，文件本身应为 UTF-8 编码的。

以下是只包含一个网址并使用所有可选标记的 Sitemaps 示例。可选标记为斜体。

```
<?xml version="1.0" encoding="UTF-8"?>
< urlset xmlns="http://www.google.com/schemas/sitemap/0.84">
  < url>
    < loc>http://www.example.com/</loc>
    < lastmod>2005-01-01</lastmod>
    < changefreq>monthly</changefreq>
```

```
< priority>0.8</priority>
</url>
</urlset>
```

Sitemaps 应:

- 以 <urlset> 开始标记开始，以 </urlset> 结束标记结束。
- 每个网址包含一个作为父标记的 <url> 条目。
- 每一个 <url> 父标记包括一个 <loc> 子标记条目。

2、XML 标记定义

以下对可用 XML 标记进行说明。

表 6-8-2 XML 标记

<urlset>	必填	封装此文件并提供当前协议标准作为参考。
<url>	必填	每个网址条目有一个父标记。剩余标记为此标记的子标记。
<loc>	必填	该页的网址。如果你的 Web 服务器需要网址的话，此网址应以协议开始（例如： http ）并以斜线结尾。此值应少于 2048 个字符。
<lastmod>	可选	该文件上次修改的日期。此日期应采用 W3C Datetime 格式。如果需要的话，此格式允许省略时间部分，而仅使用 YYYY-MM-DD。
<changefreq>	可选	<p>页面可能发生更改的频率。此值为搜索引擎提供一般性信息，可能与搜索引擎抓取页面的频率不完全相关。有效值为：</p> <ul style="list-style-type: none"> • always • hourly • daily • weekly • monthly • yearly • never <p>值“always”应当用于描述每次访问时都会改变的文档。而值“never”应当用于描述已存档网址。</p> <p>请注意此标记的值被视为提示而不是命令。尽管搜索引擎抓取工具在做决定时考虑此信息，但是它们对标记为“hourly”页面的抓取频率可能低于每小时一次，而对标记为“yearly”页面的抓取频率可能高于每年一次。抓取工具也可能定期抓取标记为“never”的页面，以便它们能够处理对这些页面的意外更改。</p>
<priority>	可选	此网址的优先级与你网站上其他网址的优先级相关。有效值范围从 0.0 到 1.0。此值不会影响你的网页与其他网站上网页的比较结果，只是告诉搜索引擎你认为你的那个网页最重要，从而它们对你页面的抓取可以按照你最喜欢的方式进行排序。

	<p>一个网页的默认优先级为 0.5。</p> <p>请注意，你为页面指定的优先级不影响你的网址在搜索引擎的结果页的排名。搜索引擎在同一网站的不同网址之间进行选择时使用此信息，因此，你可以使用此标记来增加相对重要的网址在搜索索引中显示的可能性。</p> <p>而且，请注意为你网站中的所有网址都指定高优先级对你没什么好处。由于优先级是相互关联的，只用于在你自己网站的网页之间进行选择；你网页的优先级不会与其他网站的页面优先级进行比较。</p>
--	---

3、实体转义

Sitemaps 文件为 UTF-8 编码（通常在保存文件时可以做到）。对于所有的 XML 文件，任何数据数值（包括网址）都应对下表中列出的字符使用实体转义码。

表 6-8-3 实体转义

字 符		转 义 码
& 符号	&	&
单引号	'	'
双引号	"	"
大于	>	>
小于	<	<

此外，所有的网址（包括 Sitemaps 的网址）都应编码，以便它们所在的，以及网址转义的 Web 服务器识别。然而，如果使用任何脚本、工具或日志文件来生成网址（除了手动输入之外的任何方法），通常这部分已经替你完成了。如果提交了 Sitemaps 却收到 Google 无法找到某些网址的错误消息，请查看并确保你的网址遵循 RFC-3986 URI 标准、RFC-3987IRI 标准以及 XML 标准。

这是使用非 ASCII 字符 (ü) 以及要求实体转义的字符 (&) 的网址示例：

`http://www.example.com/ümlat.html&q=name`

这是 ISO-8859-1 编码的（用于使用该编码的服务器上的托管）以及网址转义过的同一网址：

`http://www.example.com/%FCmlat.html&q=name`

这是 UTF-8 编码（用于使用该编码的服务器上的托管）且网址转义过的同一网址：

`http://www.example.com/%C3%BCmlat.html&q=name`

这是实体转义过的同一网址：

`http://www.example.com/%C3%BCmlat.html&q=name`

4、XML Sitemaps 示例

下例显示了一个 XML 格式的 Sitemaps。示例中的 Sitemaps 包含少数网址，每一个都用 <loc> XML 标记来识别。此示例中，为每个网址提供了一组不同的可选参数。

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.google.com/schemas/sitemap/0.84">

  <url>
    <loc>http://www.example.com/</loc>
    <lastmod>2005-01-01</lastmod>
    <changefreq>monthly</changefreq>
    <priority>0.8</priority>
  </url>
  <url>
    <loc>http://www.example.com/catalog?item=12&amp;desc=vacation_hawaii</loc>
    <changefreq>weekly</changefreq>
  </url>
  <url>
    <loc>http://www.example.com/catalog?item=73&amp;desc=vacation_new_zealand</loc>
    <lastmod>2004-12-23</lastmod>
    <changefreq>weekly</changefreq>
  </url>
  <url>
    <loc>http://www.example.com/catalog?item=74&amp;desc=vacation_newfoundland</loc>
    <lastmod>2004-12-23T18:00:15+00:00</lastmod>
    <priority>0.3</priority>
  </url>
  <url>
    <loc>http://www.example.com/catalog?item=83&amp;desc=vacation_usa</loc>
    <lastmod>2004-11-23</lastmod>
  </url>
</urlset>
```

你可以使用 gzip 压缩 Sitemaps 文件。压缩 Sitemaps 文件会降低对带宽的要求。请注意，未压缩的 Sitemaps 文件不得大于 10MB。

5、使用 Sitemaps 索引文件（对多个 Sitemaps 文件进行分组）

你可以提供多个 Sitemaps 文件，但提供的每个 Sitemaps 文件包括的网址不得超过 50,000 个，并且未压缩时不能大于 10MB（10,485,760）。这些限制条件有助于确保 Web 服务器不会因传输非常大的文件而遇到麻烦。

如果要列出超过 50,000 个网址，你需要创建多个 Sitemaps 文件。如果预计你的 Sitemaps 网址数量会超过 50,000 个或大小超过 10MB，应考虑创建多个 Sitemaps 文件。如果你的确提供了多个 Sitemaps，你可以将其列在 Sitemaps 索引文件中。Sitemaps 索引文件只能列出不超过 1,000 个 Sitemaps。

Sitemaps 索引文件的 XML 格式与 Sitemaps 文件的 XML 格式非常相似。Sitemaps 索引文件使用以下 XML 标记：

- loc
- lastmod
- sitemap
- sitemapindex

注：Sitemaps 索引文件只能指定与其位于同一网站的 Sitemaps。例如，
http://www.yoursite.com/sitemap_index.xml 可包含位于 http://www.yoursite.com 的 Sitemaps，但不能包含位于 http://www.example.com 或 http://yourhost.yoursite.com 的 Sitemaps。与 Sitemaps 一样，你的 Sitemaps 索引文件应为 UTF-8 编码。

6、XML Sitemaps 索引示例

下例展示了一个 XML 格式的 Sitemaps 索引。该 Sitemaps 索引列出了两个 Sitemaps：

```
<?xml version="1.0" encoding="UTF-8"?>
  <sitemapindex xmlns="http://www.google.com/schemas/sitemap/0.84">
    <sitemap>
      <loc>http://www.example.com/sitemap1.xml.gz</loc>
      <lastmod>2004-10-01T18:23:17+00:00</lastmod>
    </sitemap>
    <sitemap>
      <loc>http://www.example.com/sitemap2.xml.gz</loc>
      <lastmod>2005-01-01</lastmod>
    </sitemap>
  </sitemapindex>
```

备注：与 XML 文件中的所有值一样，Sitemaps 网址需要实体换码。

Sitemaps 索引 XML 标记定义

- 需要有 <loc> 标记并用其来识别 Sitemaps 的位置。
- <lastmod> 标记是可选标记，用于指示相应 Sitemap 文件的修改时间。它并不对应于该 Sitemap 中列出的任一网页的更改时间。lastmod 标记的值应采用 W3C Datetime 格式。通过提供最近修改的时间戳启用搜索引擎抓取工具，抓取工具将只检索索引中的 Sitemaps 的一个子集，也就是说，抓取工具只检索某特定日期之后修改的 Sitemaps。通过这一递增的 Sitemaps 提取机制，可以快速发现超大型网站上的新网址。
- <sitemap>标记封装单个 Sitemaps 的相关信息。
- <sitemapindex>标记会压缩有关文件中的所有 Sitemaps 的信息。

7、Sitemaps 文件的位置

Sitemaps 文件的位置决定该 Sitemaps 中所能包含的一组网址。位于 http://example.com/catalog/sitemap.gz 的 Sitemaps 文件可包含以 http://example.com/catalog/ 开始的任何网址，但不能包含以 http://example.com/images/ 开始的网址。

如果有更改 http://example.org/path/sitemap.gz 的许可，你应该也可以提供以 http://example.org/path/ 作为前缀的网址的信息。

http://example.com/catalog/sitemap.gz 中，被认为有效的示例网址包括：

http://example.com/catalog/show?item=23

http://example.com/catalog/show?item=233&user=3453

http://example.com/catalog/sitemap.gz 中被认为无效的网址包括：

http://example.com/image/show?item=23

http://example.com/image/show?item=233&user=3453

https://example.com/catalog/page1.html

被认定为无效的网址将不再考虑。强烈建议将 Sitemaps 放置在 Web 服务器的根目录处。例如，如果 Web 服务器位于 example.com，则 Sitemaps 索引文件应位于 http://example.com/sitemap.gz。在某些情况下，你需要针对不同的路径创建对应的 Sitemaps，例如，如果在你的组织中，安全许可对应不同目录划分上传权限。

8、验证你的 Sitemaps

Google 使用 XML 架构定义可以出现在 Sitemaps 文件中的元素和属性。你可从以下链接下载此架构：

对于 Sitemaps: <http://www.google.com/schemas/sitemap/0.84/sitemap.xsd>

对于 Sitemaps 索引文件: <http://www.google.com/schemas/sitemap/0.84/siteindex.xsd>

有多种工具可帮助你根据此架构来验证你的 Sitemaps 结构。在下面的每一个位置你都可以找到 XML 相关的工具列表：

<http://www.w3.org/XML/Schema#Tools>

<http://www.xml.com/pub/a/2000/12/13/schematools.html>

为了根据某个架构验证你的 Sitemaps 或 Sitemaps 索引文件，XML 文件需要有附加的标头。如果你使用的是 Sitemaps 生成器，这些标头已经包含其中。如果你使用不同的工具创建 Sitemaps，XML 文件中的标头应如下例所示。

Sitemaps:

```
<?xml version='1.0' encoding='UTF-8'?>
  <urlset xmlns="http://www.google.com/schemas/sitemap/0.84"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.google.com/schemas/sitemap/0.84
    http://www.google.com/schemas/sitemap/0.84/sitemap.xsd">

    <url>
    ...
    </url>
  </urlset>
```

Sitemaps 索引文件:

```
<?xml version='1.0' encoding='UTF-8'?>
  <sitemapindex xmlns="http://www.google.com/schemas/sitemap/0.84"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.google.com/schemas/sitemap/0.84
```

```
http://www.google.com/schemas/sitemap/0.84/siteindex.xsd">  
  
<sitemap>  
...  
</sitemap>  
</sitemapindex>
```

9、Google sitemaps Builder .Net V1.5 使用教程

这是一款由国外编程人员开发的 Google Sitemap 制作软件，通过该软件那些不懂 XML 语法的人也很容易制作出 xml 版本的 Google Sitemap 和纯文本链接地址。

该软件可运行于 Windows 2000 / XP / 2003 平台，必须在 .NET Framework 1.1 环境下安装，如果你的电脑没有安装 .NET Framework 1.1，你可以到微软官方网站上下载

(<http://www.microsoft.com/downloads/details.aspx?FamilyID=262D25E3-F589-4842-8157-034D1E7CF3A3&displaylang=zh-cn;>)。

Google sitemaps Builder .Net V1.5 软件可以到下面的网站地址下载：

<http://www.timev.com/download/google-sitemaps-builder.html>

Google sitemaps Builder 软件的安装相当容易，这里就不做介绍了。接下来，主要来讲解一下 Google sitemaps Builder 的使用方法。

安装完 Google sitemaps Builder 软件后，点击任务栏里的“开始”，从“所有程序”里找到“Sitemaps Builder .Net”，运行该程序。当然，安装完 Google sitemaps Builder 软件后，电脑的桌面会有其快捷方式，直接点击它可快速运行该软件。

下图是 Google sitemaps Builder 软件的操作界面：

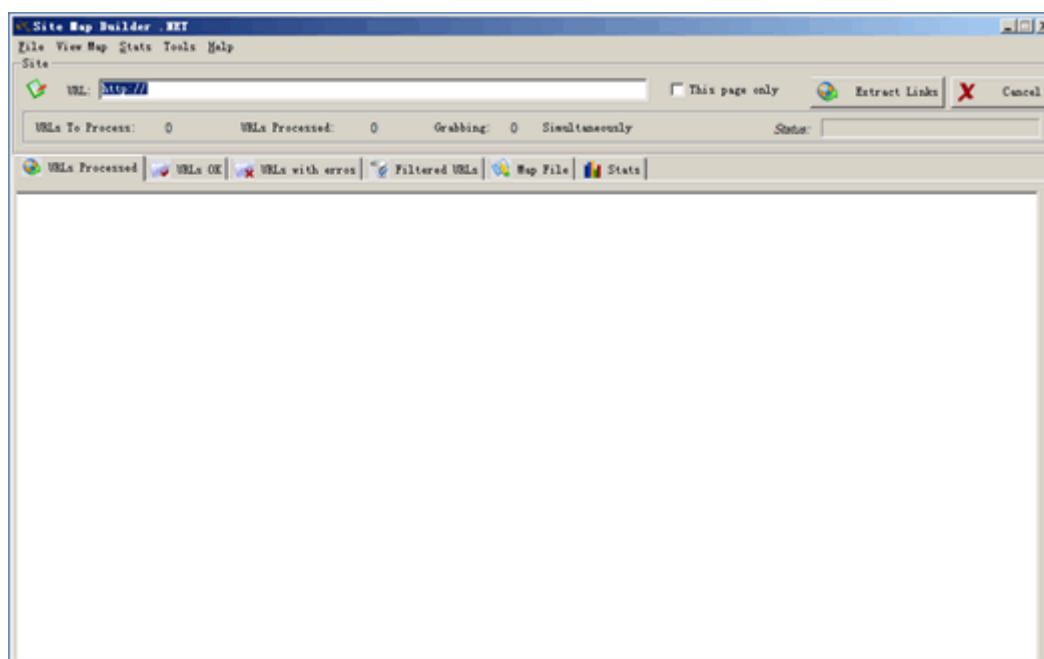


图 6-8-9 Google sitemaps Builder 软件操作界面

我们可以在 URL 输入框里输入目标网站首页地址，如下图所示：

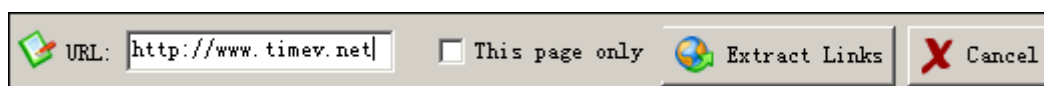


图 6-8-10 输入示例

然后点击“Extract Links”按钮，软件就会派出其机器人去目标网站上进行爬行，从而捕获目标网站的链接信息。如果你要停止该软件的机器人爬行，可以点击“Cancel”按钮。

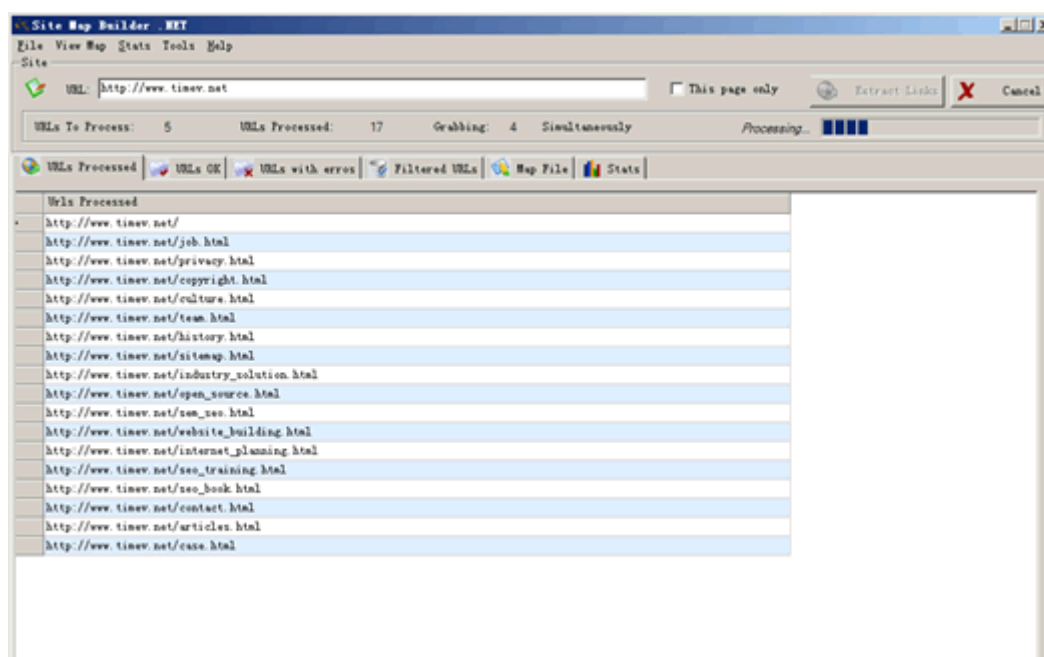


图 6-8-11 Google sitemaps Builder 软件运行过程中的截图

等 Google sitemaps Builder 的机器人爬行完整个站点后，我们就可以轻松获取 Google sitemap 文件了。

在软件界面中点击“Map File”，然后会出现下面的界面：

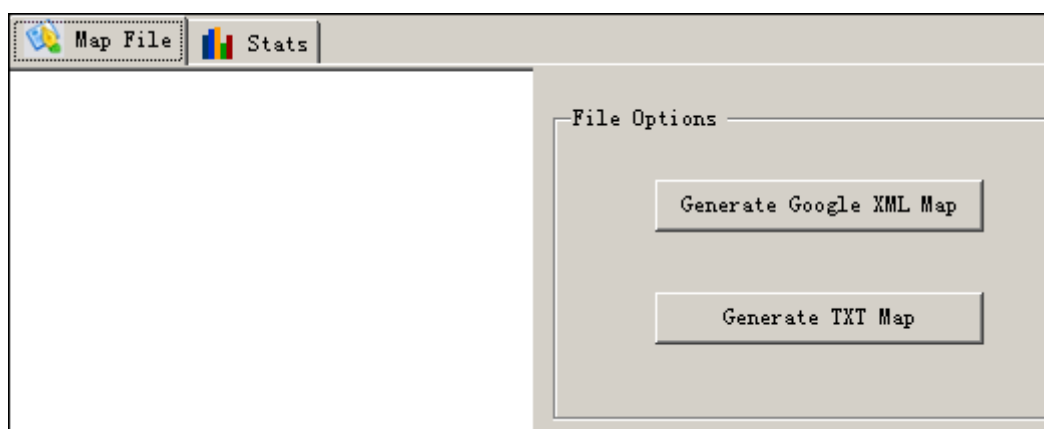


图 6-8-12 Map File 界面

当我们点击“File Options”部分的“Generate Google XML Map”按钮，我们就可以在左侧窗口获得目标查询站点的网站地图 XML 源代码。我们可以用记事本创建一个文件

sitemap.xml，然后将获得的 XML 源代码拷贝到 xml 文件中然后保存，这样我们就得到了一个 Google sitemap 文件了。

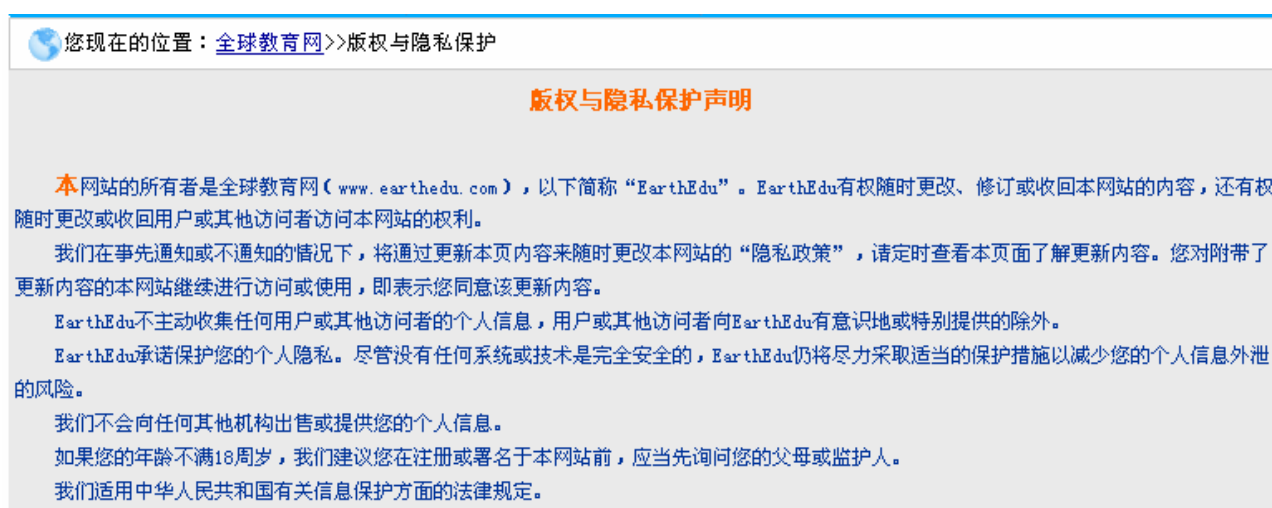
点击“Generate TXT Map”按钮，我们可以获得网站的所有网页地址，我们可以把这些网页地址制作成 HTML 文件格式的网站地图。

这款软件还有一个值得提出的功能，就是它可能帮助我们检查网站是否存在死链接。具体链接检查情况，我们可以点击软件界面上的“Stats”按钮来了解。如果你想了解你的网站上哪些 URL 不可访问，可以点击“URLs with erros”按钮来查看。

接下来，你只需要向 Google 提交的制作好的 sitemap 文件了。

七、版权与隐私

国外的网站大部分都有版权和隐私保护的说明页面，而在国内大部分网站却很少见这样的说明页面。对版权和对用户隐私尊重的网站才可以获得用户的信任。



您现在的位置：[全球教育网](#)>>版权与隐私保护

版权与隐私保护声明

本网站的所有者是全球教育网(www.earthedu.com)，以下简称“EarthEdu”。EarthEdu有权随时更改、修订或收回本网站的内容，还有权随时更改或收回用户或其他访问者访问本网站的权利。

我们在事先通知或不通知的情况下，将通过更新本页内容来随时更改本网站的“隐私政策”，请定时查看本页面了解更新内容。您对附带了更新内容的本网站继续进行访问或使用，即表示您同意该更新内容。

EarthEdu不主动收集任何用户或其他访问者的个人信息，用户或其他访问者向EarthEdu有意识地或特别提供的除外。

EarthEdu承诺保护您的个人隐私。尽管没有任何系统或技术是完全安全的，EarthEdu仍将尽力采取适当的保护措施以减少您的个人信息外泄的风险。

我们不会向任何其他机构出售或提供您的个人信息。

如果您的年龄不满18周岁，我们建议您在注册或署名于本网站前，应当先询问您的父母或监护人。

我们适用中华人民共和国有关信息保护方面的法律规定。

图 6-8-13 版权与隐私范例

第九节 用户友好性检测

我们一般通过三个指标来检验一个网站是否对于用户友好，这三个指标分别是：链接的可用性、访问速度体验和查找信息的便捷度。

一、链接的可用性

试想，一个访问者来到你的网站，点击一个超级链接，却发现浏览器只返回一个错误 404 页面。如果网页中不可用链接数过多，应该可以想像会是什么结果。一个网站如何能让访问者取得信任？最基本的一个原则就是确保网站的每个链接都可用。不要让访问者扫兴而归，是你应该做的。

二、访问速度体验

让用户有一个愉悦的访问速度体验，这是一个很理智的做法。根据大量的调查数据表明，

访问者能忍受的最大网页打开速度为 15 秒，如果你的网页打开速度能控制在 8 秒以内就最好不过了。但是，很多企业网站在这个方面却做得不尽如意。网页中充斥着大量的图片和 FLASH 动画，虽然能给人以较好的视觉效果，但这却是影响网页打开速度的最大因素。

不要因为网页的访问速度慢，而把很多用户挡在门外。

三、查找信息的便捷度

你的网站便于用户使用么？访问者是否能在你的网站上快速便捷找到他们想要的信息是检验网站可用性的检验标准。

我们可以通过做一个简单的实验来了解一下用户查找信息的便捷度：找一个对你网站不是很熟悉的人来访问目标测试站点，让他在网站里找出一个指定信息的网页（当然，这个页面应该是网站中的具体内容页面）。如果参与实验的人员通过 5 次以上的点击才找到目标页面，那就说明网站的可用性极差。

第七章 创建搜索引擎喜欢的内容

“内容为王”不仅是网站运营人员的金科玉律，同样也是 SEO 人员是否能获得显著优化效果的重要指导思想。网站的发展靠内容来支撑，但是一个网站的内容如何既能吸引访问用户，又能获得搜索引擎的喜欢呢？我们又应该如何去创造这些内容呢？在本章，我们会对这些问题进行解答。

第一节 搜索引擎喜欢什么内容

首先，我们来看看搜索引擎自己对被索引网站内容的说明和建议。

一、Google 对网站内容的说明：

向访问者提供他们要查找的信息 在网页上提供高品质的内容，尤其是主页。这是你要做的最重要的工作。如果你的网页包含有用的信息，其内容就可以吸引许多访问者并使网站管理员乐于链接到你的网站。要创建实用且信息丰富的网站，网页文字应清晰、准确地表述要传达的主题。想一想，用户会使用哪些字词来查找你的网页，然后尽量在网站上使这些字词。

以上信息摘录自 Google 网站管理员帮助中心。

二、百度对网站内容的说明：

创造属于你自己的独特内容 百度更喜欢独特的原创内容。所以，如果你的站点内容只是从各处采集复制而成，很可能不会被百度收录。

以上信息摘录自百度搜索帮助中心“给网站管理员的建议”。

另外，雅虎中文搜索和搜狗等中文搜索引擎也说明它们对原创独特内容的喜好。

第二节 内容的四大来源

很多人觉得为网站创建内容是一件很困难的事情，尤其是那些企业产品销售站点。在这里给大家介绍 4 种为网站创建内容的思路。

一、自己撰写原创文章

原创文章的内容主要包括自己对产品、服务、行业或者事件的看法。原创文章最受搜索引擎欢迎，但是自己撰写原创文章比较消耗时间和精力。

1、原创的目的

- 让搜索引擎知道你的网站是活跃的；
- 希望自己带网址链接的原创文章能被网站管理员转载或文章采集器抓取；

- 让更多的同行、更多的客户认识和了解。

2、原创对网站优化的好处

- 活跃的网站随时有新鲜和充实的内容，这样的网站搜索引擎肯定会经常光顾，因此，原创文章可以加快网站更新；
- 带网址链接的原创文章容易被文章采集器抓取或网站管理员转载，自然会产生很多优秀的外部链接。

二、使用他人的内容

如果觉得其他站内的内容不错，可以跟这些站点进行内容合作，一旦你的站点获得了他们的转载授权后，就可以使用他们的内容了。目前大多数站点都采用这种内容创建形式。由于搜索引擎并不太喜欢这些转载的内容，所以，你因该尽可能转载那些并没有被大范围转载的内容。

1、复制内容（Duplicate Content）带来的麻烦

所谓复制内容，是你撰写了一篇文章发布于自己网站的某个网页上，别人浏览到这个网页时，觉得这篇文章有价值，然后复制这篇文章粘贴到了他的网页上。复制内容最常见的手法便是转载。

互联网的信息传播力是惊人的，优秀的网页信息内容往往会被复制得四处泛滥。那么，对于毫不知情的搜索用户来说，从诸多搜索结果中，如何区分哪一个网页才是真正承载原始信息的网页呢？

如果搜索引擎没有把承载了原始信息的网页优先呈现给搜索用户，而是给搜索用户反馈了大量复制原始信息的网页，这就违背了搜索引擎的基本准则。为什么这样说呢？

一般来说，网页上的信息都是通过 Html 语言标签进行标识过的，而复制内容的人大部分都是喜欢通过“CTRL+C”和“CTRL+V”来简单地完成复制粘贴工作，这样一来，原作者精心排版过的内容很可能在简单的复制粘贴过程中出现信息失真，这是对读者一种极其不负责任的做法。

比如，原作者可能会在原始网页的文章中导出了一些超级链接到相关参考网页上，如果在复制过程中丢失了这些链接信息，很可能会增加读者对文章信息的理解障碍。有些转载文章的人并不喜欢在转载的文末留下原始网页的 URL，如果原作者对文章内容进行了更正、修改和补充的话，读者也无法了解到更新的信息内容。而一些技术类的文档，包括举例源码、数据表格等内容，几经复制后已经完全丧失了信息的准确性，读者在阅读的时候不免留下诸多困惑。

此外，这种未经许可的不道德内容复制行为大大挫伤了原创作者的写作积极性。搜索引擎判断原始网页的权威性类似于科技论文中的引用机制：谁的论文被引用次数多，谁就是权威。在理想状态下，应该是原始网页的排名始终高于转载网页，在这种理想状态达到之前，原始网页的权威性总有可能受到伤害，担心总是在所难免的。

需要注意的是，内容被转载不会在很大程度上影响原始网页的权威性的前提是，内容的原创网站处于“正常状态”，即网站能保持正常的更新，不断添加具有一定质量的内容，同时，不断获得具有一定质量的反向链接。只有这样，当网站自身的权威性提高以后，当网站被搜索引擎视为可信任网站以后，原始网页受内容复制的负面影响便会越来越小。

2、“原创”转载的内容

“原创”转载的内容也叫文章编译。如果你看到一篇文章，觉得其中的有些观点自己不赞同，你可以把原文中的观点描述句段摘录出来，然后在下面书写自己的看法；或者，觉得文章中的观点还有拓展的空间，你也可以沿着原作者的思路把原文继续拓展写下去。不过在这里，请记住一个原则，那就是你自己的文字信息内容的比率要尽量大于摘录的信息内容。

3、鼓励用户贡献内容

其实，有时候网站的访问用户的也有强烈的写作交流欲望。所以，我们可以不断完善网站的交互功能，让用户参与到网站的信息内容建设中来。

用户投稿 给网站开通一个用户投稿提交接口，让用户通过这里把他们创作的内容提交到网站编辑人员那里。

靠用户创造内容的 WEB2.0 网站 Web 2.0 的应用可以让人了解目前网络正在进行的一种改变——从一系列网站到一个成熟的为最终用户提供网络应用的服务平台。

Web 2.0 表现的社会性网络的技术包括：博客（blog）、播客（podcasting）、BT、移动博客、P2P、社交网络（SNS）、RSS、博采（blogmark）、维客（wiki）、标签（Tag）等。WEB2.0 概念的最大的特征就是网站的内容由用户来创造

如果你想了解更多关于 WEB2.0 的知识，请访问以下网站地址：

<http://bk.baidu.com/view/733.htm>

http://www.cnic.org/wiki/Web_2.0

4、视频与图片分享类网站

视频分享站点典型代表：

国外：YOUTUBE www.youtube.com

国内：土豆网 www.tudou.com

图片分享站点典型代表：

国外：Flickr www.flickr.com

国内：Yupoo www.yupoo.com

之所以把视频和图片分享站点提出来说明，是有一定的理由的。视频和图片分享网站的内容是开放的，搜索引擎可以对其内容网页进行索引，这些网站的运营者也期待能从搜索引擎获得可观的流量。

大部分的搜索引擎用户习惯使用网页搜索，我们都知道网页搜索的一个核心算法便是要基于网页的文本分析，但是搜索引擎却无法识别图片和视频所表现的信息，那么矛盾就在这里形成了。

■ 从程序开发上，视频和图片分享类网站都给用户预留了视频和图片信息描述的文本输入框。但实际上，用户贡献的文字信息内容都很糟糕。

A. 描述文字信息十分少

搜索引擎进行文本分析的时候，视频和图片的描述信息是判断某一具体网页主题的最重要依据，如果这些描述性文字的信息量远远少于该网页其他辅助信息量之和的话，搜索引擎

将不能对该网页内容做出正确的判断。随之而来的问题便是，网站中相似网页的数量增多，具体视频和图片最终呈现的网页信息内容缺乏独特性。这样的网页会用竞争优势么？让用户贡献内容，确实可以省心不少，但是每个人都会有惰性！

B. 行文主题控制性差

用户只会去按照他们的思想来对视频或产品进行描述，但没有基于用户搜索行为研究就写出来的文字信息是没有竞争力的，因为这样的网页往往在搜索引擎中的排名表现都很差。

■ 那我们应该如何来针对视频和图片分享网站给出相应的内容优化策略呢？

A. 强化内容的关联性

一般的视频和图片分享站点并不单单只有上传视频和图片功能，同时也设置了其他具有交互的功能，诸如论坛和小组等。而在这些功能板块里主要以文字信息交流为主，那么我们就应该可以考虑把这些文字内容调度到与之相关视频和图片的内容页面里去。这样的做法，很好的弥补了视频和图片页面的文字信息量不足的问题，也很好的增强了页面之间的关联性，同时也提高了该页面的核心关键词的密度。

B. 控制文本的主题集中度

用户贡献的文字信息内容，大部分跟我们部署关键词的策略要求是相背离的。这里列举一个例子，比如一个用户上传了一段关于中国功夫的短片，那么该页面的目标关键词就应该定为“中国功夫”，那我们就可以把该页面的 Title 确定为“中国功夫视频在线观看”。视频短片播放区域会有关于该视频内容的文字信息描述，但是一般用户往往对此视频给出的建议是诸如此类：“这个视频太精彩了”、“我实在是太佩服某某”和“我好喜欢他啊”。这些信息确实是发自用户内心的，但是其与“中国功夫”这一主题的关联性很差，因此这样的页面在“中国功夫”这个关键词的搜索引擎排名上也就缺乏了竞争力了。对于如何控制文本主题的集中度，这里我们给出的建议是：网站应该配备文字信息编辑，可以让编辑人员来做重要信息的编写工作，另外也可以想出一些方法来对用户进行引导，让他们贡献的信息更加具有价值。

第三节 常见问题

一、网站内容的更新

每个访问用户都有极强的好奇心，当他们离开一个网站的时候，大多数人都带这么一个疑问，我下次再访问这个网站时会浏览到什么新鲜信息？是的，正如你所想的那样，网站内容经常更新可以增强网站的用户粘度。

不过，这里要谈的是搜索引擎优化中的网站更新策略问题。

一搜索引擎机器人对一个网站爬行周期会因该站点的信息更新频率而改变，如果该网站的内容更新频率快，则搜索引擎机器人就会经常光顾这个站点，爬行也勤快多了。反之，如果站点长时间不更新的话，搜索引擎机器人也就来得少了。

搜索引擎到网站上来是为了带回新的信息，如果它经常光临你的站点，一旦有新的网页发布的话，它就会很及时地把该页面的信息捕获到搜索引擎数据库中，这就意味着搜索引擎

收录了该页面。所以，经常更新网站信息是有助提高搜索引擎对站点页面的索引效率的，做好这一点，你就比你的竞争对手多一份胜算了。

当然，如果你能每天给自己的站点更新内容，那是最好的做法。如果没有那么多的信息可以更新，怎么办？

我们给出一些实际操作中的经验做法。下面我们来看一张搜索结果的截图。



图 7-3-1 搜索结果截图

在上图中搜索结果中，黄颜色标识的区域是搜索引擎最近对网页更新的日期，一般它会滞后于我们查看搜索结果的当天日期。我们可以使用“site:yourdomain”命令（注意：domain部分要替换成自己的去 www 后的域名形式）来连续几天跟踪站点网页搜索结果日期的变化，找出那些搜索结果日期连续几天都刷新的页面。一般来说，网站的首页和栏目页面的搜索结果日期刷新频率较快，一旦网站有新的页面发布，在这些日期刷新较快的页面上做新页面的文本链接，有助搜索引擎快速收录这些新页面。

这里顺带提出一个大中型门户站点在页面更新中所需要注意的一个问题：在网站一次性发布大量页面，如果页面数量级超过十万或者百万的话，就会引起搜索引擎的怀疑。

Google 的搜索引擎反作弊工程师 Matt Cutts 认为，上百万个网页同时上线，不一定有问题，但绝大多数情况下都是可疑的。所以他建议，要是真有这么多个网页，最好尝试软性逐渐发布，几千页几千页地上线，可能会好一点。

二、内容匮乏怎么办

对于中小型企业网站来说，网站内容匮乏现象普遍存在。这些网站的内容规划大同小异，栏目设置也是让网站策划者耳熟能详的，诸如关于我们、产品或服务介绍、企业文化和联系我们等，一般这样的站点的网页总数不会超过 20 页。

搜索引擎会根据一个网站被其收录的页面数来评定一个网站规模，网站的规模越大，它在搜索引擎所获得的权重就会越大。当然，相对于竞争对手来讲，这个网站将更加具有关键词排名优势。

互联网上没有信息，也可以创作出内容。其实大部分行业都有很多话题来写，就算是专业性很强、一般大众并不熟悉的产品，也同样可以创作出很多内容。

点石互动成员 ZAC 曾经就此问题发表过他的看法，十分具有参考价值。以下是他给出的围绕一个很冷门的产品而展开的内容拓展建议。

- **从产品历史和沿革出发** 这个机器是怎么发明的？从简陋到复杂有什么沿革过程？发明人和改进者都是谁？有些什么具体贡献？在国际和国内都是怎样延续到

今天的？怎样引进国内的或怎样介绍到国外的？得过什么奖？如果碰巧这个机器就是你们公司发明的，那就有更多可写的了。

- **从制造机器的人出发** 研发团队都是由哪些人组成的？这些工程师的背景，经历，发表过什么论文？有什么学术成就？有没有从大学什么的请顾问？直接生产的工人需要掌握哪些技术？生产过程怎么组织的？
- **从原材料出发** 机器的生产都需要什么材料？是哪种钢还是哪种铁？还是其他什么金属？这些材料哪儿生产的最好？为什么？用什么零配件？上游供货商情况？为什么选这些供货商？
- **从客户出发** 都有哪些客户使用你的机器？你的机器给他们带来了什么样的好处？使他们的盈利增长了多少？客户们都是怎样夸奖你们或批评你们的？这些客户所在的行业也可以简单介绍一下。
- **从本行业出发** 行业新闻，竞争对手动向，有没有什么调查机构对你的行业做过评比调查等。把你的产品和竞争对手的产品做一个比较，当然不能失去客观。你的产品比竞争对手的好在哪里？国际动态，展销会，研讨会内容记录...
- **从产品用途出发** 产品使用情况介绍，使用窍门，应用领域，可以帮客户做些什么？怎样维护保养？怎样排除故障？有什么替代产品？
- **从产品技术出发** 机器的工作原理介绍，性能指标，统计，包含了哪些创新科技？用到什么实验设备？检测设备？生产工序是怎样的？有什么国际标准？行业标准？使用安全事项？怎么防假冒伪劣产品？
- 等等等等。

只要开阔思路，任何产品都可以写出很多相关内容。像非标淬火机床这类很冷门的产品，都可以至少写出几十篇文章。如此一来，与竞争对手相比，你就已经占据了很大优势。对其他人们更熟悉、竞争也更多的行业和产品，就能写出更多的内容了。

三、采集系统

很多网站所有者并不希望自己在站点的内容建设上投入太多的精力和时间，于是他们让技术人员编写了一套程序系统，通过该系统他们很容易把目标网站上的网页抓取到自己的数据库里，然后在自己的站点网页上发布这些信息，这样的程序系统便是采集系统了。

通过以上讲解，我们可以得知，通过采集系统来获得其他站点信息的这种不劳而获行为极其不道德。

1、采集系统有何弊端呢？

- 站点的页面都是按照同一程序规则来输入的，页面布局风格千篇一律，这样极有可能会产生大量相似页面；
- 如果系统采集了 100 万条数据，那么接下来它会在某一短时间内输入 100 个采集信息的承载网页，这样会让搜索引擎对该站点产生怀疑。如果 Google 察觉到这种行为，该站点极有可能会进入 Sandbox；
- 采集行为和文章转载的做法性质差不多，这种复制内容的行为会影响搜索引擎对原始网页的权威性的正确判断；
- 采集数据失真，因为在采集的过程中丢失了原始网页中的 Html 标识，影响搜索用户体验。

2、搜索引擎如何看待网站数据采集行为？

从 05 年开始，大量的站点采用采集系统来架构，这些站点所产生的页面大量充斥着搜索引擎的搜索结果，很多采集站点截取了原本属于原始网页的搜索流量，这极大地影响了搜索引擎排名的公平公正性，同时也大大降低了互联网用户的搜索体验，所以网页数据采集行为并不受搜索引擎的欢迎。

Google 官方博客曾发表一篇名为“Site content and use of web catalogues”的文章，在该文章中有这么一句话：

Google is willing to take action against domains that try to rank more highly by just showing scraped or other autogenerated pages that don't add any value to users.

从上文看来，Google 不再对采集行为坐视不管了。

与此同时，百度和雅虎中文搜索引擎也在其给网站管理员的帮助说明中指出，它们不会对采集系统产生的页面进行收录。到目前为止，我们依然可以在各大搜索引擎的搜索结果中看到很多采集系统输入的页面，至少网站所有者应该明白采集行为会增加网站在搜索引擎中的风险。如果搜索引擎察觉你的站点有采集行为，就很可能对你的收录页面数据进行删除。

四、商城程序的内容问题

近几年来，SEO 持续升温，许多商城系统开发商也日益重视其程序对于搜索引擎的友好性了。国内一些商城系统，诸如 Shpoex，已经实现了 URL 伪静态化、CSS+DIV 重构和系统自动输出 Google sitemap 功能。

尽管软件开发商做了如此多的工作，但是其系统用户仍然有不少的抱怨，因为他们使用了这些优化过了的商城程序后，网站依然不能在搜索引擎里有好的排名表现。如果进行理性分析的话，这错并不在商城系统软件开发商，他们制作了最优化的程序，只是用户没有把这个程序用好而已。

限制商城程序在搜索引擎中表现的一个最大的原因就是缺乏内容。一个商城网站中，大多数的网页都是介绍产品具体信息的，在该页面中我们应该向用户介绍产品的价格、性能、质量和其他相关信息。但实际情况如何呢？很多商城网站管理员或编辑在介绍产品信息部分时惜字如金，简单潦草说明了事。大部分商城网站的产品展示页面中产品介绍部分的信息相当简单，通常是不超过 20 个字的产品描述外加一个产品图片。当搜索引擎对这些产品介绍页面进行文本分析的时候，产品介绍的文字信息相对与整个网页的所有文字信息来说，所占的信息比率十分少，这就会造成搜索引擎不能正确对该页进行关键词定位识别，而是将该页面与产品无关的一些导航、版权和其他辅助信息判为核心信息。这是因为，商城网站的产品页面中导航、版权和其他的辅助信息基本是一样的，在产品文字信息不足的情况下，它们占据了整个页面信息的大部分，如此一来，网站中的产品介绍页面相似程度就高了。这也是许多商城网站在搜索引擎里出现大量补充材料的原因。

一个网站被搜索引擎收录的页面中出现大量补充材料的话，该网站在搜索引擎中的权重会被降低，随之而来的负面影响便是已有的关键词排名会出现下滑趋势。所以，如果要想让那些已经被优化好了的商城系统对你的在线生意真正起到帮助，还是建议商城网站的管理员或编辑勤快一点，在网站的产品信息建设上投入足够的精力和时间。

第八章 提交你的网站

网站建好之后，并不意味着工作已经结束了，没有人访问的网站是没有意义的。接下来，你所要做的事情便是如何向搜索引擎和其它网站提交你的网站信息了。

第一节 向搜索引擎提交网站

一、常用搜索引擎登录入口

百度网站提交入口：http://www.baidu.com/search/url_submit.html

谷歌网站提交入口：http://www.google.cn/intl/zh-CN/add_url.html

雅虎中文搜索网站提交入口：http://search.help.cn.yahoo.com/h4_4.html

搜狗网站提交入口：<http://db.sohu.com/regurl/regform.asp?step=regform&class=>

爱问：http://www.iask.com/guest/add_url.php

有道：<http://tellbot.yodao.com/report?keyFrom=help>

二、提交注意事项

- 确保提交网站地址格式的正确性，一般搜索引擎建议的网站地址为包含“http://”的完整网址，例如 <http://www.timev.com/>；
- 只需要提交网站的首页便可，不需要一个网页一个网页进行提交；
- 有些搜索引擎登录需要提交站点的简要描述，请注意按照实际情况书写，不建议出现与现实严重不相符的夸大信息。
- 搜索引擎对提出登录请求网站的收录周期一般为一个月。向搜索引擎提交了自己的网站，如果一个月后依然发现搜索引擎没有收录该站，可以继续向搜索引擎提交登录申请。
- 搜索引擎更为喜欢内容健康的站点。如果你的站点展示的信息内容与我们所处社会的基本道德相违背，那搜索引擎将不会对你的站点进行收录。
- 无论如何也要确保你的站点正常运行。等你向搜索引擎提交了登录网站地址，搜索引擎决定派出它的搜索机器人去对你的站点进行爬行索引，如果你的站点在搜索引擎机器人造访的时候打不开，其后果可想而知。

三、登录搜索引擎的最佳办法

大部分的网站所有者都期望自己的网站能有 100%的把握登录搜索引擎，当然登录所花费的时间越短越好，最好是 1-3 天，而不是搜索引擎所提示的需要一个月！

在实际的 SEO 项目操作中，我们能不能达到这样的搜索引擎登录效果呢？当然能！我们将给大家介绍一个网站登录搜索引擎的最佳办法。

事实上，邀请机制同样也适用于搜索引擎。其实你并不用通过搜索引擎登录入口提交网站的 URL 这一种方法来登录搜索引擎，最便捷和最快的登录搜索引擎的方法应该是让别的网站来“邀请”你的站点进入搜索引擎。

“邀请”的实际操作方式为：寻找一个已经成功登录搜索引擎的网站，让该网站的管理员在其网站首页给你的站点做一个纯文本的超级链接。

向你的站点发出“邀请”的网站应该具有如下特性：

A. 内容原创权威

搜索引擎对该类型站点非常信任，因此其推荐的网站很容易被搜索引擎采纳接受。

B. 经常被搜索引擎机器人访问

搜索引擎机器人对一个站点的爬行记录会留在服务器的日志里，当然，你是不可能获得链接导出站的服务器日志信息。这里我们可以给大家推荐一个经验操作方法：你可以使用“site: 链接导出站点的域名”（诸如 site:timev.com）来跟踪观察该网站首页在搜索引擎索引数据，如果在搜索结果中该网站首页 URL 后是跟有日期信息，并且该日期信息在连续几天都有更新，那这个网站给出的链接就十分理想了。



图 8-1-1 黄色标识为网页最近一次被搜索引擎更新数据的日期

接下来，你只需等待。一般来说，在 1-3 天内，你的网站就会成功登录搜索引擎了。

第二节 登录分类目录站

分类目录站点主要是把搜集而来的优秀站点按照一定的分类标准进行归类，互联网用户通过这些站点的帮助能更好的获取优质的信息。分类目录站点不同于搜索引擎，虽然其可用数据远远少于搜索引擎的索引数据，但是陈列于分类目录中的站点都具有很高的可访问性，因为每个网站都是经过网站的编辑人工审核后才可以发布到相应的类别中。

■ DMOZ (<http://www.dmoz.org>)

DMOZ 源于 Directory.Mozilla.org 的简写，是一个支持多语种互联网网站目录社区，按照固定的目录体系策略开放编辑，由志愿者组成编辑和维护，由 AOL 所有。DMOZ 是互联网的一个重要开放目录项目（Open Directory Project，简称 ODP）。

关于如何登录 DMOZ 的详细介绍将会在后面章节有介绍。

■ Yahoo (<http://www.yahoo.com>)

Yahoo 开创了分类目录导航搜索的新时代。作为著名的分类搜索网站，Yahoo 在很长一

段时期几乎成了“搜索”的代名词，它的分类体系和搜索模式对互联网的发展产生了巨大而深刻的影响。

如果你想了解雅虎分类目录的详情，可以访问下面的网站地址：

<https://ecom.yahoo.com/dir/submit/intro/>

雅虎分类目录每天要受理来自全球的大量网站的登录申请，所以要想成功登录雅虎目录，你必须有足够的耐心和毅力。当然，如果你愿意向雅虎支付 229 美元的话，你的站点将会得到雅虎分类目录的编辑优先审核，这为你赢得了时间，但是雅虎分类目录并不确保一定会收录你的站点。

■ 搜狗 (<http://www.sogou.com/dir/>)

搜狗分类目录前身为搜狐中文分类搜索引擎，创建于 1998 年 2 月。在 2000 年 6 月百度正式推出中文搜索引擎，2000 年 9 月 Google 提供中文搜索以前，搜狐是唯一能与雅虎中国抗衡的中文分类搜索引擎，占据着中文搜索的半壁江山。由于搜狐的本土优势，它的分类体系、立类原则和类目设置更符合中国网民的检索习惯，也为国内分类搜索网站树立了典范。

目前搜狗分类目录不再接纳新站登录服务，如果你的站点想登录其分类目录，必须支付一定的服务费用，而且所支付的费用是按年收取。

■ 网易搜索 (<http://dir.so.163.com/>)

2000 年 9 月 14 日网易公司在国内率先推出了 ODP，<http://search.163.com> 做为开放式目录的示范和“中国最大的开放式引擎”，受到众多网民的瞩目和热情参与。

在功能齐全的分布式编辑和管理系统的支持下，网易已拥有近万名义务目录管理员，创建了一个拥有多达一万个类目，超过 25 万条活跃站点信息，日增加新站点信息 500-1000 条，日访问量超过 500 万次的专业权威的目录查询体系。

不过，可惜的是，2006 年下半年网易 ODP 就已经停止运营了，目前其 ODP 目录编辑管理平台已经访问不了。

■ 其他分类目录站点：

新浪分类目录 <http://dir.iask.com/>

北极星 <http://www.beijixing.com.cn>

第三节 登录网址导航站

网址导航站提供着互联网中最基础的服务，只要你会使用鼠标便可通过网址导航站轻松访问大量的知名站点。网址导航站一般有着数目巨大的访问用户，因此，不可忽略其对于自己网站的推广作用。

网址导航站上陈列出来的网址信息一般是互联网用户推荐的，你也可以向这些网址导航站提交你的站点。以下是我们推荐登录的网址导航站以及提交入口：

Hao123 <http://www.hao123.com/login.htm>

265 上网导航 <http://feedback.265.com/support.php?OneSort=综合网站&OneType=2>

9991 网址大全 <http://bbs.9991.com/forumdisplay.php?fid=27>

向 Hao123 和 9991 网址大全提交网站，事先需按照以下格式准备好网站的相关信息：

- 网站名称;
- 网站地址;
- 网站描述;
- 参考分类;
- 提交时间;
- Alexa;
- 提交人;
- 联系方式。

265 上网导航对提交网站有更多的要求，以下是其收录网站的原则：

- 遵守国家相关法律法规，内容积极健康；
- 没有任何不良信息，无任何非法链接；
- 内容丰富，更新及时，具有一定访问量；
- 在同类网站中排名居前（搜索引擎排名居前）；
- 无过多广告、弹出提示，无恶意代码、病毒；
- 非企业公司、个人商业网站；
- 有 ICP 证、非经营性网站备案；
- 科教文化、机关组织、公益性网站优先；
- 已在网站主页放置 265 上网导航的文字或 LOGO 链接。

以上三个网站为目前中国最有影响力的网址导航站。一般来说，如果你的站点能顺利登录到以上三个主要网址导航站，就会有大量的来自网址站的外部链接指向你的站点，因为其它很多网址导航站都采用了他们的数据。

随着互联网网站数目的增多，三大主要网址站受理的登录申请越来越多，站点不容易被审核通过，同时网站登录过程的时间周期也大大变长了。不过，这些站点已经开辟了收费登录服务，虽然这项服务并未在其站点上标明。如果你向这些网址站公司支付一定的登录服务费用，就会大大减少你网站登录所需的时间，并且登录成功的机会很大。

第四节 登录垂直行业门户

向专业的分类信息目录或垂直行业门户发布自己网站的信息，也是一种不错的增加外链数目的方式。

但是，如何来寻找这些我们发布信息的网站呢？搜索引擎是最好的工具。如果你是一家“螺丝”生产商，可以打开一个主流中文搜索引擎（可以是百度、谷歌和雅虎中文搜索的任意一个）使用“螺丝”这个关键词进行搜索，你会在搜索引擎反馈的搜索结果中找到大量可以免费发布产品信息的螺丝行业垂直站点。

在向这些站点发布你的产品信息之前，请准备相应的资料。一般来说，刊登你的产品需要如下信息：

- **公司介绍** 包括公司名称、主营业务、规模以及诚信凭证等；
- **产品信息** 包括原材料、性能、应用范围和质量保证等；
- **联系方法** 包括地址、邮编、电话、传真和电子邮件。

其实，向专业的分类信息目录或垂直行业门户发布自己的网站信息，也是提高网站在搜

搜索引擎中的曝光率的一种方法。采用这种方式能快速增加网站的外部链接数目，但是需要投入大量的精力和时间。另外，通过该种方式获得的外部链接只能在数量上的增长，但链接的相关性很差。你在发布产品信息的编辑框里并不能有针对性做关键词链接，网址只能在规定的输入框内填入，因此，网站获得的链接往往只是缺乏关键词锚文本的单纯 URL 链接。

第九章 链接策略（一）

第一节 链接的基本知识

超链接是指从一个网页指向一个目标的连接关系，这个目标可以是另一个网页，也可以是相同网页上的不同位置，还可以是一个图片、一个电子邮件地址、一个文件，甚至是一个应用程序。而在一个网页中用来超链接的对象，可以是一段文本或者是一个图片。当浏览者单击已经链接的文字或图片后，链接目标将显示在浏览器上，并且根据目标的类型来打开或运行。

按照使用对象的不同，网页中的链接又可以分为：文本超链接、图像超链接、E-mail 链接、锚点链接、多媒体文件链接、空链接等。其中，可以被搜索引擎识别的链接有文本超级链接、图像超级链接和锚点链接。按照链接路径的不同，网页中超链接一般分为三种类型：内部链接、锚点链接和外部链接。

第二节 为什么链接很重要

链接是搜索机器人爬行的路径，链接分析是搜索引擎排名算法中一个很重要的部分，一个网站如果能获得其它许多网站给出的、能被搜索机器人识别的超级链接，搜索引擎就会认为该网站是受欢迎的。受搜索引擎欢迎的网站一般都会在搜索结果里有好的排名表现，这就是链接之所以很重要的原因。

一、超链分析技术

超链分析技术是新一代搜索引擎的关键技术。在学术界，一篇论文被引用得越多，说明其越好，学术价值就越高。类似的，超链分析技术就是通过分析链接网站的多少来评价被链接的网站的质量，这保证了用户在使用搜索引擎搜索时，越受用户欢迎的内容排名越靠前。这一技术的运用，大大地方便了网民对互联网资源的查找和利用。

如果一个网站给另外一个网站导出了链接，搜索引擎认为，导出链接的站点给了被链接网站一个信任投票。如果某个网页的链接质量很高，则该网页就会有比较好的关键词排名，尽管链接的数量并不一定很多。如何判断一个指向自己站点网页的链接是否有价值呢？一般来说，影响链接的因素有网页级别、信息更新的频率和内容的相关性。

二、网页级别

在这里主要给大家介绍两个网页级别评定参数：Google PR 和 Sogou Rank。

1、Google PR

什么是 PR 值？ PR 的英文完全拼写为 PageRank，它是 Google 的搜索引擎自然排名算法中的一部分。PR 值有 10 个等级，1 至 10，一个网页的等级越高那说明 Google 对该网页的评价越高。

如何查询一个网站的 PR 值? 可以到 Google 的官方网站去安装一个 Google toolbar, 然后在访问网页的时候就会发现, 浏览器的菜单栏部分会有一个 PR 值显示。当然, 现在也有很多网站提供在线查询 PR 值的功能。

PR 值在网站中是如何分布的? 一个网站中包含很多网页, 每个网页 Google 都会给它评分, 即每个网页都会有 PR 值。一般来说, 网站首页的 PR 值最高, 然后 PR 的数值会随着网站导航或者目录层次的加深而变小。一个比较形象的数学不等式表示如下: 首页 PR > 二级栏目页 PR > 三级栏目页 PR > ……。但是注意, 以上所说的 PR 值分布适用于一般情况, 在网站中也会出现内页的 PR 值大于首页的情况。

影响网页的 PR 值的因素有哪些? A. 网站首页的 PR 值, 如果首页的 pr 值高那其传递到下级页面的 pr 也会不错; B. 该网页的内容质量, 原创内容最佳; C. 反向链接 (back link), 当网站内页的反向链接数目大于首页的反向链接数目的时候, 就可能会出现内页的 PR 值大于首页的 PR 值。

网页 PR 值对搜索引擎排名有帮助么? 这是每个 SEO 初学者都比较关心的问题。早些年, 网页的 PR 值对网页的搜索引擎自然排名结果确实有很明显的帮助。但随着搜索引擎算法的不断完善, 现在搜索引擎已经逐渐淡化 PR 值对网页排名的影响力了。怎么理解这句话? 即 PR 值已经不再是影响搜索引擎排名的主要因素, 但对排名还是有帮助的。

2、Sogou Rank

什么是 Sogou Rank? Sogou Rank 是搜狗衡量网页重要性的指标, 它不仅考察网页之间的链接关系, 同时也考察链接质量、链接之间的相关性等特性。Sogou Rank 是机器根据其算法自动计算出来的, 值从 0 至 100 不等。Sogou Rank 值越高, 该网页在搜索中越容易被检索到。

如何提升 Sogou Rank? 为了提高页面的 Sogou Rank, 你需要使页面清晰易读, 让更多同领域的高 Sogou Rank 站点来链接你的页面。同时你需要慎用对外链接, 尤其是对低质量站点的链接 (如果链接到垃圾站点, 将极有可能降低你的 Sogou Rank)。

Sogou Rank 是影响页面排名的一个重要因素, 但不是全部因素。你同样需要努力去丰富自己页面的内容, 给页面选择尽可能简洁明了的标题, 拒绝向恶意堆砌关键词的垃圾页面提供链接。

三、网站的更新频率

网站的信息更新频率包括两个方面: 站点自身的信息更新频率和站点搜索结果信息的更新频率。

站点自身的信息更新频率 一般是指单位时间内承载发布信息的网页产生 (也即是网站中新页面的出现) 或变更的速度。在提到站点自身信息更新的时候, 我们也应该考虑网站已存在的网页信息的变更修正, 尤其是网站中的主要页面 (如首页和栏目页), 这是比较让人容易忽视的。

站点搜索结果信息更新频率 指单位时间内, 搜索引擎对站点网页搜索结果信息的刷新速度。判断网页搜索结果展示信息刷新的三个因素分别为: 标题、随机摘要描述和日期。其中的任何一个因素发生变化, 我们便视其搜索结果发生更新。通常情况下, 网页搜索结果中的日期变动是最容易让人察觉的。

一般来说, 站点自身信息的更新频率是影响站点搜索结果信息更新频率的主要因素。理

想情况下，网站搜索结果中的信息更新频率应该和站点的信息更新频率同步，一旦网站上有信息更新，该网站在搜索引擎中的网页所对应的网页索引信息就会作出相应的更新。在这样的理想情况下，搜索引擎对网站信息更新行为有着更为灵敏的察觉性，一旦有新的网页被发布出来，搜索引擎机器人会很迅速地把该网页的信息捕捉进搜索引擎的数据库，这么一来，搜索引擎对网站的收录将不会是什么大问题了。

而实际上并非如此。很多中小型网站很不重视网站信息的更新，有的甚至好几个月或者一年多不会为网站增加网页，这样会造成搜索引擎机器人不会经常来光顾网站。如果某一天，一个经常不更新的网站发布了一个新网页，真的不知道搜索引擎机器人什么时候会再次光临，并把新网页的信息带回搜索引擎的数据库。其实，网站的收录问题困扰着很多的网站所有者。

如果一个信息更新不勤快的站点发布了一个网页，该网页迟迟不能被搜索引擎收录，这是可以被理解的。毕竟在以前很长一段时间内，搜索机器人造访过这个站点后都是空手而归，后来它就相应调长了自己对这个站点的回访周期。如果这个站点突然某一天有信息更新了，搜索机器人也不能快速察觉。但是，如果有一个信息更新频率很高的站点给了这个网站一个外部导出链接呢？搜索引擎蜘蛛访问此信息更新频率很高的站点时，发现了那个导出链接，然后顺着链接爬行到了信息更新频率慢的网站，便带回了那个新发布的网页信息。

基于此分析，网站的信息更新频率是可以传递的。

四、内容的相关性

如果你网站的内容主题是关于网站建设的，而另外有一个网站是发布网页制作教程的，这两个网站的主题就十分相关了。来自这个网页制作教程内容的站点导出给你网站的链接，对于提高你站点搜索引擎排名的作用要远远大于一个关于啤酒酿造的站点指过来的链接。

两个都是关于网站建设主题的站点内容关联性是最大的，但实际上，我们发现这些类型的站点往往都是竞争对手，除非你是同时拥有这两个站点的人。期望你的竞争对手给你导出一个指向你网站的外部链接，几乎是不可能的事情。不过，促成这样内容相关性很高的站点互相导出外部链接给彼此的可能还是存在的，原因在于，他们所面对的市场因区域或用户群体而被细分，彼此并不形成竞争关系。

第三节 网站内部页面的链接策略

网站内部页面之间的链接规划在整个 SEO 过程中也是十分重要的，因为我们不仅要向用户和搜索引擎呈现一个严谨的网站逻辑结构，同时也应该让搜索引擎能够准确识别页面关键词。

一、给重要的网页更多的关注

一个网站中并不是所有的页面都很重要，每个页面承载信息的重要性不同，因此也应该在网站中体现出每个网页的层次感。整个网站中如果一个页面获取其他页面的链接数比较多，那搜索引擎就会认为这个页面在网站中最为重要，就会把它在整个网站中的权重加强。基于以上思路，我们也就不难明白，为什么一个网站的首页经常会在搜索引擎中取得好的排名了。

二、相关链接（人工和自动）和频道关联

增强网页之间内容的相关性，不仅可以提升具体网页的关键词密度，同时也可以为用户提供更多的相关信息，因此，成为凸现目标部署关键词搜索引擎排名的一个很有效的方法。

如果你的网站是基于 CMS 系统（Content Management System 的缩写，意为“内容管理系统”）构建的话，那你可以通过站内关键词和 TAG 标签来有效关联同一主题内容的网页。当然，我们也是可以通过人工来做内容相关性的操作的。程序实现更为省心和便捷，但与人工相比，它缺少灵活性和人性化，不过人工操作是要投入很多时间和精力。

网站中内容的关联最常见的表现形式是文章正文中的关键词导出链接和针对该文章的相关列表导航。所以，在网站搭建的时候一定要规划好内容的关联实现功能。

三、链接的导出

为访问用户提供尽可能多的有价值信息的网站才是好网站，这是搜索引擎给出的建议。如何给用户提供有价值的信息呢？这里的信息内容不仅仅存在于你的网站上，同时也应该包括其他网站上的内容。所以，我们要尽可能向其他站点导出链接。

但是，你的导出链接所指向的网站必须要满足以下要求：

- 没有搜索引擎排名作弊行为；
- 网站服务器不会经常宕机而造成网页无法访问；
- 网站的内容权威。

第十章 链接策略（二）

第一节 登录开放目录 DMOZ

一、为什么登录 DMOZ 如此重要

网站被 ODP 收录后，一般两周到几个月后会在共同使用 ODP 数据库的合作伙伴网站上列出，如 AOL Search、DirectHit、HotBot、Google、Lycos、Netscape Search 等。ODP 对数据每周更新一次，但是合作伙伴有其自己的更新规则。

你只需要向 DMOZ 提交登录申请，一旦网站登录成功，你不仅会得到 DMOZ 站点导出的链接，而且会获得数目可观的 ODP 数据库合作伙伴网站指向你网站的链接。这些导出链接的站点都是被搜索引擎认为权威和可信任的，因此这样的外部链接对于你的站点来说是相当具有价值的。

虽然成功登录 DMOZ 并非一件很容易的事，但鉴于其对站点能带来高质量的外部链接，很多网站管理员在锲而不舍地向 DMOZ 提交自己的网站登录申请。

二、DMOZ 开放目录登录实战攻略

1、第一步：请仔细阅读 DMOZ 的登录建议

DMOZ 官方网站对提交登录申请的站点有严格的要求，请确保你的网站与其要求遵循的规则相符合。DMOZ 的具体网站登录要求如下：

- 不要提交镜像网站。镜像网站虽然内容相同，但是 URL 地址不同；
- 不要提交与已收录内容相同或相似的 URL，因为这将不方便用户使用。大量重复地提交内容相同或相似的网站可能会导致这些或所有相关网站被淘汰或删除；
- 不要隐藏并重复提交相同的 URL，如：<http://www.dmoz.org> 和 <http://www.dmoz.org/index.html>；
- 不要提交带有多个地址的网站；
- ODP 拒收任何带有非法内容的网站。这些非法内容包括少儿不宜的色情描述、诽谤；侵犯知识产权的内容；鼓吹、发动或煽动非法活动（如欺诈、暴力）的内容等；
- 不要提交正在更新的网站，可在更新完毕后进行提交。不完整的、显示“正在更新”或包含无效图片、坏链接的网站不适合向目录提交；
- 带有色情内容的网站需向相应的成人目录下提交；
- 非英语网站需向相应的世界目录下提交；
- 不要提交含有大量会员链接的网站。

2、第二步：检查你的站点否已经被收录

在开放目录的首页（www.dmoz.org）查找将要提交的网站是否已被收录，这将节省双方的时间。如果你的站点确实内容十分优秀，一旦 DMOZ 的编辑们发现了它或者是互联网用户向 DMOZ 网站推荐了它，DMOZ 会主动收录的。

3、第三步：选择合适的分类提交网站

开放目录中的分类十分繁多，建议仔细选择一个最恰当的目录进行提交，不当或无关目录下提交网站会被拒绝或移除。选择一个合适的目录提交站点，有助于提高站点的登录效率。注意，有些目录没有“加入 URL”或“更新 URL”的提示链接，那么说明这些目录不接受提交，可以另选别的相关目录进行提交。

4、第四步：认真填写网站的登录信息

如果已经选择好了一个合适的目录，你可以到该目录下，点击网页界面上方的“加入 URL”超级链接，然后按照提示步骤顺序完成提交。

在提交网页界面填写网站描述信息的时候，请确保针对网站的描述内容应该简明、准确，不可夸大其词，一旦 DMOZ 编辑发现所提交网站含有非客观的夸大描述，该网站将被延迟或拒绝收录。

另外，有些网站管理员通过某些软件来自动向 DMOZ 提交登录请求。这里特别说明的是，通过自动提交的网站在收录后若被删除，该网站管理员将不会得到任何提示通知。

三、一些对你有帮助的建议

- 提交后，DMOZ 的编辑会复查你的网站以决定是否对其最后收录。决定 DMOZ 是否收录的因素很多，因此一个网站从提交到审核的过程一般需要大约几周或更长的时间；
- 每次向 DMOZ 登录网站的时候只能提交一个 URL，对相同的或相关的网站多次提交可能导致拒收或删除。另外，也不建议对相同的 URL 进行多次提交；
- 如果所提交网站三个星期后仍没有被 DMOZ 网站显示出来，可尝试再次提交或向负责提交目录的编辑人员发送邮件了解情况；
- 找到网站收录下的目录，填写“更新 URL”表格，可对已收录网站或网页的标题、描述等进行更改。若要使网站更改所在目录，可向编辑发出邮件阐明修改理由。

第二节 交换链接

交换链接最原始的意图是获得其他站点的流量导入，SEO 出现后，链接交换的性质就发生了改变，因为来自其他站点的信任投票更具有价值。

一般来说，新站很难获得其他老站的导出链接。所以，我们跟其他网站管理员交换链接的时候是需要技巧的。与别人交换友情链接之前，应该先表诚意，首先在自己的站点做好别人网站的链接，然后再给被链接的网站所有者发一封电子邮件。在电子邮件正文中，一定要把你的网站详细描述清楚，一定要体现出网站的独特之处。以下是可以参考的电子邮件正文范本：

尊敬的某某网站管理员，你好！

我是模板聚合（www.ecstudio.cn）网站的管理员，很高兴能通过这封电子邮件向你表达我想与你进行站点互换友情链接的意向。

模板聚合站点创建于 2006 年 2 月 16 日，它为广大 WEB 标准推崇者提供可以

参考研究的采用 CSS+XHTML 重构的优质网页模板下载服务。

我可以确保模板聚合站点没有采用针对搜索引擎排名作弊的过激做法,已经跟它互换链接的站点都是内容健康且受搜索引擎喜欢的站点。

为了先表诚意,我已经在我的站点上给你的站点做上了纯文本超级链接,你可以访问网站地址 www.ecstudio.cn 查看。

如果你对我的站点感兴趣,并愿意在你的站点上做上它的纯文本链接,以下是我给出的链接代码供你参考: `模板聚合`

最后,祝福你站点的发展越来越好!

模板聚合网站管理员: 某某

网站友情链接注意事项:

- 不要与那些发布有反动、色情和博彩等违背社会道德信息的网站交换链接;
- 来自被搜索引擎开除的站点的链接对你的站点没有益处,因为它们的信任投票在搜索引擎认为是无效的。也不建议给作弊站点导出链接,因为这样会让你的网站容易无辜受到搜索引擎惩罚牵连;
- 与你交换链接的站点必须在搜索引擎里有良好记录,如果该站点充斥着一些搜索引擎排名作弊行为,千万不要与它交换链接,尽管它目前还没有被搜索引擎开除;
- 尽量与那些与你网站内容相关的网站交换链接,建议不要和那些与你站点无关的站点交换链接;
- 不要与那些喜欢泛滥交换友情链接的站点交换链接,否则会导致导出给他的站点权值比他导入给你的要大,不划算;
- 花点时间认真分析一下即将与你交换链接的站点是否已经加入链接工厂,如果是的话,最好不要跟其交换链接;
- 没有针对性地建设自己外部链接的网站不值得你与其交换链接。比如说,这个网站是销售女性时尚小饰品的,而它先后跟一些诸如网站建设、白酒鉴赏和汽车零部件制造等类型的网站交换了链接;
- PR 值高的外部链接导入可遇不可求。当然,如果你能获得高 PR 的外链,那就最好不过了;
- 如果在一个 PR 值高但内容更新慢的网站,与一个 PR 值低但内容更新速度快的网站之中选择一个交换链接,那么选择后者。因为网站信息更新频率的传递比单纯 PR 值的传递更为重要。

一个网站刚开始运营的时候,并没有多大机会来选择链接交换伙伴,在这个阶段链接数量是主要追求目标,所以只要信息内容健康并且没有作弊行为的网站都可以是链接交换目标对象。当网站发展到一定规模的时候,就应该对那些想跟你的站点交换链接的网站进行筛选了。

第三节 购买链接

搜索引擎没有出现之前,互联网上就出现了购买文字或图片广告链接的行为,这是一种最原始的网站推广方式。

以 Google 为代表的第二代基于链接分析的网页搜索引擎出现后,有人通过购买大量高

PR 值网站的外链，使自己的网站排名能在短时期内获得提升。搜索引擎非常不欢迎这种行为，因为这有悖网站排名的公平性，一旦那些有链接买卖的网站被发现，搜索引擎会对其做相应惩罚，诸如降权或者从其数据库中删除网站数据。

其实，合理的购买链接是可以避免被搜索引擎的惩罚的。在购买链接的过程中请注意以下建议：

- 购买的链接应该来自专业权威站点；
- 购买链接的网站应该具有独特原创内容并且经常有内容更新；
- 注重链接质量，而不是数量；
- 购买链接应该是一种阶段性规划行为，不可在短时间内给你的网站购买大量链接。

第四节 链接诱饵

链接诱饵是一种很巧妙的获取外部链接的做法，近几年引起 SEO 技术爱好者和从业人员的广泛关注。链接诱饵的页面往往具有吸引大家眼球的内容，因此执行施放诱饵的人必须有很高的创意策划能力。合理规划链接诱饵页面，能够让网站获得数量极其庞大的反向链接，这不仅可以提高网站的曝光度，提高网站的访问量。更重要的是，这些链接是自然获得的，而且具有较强的文本相关性，能够有效地提高网站的链接广度，从而提高网站在搜索引擎中的排名。

一、如何制造链接诱饵

撰写专业新闻 行业或领域内的新闻，当然，不是那种类新浪的大路新闻，而必须是某种网站独有的内容，比如说对特定事件具有独特视角的分析等。

提供资源下载 提供可供其他人使用的资源，包括信息、内容、工具等。

发挥幽默天赋 人们都希望能看到些轻松搞笑的内容，这方面不必多言。

你的观点是什么 反驳其他人的论点一般要把握两个前提：1、事件本身值得探讨，鸡毛蒜皮的小事恐怕没多少人关注；2、反驳对象在该领域内具有一定知名度。

有目的性的攻击 与批判他人观点相比，有目的性的攻击做法更甚一步，很多时候通过强烈的个人攻击迫使对方不得不回应。当然，被攻击的对象是同行业或领域内的知名人士。近来这类有目的性攻击的做法似乎越来越普遍，这也是链接诱饵受到不少人质疑的主要原因。客观说来，这种策略存在先天上的问题：1、即使这个网页可以获得一定数量的链接，但更容易破坏整个网站的形象，让读者对网站的其他内容失去兴趣；2、受攻击者就算仅仅出于不鼓励这类攻击诱饵的考虑也会在回应时使用“nofollow”链接。

二、对链接诱饵应如何看待

事实上，从链接诱饵现身之初起，便有人提出质疑：建立一个目标旨在捕获链接的网页，是否有黑帽 SEO 的嫌疑？对这个问题要从两方面来看。

首先，在一般意义上说，链接诱饵本身并不存在任何问题，与早期的网站建设包括 SEO 策略相比并没有任何出格之处。即便主要目的是为了得到链接，也要通过向用户提供独有的、有价值的内容或信息来实现，这是链接诱饵能否成功的基础，同样，也是网站建设能否成功的基础。

其次，是否黑帽也取决于在链接诱饵上究竟要走多远。如果过于沉溺于链接诱饵，比如说滥用有目的性攻击方法，为了得到链接不惜恶意攻击他人，恐怕就偏离了 SEO 的正确轨道。当然，在吸引眼球方面也许有同样有效，就象娱乐圈的恶炒，就像那位成为名人的宋大嘴。

就象 Google 搜索引擎反作弊工程师 Matt Cutts 在“SEO Advice: Linkbait and Linkbaiting”中指出的那样，链接诱饵本身并不具有先天上的负面内涵，而是取决于网站建设者如何应用，如同其他很多 SEO 技术与策略一样。

第五节 外链建设中的几个误区

一、外部链接数目增长的自然化

一个网站的外部导入链接数目越多，其获得搜索引擎排名的胜算也就越大。但是，有时候也会弄巧成拙。当很多网站管理员了解网站外部链接对其站点的搜索引擎排名的作用后，他们势必会花大量的精力来着手进行自己站点的外部链接建设工作。于是，我们就经常看到这样的情况，有些网站管理员通过各种手法让一个站点在短时间内获得了大量的外部导入链接，但是事后有些人抱怨说，这种做法并没有让他们的网站获得好的排名，反倒是以前已经获得的好的关键词排名位次开始下滑，甚至有些网页已经在搜索结果中杳无踪影了。

一个网站的外部链接数目增长应该是自然化的，随着网站的发展，其获得的外联链接数目会呈现阶段性增加。这里给出的建议是，不要在短时间内给你的网站增加大量的外部链接，这种做法会引起搜索引擎的怀疑，因为搜索引擎会认为你这样的做法有悖其排名公正原则。

二、不要忘记你最能做好的事情

网站的外链建设工作是网站管理员必须做的，但是与其他工作相比并不是最重要的，如站点的内容建设。网站的外链的增加确实对网站的搜索引擎排名效果能起到很显著的效果，但是由于其具有不可操纵性，成功的链接互换背后不仅仅有时间和精力投入，同时也存在运气因素。为什么我们不投入更多的时间着手自己的网站内容建设工作呢？这里撇开站点内容更新有助于增加网站用户的忠诚度的益处不谈，我们来分析网站内链对于 SEO 的作用。

与外链相比，网站的内链具有以下特点：

- 在一个网页导出一个指向另一个网页的链接是一件很容易的事情，不用四处费力求人；
- 合理内链部署规划让网站中的网页更加具有层次感。你的网站中哪些网页将会参与哪些关键词的排名，只需从站内其它关联页面上多向其调度些链接过来。通过网页所获得的站点链接数的多少，搜索引擎很容易识别哪些页面在网站中是重要的；
- 增加一个网页的内链数目方法很简单，只需要在网站中多建立一些与其内容相关的网页，并将新建的网页内容的关键词链接导入过来即可；
- 内链的合理使用有助于集中网站内容主题，从而使该主题中的核心关键词在搜索引擎中更加具有排名优势；
- 站点中的网页间互链有助提高搜索引擎对网站的爬行索引效率。

我们应该在自己能做好的事情上做出更多的努力，毕竟优质的外部链接可遇不可求，数

目巨大的质量不高的链接对提升你网站的搜索引擎排名帮助不会很大,反而会让你的网站招致搜索引擎的怀疑。

三、不宜使用软件群发

通过软件群发而获得大量外部链接的做法确实可以在短期内提升一个站点在搜索引擎中的关键词排名。但是这种做法存在极大的风险,一旦被搜索引擎察觉,获得大量外部导入链接的网站将会受到搜索引擎的惩罚。

很多网站开辟了与用户互动的功能(留言板、博客和论坛等),用户可以对网站发布的信息进行留言评论,SEO 人员通过群发软件可以自动向这些网站群发可以留有规划关键词的链接信息。而这些明显企图通过外链数目提高搜索引擎排名的行为是不道德的,他们发出的评论信息大都没有任何价值。

通过软件群发可以有效提升站点的外部链接数量,但是链接的相关性却很差。由于此做法存在被搜索引擎开除的风险大,所以并不提倡采用软件群发的方式来进行网站外链建设。

第十一章 数据监测和分析

第一节 网站流量数据统计与分析

一、什么叫网站流量及网站流量统计分析

通常说的网站流量 (Traffic) 是指网站的访问量, 是用来描述访问一个网站的用户数量以及用户所浏览的网页数量等指标, 常用的统计指标包括网站的独立用户数量、总用户数量 (含重复访问者)、网页浏览数量、每个用户的页面浏览数量、用户在网站的平均停留时间等。

网站流量统计分析, 是指在获得网站访问量基本数据的情况下, 对有关数据进行统计、分析, 以了解网站当前的访问效果和访问用户行为并发现当前网络营销活动中存在的问题, 并为进一步修正或重新制定网络营销策略提供依据。

网站访问统计分析的基础是获取网站流量的基本数据, 这些数据大致可以分为三类, 每类包含若干数量的统计指标。

二、网站流量指标

网站流量统计指标常用来对网站效果进行评价, 主要指标包括:

1、独立访问者数量 (Unique Visitors)

独立访问者数量, 有时也称为独立用户数量, 是网站流量统计分析中另一个重要的数据, 并且与网页浏览数分析之间有密切关系。独立访问者数量描述了网站访问者的总体状况, 指在一定统计周期内访问网站的数量 (例如每天、每月), 每一个固定的访问者只代表一个唯一的用户, 无论他访问这个网站多少次。独立访问者越多, 说明网站推广越有成效, 也意味着网络营销的效果越有效果, 因此是最有说服力的评价指标之一。

相对于页面浏览数统计指标, 网站独立访问者数量更能体现出网站推广的效果。从独立访问者数量的变化, 大致上可以看出你所实施的网络营销效果如何。一般流量分析软件都会按日期画出图像, 非常直观, 可以很清楚的看到 SEO 的总体效果。另外, 从独立 IP 数也可以看到流量和某些特定努力的关系。比如说如果你的网页被放在 365key 的首页一两天的话, 可以看出这些社会化书签的影响力有多大, 你也就知道这类社会化搜索、书签、网摘、标签等系统现在是很重要的流量来源。

下图为某站点的独立用户访问量数据统计, 统计后台为 CNZZ。

时间段分析 - 从 2007-08-01 到 2007-08-29 的统计
日访问量分布

日期	PV访问量	独立访客	唯一IP
2007-08-29	2171	1105	1103
2007-08-28	3384	1680	1681
2007-08-27	3935	1995	1983
2007-08-26星期日	3273	1549	1551
2007-08-25星期六	3242	1556	1548
2007-08-24	3658	1650	1661
2007-08-23	3680	1757	1740
2007-08-22	3234	1649	1654
2007-08-21	3606	1706	1703
2007-08-20	4264	2091	2096
2007-08-19星期日	3347	1600	1614
2007-08-18星期六	3689	1598	1591

图 11-1-1 独立用户访问量统计

2、重复访问者数量 (Repeat Visitors)

重复访问者数量反映了站点用户的忠诚度，站点用户的忠诚度越高，重复访问者数量越高。

3、页面浏览数 (Page Views);

在一定统计周期内所有访问者浏览的页面数量。如果一个访问者浏览同一网页三次，那么网页浏览数就计算为三个。页面浏览数常作为网站流量统计的主要指标。

不过，页面浏览数本身也有很多疑问，因为一个页面所包含的信息可能有很大差别，一个简单的页面也许只有几行文字，或者仅仅是一个用户登录框，而一个复杂的页面可能包含几十幅图片和几十屏的文字。另外，即使是同样的内容，在不同的网站往往页面数不同，这取决于设计人员的偏好等因素。例如一篇 6000 字左右的文章在新浪网站通常都放在一个网页上，而在有些专业网站则很可能需要 5 个页面，对于用户来说，获取同样的信息，新浪网的网站统计报告中记录的页面浏览数是 1，而别的网站则是 5 个。而且 Page Views 作弊也非常容易，因而很多针对 Alexa 排名的作弊重要手段之一就是采用各种办法刷 PV 值。

4、每个访问者的页面浏览数 (Page Views Per User);

每个访问者的页面浏览数是一个平均数，即在一定时间内全部页面浏览数与所有访问者相除的结果，即一个用户浏览的网页数量。这一指标表明了访问者对网站内容或者产品信息感兴趣的程度，也就是常说的网站“粘性”。

每个用户的页面浏览数量反映了用户从网站获取信息的多少,一般来说,这个平均数越高,说明用户获取的信息量也越大(一个例外情况是,网站提供的信息对用户有价值,但用户获得信息不方便而造成平均页面浏览数过大,如需要多次点击、查找信息不方便,每个页面的信息量过小等)。通过对每个访问者的页面浏览数变化趋势分析,如果发现这一数据基本保持稳定,那么当与网站页面浏览数进行对比分析时,页面浏览数的变化趋势就反映了网站总体访问量的变化,如果平均页面浏览数有较大变化,则需要对网站独立用户数、网页浏览数等指标进行比较分析才能发现网站访问量变化的真正趋势。因为如果每个用户平均页面浏览数增加,即使独立用户数量没有增长同样会使得总的页面浏览数增加,反之,如果独立用户数保持稳定,但平均页面浏览数下降了,也会造成网页浏览数量的减少。因此单纯从网站页面浏览数的变化情况还不足以说明网站的总体访问量变化趋势,需要与独立用户数、每个用户的平均页面浏览数量等进行比较分析。

5、某些具体文件或页面的统计指标,如页面显示次数、文件下载次数等。

这样指标主要是针对具体每个文件的浏览和下载量。在流量统计中,受访页面统计数据可以反映具体页面的来访情况。通过这些具体的页面统计指标的分析,可以迅速看出最近用户的访问热点,也可以访问站点的那些页面及其对应的关键词在搜索表现较好。

受访页面分析		导出此表数据	
	访问页面(页面动态参数忽略)	次数	比例
1	http://www. [REDACTED] com/CateGory/15	46	2.5%
2	http://www. [REDACTED] com/userpic/	44	2.4%
3	http://www. [REDACTED] com/	42	2.3%
4	http://www. [REDACTED] com/login.aspx	33	1.8%
5	http://www. [REDACTED] com/CateGory/9	32	1.8%
6	http://www. [REDACTED] com/brand	30	1.6%
7	http://www. [REDACTED] com/CateGory/18	27	1.5%
8	http://www. [REDACTED] com/CateGory/29	26	1.4%
9	http://www. [REDACTED] com/SearchResult.aspx	25	1.4%
10	http://www. [REDACTED] com/brand/3499	24	1.3%

图 11-1-3 受访页面统计

三、用户行为指标

用户行为指标主要反映用户如何来到网站、在网站上停留了多长时间、访问了那些页面等,主要的统计指标包括:

1、用户在网站的停留时间

一个用户在网站上停留时间（在线时间）的长短，反应出一个网站的粘度和吸引用户的能力。一般情况下，用户在某个网站停留时间越长，反映改站点的内容越吸引人，用户粘度越高。但这样的说法也不是绝对正确地，比如 Google 曾经有一个追求目标就是让用户在 Google 停留的时间尽可能短。因为用户每次在 Google 搜索上停留越短，说明客户通过 Google 找到答案越迅速，Google 的搜索质量越高。

2、用户来源网站

通过用户来源网站（也叫“引导网站”）统计，可以了解用户来自哪个网站的推荐、哪个网页的链接，可以看出部分常用网站推广措施所带来的访问量，如网站链接、分类目录、搜索引擎自然检索、投放于网站上的在线显示类网络广告等。一般后台能反应用户来源的几项统计数据有搜索引擎来源统计，搜索关键词统计，来路统计。

访问来源分析 (说明:来源如果为“本站”,则指的是您自己网站的站内链接,并非指cnzz.com) 导出此表数据			
来源站点 (页面动态参数忽略)	站点信息	次数	比例
本站(您的网站)		801	37.6%
baidu.com	R IP S	525	24.6%
google.cn	R IP S	508	23.8%
soso.com	R IP S	81	3.8%
直接输入网址或书签		79	3.7%
search.cn.yahoo.com	R IP S	40	1.8%
google.com	R IP S	21	0.9%

图 11-1-4 某站点外部流量来源站点

来源页面分析 导出此表数据		
来源页面	次数	比例
本站页面	801	37.6%
直接输入网址或书签	58	2.7%
http://www.soso.com/q?w=%c7%bd%d6%bd%c6%b7%c5%c6@sc=adr@ch=th.u@gid=QWv iNDm0tQewURfHY3DSiB017wc30g00	16	0.7%
http://www.google.cn/search?complete=1&hl=zh-CN&inlang=zh-CN&newwindow= 1&q=%E5%A2%99%E7%B8%B8%E5%93%81%E7%89%8C@meta=@aq=t&oq=%E5%A2%99	9	0.4%
http://www.google.cn/search?complete=1&hl=zh-CN&ie=GB2312&q=%C6%B7%C5% 6%C7%BD%D6%BD@meta=@aq=t&oq=%C6%B7%C5%C6%C7%BD	8	0.3%
http://www.baidu.com/s?kw=@sc=web@cl=3@tn=sitehao123@ct=0@pn=@rn=@lm=@i e=gb2312&rs2=@myselectvalue=@f=@pv=@z=@from=@word=%B4%BA%CO%BC%BF%D5%B 5%F7%FR-70LW%2FBFD%B3%F6%CF%D6E4%CA%7%CA%B2%3%B4%D4%AD%D2%F2	7	0.3%
http://www.google.cn/search?complete=1&hl=zh-CN&inlang=zh-CN&newwindow= 1&client=aff-bitcomet@channel=85851@hs=cwX@affdom=85851.com@q=%E7%94%B 5%E8%A7%86%E6%9C%BA%E5%93%81%E7%89%8C@meta=@aq=t&oq=%E7%94%B5%E8%A7%86 E6%9C%BA	6	0.2%

图 11-1-5 来路统计

3、用户所使用的搜索引擎及其关键词

从流量分析软件中可以很清楚的看到，用户是通过搜索哪些关键词来到你的网站的，这可以辅助你对关键词实际优化情况有个大致了解。

另外一个很重要的一点是，从这些关键词中你可以扩展出很多可以增加的内容。有很多用户会搜索一些你很难想到的关键词，可能你的文章里偶尔提到过，恰恰是这些你想不到的，也许还比较冷门的关键词，能给你带来很多流量。发现这些关键词就可以适当的把自己内容发展方面的策略做一些调整。

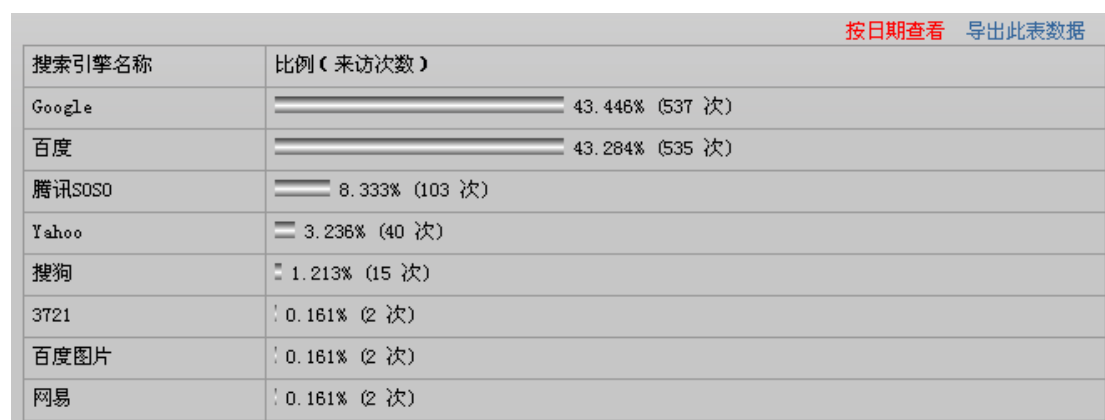


图 11-1-6 用户所使用的搜索引擎

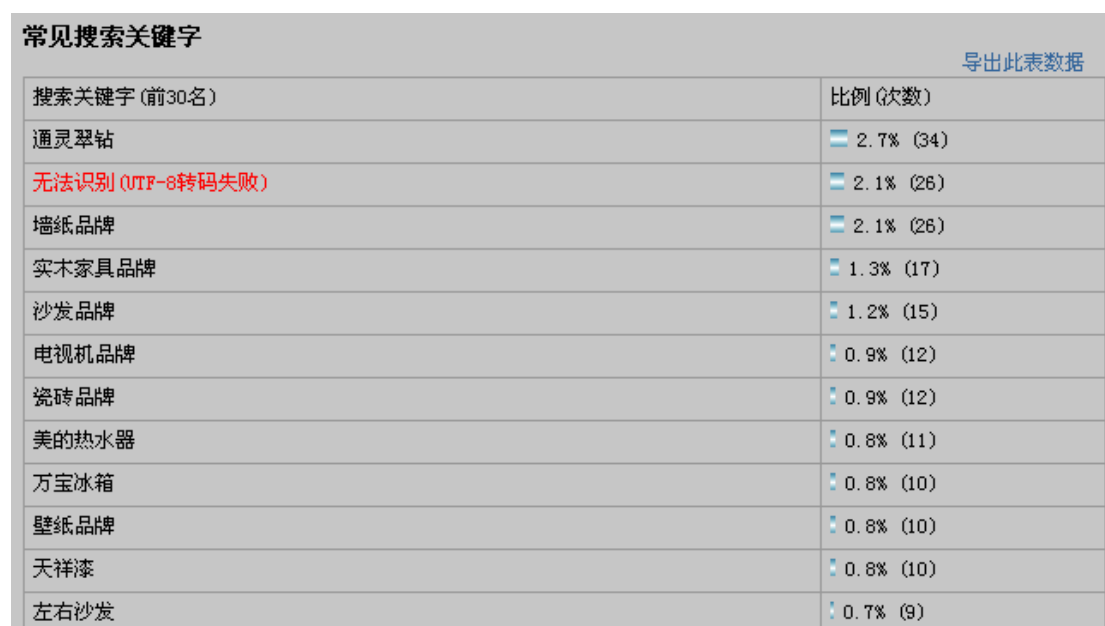


图 11-1-7 用户所使用的关键词和搜索引擎

四、用户浏览网站的方式

用户浏览网站方式的相关统计指标主要包括：用户上网设备类型，用户浏览器的名称和版本，访问者电脑分辨率显示模式，用户所使用的操作系统名称和版本，用户所在地理区域分布状况等。

第二节 中文常见流量统计系统介绍

一、51.la 统计服务

这个可以说是国内最经典的统计服务了。“我要啦”的功能是所有统计服务中比较丰富的，连不是很重要的屏幕颜色屏幕分辨率都可以查到。不过比较实用的功能还是关键词分析功能，可以通过这一功能了解到访客是通过搜索哪些关键词找到了你到网站。另外网站排名、SEO 数据分析等对于了解网站的概况也很有用处。51.la 的缺点就是有少数时间会对页面载入速度有一定的影响，毕竟它要统计的功能太多了。此外，51.la 的统计代码为了保证绝对有效，连客户端不支持 javascript 的情况都考虑到了。

二、网站管理员统计

网站管理员统计在速度和稳定性上比 51.la 还胜一筹，功能上只是比 51.la 略少一些，但核心的统计功能可以说是毫不逊色，这也是不少网站管理员倾向于网站管理员统计的原因。其实准确地说，国内的统计服务在功能上没有哪个会特别突出，也没有哪个特别糟糕。他们提供的网站管理员工具也非常实用，主要也是关于 SEO 方面的东西。比较不舒服的是它“记住用户”的功能不是很全面，有时候还是要再次输入密码来查看统计数据，这点上不如 51.la 方便。

三、ItSun 统计

在优势劣势上感觉和网站管理员统计类似，不过也有少数时候不稳定，甚至还会出现“服务器正忙，请稍后查看统计数据”这样的提示，功能上也比上面两个统计要少。

四、51Yes 网站流量统计

速度飞快，页面载入速度竟然高达 130kb/s！功能上似乎和前两个统计系统也没有太大的差别。服务也非常稳定，使用这个统计系统的大网站也不在少数。没有“记住用户登录”的功能，统计图标有点难看，是个小小的缺憾，不过整体给人的感觉非常好，是个很值得推荐的统计服务。

五、武林榜、太极链

武林榜统计服务，和太极链，虽然也是比较老牌的服务（尤其是武林榜），但现在的速

度都不是太理想，统计功能倒也不是太逊色，不过使用这两个服务的网站现在好像已经不像从前那么多了。

六、一统天下统计服务

主要是个付费统计服务商，也提供少量的免费统计服务，有不少大网站使用他们的统计，不过对于一般的 Blog 和小网站来说，花钱去买一个统计服务似乎就没什么必要了。

七、趋势流量分析系统

不是一个专门的统计网站，是吾爱博客网站管理员提供的一个统计服务，功能上一般，只是速度上比较快，所以写出来作个参考。不过目前这个服务的用户数量好像不是很多，不知道用户多了还能不能保持这个速度，这几乎是它唯一的优势。

下面是一些国外的统计服务，如果你的读者有不少生活在外国地区的，国内的统计代码对他们来说可能就比较慢了，应该适当的考虑一下他们的感受。

八、Google Analytics

最麻烦的问题莫过于拿到它的邀请码了，其实它如此吸引人也正是这个原因。GoogleAnalytics 是 Google 收购 Urchin 推出的服务，原来好像是要 199 美元一个月的，现在只要你的每月 PageView 在 500 万以下就可以免费使用。

但使用过后你就会发现这个东西不过如此，功能还不如国内很多不太知名的流量统计服务。而且 Google Analytics 的统计不是实时更新的，最多只能看到两个小时以前的统计结果，如此滞后的统计让人很不满。只是 Google Analytics 的统计浏览模式设计得比较直观大方，看起来很舒服，它的统计代码在速度上也不错，这一点对某些网站和 Blog 来说还是有点吸引力的。

九、Measure Map

Measure Map 主要应用在 blog 链接查询，通过 Measure Map 可以更方便地了解有多少 blog 链接自己的 blog。注册好像更困难，但似乎和其他的统计服务有着一定的区别，比较新颖。

十、Statcounter.com

国外的老牌统计服务商，一个非常不错的统计系统，不仅提供了很多细节上的统计数据，最主要的是其导入速度比较快，不会因此拖慢了网站的速度。不过在国内的速度还不是特别理想。

第三节 流量统计分析实例

在前面我们提到了流量统计能读出的统计指标并介绍了的流量统计分析系统。接下来就以某旅游站点的后台流量统计为例，来介绍流量统计分析。

说明：该网站访问流量数据分析基于第三方统计系统“我要啦”(<http://www.51.la>)而进行。

一、典型性数据采集抓取

	访问量	比例	浏览量	比例
2006-12-19 (上周二)	27753	16.9%	103343	17.8%
2006-12-20 (上周三)	24633	15.0%	84021	14.5%
2006-12-21 (上周四)	28345	17.3%	101383	17.5%
2006-12-22 (上周五)	27211	16.6%	95634	16.5%
2006-12-23 (上周六)	26918	16.4%	89744	15.5%
2006-12-24 (本周日)	23965	14.6%	85213	14.7%
2006-12-25 (本周一)	4965	3.0%	20404	3.5%

图 11-3-1 最近七天分析表

在最近七天分析表中，选择数据最为偏中的 2006 年 12 月 23 日最为分析基础。

二、IP 与 PV

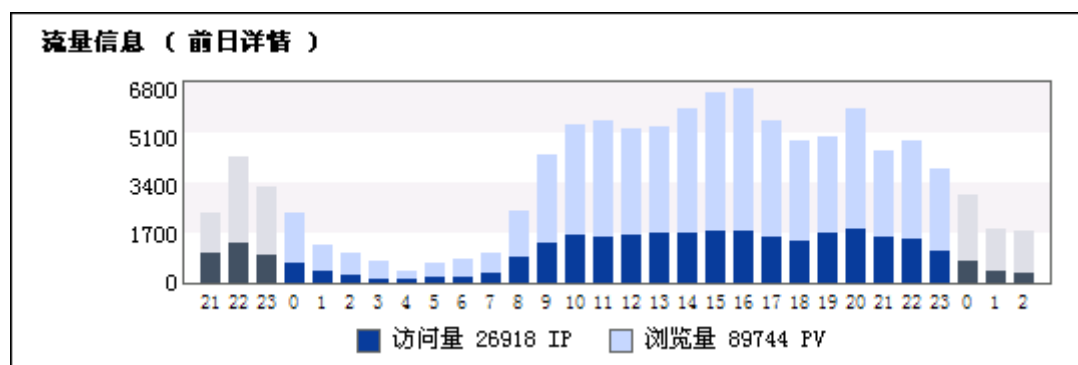


图 11-3-2 IP 与 PV

注：IP=26918 PV=89744 $R = PV/IP = 89744/26918 = 3$

以上的 R 值可以告诉我们，每个访问久游网站的用户大约只浏览 3 个页面。分析说明用户的粘度不是很好，有待提高。

以上结论亦可以通过下面的用户访问回头率截图数据看出：

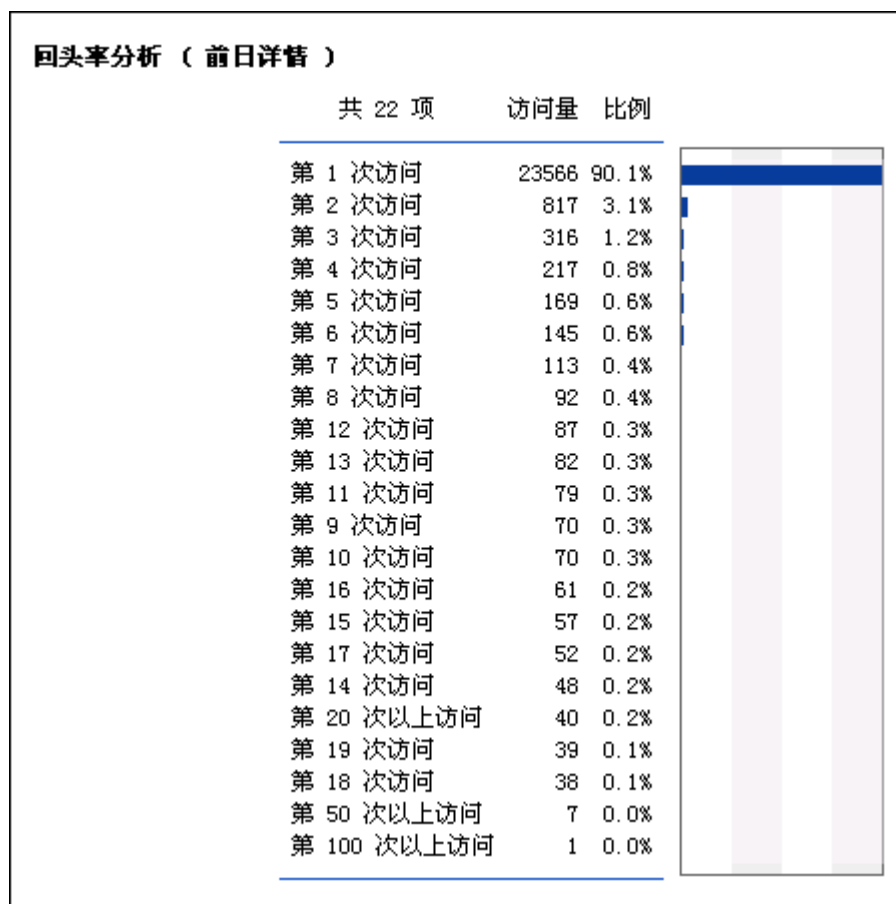


图 11-3-3 回头率分析

90.1%的用户只访问一次，而最多访问 6 次的用户只占总用户人数的 0.6%。

三、网页访问入口分析

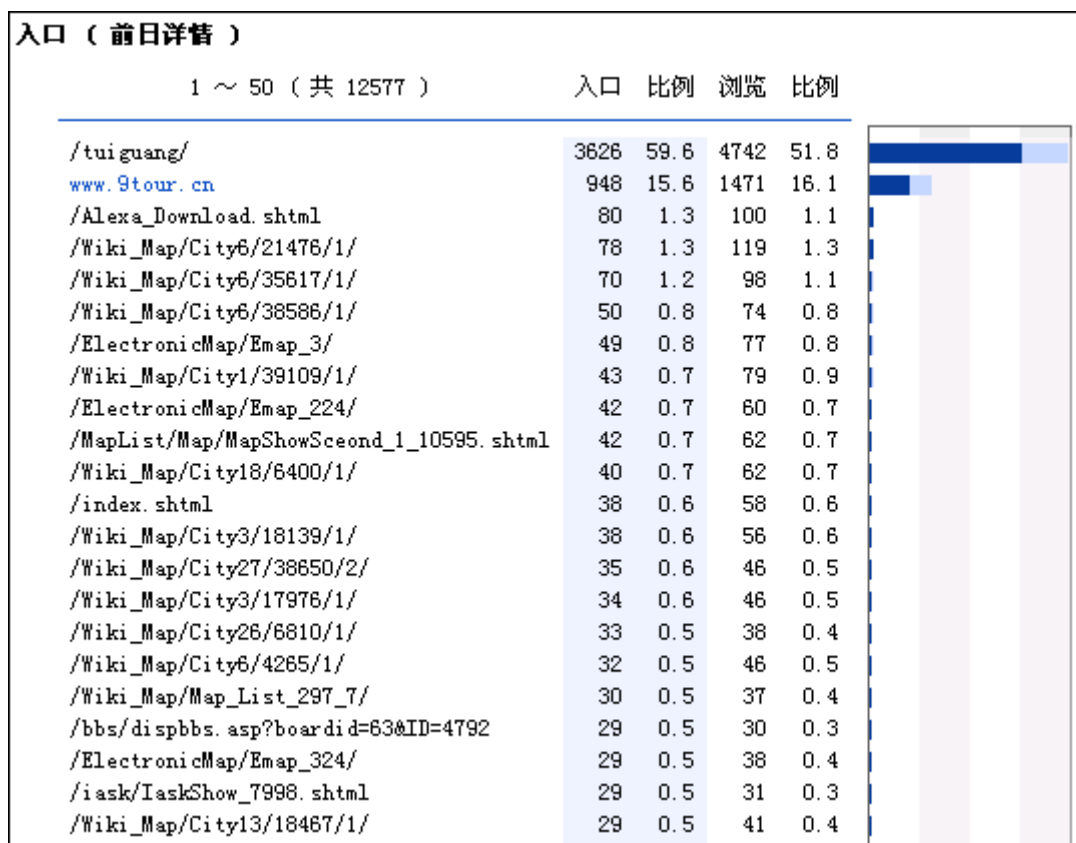


图 11-3-4 入口分析

通过上面数据可以看出，目前用户访问最多的页面为推广部分，然后才是久游首页。这个对于站点的发展是很不利的，推广只是适合阶段性短期计划，如果推广部分终止，该网站的流量将会大幅下降。

其他部分的流量主要来自于旅游景点地图信息页面。这反应了一个很严重的问题，地图信息页面站整个站点的比率比较小，却承载了很多流量访问，而对于那些占网站大部分比率的新闻资讯、景点介绍等页面却没有发挥导入流量的作用。那就说明，久游网的网站策划和内容编辑人员花费了很多精力来做一些收不到回报的事情。

四、搜索引擎流量导入

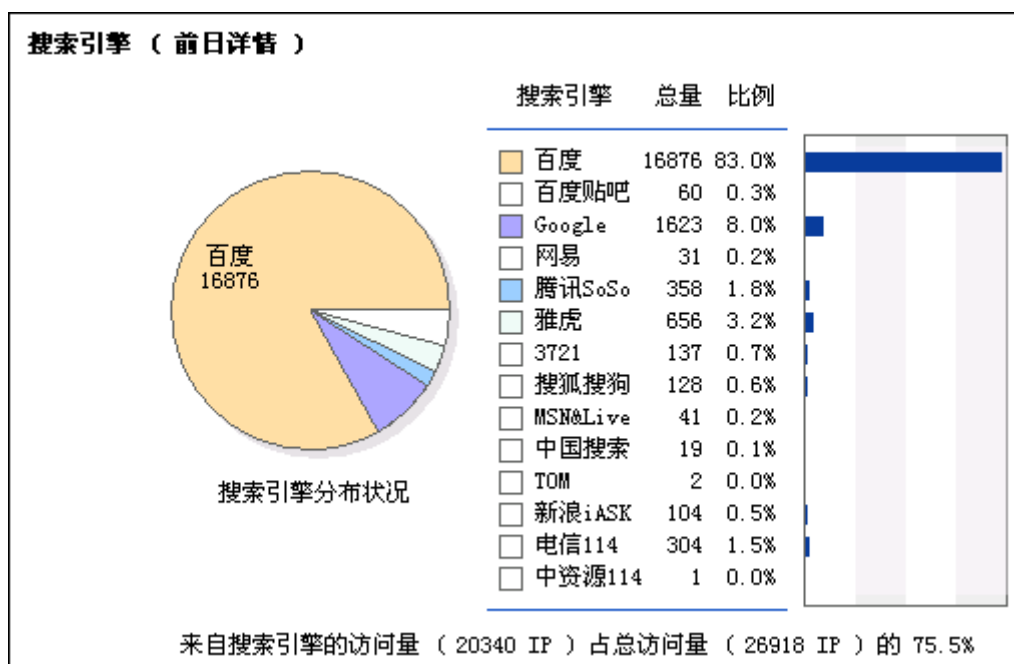


图 11-3-5 搜索引擎流量导入

通过以上搜索引擎分布状况图中的数据，我们可以看出目前从百度导入的流量最大，比率高达 83%，而导入流量身居第 2 的 google 只贡献了 8%，第 1 与第 2 的数据比竟为 10. 这种情况是很不正常的，风险系数很大。如果久游在百度里的表现变差或者因过度优化被惩罚，那就会给整个网站的流量带来重创。

降低流量风险系数的一个方法就是提升网站在其他搜索引擎里的排名表现。

五、搜索引擎关键词分析



图 11-4-1 关键词分析

通过上述数据，我们可以看出关键词构成形式几乎都为“地区+地图”，如果我们事先不知道久游是个旅游门户网站的话，肯定会以为它是提供地图服务的站点。久游站点是以旅游为中心，主要流量导入关键词应该以旅游景点名称导入才为正常。而实际上，我们在关键词的列表里很少看到旅游景点的名称出现。如何把关键词的类型往旅游景点上转变是网站策划和内容编辑人员接下来所需要做的一个重要工作。

第四节 行业热点跟踪

一、Yahoo 风向标：



图 11-4-1 Yahoo 风向标

雅虎风向标是基于雅虎网页搜索的每天上百万的用户查询日志和点击日志，折射社会热点，为用户提供关键词社会关注度查询服务的产品。通过雅虎风向标，你可以通过“关键词查询”了解网络用户对某个事务的关注程度，也可以通过“流量查询”了解当前网络的热门站点，各种“榜单查询”能够快速展现社会和行业热点。

二、百度风云榜：

跟 Yahoo 风向标类似，百度风云榜也是基于大量搜索用户的查询日志和点击日志，折射社会热点，为用户提供关键词社会关注度查询服务的产品。只是由于百度搜索的市场地位，使得该榜含金量更高。如下图：

上升最快风云榜

[搜索风云榜首页](#)

上升最快风云榜

排名	关键词	今日搜索	今日排名	昨日排名	排名上升	相关链接
1	华晨a1	1792	140	4540	4400	评论 新闻
2	黄鲨	1520	168	4498	4330	评论 新闻
3	中国泛蓝联盟	1433	189	4466	4277	评论 新闻
4	傅华	1031	297	4304	4007	评论 新闻
5	凶域计中计	883	595	3905	3310	评论 新闻
6	顺昌地震	877	627	3873	3246	评论 新闻
7	大唐西域记	2804	75	2697	2622	评论 新闻
8	电视剧接触	1077	280	2754	2474	评论 新闻
9	鄢凌	699	1090	3455	2365	评论 新闻
10	达子的春天21集	1028	300	2313	2013	评论 新闻

图 11-4-2 百度风云榜

百度风云榜目前已经成为 SEO 和网站管理员和媒体编辑必看的风云榜。根据该榜针对性地为站点补充内容，制作专题，能迅速为站点带来大量的访问量。

除此之外，百度还有一个很好的产品，叫百度指数。

三、百度指数

1、百度指数

百度指数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，用以反映不同关键词在过去一段时间里的“用户关注度”和“媒体关注度”。你可以发现、共享和挖掘互联网上最有价值的信息和资讯，直接、客观地反映社会热点、网民的兴趣和需求。百度指数可以查询到具体某一关键词的页面搜索次数。百度指数是目前最重要的关键词分析工具之一。



图 11-4-3 百度指数

很多人都通过比较百度指数和百度风云榜，认为百度指数不能反应实际的搜索次数。有时候比如某明星风云榜显示搜索量有好几万，而在百度指数里却只有几千。

这种差别的根源在于风云榜统计了网页搜索、图片搜索、mp3 搜索等所有途径的搜索数据。而百度指数只统计了页面搜索。

2、百度指数官方问答

什么是百度指数？ 百度指数是用以反映关键词在过去 30 天内的网络曝光率及用户关注度，它能形象地反映该关键词每天的变化趋势。

如何注册百度指数？ 百度指数系统的注册必须通过邀请来进行。即你必须通过一个已经在使用百度指数系统的朋友邀请你，你才有机会注册使用百度指数。

什么是邀请注册序列码？ 邀请注册序列码就是你的朋友邀请你注册百度指数系统时系统发送给你的随机的字符串。你在注册的时候只有正确填写了邀请注册序列码才能完成注册。

什么是榜单？ 榜单就是你要监测的一组关键词的集合。在这组关键词中，你可以输入不大于 100 个关键词来监测。其中每个关键词还可以配置 10 个以内的同义词。

什么是同义词？ 在录入榜单监测词表的时候，有的词有在搜索的时候有同义词，如百度，那么 baidu 就是他的同义词，百度搜索引擎也是它的同义词，如果输入了同义词，搜索量和新闻数量会自动归并在一起计算。

百度指数是如何计算的？ 百度指数是综合反映该关键词在过去 1 天用户对它的关注和媒体对他的关注的一个参考值。任意关键词的百度指数都是该关键词在比较期的数值/该关键词在基期的数值。比较期的数值和基期的数值是通过当天的用户搜索量和百度新闻中过去 30 天相关的新闻数量相比得来。

网络曝光度，用户关注度是指什么？ 用户关注度反映的是百度数千万用户的关注，是来自对百度每天 1 亿多次搜索请求的分析获得的。网络曝光度指在线媒体对关注，通过过去 30 天百度新闻中与该关键词相关的新闻的数量来计算获得。

后台如何监测一个关键词？ 如果你建立了榜单，直接输入要监测的关键词，监测过程由系统自动完成，并计算出结果反馈给你。结果以数据柱形图、描述和最近 30 天变化曲线

三种形式展现。

如何设置提醒？ 你登录以后，会在顶部导航栏看到提醒设置，在这里可以开启或关闭整个榜单的提醒，或按照自己的需求配置提醒。也可以通过添加关键词来配置关键词的提醒。

如何查看曲线图？ 登录以后，会在查阅榜单监测结果列表那里看到每个关键词后面的查看链接，点此打开新窗口，即可查阅该关键词的监测结果曲线图，图表的右侧是该关键词今日百度指数，网络曝光度，用户关注度的具体值和与昨日变化比率及柱形图。点击图表上方的关键词提醒设置可以针对该关键词设置提醒，如果该关键词监测的参数达到你的变化要求即可提醒你。点击榜单关键词比较可以比较该关键词与别的关键词的百度指数变化曲线。

如何比较榜单关键词变化趋势？ 榜单关键词比较是指同一榜单的关键词中任意两个的百度指数变化曲线的比较。注意比较的关键词必须是在同一榜单内。在我的指数页面或在关键词曲线图液面都能看到与其它关键词比较的功能。

第十二章 高端技巧

第一节 搜索行为分析

一、什么是搜索行为分析

用户在使用搜索引擎时，动辄返回上千条返回记录，但里面却充斥着大量与用户需求无关的网页，这对用户从搜索引擎中获取知识带来极大的不便。因此有必要对用户的搜索行为进行分析，有针对性地返回用户所需要的记录。这种基于大量统计数据，针对用户搜索的特点，搜索习惯和搜索关键词和点击习惯等要素进行分析的工作便是用户搜索行为分析。

搜索行为分析主要有两种类型：“群体行为分析”和“个性化搜索分析”。现在许多搜索引擎推出的“热门关键词”服务就是一个群体行为分析的典型例子。“热门关键词”是通过以往的查询记录对有相似兴趣的用户进行聚类，同时使用查询关键词扩展技术，尽量提供用户使用最多的关键词，用户可以直接点击这些关键词连接。由于人们的从众心理，“热门关键词”往往也是你想查询的关键词。个性化搜索则是通过积累用户的搜索个性化数据（也就是用户搜索日志），将使用户的搜索更加精确，更符合每个用户的需求。

二、为什么要进行用户搜索行为分析

在 SEO 中，用户的搜索行为分析非常重要。因为只有更清晰地了解用户的搜索行为，才能避开激烈的竞争，选择与目标搜索用户最匹配的搜索关键词，并将最终搜索结果以用户最喜欢的方式呈现出来。

如果不对用户的搜索行为有深入分析，则无法正确地选择最有价值的关键词，无法正确地跟踪，提升站点在搜索引擎上的表现。有人甚至提出，经过用户搜索行为分析，而实施的 SEO 策略，等于变性地提升站点的排名。以符合搜索行为方式呈现的结果，即使排名靠后，但可能吸引的点击比相对靠前的仍占优势。

三、用户搜索行为的几个特点

1、用户浏览的选择性

用户每次搜索时，搜索引擎都会返回成百上千个查询结果，这些结果当中有一部分是质量较高的结果，有一部分是质量很差的查询结果，用户根据结果的快照或摘要进行判断，浏览他所需要的结果。因此，如果用户点击一个查询结果，就可认为用户视此查询结果质量较高；被用户多次点击浏览的页面无疑就是用户认为质量较高的页面，这说明用户的浏览行为具有选择性。

2、用户浏览的局部性

用户点击的网页地址相当集中。大部分用户点击都落在前面几页，拿天网搜索为例。第一页的用户点击占总点击的 47%，而前面 5 页的点击占到总点击的 75% 以上，不到总量 1/3

的页面的点击次数占到总点击次数的 2/3，这表明用户点击 URL 具有很强的局部性，用户点击的局部性提醒我们注意排在前 5 页的查询结果，保证排在前几页的查询结果都是高质量的查询结果。

3、用户点击率的问题

由于网页存在的时间越长，累计下来的访问次数可能越多，故网页被访问的次数不能很好地反映一个网页内容的质量。所以，应使用页面的用户点击率来反映页面的质量。用户点击率是页面被访问次数和页面被搜索次数之和。虽然每次用户的点击都是在某查询项下的点击，但研究表明，在大部分的查询项下，URL 的点击频率和在所有查询项下 URL 的总点击频率基本一致。因此，在计算用户点击率时就不必考虑该点击次数是在什么查询项下的点击次数。

四、搜索行为分析的重点对象

用户搜索行为分析包括的内容非常丰富。常见的分析对象有以下几种。

1、查询词分布情况统计

我们必须了解用户在查找某一信息时，使用什么样的关键词或关键词组合，这些关键词和关键词组合有什么特点，核心关键词和长尾关键词的分布情况等。

2、翻页情况统计

翻页情况统计即分析用户在什么样的情况下考虑翻页以及翻页人数的比例。以及分布在不同页面的点击情况。

3、搜索结果文字描述对点击的影响

主要探讨搜索结果（包括标题，描述，网址，补充材料，快照等）的文字描述，文字长短和文案写作水平，对于用户点击行为所产生的影响。

4、搜索结果不同位置的点击分布

一般而言，排名越靠后，点击量越少，但具体到某个位置和某个页面的点击变化增减情况，所占比例的研究是非常有意义的。研究结果对于 PPC 竞价计划有一定的指导作用。

5、用户对于不同搜索引擎的习惯

主要研究不同地域，不同教育、学历或收入状况的用户在选择搜索引擎上的不同，什么样的情况下用户会转而使用另一搜索引擎等。

五、搜索行为变化情况

1、美国某研究机构统计的 2002-2006 年搜索引擎用户行为变化情况

A. “当你在一个搜索引擎实施一次搜索并查看结果时，在点击结果前，你一般会察看大约多少条结果？（单选）”

回 答	2002.11	2004.04	2006.04
很少几条	16%	24%	23%
第 1 页	32%	36%	39%
前面 2 页	23%	20%	19%
前面 3 页	10%	8%	9%
超过 3 页	19%	13%	10%

B. “当你在一个搜索引擎实施一次搜索，并没有找到你所要的，在什么情况下，你一般才会修正你的搜索或转向另一个搜索引擎？（单选）”

回 答	2002.11	2004.04	2006.04
在查看前几条结果后	14%	23%	16%
在查看第 1 页之后	14%	19%	25%
在查看前面 2 页后	28%	26%	27%
在查看前面 3 页后	22%	15%	20%
在查看超过 3 页后	22%	17%	12%

C. “当你在一个搜索引擎实施一次搜索，并没有找到你所要的，一般情况下你会更可能做什么？（单选）”

回 答	2002.11	2004.04	2006.04
输入更多几个词以使搜索目标更好	68%	-	82%
换另一个搜索引擎并输入与第一次相同的关键词	27%	-	13%
放弃	3%	-	3%
换不同的搜索引擎并输入与第一次不同的关键词	2%	-	2%

D. “请表明你有多赞同/反对下面的陈述：当看到一家公司在搜索引擎的结果里列在第一位时，会让我觉得这家公司在它的领域里是顶级的。（单选）”

回 答	2002.11	2004.04	2006.04
赞同	33%	-	36%
不确定	N/A	-	39%

反对	67%	-	25%
----	-----	---	-----

2、通过上述数据，可以得出如下一些结论

90%的用户的点击出现在搜索结果前3页。如果在前三页找不到所需要的搜索结果，将近90%的用户会修正搜索关键词或者转向另一搜索引擎。

对搜索引擎营销者来说，这很重要：要确保他们的网站在搜索结果的第一页被找到，或者最低限度也得在搜索结果的前3页内，以便被搜索引擎的用户看到。只有10%点击在第三页之外的结果。

同样也可以看出，过去四年来，搜索引擎的用户信任有所提升，使用更多关键词的搜索也有所增多。因此，营销者需要扩大关键词的定向范围。同时在品牌传播效果方面，对于那些出现在自然搜索结果前列的公司，被认为有更多品牌公信力，说明搜索引擎也有助于实现品牌目标。

3、上面数据最终总结出用户搜索行为的三大变化趋势

搜索引擎得到了用户忠诚度 82%的用户在第一次搜索不成功时，并不更换搜索引擎，而是用更多的关键词重新搜索。搜索引擎技术方面的差异性似乎已经不如品牌差异性那么明显。

用户愈加采用长尾查询 用关键词更多、更长的查询。对营销者来说，不仅要定向到简单关键词，还有更多关键词的词语；对于SEO来说，这是一个好消息。相比那些竞争激烈的核心关键词，长尾关键词的排名要容易得多。

用户选择不同的搜索引擎完成不同的搜索需求 尽管搜索引擎得到了用户忠诚度，但是更多的人还是习惯用不同的搜索引擎完成不同的搜索需求。这也反映了不同的搜索引擎的品牌有不同的涵义。比如百度更娱乐化，而Google相对更有商业价值。人们习惯用百度搜索MP3，用Google搜索国际论文。随着不同领域的垂直搜索的诞生，这一趋势更为明显。

六、用户的关键词搜索行为

1、各种各样的关键词组合

关键词的组合是无穷无尽的，不同的人有不同的搜索习惯。想确定不同信息内容对应的检索关键词，搜索必须对该内容有一定的了解，确定一些核心关键词，然后借助于搜索引擎本身或其他关键词工具来对核心关键词进行扩展。比如我们要了解人们关键词“劲舞团”究竟有多少组合，可以通过百度的关键词扩展来分析。



图 12-1-1 关键词组合

跟上面类似，我们同样可以通过 Yahoo，Google 本身的提示，来扩展关键词，来了解不同的关键词组合。

这里，我们有必要指出的几种特殊的关键词组合：

加引号 有一些搜索用户，喜欢在关键词前面加引号。因而对于一些激烈的关键词，我们可以尝试优化带引号的关键词组合。

加书名号 比如“劲舞团”，很多人也会搜索“《劲舞团》”，“《劲舞团》”相对而言竞争激烈程度要大幅度下降，更容易从改关键词获得有效流量。

倒过来 比如对于深圳的一家翻译公司，“深圳翻译”、“深圳翻译公司”等可能根本就找不到他们，但如果客户搜索的是“翻译 深圳”，结果就会不同。所以，针对这种倒过来的关键词，只要稍微处理，也能取得出其不意的效果。

汉语拼音 如广州某装饰企业站点中，通过流量统计系统，发现有不少用户是通过搜索“jiatingzhuangxiu”这样的关键词找到这家企业的。显然，很多人有时候忘记了切换输入法，直接将拼音输入到搜索框，然后顺手就回车进行了检索。而且 Google 曾经有一段时间，在搜索框下面提醒用户可以只输入汉语拼音来检索。这也充分证明了汉语拼音，也是值得重视的关键词。

2、用户也会犯错

用户搜索时候,经常会由于各种原因输入错误的关键词进行检索。常见的错误几种可能:

语文错别字 可能是由于用户的语文知识水平有限,比如很多人会输入“爱滋病”代替“艾滋病”去搜索。而且能区别艾滋病与爱滋病的人并不多。从百度指数来看,艾滋病和爱滋病的检索量也非常接近。但从竞争程度上来看,爱滋病毕竟是错误的写法,信息量要小得多。从搜索引擎营销角度来说,针对“爱滋病”进行一些处理,能避开激烈的竞争,而取得的效果则比较接近。

关键词	用户关注度
艾滋	283 ↓ -12%
艾滋病	1310 ↓ -11%
艾滋病 症状	93 ↓ -28%
艾滋病的症状	465 ↓ -18%
艾滋病症状	589 ↓ -5%
艾滋女生日记	85 ↑ +29%
爱滋	194 ↓ -22%
爱滋病	999 ↓ -13%
爱滋病的症状	175 ↓ -7%
爱滋病试纸	61 ↑ +0%
爱滋病图片	102 ↓ -9%
爱滋病症状	297 ↓ -5%

图 12-1-2 语文错别字

输入法错误,比如奇侠传输入错误 由于汉字输入法的不同,也会诞生大量的输入法错别字。比如采用搜狗输入法的很容易把“七侠传”打成“奇侠传”,使用智能 ABC 的可能把“下载”变成“下栽”等。在《虹猫蓝兔七侠传》大热的时候,能针对“虹猫蓝兔奇侠传”进行处理,也能收到非常好的效果。

记忆有误,比如 waga 由于很多搜索用户是在别的地方听说某一东西或者见过几次改关键词,但记忆不太清楚,因而会输入很多形似或音似的错词来进行搜索。比如很热的哇嘎(Vagaa),到了中国,很多人使用 Waga 代替 Vagaa 去搜索。而相比“Vagaa”,“Waga”的竞争程度要小得多,但流量依然客观。

3、越长尾关键词针对性越高

用户的所有搜索行为中,20%的关键词产生了80%的访问量。是不是另外20%的访问量就不用关注了?我们必须注意到的一点是:另外20%访问量的关键词转化率可能更高!

在实际的关键词组合分析中,多数产品的固定名称只有几个,而根据不同用户的搜索习惯、搜索行为衍生出的100个甚至500个关键词,尽管每个月只有少数几个人搜索了它们,

而且还分布在不同的时间、不同地域中，但就是这些看起来微不足道的流量，汇聚成了用户每月 60% 以上的订单，众多小市场可以汇聚成与主流大市场相匹敌的市场能量。

所以在关键词分析中，尤其是产品类的传统制造业企业站点中，长尾关键词是绝对不能忽略的。

比如通过百度指数和经验分析，发现机票类站点推广上，选择大量包括国内 140 多个目的地、国际 152 个地区，900 个目的地的 50,000 余个长尾关键词进行推广，总体流量上跟“特价机票”、“机票”等核心关键词差不多，但费用大幅度下降，而客户转化率则更高。具体而言，长尾关键词的扩展形式无穷无尽。不同类型关键词的扩展方式不同。一般情况下长尾关键词需要通过关键词分析工具来获得，但大致的常见关键词长尾可包括以下几种：

相关关键词词根排列组合而成 比如围绕“贞观长歌”这部电视剧，类似的扩展可围绕演员名称、剧情、下载、在线视频、片尾曲、剧集介绍等排列组合而成。前面提到的机票类关键词长尾也是这种方式。

添加一些语气助词 尽管大部分的人更喜欢用精简的关键词进行检索，但还是有不少人用带有语气助词的更接近口语的语句进行检索。比如“4月6日火箭跟公牛的比赛录像”，“艾滋病的治疗”，“治疗的口腔溃疡的偏方”，“贞观长歌的剧情介绍”，“从南京到厦门的机票”等等。

问答式的关键词 健康医药类关键词长尾中，这一类型居多。比如“什么是艾滋病”，“苏丹红是什么”，“少年为什么有白发？”等。

第二节 关键词分析

一、什么是关键词

关键词，就是你输入搜索框中的文字，也就是你命令搜索引擎寻找的东西。所以关键词的内容可以是：人名、网站、新闻、小说、软件、游戏、星座、工作、购物、论文、关键词，可以是任何中文、英文、数字，或中文英文数字的混合体。例如，你可以搜索“大话西游”、“windows”、“911”、“F-1 赛车”。

关键词，你可以输入一个，也可以输入两个、三个、四个，你甚至可以输入一句话。例如，你可以搜索“爱”、“美女”、“mp3 下载”、“游戏 攻略 大全”、“蓦然回首，那人却在灯火阑珊处”。

二、如何挑选正确的关键词

1、正确关键词的标准

关键词的研究和选择是搜索引擎营销活动中至关重要的一步。SEO 工作是围绕产品关键词进行的，因此关键词的选取事关整个 SEO 工作能否有效的开展；众所周知，大多数人在网上寻找信息都是从搜索引擎开始。我们都是通过输入关键词段来寻找想要的信息。目前，大多数人搜索时使用平均 2-5 个关键词。

因此，选择恰当的关键词对于优化网页内容变得越来越重要。因为搜索引擎主要提供与关键词有关的内容，比起那些没有用适当的关键词来描述其产品或服务的网站，网页上如果有相关的关键词，该网站的排名会更靠前。

如果关键词选择缺乏合理性，可能导致如下情况：

- **关键词热度不够：**很少人搜索设想的关键词，即便关键词排名再高，引入流量也不会大。
- **关键词竞争激烈：**SEO 工作难度巨大，花费大量时间和精力却长时间达不到预想的目标
- **直接影响网站内容。**

关键词选取在网站策划阶段就应该考虑进去，网站定位、栏目设置、产品所在行业的特点、目标群体所在区域等因素都或大或小影响关键词的选取。

2、选择关键词的基本原则

■ 关键词不能过于宽泛

大家都知道，“手机”这个关键词每日的搜索量巨大，如果能在该关键词上取得好的排名，肯定能引入不错的流量进而可以提高在线销售的转化率。可想而知，这个关键词的竞争将非常激烈，一个以手机在线销售的网上店铺值得去为这个词拼的头破血流吗？

这些通用关键词，例如：“旅游”、“广告”、“机票”等的竞争者数不胜数，可能你有钱投入，一定要获取好排名，但你也要想到，会有很多人拿出更多的钱跟你竞争，也许最终你会很不划算，还是忘记通用关键词吧；即便你能排到前面，确实带来了不小的流量，但是由于搜索通用关键词的用户的目的并不明确，这些流量并不具有很强的目标性，而且用户所在地区未必含概在既定的市场范围，其订单的转化率也很低，实际上是花费了更多的成本却没有得到相应的回报。应该将区域、品牌等因素都考虑近来，譬如：“北京手机”、“三星手机”等，这样的关键词的竞争压力将小很多。

■ 关键词也不能过“冷”

过“冷”的关键词取得好排名要容易的多，但是引入的用户量将会小；千万别以公司名称为做主要关键词，几乎没有人搜索你的公司名，即便你有一定的品牌知名度。SEO 的目的是让不知道公司及产品的人转化为客户。

■ 调查用户的搜索习惯

如果你打算买一部手机，打算了解即将购买的产品，那你会通过 Google 搜索什么词汇来获取信息呢？

网站设计者、经营者由于过于熟悉自己的产品及所在行业特点，在选择关键词的时候，容易想当然地觉得某些关键词是重要的、是用户肯定会搜索的，但实际上，用户思考方式和商家不一定一样。比如说一些技术专用词，例如一些笔记本的具体型号等，普通客户也许并不清楚，也不会用它去搜索，但卖产品的人却觉得这些词很重要，具体型号可能会决定不同的价格。

所以在选择关键词的时候，应该做一下调查，比如问问朋友网友，如果要搜索这类产品，他们会用什么词来搜索。最有效率的关键词就是那些竞争网页最少，同时被用户搜索次数最多的词。

■ 关键词要和网站内容息息相关，网站内容体现着产品关键词

可以基于关键词来提供内容，也可以根据现有内容来提炼关键词。不论如何，内容都必

须覆盖你所需要考虑的目标关键词。

3、挑选关键词

如何挑选关键词是网页优化中最重要的步骤之一，同时也是多数人容易忽视的问题。人们往往随随便便想出几个关键词，将它们用在网页中，然后把这些页面提交到搜索引擎，以为就能得到好的排名。殊不知此举通常是不会为你带来理想的效果的。

■ 第一步 确定核心关键词

坐下来，打开你最喜欢的文字编辑工具软件，再在另一个窗口中打开你的网页，从头到尾将你的网页读一遍，然后停下来思考下面的问题：

- 我的网页是有关哪方面的内容？
- 文中哪些词能够准确地描述整个网页的内容？
- 当某个人在搜索引擎上查找类似的信息时，他可能使用的关键词有哪些？

对于单一产品的网站，核心关键词就是这类产品的名称；对于有多种产品的网站，每种产品的名称都是核心关键词。

■ 第二步 核心关键词定义上的扩展

首先，很多产品不只有一种名称（如电饭锅又叫电饭煲，照相机又简称相机等）；

其次，如果产品不是针对单一地区的话，还要考虑产品在不同地域的名称问题（如鼠标在台湾等地也叫滑标或游标，米线四川人叫米粉等）；

最后，将核心关键词进行组合，与厂商、品牌、原产地（或盛产地）组合（如格力空调、三鹿奶粉、涪陵榨菜等），与产品特性组合（如光电鼠标、真皮皮带、玻璃茶环等），进行多重排列组合（如彩屏和弦手机、数字高清电视、真皮单人沙发）。

■ 第三步 站在客户角度扩展关键词

这需要模拟客户思维，假定自己是一个准客户，要购买某种产品，一般会用什么样的词去搜索。完成这一步需要从多方面着手，可以根据客户反馈的资料中去发现，可以向客户、销售人员、代理商咨询，也可以让多人模拟客户思维，然后将他们搜索的词进行整理——毕竟，每个人的习惯都不一样。

■ 第四步 利用关键词工具

使用搜索引擎工具，将按上述步骤扩展出来的关键词列表进一步扩展、完善。

Google 关键词工具：<https://adwords.google.com/select/KeywordToolExternal>

百度相关搜索：<http://d.baidu.com/rs.php>

在这些工具中，输入列表中的关键词，查询与这个词相关的系列关键词或关键词组合，工具会按热门程度（搜索量）进行排序。这一步中，我们可以做以下工作：

- 参考查询结果中的“被搜索次数/搜索量”，在关键词列表中进行标注，供关键词部署时参考；
- 在搜索出来的关键词列表中，选择最适合于某一类产品或某一件产品的词或组合，对关键词列表进行扩充；
- 将关键词列表中查询结果没有的词或组合进行筛选，这些词要么搜索量很低没有实际

意义，要么本身组合错误。

■ 第五步 参考竞争者的关键词

可以浏览竞争者的网站，看看是否有没你还注意到的关键词，这样可以更好的完善关键词列表。

以你关键词列表中的任何一个词搜索，结果反馈里的网站都是竞争者，建议研究前 10 名（不包括付费推广的网站）作为竞争对手。当然，也不需要列表中的词一一搜索，找出竞争者的网站，再一一去看，可以适当挑选比较核心的关键词进行搜索，再从结果中选择部分网站进行研究。

虽然，认真按上述步骤所创建出来的关键词列表，就基本上没有多少关键词不包含在内了，但百密也可能有一疏，也许竞争者网站中就暗藏着一个你没有注意到的、有效关键词指数（KEI, Keyword Effectiveness）极高的词。

当你逐步完成了这五步之后，你会发现你拥有了一个准确完善的关键词列表了，这将为网站优化的关键词部署及整个网站优化的过程提供重要的参考。

■ 第六步 关键词长尾

在关键词搜索中，长尾效应是十分明显的，有差不多 20%-25% 的搜索都是以前从没有被搜索过的。也就是说用户会搜索各种各样稀奇古怪的关键词，而且这种搜索数量巨大。所以在优化网站关键词的时候，主要的目标关键词是一方面，这些长尾关键词也是非常重要的一方面。

做关键词研究的时候，应该列出几十上百个相关关键词组合，不过，主页只适于两到三个关键词优化，剩下的相关关键词完全可以融入到其他网页中，比如针对每一个关键词专门写一篇文章。

■ 第七步 挖掘日志文件 logfile

网站有了一定流量后，观察研究网站的日志文件是所有搜索引擎优化人员必做的功课，其中一项原因是看用户都是通过哪些关键词搜索来到你的网站。有的时候会看到让你很惊奇的关键词。

通过研究这些已经来到你网站的搜索，你可以知道用户对什么感兴趣，你的哪类文章需要加强，以及哪些关键词有更高的潜力。找到这些有潜力的关键词后，可以进行扩展，然后针对这些关键词写网页内容。

■ 第八步 词干技术

这只适于英文网站。所谓词干技术是指对同一个词干所衍生的不同的词，搜索引擎都会认为是同一个意思。

比如说 work 这个词，假设它是你的关键词，那么在你的网页文章中，应该交替出现 work 的不同形式，比如 working, worker, worked 等等，对名词还有单复数的变化等。你的网页中应该自然融入这些不同的源于相同词干的词。

三、竞价关键词的选择

1、什么是关键词竞价

关键词竞价是搜索引擎的增值服务。企业可以参与搜索引擎任何一个关键词的竞价，获得任何一个想要的排名。通过自然排名，还可能在 Google、百度等搜索引擎拥有前 20 甚至前 3 的位置，让你的买家一眼看中。

2、什么是关键词？竞价时该如何选择关键词？

关键词就是客户搜索产品时最可能用的名称，例：轴承、玩具、工艺品等。

参加关键词竞价是为了让供求信息海量信息中取得领先排名。

一般买家搜索时只用产品名为关键词，所以建议你选择关键词时只用产品名，去掉所有修饰词。比如“带不干胶的手提包”，只用“手提包”。

3、竞价关键词选择技巧

寻找大量不太热门的关键词 当然这些关键词也要和你的产品和服务相关。如果只是把目标定在行业里最热门的几个关键词，竞标的价钱会很高。如果你能寻找到大量相关的但不太热门的关键词，平均每个关键词的价钱很低。但是由于你可以寻找大量的这类关键词（通常在几百上千以上），那么你能得到的流量不会比从热门关键词来的少，有可能更多，但是平均价格却要低得多。这是 PPC 排名竞价搜索引擎广告服务最重要的关键。

关键词的选择和竞价的高低，应该以投资报酬率为成功与否的标准 有的时候，某个关键词可能竞价很高。但是如果你所卖的服务本身价值很高，那么你的投资报酬率可能依然很好。

对所有的关键词要进行完整跟踪分析 对所有关键词价格、点击率、转化率应该做完整的跟踪和分析。只有这样，才能知道哪些关键词带给你最高的点击率，哪些关键词带给你最高的投资报酬率。经常的情况是，一些不太热门、价格也不高、但是相关性很高的关键词投资报酬率最高。因为访客所要寻找的东西非常具体，非常明确。

不停地测试比较 究竟哪些关键词带来了销售额等等。所有这些你都应该跟踪，并适当做一些变化，来达到最好的效果。有的时候，我们主观认为可能效果最好的组合，实际上并不是最好的方法。实验数据才是最后的依据。

4、中文关键词选取工具

- **百度搜索词频** 百度曾列出关键词每日搜索量所在的数量区段，后来改为只列出相对比例；
- **Sohu 搜索词频** 列出该关键词在 SOHU 上的搜索次数区段；
- **Google 关键词工具** 输入关键词可以生成大量相关关键词作为参考，并且可以查询相关关键词受欢迎程度（即搜索热度）、在 Google 上获得相应排名所需的费用以及搜索量趋势。

5、英文关键词选取工具

- Overtrue 提供的搜索词频工具（已经被 Yahoo 收购）；
- Google 关键词工具也同样可以用于英文关键词选取。

四、关键词竞争程度分析

1、关键词竞争程度衡量指标

我们的目标是找出搜索次数相对较多同时竞争还小的关键词，以达到最高的效能。关键词竞争程度有以下两个衡量指标：

搜索结果右上角列出某个关键词返回的总相关网页数 这个数字大致反映了与这个关键词相关的网页数，而这些网页都是你的竞争对手。但一定要注意页面总数只能参考，有时候页面数很少，但竞争激烈；有时候虽然页面数很多，但由于搜索关键词过于宽泛，而无实际商业价值，这词不管从优化角度和竞价角度竞争都不激烈。

该关键词在某些搜索引擎提供的竞价排名广告中排名前三位所需花费的费用 可以使用雅虎的 View Bids Tool 查询某个关键词需要付的价钱，Yahoo 这套 PPC 查询工具支持中文。还可以使用 Google 提供的关键词扩展工具、百度的关键词价格查询工具，或者 Sohu 的关键词价格查询工具，这些关键词的价格比上面说的竞争网页数更能说明竞争程度。

2、计算关键词效能

有了关键词的竞争程度和被搜索次数，就可以计算出哪些关键词效能最高。

当然最简单的计算方法就是搜索次数除以竞争程度。在这个公式里面还可以做适当的变化。比如在竞争程度中，如觉得 PPC 价格重要性更高一点，可以把被除数改为：

$$0.4 \times \text{搜索次数} + 0.6 \times \text{PPC 价格}$$

也就是给总搜索次数和 PPC 价格不同的权重，当然这个权重各占多少就是你主观判断和偏好了。

关键词选择：就是选择效能最高的两到三个关键词作为你主页的目标关键词，其他相关关键词作为辅助。

3、实例分析

以某油画类站点为例。该站点目前搜索带 China 类的关键词已经靠前，那么如何为其下阶段的优化来确定重点处理关键词，以追求较高的搜索量和转化率，而竞争程度相对较低？

Overture 给出了关键词的大致搜索次数。而 Google 提供的关键词工具，同时给出了改关键词的广告竞争程度和搜索热度。

我们需要重点注意的关键词为：广告竞争程度激烈且搜索量较大的关键词。广告的竞争程度越大，说明改关键词与具体的产品销售更相关。而搜索量越大，意味着更多的潜在客户在使用该词搜索，寻找他们需要的油画。尽量避开搜索量大，但与产品购买无关的关键词。比如“oil painting technique”。

从上面两个列表中可以很清晰地了解具体的关键词扩展形式和搜索次数及广告竞争程度。特别是 Google 提供的数据，对于 SEO 来说非常重要。

从长远来说，那些竞价竞争程度激烈并且同时搜索量也很大的词，是长期的 SEO 战略必须考虑的。因为竞价程度激烈说明该词是符合搜索需求，有一定的转化率，否则不会那么多同行去竞价，竞价越激烈说明该词的综合商业价值越高。另一方面这类词搜索量很大，意味着 SEO 后将会获得很好的回报，节省大量的广告费用。

从短期 SEO 策略来看，则需要根据实际情况，选择有一定搜索量，跟产品销售相关，

同时关键词包含单词越多越好，因为一般情况下，越长的关键词竞争相对更激烈。
加粗部分为推荐的下阶段重点关键词。

表 12-2-1 overture 提供的关键词月度搜索量

每月搜索量	关键词
55666	oil painting
6899	oil painting technique
3093	oil painting reproduction
2350	oil painting picture
2319	floral oil painting
2318	oil painting on canvas
2235	oil painting instruction
2128	painting in oil
2076	original oil painting
1772	landscape oil painting
1554	oil painting portrait
1328	oil painting artist
1159	art oil painting
1056	oil painting for sale
953	oil painting secret from a master
904	oil painting book
781	fine art oil painting
658	abstract oil painting
648	beautiful color in light oil painting pastel skin tone watercolor
621	wholesale oil painting
611	nude oil painting
608	antique oil painting
597	fill your oil painting with light color
584	from oil painting picture
554	oil painting supply
522	china oil painting
494	flower oil painting
428	still life oil painting
422	oil painting lesson
399	oil painting frame
393	oil painting usa
385	oil painting for the serious beginner
383	oil painting from photo
383	oil painting gallery

379	oil painting italy
371	cat oil painting
359	custom oil painting
341	oil painting for beginner
332	chinese oil painting
310	framed oil painting
306	adam new oil painting york
305	bronx new oil painting york
305	buy oil painting
303	adirondack new oil painting york
303	alma new oil painting york
300	oil painting material and technique
299	france oil painting
291	addison new oil painting york
284	astoria new oil painting york
284	get oil painting
278	painting with water soluble oil
269	artist book material oil painting technique today
269	best oil painting
268	angola new oil painting york
266	oil figure painting
266	oil painting photo
265	flower fruit in oil painting sumptuous vegetable
263	seascape oil painting
261	bombay new oil painting york
259	cleaning oil painting
258	de oil painting
256	famous oil painting
253	oil painting restoration
247	dog oil painting
244	acrylic effects elements in oil painting watercolor weather
243	horse oil painting
240	free oil painting lesson
233	english oil painting
232	oil painting class
231	brookfield new oil painting york
230	animal oil painting
229	oil painting kit
227	modern oil painting
226	painting the beauty of flower with oil
222	western oil painting

220	albany new oil painting york
219	alton new oil painting york
216	oil painting tip
213	angelica new oil painting york
211	acrylic and oil painting
209	accord new oil painting york
207	it oil painting
204	atlanta new oil painting york
203	alpine new oil painting york
203	bradford new oil painting york
203	brooklyn new oil painting york
202	akron new oil painting york
202	alabama new oil painting york
201	ashland new oil painting york
201	bayside new oil painting york
195	almond new oil painting york
192	discount oil painting
189	contemporary oil painting
188	oil pastel painting
187	bob ross oil painting
185	painting in oil
182	oil painting video
181	bloomfield new oil painting york
180	brentwood new oil painting york
177	brunswick new oil painting york

Keywords	February Search Volume	Advertiser Competition	Match Type: ?
oil painting			Add >
oil paintings			Add >
oil painting reproduction			Add >
oil painting gallery			Add >
oil paint			Add >
oil based paint			Add >
oil painting techniques			Add >
how to oil paint			Add >
original oil paintings			Add >
oil paintings for sale			Add >
oil painting classes			Add >
oil painting lessons			Add >
oil painting tips			Add >
oil painting artist			Add >
original oil painting			Add >
oil painting reproductions			Add >
abstract oil paintings			Add >
reproduction oil paintings			Add >
oil painting restoration			Add >
oil painting artists			Add >
painting with oil			Add >
landscape oil paintings			Add >
oil painting supplies			Add >
art oil painting			Add >
painting oil			Add >
landscape oil painting			Add >
abstract oil painting			Add >
cleaning oil paintings			Add >
new century oil painting			Add >
oil painting how to			Add >
oil painting portrait			Add >
art oil paintings			Add >
oil painting tutorial			Add >
oil painting frames			Add >
oil painting class			Add >
famous oil paintings			Add >
antique oil paintings			Add >
chinese oil paintings			Add >
artist oil paint			Add >
oil painting canvas			Add >
oil painting on canvas			Add >
oil paintings on canvas			Add >
how to paint with oil			Add >
oil paints			Add >
oil painting for sale			Add >
oil paint brushes			Add >
oil painting tutorials			Add >
latex paint over oil			Add >
learn oil painting			Add >
oil painting for beginners			Add >
oil painting courses			Add >
oil painting medium			Add >
paint with oil			Add >
framed oil paintings			Add >
china oil painting			Add >
chinese oil painting			Add >
artist oil paintings			Add >
oil paintings china			Add >
antique oil painting			Add >
wholesale oil paintings			Add >
oil painting instruction			Add >
nude oil paintings			Add >
oil painting portraits			Add >
canvas oil paintings			Add >
oil painting workshops			Add >
history of oil painting			Add >

图 12-2-1 Google 提供的关键词分析工具

五、关键词查询工具

关键词查询工具可以帮助你选择关键词，避开激烈的竞争，独辟蹊径。

百度搜索关键词搜索频率查询：

<http://www2.baidu.com/inquire/dsquery.php>

网易搜索关键词搜索次数及价格查询：

http://adpsearch.163.com/find_price.php

搜狐搜索关键词搜索次数/类目固定排序报价查询：

http://db.sohu.com/regurl/pv_price/query_consumer.asp

新浪搜索关键词价格查询：

<http://chaxun.sina.com.cn/>

第三节 关键词密度

一、什么是关键词密度

关键词密度 (Keyword Density) 与关键词频率 (Keyword Frequency) 所阐述的实质上是同一个概念，用来量度关键词在网页上出现的总次数与其他文字的比例，一般用百分比表示。相对于页面总字数而言，关键词出现的频率越高，关键词密度也就越大。举个例子，如果某个网页共有 100 个词，而关键词在其中出现 5 次，则可以说关键词密度为 5%。

二、正确理解关键词密度

我们给出的例子只是一种理想的简化方式，事实上，在我们的网页中，还有很多的 HTML 部件，如 meta 标签中的 Title、Keywords、Description，如图像元素的 Alt 文本、HTML 文件中的注释文本等，这些在计算关键词密度时也都要考虑在内的。继续使用上文的例子，即网页内容为 100 个词，但同时，页面的 Title 有 10 个词，Description 有 20 个词，图像的 alt 则用了 70 个词，HTML 注释用了 10 个词，这样，页面中的文本字数即为 $100 + 10 + 20 + 70 + 10 = 210$ 。

同样地，在计算关键词频率时，也不仅仅指网页可见部分中关键词出现的次数，也包括在上文中提到的非可见部分中出现的次数。同样以上文的例子来说，如果关键词在页面 Title 中出现 3 次，在 Description 中出现 5 次，在 Alt 文本中出现 30 次，在注释中出现 2 次，那么总的关键词出现次数即为 $5 + 3 + 5 + 30 + 2 = 45$ 。这样，这样关键词密度即为 $45/210 \times 100\% = 21\%$ 。当然，搜索引擎在算法上要比这复杂得多，但基本策略与此近似。

三、适当的关键词密度

不同的搜索引擎，包括 Google、Yahoo 和 MSN Search，在对关键词密度的算法上数学公式有所差别，其接受的最佳关键词密度也不尽相同。就过度优化如关键词 Spam 而言，不同的搜索引擎在采取惩罚前的容忍级别也不尽相同。一般说来，在大多数的搜索引擎中，关键词密度在 2%~8% 是一个较为适当的范围，有利于网站在搜索引擎中排名，同时也不会被

搜索引擎视为关键词堆砌。

四、关键词位置和形式

对页面的排名有影响的不仅仅是关键词密度，关键词的位置和形式也非常重要。下面以国内著名的旅游站点九游网和通用旅游网为例进行阐述。

1、关键词位置

为了确保最基本的关键词密度或内容相关性，关键词更多的时候应该分布在页面中的固定内容上，而不是可变因素上（比如评论）。页面最上面的关键词布局，就是稳定出现、而不是可变的。这样的布局至少保证了页面一开始，关键词至少会出现一次。

统计：共有1099位网友浏览了平遥古城景点主页，当地天气：晋中 晴-晴 5℃/-4℃ 微风转西北风3-4级

图 12-3-1 页面最上面的关键词布局

关键词最好出现在页面的开头或结尾以及连续段落文本的中间。总体而言，九游网大部分页面的关键词密度已经达到了要求，但大部分关键词都不是出现在连续段落文本里面。人为制造关键词密度的痕迹明显。

通过搜索引擎模拟程序测试，正文最开始的内容中，干扰信息和与关键词无关内容占据较多。下图中蓝色背景的文字，占据了最重要的位置，对于搜索引擎对文章内容的影响起到了干扰作用。

Spidered Text :
平遥古城 - 晋中景点 - 山西景点 - 九游网 - 您的出行资讯专家 www.9tour.cn
统计：共有1099位网友浏览了平遥古城景点主页，当地天气：晋中 晴-晴 5℃/-4℃ 微风转西北风3-4级会员注册 会员登录 设为首页 当前位置：首页 >> 目的地指南 >> 山西景点 >> 晋中景点综合搜索 目的地景区景点城市地图宾馆酒店新闻资讯风景图片电子地图游记攻略火车站列车车次百度搜索 Google 景点首页景点介绍 门票价格联系方式开放时间周边中心城市景区荣誉自驾攻略景区玩法最佳旅游时间 附近景区景点 美食小吃 投诉电话 交通指南 购物指南 住宿情况 民俗节庆 娱乐指南 特别提示 电子地图游记攻略旅游地图旅游图片 平遥古城 我要点评 我要提问 发表文章 上传图片 平遥古城1986年国务院公布为第二批国家历史文化名城。目前全县境内拥有全国重点文物保护单位3处，省级重点文物保护单位6处，县级文物保护单位90处。其文物古迹之多，价值之高，内容之丰富，为全国县级城市中所罕见。在这座古城的珍贵文物中，有建于北汉天会七年（983年）的镇国寺万佛殿；有始建于北齐武平二年（571年）的双林寺；有始建于唐显庆二年（657年）的清虚观；有建于北汉天会七年（983年）的文庙大成殿；也有建于清康熙十年（1671年）的惠... [景点介绍][点评景点,得积分] 最新景区点评更多点评>> wlfeng0923 在2006-11-19

图 12-3-2 干扰信息

2、关键词的形式

搜索引擎越来越喜欢接近自然的优化方式，而对于自然的理解，最简单的就是，关键词在同一个页面出现的形式不可能是千篇一律的，文字、字体、格式和链接，都不可能一模一样，这才更接近一个页面自然的状况。

而目前 9tour.cn 景点页面的关键词布局，在同一页面中，恰恰大部分情况，关键词出现的位置、字体、格式、链接都是一样的。

比如文章最开头的关键词和评论及图片名称上的关键词形式，没有任何差别。而且还刻意为了提升关键词，将关键词跟周围的文字区分开来。



图 12 3-3 单一的关键词形式

相反，在 9tour.cn 地图搜索中，为什么虽然看起来关键词密度更高，但实际上统计任何关键词密度比景点还低呢？这主要是因为在地图相关页面中，关键词的变化性处理得不错。以攀枝花市地图 (http://www.9tour.cn/Wiki_Map/City26/41401/1/) 这一页面为例，围绕“攀枝花市地图”，先后在页面内容中出现了“攀枝花地图”、“攀枝花市”、“攀枝花地图”、“攀枝花”、“攀枝花地图册”、“攀枝花西区地图”、“攀枝花仁和区地图”、“攀枝花东区地图”、“攀枝花市地图”、“攀枝花市市区地图”、“攀枝花地图”、“四川攀枝花城区地图”、“四川攀枝花市交通地图”、“四川攀枝花市市区地图”、“四川攀枝花行政区划图”等多种方式。不同关键词的出现位置及链接也有所变化。总体在页面中的分布均匀。

对比竞争对手，如 51766.com 中平遥古城的页面 (<http://www.51766.com/img/pygc/>)。该页面整体检查，没有任何堆积关键词的嫌疑，也没有任何人为制造关键词密度的嫌疑。相反，整个页面内容几乎全部围绕“平遥”和“平遥古城”做文章，总体上是最相关的。全文不到 2500 个字中，包含平遥的关键词有“平遥民俗表演”、“平遥小吃”、“平遥古城”、“平遥”、“平遥大院”、“平遥县”、“山西平遥古城”、“平遥国际摄影大展”、“平遥火车站”、“平遥列车时刻表”、“平遥古城墙”、“平遥城隍庙”、“平遥古民居”、“山西平遥订房”、“平遥东南海宾舍”、“平遥丽泽苑国际酒店”、“平遥云锦成客栈”、“平遥四盛庆民俗客栈”、“平遥古城自由游”、“平遥琉璃”、“平遥水陆画”、“平遥木雕神像”、“平遥交通”、“平遥古城出租车”、“平遥古城乘车交通指”、“平遥长途汽车营运表”、“平遥火车站列车时刻表”、“平遥古城游路线指南”等。

对于搜索引擎来说，当检索到这么多相关的信息，而且一切出现的位置又显得非常自然的时候，没有理由不相信这是一个关于“平遥古城”最详细介绍的页面之一。



图 12-3-4 51766 中平遥古城的页面

五、Google 和 Yahoo 的区别

Yahoo 非常看重站点内容的相关性，Google 更喜欢拥有原创内容的站点。对于 Google 来说，网站管理员和 SEO 们将关键词密度优化在 2%~3% 左右比较合理，而 Yahoo 似乎还需要更高的关键词密度。对于 Yahoo 来说更合理更有效的关键词密度是否会导致 Google 认为你的站点是关键词堆积呢？幸好，两种算法之间还有其他的一些不同的特点，可以帮助解决这个问题，使得进行 Yahoo 优化而不受 Google 惩罚。

Yahoo 很大程度上是以语言为基础的，需要我们更关注语言文字的多样性。如果同一关键词或短语的可以利用四种不同说法，某一说法对应词的密度在 2% 左右，那么我们就可以利用这种差异来提升 Yahoo 的关键词密度到 7% 至 8%，而又保证在 Google 上也在合理范围内。这样，对于 Google 来说，你等于同时针对了四个不同的关键词进行了优化并给予了 Google 认为合理的密度。由于可以使用各种词根进行关键词扩展，也可以轻易地保证文案写作非常流畅自然。

六、关键词密度的基本原则

通过上面的实例分析，我们可以总结出关键词密度和布局的一些基本原则：

- 关键词必须合理地融入内容，文案写作上必须确保关键词的出现是为了配合行文需

要。

- 关键词密度必须合理。
- 虽然整体关键词密度不高，但也需要防止局部密度过高。
- 如果是为了用户体验，即使关键词超过标准，也是合情合理，不会被惩罚的。如下图新浪于丹专题。

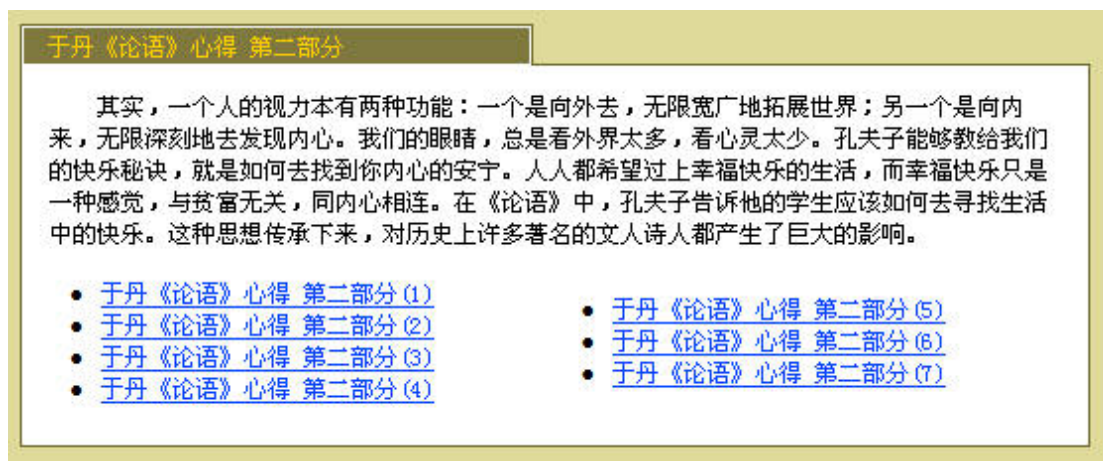


图 12-3-5 新浪于丹专题

- 围绕核心关键词的表现或组合形式应该多样化，而不能过于单一。

七、如何增加关键词密度

1、Meta 和 Title

在 Meta 上和标题上要适当地嵌入关键词。基本原则是：能准确概括页面内容，简洁而流畅。

2、内容

对于搜索引擎而言，“内容为王”的意思，就是丰富的文本中频繁出现你的关键词。一个基本的原则是：关键词布局配合文章的行文来定，保证良好的用户体验。感觉关键词出现是为了行文的需要，而不是刻意堆积。

3、链接

如果正文中有大量超级链接，每个链接文字都写成“点此查看”之类，搜索引擎就极有可能误以为这个大量出现的“点此查看”四个字就是你的关键词。因此，一页中出现多个链接的时候，最好把链接标题直接写出来，用不同的色彩或下划线来标示链接。

4、图片

图片对于搜索引擎来说犹如空白，但你可以通过在图片的代码中运用 ALT 标签使之获得改观。ALT 属性里运用你的关键词，搜索引擎也可以抓取到它们。

八、关键词密度工具

http://tool.chinaz.com/SEO/Key_Density.asp

http://www.timev.com/tools/key_density.html

第四节 信息干扰现象

一、信噪比与干扰信息

关于信噪比，网上广泛流传的定义是指网页中的文本内容部分与生成这些文本的 html 标签内容的比率。我们也可以理解成，网页内容信噪比的定义是网页有用信息跟干扰信息之比。所以理解信噪比的关键就在于什么是有用信息，什么是干扰信息。

搜索引擎对于其所抓取的内容会有一个基本的判断，判断该页面内容对应的主题及所有可能的关键词，并将其进行排序。是有用信息还是无用信息，对于搜索引擎来说，是针对具体关键词而言的。所以网页的信噪比跟具体某一关键词的信噪比应该有所区分。具体关键词的信噪比是网页中对该关键词检索有利的信息和干扰信息之比。而网页的信噪比为该页面核心关键词的信噪比。

网页中如果存在搜索引擎认为与某一搜索词没有任何相关性的内容(即搜索引擎无法判断该段内容与搜索关键词之间的关系)，那么这些内容即为该关键词的干扰信息。这些干扰信息可能包括 flash、图片，及与关键词没有相关、却十分冗长的段落文本等。同理，网页的干扰信息即为网页核心关键词的干扰信息。

二、信息干扰的副作用

干扰信息的存在，一方面增加了站点的代码复杂程度和文件大小，导致网站访问速度变慢；另一方面也稀释了关键词密度，影响了搜索引擎对于站点主题的提炼和核心内容的理解。特别是一些大型站点，过多的干扰信息，导致有效信息比例低下，如果页面的主体内容并不丰富，这些页面之间相似程度太高，就会造成大量页面类似。

为了让搜索引擎更好的理解整个页面内容的主题，对页面进行相关性判断，进而给予合理的排名，必须在整个页面的编排中，突出有效信息，降低干扰信息。

三、如何减小信息干扰

减低干扰信息一个基本原则是，除了必要的上级和重点页面文本导航外，尽量突出站点的重要内容，非重点内容可以采用搜索引擎无法识别的技术来调用，比如框架和 Js 等。

在对站点进行降干扰处理后，一定要通过类似的搜索模拟程序进行检测，看看搜索引擎抓取的文本内容是不是重要的核心内容。

第五节 重定向

一、什么是重定向

网站建设中时常会遇到需要网页重定向的情况：比如因网站调整，网页目录结构发生改变，网页需要被移到一个新地址，又或者，网页扩展名改变，如因应用需要把.php改成.html或.shtml，在这种情况下，如果不做重定向，用户收藏夹或搜索引擎数据库中旧地址只能让访问客户得到一个404页面错误信息，访问流量白白丧失；再如某些注册了多个域名的网站，也需要通过重定向让访问这些域名的用户自动跳转到主站点，等等。

下面详细介绍几种常见的重定向。

二、301重定向

1、什么是301重定向？

301重定向也称301转向，或者301跳转。当用户或搜索引擎向网站服务器发出浏览请求时，服务器返回HTTP数据流，其中头信息(header)里包含某种状态码，301重定向即状态码中的一种，表示本网页永久性转移到另一个地址。

其它常见的状态码还包括：200一切正常，404网页找不到，302暂时转向，等等。

2、为什么要做301网址转向？

除了前面介绍过的网址规范化问题外，还有很多需要做网址转向的情形。比如，为保护版权，你注册了多个域名：

company.com
company.net
company.org
company.com.cn
company.cn
等等。

很自然地，这些域名全部指向一个网站。如用company.com为主域名，其它域名就可以转向到company.com。

或者你注册了公司全称域名longcompanyname.com，但太长，你也注册了缩写域名方便用户记住lcn.com，其中一个做主域名，另一个就可以转向到主域名。

3、301转向的意义

页面永久性移走(301重定向)是一种非常重要的“自动转向”技术。

4、301重定向可促进搜索引擎优化效果

从搜索引擎优化角度出发，301重定向是网址重定向最为可行的一种办法。当网站的域名发生变更后，搜索引擎只对新网址进行索引，同时又会把旧地址下原有的外部链接如数转移到新地址下，从而不会让网站的排名因为网址变更而收到丝毫影响。同样，在使用301永久性重定向命令让多个域名指向网站主域时，亦不会对网站的排名产生任何负面影响。

网址转向方法主要包括：301 转向，302 转向，JavaScript 转向，PHP/ASP/CGI 转向，META REFRESH 网页 META 刷新等。302 转向可能会有 URL 规范化问题。其它方法都是常用的作弊手法，当然不是说不可以正当地用，方法本身没有错，但被作弊者用多了，搜索引擎对这些可疑的转向都很敏感。

当网页 A 用 301 重定向转到网页 B 时，搜索引擎可以肯定网页 A 永久地改变位置，或者说实际上不存在了，就会把网页 B 当作唯一的有效目标。好处是，第一，没有网址规范化问题，第二，也很重要，网页 A 的 PR 网页级别会传到网页 B。

很多时候其它网站会链接到

`http://domainname.com`

`http://www.domainname.com/index.html`

`http://domainname.com/index.html`

而不是你希望的

`http://www.domainname.com`

这样链接的 PR 也会传递到所有 URL，如果你用 301 转向把前三个 URL 转到

`http://www.domainname.com`

PR 也就集中在网页 `http://www.domainname.com` 了。

5、IIS 服务器实现 301 重定向

- 打开 internet 信息服务管理器，在欲重定向的网页或目录上按右键；
- 选中“重定向到 URL”；
- 在对话框中输入目标页面的地址；
- 切记，记得选中“资源的永久重定向”；
- 当然，最后要点击“应用”。

6、Apache 服务器实现 301 重定向

相比较来说，Apache 实现起来要比 IIS 简单多了。在 Apache 中，有个很重要的文件 .htaccess，通过对它的设置，可以实现很多强大的功能，301 重定向只是其中之一。

```
redirect permanent / http://SEO.highdiy.com/:
```

将目录下内容重定向到 `http://SEO.highdiy.com/`

```
redirect permanent /old.html http://SEO.highdiy.com/new-url/:
```

将网页 `old.html` 内容重定向到 `http://SEO.highdiy.com/new-url/`

通过合理地配置重定向参数中的正则表达式，可以实现更复杂的匹配。有兴趣的朋友可参考 Apache 手册。

其他方法还有：

7、PHP 下的 301 重定向

```
<? Header ( "HTTP/1.1 301 Moved Permanently" ) ;
```

```
Header ( "Location: http://SEO.highdiy.com" ) ;?>
```

8、ASP 下的 301 重定向

```
<%@ Language=VBScript %>
```



```
<% Response.Status="301 Moved Permanently" Response.AddHeader "Location", "http://SEO.highdiy.com">
```

9、ASP .NET 下的 301 重定向

```
<script runat="server">
private void Page_Load (object sender, System.EventArgs e)
{
Response.Status = "301 Moved Permanently";
Response.AddHeader ("Location", "http://SEO.highdiy.com");
}
</script>
```

三、302 重定向

302 代表暂时性转移 (Temporarily Moved)，前些年，不少 Black Hat SEO 曾广泛应用这项技术作弊，后来，各主要搜索引擎均加强了对应的打击力度，像 Google 前些年对 Business.com 以及近来对 BMW 德国网站的惩罚。即使网站客观上不是 spam，也很容易被搜索引擎容易误判为 spam 而遭到惩罚。

1、网站重定向成祸端（从 Business.com 遭封看 302 重定向）

Business.com 是网上最大的商业搜索引擎和分类目录，以专业提供商业信息而享负盛名，包括近 19 万网页。若以“business”为关键词在 Google 中进行搜索，该网站名列第一。不过在 9 月 5 日，Business.com 遇到一件蹊跷之事：它的首页由 PR8 变成了 PR0，而且 Google 搜索结果中找不到首页。好在被“蒸发”的只是首页，并且庆幸的是第二天首页又回到了 Google 的搜索结果，但 PR 仍旧为 0。

Business.com 的问题出在它的重定向命令上。该网站让 business.com 跳转到 <http://www.business.com/>，这种重定向本该是永久性的。而 302 则属于临时性重定向，只有当一个网站或网页在 24 到 48 小时之内临时移到其它位置的情况下才能使用该命令。但 Business.com 却错误地使用了“HTTP/1.1302 Object Moved”状态码。

其实网站重定向极为普遍，譬如不满意原来的域名而申请了一个新域名；买下容易被人错拼的域名，防止客户因为拼错 URL 而找不到网站，等等。可是，很多人却会由于使用了错误的重定向状态码而遭“灭站之灾”，就像 Business.com。尽管他们的重定向理由充分合理，然而若使用不当，就可能被 Google 误认为是利用多个域名指向同一网站，那么你的网站就会被封掉，罪名是“利用重复的内容来干扰 Google 搜索结果的网站排名”。大多数使用错误重定向参数的网站没 Business.com 这么幸运，一个小小的重定向就可能使网站前功尽弃，只能从头来过：重新申请新域名，重新发布新网站，等等。记住：Google 绝不会同情任何人，即使是无心犯下的错误。

2、302 重定向和网址劫持 (URL hijacking)

从网址 A 做一个 302 重定向到网址 B 时，主机服务器的隐含意思是网址 A 随时有可能改主意，重新显示本身的内容或转向其他的地方。大部分的搜索引擎在大部分情况下，当收

到 302 重定向时，一般只要去抓取目标网址就可以了，也就是说网址 B。如果搜索引擎在遇到 302 转向时，百分之百的都抓取目标网址 B 的话，就不用担心网址 URL 劫持了。

问题就在于，有的时候搜索引擎，尤其是 Google，并不能总是抓取目标网址。比如说，有的时候 A 网址很短，但是它做了一个 302 重定向到 B 网址，而 B 网址是一个很长的乱七八糟的 URL 网址，甚至还有可能包含一些问号之类的参数。很自然的，A 网址更加用户友好，而 B 网址既难看，又不用户友好。这时 Google 很有可能会仍然显示网址 A。

由于搜索引擎排名算法只是程序而不是人，在遇到 302 重定向的时候，并不能像人一样的去准确判定哪一个网址更适当，这就造成了网址 URL 劫持的可能性。也就是说，一个不道德的人在他自己的网址 A 做一个 302 重定向到你的网址 B，出于某种原因，Google 搜索结果所显示的仍然是网址 A，但是所用的网页内容却是你的网址 B 上的内容，这种情况就叫做网址 URL 劫持。你辛辛苦苦所写的内容就这样被别人偷走了。

302 重定向所造成的网址 URL 劫持现象，已经存在一段时间了。不过到目前为止，似乎也没有什么更好的解决方法。在正在进行的大爸爸数据中心转换中，302 重定向问题也是要解决的目标之一。从一些搜索结果来看，网址劫持现象有所改善，但是并没有完全解决。

四、404 重定向

1、404 设置

在 IIS 中设置 404 网页：点击网站属性→自定义错误信息→找到 404 了→双击→选“URL”→在下面输入自定义的 404 页面 URL。

一般而言，从技术上设置 404 错误导航并不复杂，关键是 404 错误页面本身的设计。

比较理想的情况下是给出出错提示，列出站点主要内容导航，及给予用户一个错误 URL 的提醒，同时方便用户浏览页面，增加用户友好程度。这一点可以学习 dell.com（如下图）。对于已经不存在的目录可以用 robots 文档限制搜索引擎抓取。

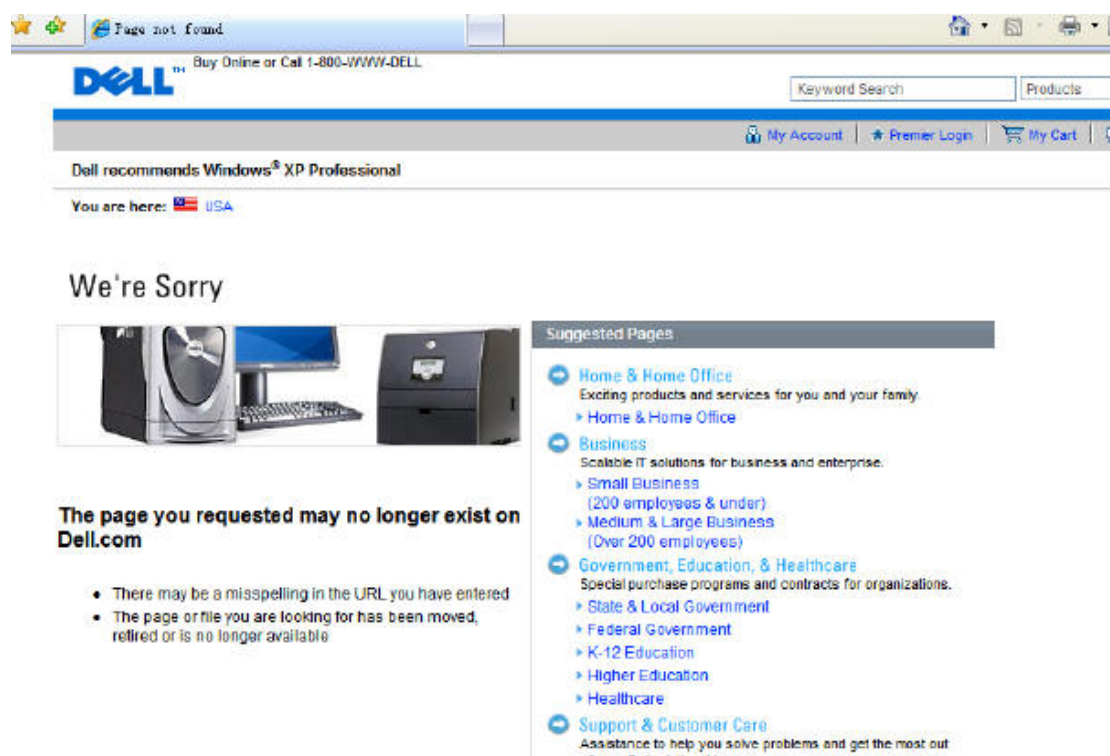


图 12-5-1 Dell.com 的错误页面导航

2、可能导致错误的 404

定义 404 错误页面，切记还需要禁止搜索引擎对该错误页面进行抓取。具体见 robots 协议。因为 404 错误不能直接定向到站点的每一页面，而是需要按照上面的参考案例精心设计页面。以下抓图是某站点，因为错误页面直接重定向到站点首页，而站点首页是没有禁止收录，导致搜索引擎重新收录了这些页面。产生了大量类似页面。

在 yahoo 收录的页面中，我们发现了 yahoo 收录了大量已经不存在的页面，而且收录的这些页面跟首页都是一模一样的。



图 12-5-2 Yahoo 收录页面

这些不存在的页面都定制了 401 错误页面，而这个错误页面恰好就是 9tour.cn 的首页。这些页面以前都存在，在变成 404 错误页面后，Yahoo 继续把定义的错误页面当作了内容抓取。这样的后果是，对于 Yahoo 而言，该站点存在大量完全一样的页面，有非常严重的堆积页面的嫌疑。

五、其他重定向

1、JS

这一代码被黑帽 SEO 广泛使用。JS 的主要作用是将一作弊页面优化上去后，迅速重定向到目标页面，由于搜索引擎对 JS 的识别有限，这种代码能避免被搜索引擎惩罚。

一般情况下，这种代码有两种方法：一种是直接将 Javascript 代码放在页面的 Html 代码中，另一种是将 Javascript 代码放到外部 JS 文件，而在 Html 代码里调用该 JS 文件了。

这种 Javascript 代码种类非常多。下面是一段 Javascript 代码。我们将其存放在跟页面处于同一目录的 abc.js 文件里。

```
bName = navigator.appName;  
if ((bName == "Netscape") ||  
(bName == "Microsoft Internet Explorer")) window.location="http://www.abc.com";
```

而在 Html 代码中，可以通过下面这样的代码来调用这个外部 js 文件：

```
<script language="javascript" src="abc.js"></script>
```

2、Flash

制作一个 swf 动画文件，该动画文件没有什么功能，一打开就是直接转向某一 URL。之前由于搜索引擎对于 swf 文件识别有限，大部分的博客、免费空间等禁止编辑 Html 文件，但允许插入动画，因而这种跳转也被黑帽广泛使用，将他们的免费空间、博客转向到目标客户站去。

3、meta fresh

在 2000 年前比较流行，不过现在已很少见。meta fresh 通过网页中的 meta 指令，在特定时间后重定向到新的网页，如果延迟的时间太短（约 5 秒之内），会被判断为 spam。

```
<meta http-equiv="refresh" content="2; URL=http://www.root.net">
```

这段代码可以使当前某一个网页在 2 秒后自动转到 <http://www.root.net> 页面中去，这就是 meta 的刷新作用，在 content 中，2 代表设置的时间（单位为秒），而 URL 就是在指定的时间后自动连接的网页地址。

第六节 垂直搜索引擎优化概览（图片搜索、移动搜索等）

一、图片搜索

- 建议 Alt 直接采用图片关键词；
- 建议同一景点图片集中在一起，每张图片显示大图的同时，还需要显示更多同一景点名称的其他图片（最好的图片分享网站 Flickr 采用这种方式），如下图；

普陀山海边

页面功能: 评论 | 大图 | 收藏 | 上传图片 | 图片大小(1760x1168) | 更多普陀山海边图片

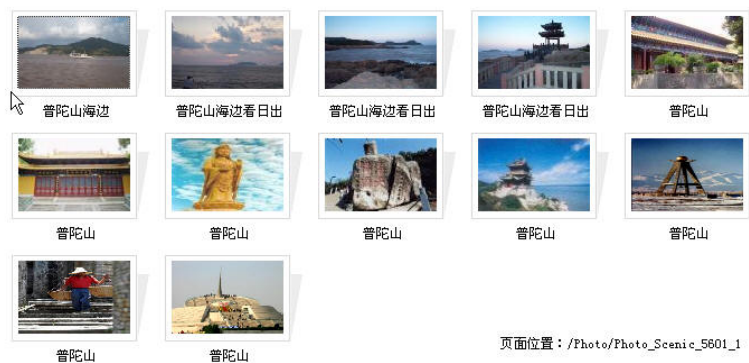


整理您的相片 下载管理软件 [HOT](#) 页面位置: /Photo/ShowPhoto_392_61364/

复制地址: http://www.9tour.cn/Photo/ShowPhoto_392_61364/ 奇虎更多图片

人气指数: 157 次 上传时间: 2006-11-17 9:40:53 所属景点: 普陀山 感谢会员 [yangege](#) 提供以上图片

图 12-6-1 大图与小图的组合



页面位置: /Photo/Photo_Scenic_5601_1

页数: 1/1 首页 上一页 下一页 尾页 共有 12 幅图片资料

图 12-6-2 同一景点图片集中摆放

我们建议将图 12-6-2 所示的页面内容同时增加到图 12-6-1 所示页面下面。这些将提高搜索引擎对九游图片的抓取能力和图片搜索排名表现。

- 建议每张图片用链接方式带出图片名称，并尽量在每张图片上也加上链接。



图 12-6-3 图片下增加文本说明

- 建议图片页面关键词密度可以适当增大，比正常文字页面关键词增加一倍；
- 图片页面的 title 一定要包含关键词；
- 图片格式尽量为.jpg；
- 每张图片上传时尽量加上水印，或者进行稍微的调整。目标之一是保证本站的图片在大小，尺寸上跟其他站点不相同。搜索引擎不喜欢一再抓取一些大小、格式一模一样的图片。

二、Flash

大部分搜索引擎无法识别 Flash 中的信息，Google 也只能检索到部分 Flash 文件中的内嵌链接。

Flash 优化可以从以下三个方面来考虑：

- 做一个辅助的 Html 版本；
- 将 Flash 内嵌 Html 文件；
- 付费登录搜索引擎。

三、视频搜索

随着视频网站的蓬勃发展和视频内容资源不断丰富，如何进行视频网站内容的搜索引擎优化成为一个倍受关注的问题。研究发现，基于文字信息的网站内容搜索引擎优化的思路，对于视频内容检索同样具有参考意义。

1、关键词

- **关键词列表** 列出与视频内容相关的关键词列表；
- **了解关键词被搜索的情况** 利用相关工具了解用户通过哪些关键词找到这些视频，以及现有的关键词列表被搜索的情况。如果你的关键词搜索量低或者根本没有被搜索的话，就要考虑用更热门的关键词来描述你的视频内容；
- **上传优化后的新鲜视频；**
- **创建一个链接数适当的视频网站地图。**

2、视频内容

- 将丰富的关键词自然融入视频内容中，描述简洁流畅，做到用户喜欢与搜索引擎友好并重；
- 注意 title、meta 的编写，也要在链接锚文本上做文章；
- 将公司的品牌商标或网址插入在视频内；
- 以与视频内容紧密相关的关键词作为视频的名称；
- 使用主流播放器支持播放的格式，并保证网站播放视频的流畅；
- 添加相关的热门内容和新闻素材。

3、内容推广和跟踪

- 确保视频及视频地址的有效性；
- 还可以将你的视频的部份精彩剪辑批量地上传到 YOUTUBE 等流量大的网站上，

这样你的视频也会被搜索引擎快速地收录和在 YOUTUBE 上得到宣传;

- 要用分析软件测量视频的观看人数等等,并通过相关数据进一步调整优化视频网站的内容。

四、避免搜索中的搜索

现在很多垂直搜索或者所谓的元搜索,为了从 Google、百度那里分一点流量,只是把自己的搜索结果静态化了。这种实际不创造内容,想办法出现在别的搜索引擎索引里的搜索,是迟早会被搜索引擎抛弃的。搜索引擎不希望自己的搜索用户看到的是另一个搜索引擎的索引页面,那样用户体验太糟糕了。

第十三章 几个重要的问题

第一节 搜索引擎排名作弊

一、什么是搜索引擎作弊？

1、什么是搜索引擎作弊

搜索引擎作弊，或 SEO 作弊，SEO spam，搜索引擎垃圾等，是指为了提高在搜索引擎中展现机会和排名的目的，欺骗搜索引擎的行为。

2、SEO 与搜索引擎作弊的区别

最大的区别在于，SEO 遵循搜索引擎制定的一些规则，维护良好的搜索体验。搜索引擎优化更侧重改善站点本身的结构和内容，使得搜索引擎更容易理解站点，使得站点的质量提升，更能满足搜索引擎使用者的搜索需求。而搜索引擎作弊更多的是违反搜索规则，为了获取排名和流量，抛弃了用户体验，也影响了搜索引擎的使用体验。

3、作弊与非作弊是个渐变过程

网页作弊的多样性和搜索引擎优化的各种策略使得作弊呈现一个渐变的趋势，而不是一个非 0 即 1 的问题。搜索引擎会给网页的某些参数进行评分，在一定的分值内，搜索引擎会认为在合理范围内，而超过一定分值的网页则可能彻底封杀（不收录），或实施其他的进行惩罚性减权（比如排序往后）。

当然，在目前 SEO 市场及不规范的情况下，搜索引擎优化手段参差不齐，有的优化纯粹就是作弊，这值得引起高度重视。

二、搜索引擎排名作弊的常见形式

1、隐藏文本或隐藏链接

一般指网页专为搜索引擎所设计，但普通访问者无法看到的文本内容或链接。在形形色色的隐藏技术中，最常见的就是把文本或链接文字的字体颜色设置为与背景色相同或十分接近。

隐藏文本内容 (Invisible/hidden text) 意欲在不影响网站美观的前提下通过包含大量关键词的网页提高关键词相关性得分，从而达到改善搜索引擎排名的目的。

隐藏链接 (Invisible/hidden links) 意欲在不影响网站美观的前提下通过在其它页面添加指向目标优化页的隐形链接，通过提升链接得分而改善搜索引擎排名。

现在大多数搜索引擎都能检测隐藏技术，并视为作弊，包含隐含文本的网页面临被搜索引擎降低排名甚至删除列表等惩罚的风险。虽然在 Google 上不乏使用隐形技术而侥幸逃脱的网站，但多数人还是认为不值得冒这个险。其实通过添加可视文本内容并保证一定的关键词密度即可达到相同的优化效果。

2、恶意更换页面内容

一般发生于先向搜索引擎提交一个网站，等该网站被收录后再以其它页面替换该网站。“诱饵行为 (Bait-&-Switch)” 就属于此类偷梁换柱之举：创建一个优化页和一个普通页，然后把优化页提交给搜索引擎，当优化页被搜索引擎收录后再以普通页取而代之。

3、误导性或重复性关键词

误导性关键词 (Misleading Words) 在页面中使用与该网页毫不相干的误导性关键词来吸引查询该主题的访问者访问网站。这种做法严重影响了搜索引擎所提供结果的相关性和客观性，为搜索引擎所深恶痛绝。

重复性关键词 (Repeated Words) 这种作弊技术也被称为“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出现的关键词的高度关注来对关键词进行不合理的重复。类似的其它做法还包括在 HTML 元标识中大量堆砌关键词或使用多个关键词元标识来提高关键词的相关性。这种技术很容易被搜索引擎察觉并受到相应惩罚。

4、隐形页面 (Cloaked Page)

对实际访问者或搜索引擎任一方隐藏真实网站内容，以向搜索引擎提供非真实的搜索引擎友好的内容提升排名。典型的方式为使用 Frame 技术，通过调用另一页面，隐藏实际页面的内容。

5、欺骗性重定向 (Deceptive redirects)

指把用户访问的第一个页面 (着陆页) 迅速重定向至一个内容完全不同的页面。在 SEO 从业人群中，有一部分人很擅长使用群发工具在互联网上到处留下带有其目标关键词的垃圾链接信息，并且目标网页上大量堆积关键词，一旦某一个页面快速获得好的排名，他们将会向客户出售或者租用这个页面，而把它转到客户的站点页面。对于上述做法，搜索引擎是不赞同的。不过，如果不是恶意欺骗性重定向，比如目前被很多站点使用的根据客户 IP 地址，自动转向对应的本地频道首页的方式，是被搜索引擎认可的。

6、鬼域 (Shadow Domain)

这是最常见的欺骗性重定向技术，通过欺骗性重定向使用户访问另外一个网站或页面。一般利用 HTML 刷新标识 (Meta Refresh) 来实现。大多搜索引擎都不会索引使用该标识的页面。如果你使用刷新标识来检测浏览器或分辨率，那么就请使用 Javascript，且尽可能延长重定向时间。如果一定要使用重定向功能，那么请务必确保刷新周期不少于 10 秒钟。

还有一种情况就是当用户打开一个网站，该网站声称其网站已移至新域名下，并请用户点击新域名链接进入网站。但当用户进去后才发现，这个链接是一个“会员”链接。这也属欺骗性重定向行为。

7、门页 (Doorway Page)

也叫“Bridge/Portal/Jump/Entry Page”，是为某些关键词特别制作的页面，专为搜索引擎设计，目的是提高特定关键词在搜索引擎中的排名所设计的富含目标关键词的域名，且重定向至另一域名的真实网站。搜索引擎的 Spiders 往往忽略对那些自动重定向到其它页的页面的检索。

8、复制的站点或网页

最常见的当属镜象站点（Mirror Sites）。通过复制网站或网页的内容并分配以不同域名和服务器，以此欺骗搜索引擎对同一站点或同一页面进行多次索引。现在大多数搜索引擎都提供有能够检测镜象站点的适当的过滤系统，一旦发觉镜象站点，则源站点和镜象站点都会被从索引数据库中删除。

9、作弊链接技术或恶意链接（Link Spamming）

链接工厂（link farms） 指由大量网页交叉链接而构成的一个网络系统。这些网页可能来自同一个域或多个不同的域，甚至可能来自不同的服务器。一个站点加入这样一个“链接工厂”后，一方面可得到来自该系统中所有网页的链接，同时作为交换它需要“奉献”自己的链接，藉此方法来提升链接得分，从而达到干预链接得分的目的。如今搜索引擎发现它们已然不费吹灰之力。只要发现，不管属有意还是无意，“链接工厂”中的所有站点都难逃干系。

群发链接（针对 BBS, blog 等） 指为了提升站点的外部链接，采用人工或者程序的方式在其他站点的文章评论、留言板，或者 wiki、Blog 上留下大量指向垃圾站点的链接。这种方法也被 Google 形象地描述为“Google 炸弹”。目前各搜索引擎对于群发的行为都给予了严格的惩罚。

10、其它

日志欺骗行为 通过对一些页面等级较高的站点进行大量的虚假点击以求名列这些站点的最高引用者日志中，从而获得它们的导入链接。Google 已然意识到这种行为并将对利用这种途径获得导入链接的站点采取相应措施。

门域（Doorway Domain） 专为提高特定关键词在搜索引擎中的排名所设计的富含目标关键词的域名，然后重定向至其它域名的主页。由于搜索引擎一般忽略自动重定向至其它页的页面的检索，所以不提倡使用这种技术。

群发链接作弊 目前的主流搜索引擎在考虑结果排序时仍然会参考链接关系，虽然它只是上百个参数中的一个。但是，由于业界对链接关系的公关宣传效果，使得网站管理员们在制作网站时舍本逐末，纷纷采用基于链接的作弊方式试图提高网站在搜索结果的排名。为保证用户的搜索体验，大部分搜索引擎已经对这些通过“人造恶意链接”获得排名的网站，包括对通过留言本、论坛、以及 Blog 等群发 URL 试图获得排名提高的网站进行了降权或对这些链接不计于任何权重，以遏制群发作弊软件的泛滥，还互联网一个干净公平的“搜索环境”。

三、常见的作弊载体

1、博客

博客得出现，为搜索引擎作弊重新营造了一个天堂。作弊者在大量博客网站申请自己的博客，然后在博客页面堆砌大量关键词，所有关键词都链接到自己的页面，或者干脆自动弹出自己的网站窗口。

博客诞生早期，Blog 的 PR 值普遍较高，在百度和 Google 的权值都很大。搜索引擎非常青睐这些 Blog 群，它们更新速度非常快，搜索结果排名比较靠前。如果你申请一个 Blog，

发表的内容有的会出现在 Blog 站的首页，根据权值传递说，那么你发表的内容在搜索引擎的权值比较高，出现在结果前面几页的几率非常大。用 Blog 作弊还有一个好处就是不必太担心被搜索引擎把自己的域名屏蔽。因为搜索引擎不大可能封掉比如 blogchina.com 这样的域名，有些 Blog 的地址是 domain.com/user 这样的形式，没有一个搜索引擎智能到自己可以辨识它而封掉这个目录。简言之，搜索引擎目前对它的惩罚方法不够。

随着时间的推移，搜索引擎已经意识到这些危害，对于大部分博客的用户空间的权重降低，对于权重的分配和惩罚细分到站点的每个栏目或目录，有力的抑制了直接使用博客空间能优化关键词排名的行为。

2、免费空间和免费域名作弊

比如早期网易的个人空间，能给个 ***.nease.net 的域名，其他类似的还有 51.net 等免费空间提供商。除此之外，提供二级域名和空间的博客、日志、商城等也是作弊者的重要平台。

这样平台的共同点是提供相对独立的免费空间和 URL。申请免费，使得成本几乎可以忽略。同时这些空间往往给予申请用户比较大的控制权（比如可以在页面添加和修改代码等），使得作弊者可以更轻易地将流量引导到自己的其他站点或投放联盟广告代码等。

第二节 搜索引擎的反作弊

一、为什么搜索引擎要反作弊

对于上述提到的搜索引擎作弊行为，几乎所有搜索引擎都是反对的。SEO spam 的破坏性是相当巨大的，它不仅直接破坏了搜索引擎网站赖以生存的关键所在：每日数亿计的用户搜索体验，同时也破坏了搜索引擎获得收益的主要来源：竞价排名以及相关广告。

搜索引擎反作弊，类同于互联网的病毒与反病毒。因此，作为搜索引擎，对待垃圾和作弊网站的态度是高度地一致：惩罚或者删除。几乎所有，的搜索引擎官方网站，也公布了一些原则，对作弊行为进行鉴定，并给出了对应的惩罚措施。比如 2005 年 3 月 26 日凌晨 4 点，忍无可忍的 Google 突然发力，将广大批 spam 站点清除。

二、搜索引擎对待 SEO 与作弊的态度

对于 SEO，目前主流搜索引擎都是欢迎的，Google、Yahoo、Msn 也经常积极地参与 SEO 的相关活动。在以搜索引擎营销和 SEO 行业交流为主的全球搜索引擎战略大会，得到了整个搜索业界的支持，SEO 专家也有机会同搜索引擎进行面对面的沟通。

特别值得注意的是，目前主流搜索引擎比如 Google、百度、Yahoo 都拥有自己的反作弊小组，专门应对搜索引擎作弊，维护良好的搜索体验，同时这些反作弊的小组也跟 SEO 保持着良好的关系。比如 Google，其反作弊小组负责人 Matt Cutts，经常参加各种 SEO 活动，解答 SEO 的一些疑问。而国内的 Yahoo，也跟国内多位知名 SEO 保持着密切交流。

对于合理的 SEO，所有的搜索引擎都是欢迎的，而对于搜索引擎作弊，搜索引擎和 SEO 都是反对的。正是双方对于反作弊的共识，所以 SEO 与搜索引擎之间才能保持密切的交流。

三、主要搜索引擎反作弊的一些规定

1、百度对作弊的问答

■ 哪些行为会被百度搜索引擎认为是作弊？

答：以下行为都可能被认为是作弊：

- 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；
- 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；
- 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；
- 故意制造大量链接指向某一网址的行为；
- 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）；
- 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；
- 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）。

■ 作弊网页会被怎样惩罚？

答：根据网页的作弊情况，会受到不同程度的惩罚。

- 较轻微的惩罚，是系统在给网页排名时略为降低该网页的排名；
- 较常见的惩罚，是作弊网站的某些关键词（或所有关键词）的排名降到其它任何网站之后；
- 较严厉的惩罚，对于作弊行为和后果特别恶劣的网站，将从百度中消失；
- 特别补充，含恶意代码或恶意插件等用户极端反感内容的网页，虽然不是作弊，但只要用户投诉，一经确认也会被百度拒绝收录。

■ 我发现一个网站有作弊行为，如果我告诉百度，百度会惩罚它吗？

答：只要这个网站真的有作弊行为，一定会被惩罚的。百度为用户提供了举报专用网址，请百度登录吧或发送邮件到 spam@baidu.com 举报作弊网站。

■ 我是百度客户和搜索联盟成员，可以获得优待吗？不要惩罚我的网站好吗？

答：王子犯法，与民同罪。百度没有为惩罚作弊网站准备两套标准，百度对所有网站是一视同仁的，既不会因为哪个网站与百度有关系而不惩罚，也不会因为哪个网站与百度无关而加重惩罚。

■ 作弊网站的惩罚可以解除吗？

答：可以的。首先请确认你的网站上已经没有作弊行为，然后请到百度登录吧或发送邮件到 spam@baidu.com，按照下面的内容进行说明，一般在一个月后可以被解除。

你应在申请解除惩罚的邮件中说明以下内容：你的网站地址；你什么时间发现网站在百度消失的；你确认使用了何种被百度确认的作弊方法；你什么时间删除了网站中的作弊内容。

■ 我对作弊行为和惩罚有自己的看法，能与百度对话吗？

答：当然，到百度登录吧或发送邮件到 spam@baidu.com，发表任何意见与建议。百度可能做不到回复每一位用户的发言，但一定会做到阅读每一位用户的发言，并作为改进百度搜索服务的参考。

2、Google 对于反作弊的原则：

■ 反作弊天条一：善意原则

“假设网站设计意图是好的”，很像英美法系“无罪设定”模式在 Google 内部反作弊的复制，可称为“善意原则”，这是 Google 在施行反作弊技术时的起点。在确定某一网站是否有作弊嫌疑时，Google 会花很长时间，走很多流程，收集证据，来辨别这些网站是否真有所谓“恶意”。没人会武断地给某一个网站下结论，轻易冤枉任何网站。这是一个挺让人愉快的行为方式：不要错杀。因为，Google 相信，错杀一个网站，就是对用户有价值信息的损失。

■ 反作弊天条二：坚持自动、机器判断的原则

这一点同上面讲到的“善意”原则是相辅相成的。人们总愿意相信自己是细心与理性的，但很多时候，机器也许比工程师更加理性与细心。

在搜索结果作弊上，也许有一万种方法，而工程师能做的，就是把其中的规律总结出来，然后搞出一些智能算法并不断完善，剩下的，交给机器。似乎也只能这样做，如果反作弊要纯粹依赖手工，Google 的员工人数恐怕将达到天文数字。

统一标准，从而保证一个高质量稳定的网页结果，这是多数人都愿意看到的。因为，作弊结果比率甚至百分之零点的升高，对用户体验都是很大的伤害。在 Google 中国工程院从有了第一个工程师，就开始了反作弊的工作。

3、Yahoo

如果你已经在 Yahoo!完成网站登录申请，却还是找不到你的网站，可能是以下几种情形：

- 若你的网页内容触犯隐私权等法律相关问题，经网站所有者申请后，相关人员才能进一步处理相关事项；
- 非刊载或连结色情信息的网站，如贩卖色情影音产品、成人聊天室、连结色情广告等不予收录；
- 非刊载买卖枪械、贩毒等非法信息的违法网站；
- 非连结后开启多个窗口，且执行关闭时会开启更多窗口的网站；
- 非刊载盗版或其它违反智能财产权信息的网站：未提供下载非法 MP3、贩卖大补帖等相关信息；
- 非提供赌博信息网站；
- 非建置不完全的网站；
- 非网站实际内容与登录网站名称或网站描述不符的网站；
- 非重复登录的网站，网站内容重复的认定包括：相同网址（URL）再次登录者；相同网站内容但以不同网址（URL）、网站名称而再次登录者；公司或个人因同样业务需求而以不同网址或不同内容再次登录者；

- 必须能被正常连结或浏览内容的网站，且网站要能在一般常见浏览器上正常浏览（例如 Netscape、Internet Explorer）；
- 网站必须每天 24 小时正常联机运作；
- 非网站内的个别网页数据或单页网页数据；
- 非提供自杀等负面信息的网站；
- 网站内容为关于某地区的信息时，网页上需提供详细地点数据：例如某台中市面包店网站，则需提供面包店地址；
- 网站必须是公开且非封闭性，不须账号密码登入；

请注意并非你所提出申请的网页都会被删除，相关人员会针对你的申请作进一步处理。

四、搜索引擎惩罚站点事件举例

■ 2005 年 3 月 26 日 Google 对中文搜索结果做大面积清理

该次清理中，依赖中文 SEO（搜索引擎优化）技术的网站或内容均受到影响，中文搜索结果中的二级域名数量和广告招租等字样急剧减少，包括前段时间很多 SEO 喜欢使用的 51.net 和 91x.net 都纷纷被杀，网易的免费二级也被删除了不少，整个 Google 中文搜索的结果数量回到上一年度年底的时候。

■ Google 封杀宝马

宝马公司一个“门径网页”上，“gebrauchtwagen”（德文，意为“半新汽车”）出现 40 多次，极大满足了搜索软件寻找“与键入词相匹配内容”的要求。而事实却是，这一单词在真实网页上只出现了两次。凭借这一“门径网页”，宝马阻止了搜索引擎在德国网站中查找其他相关信息，确保想查询二手“宝马”车信息的用户进入到宝马公司网站，而不是其他的汽车销售网站。

不少专业网站会用这类技术提高自己的吸引力。不过，这类技术中的一些手段实属“居心叵测”，加之赌博网站和色情网站广为使用，因而名声欠佳。宝马公司这一丑闻曝光后，Google 的工程师马特·库特斯在自写“博客”网页上证实了这一传闻。他指出，宝马公司这种做法违背了网络搜索的原则。Google 网页管理原则规定，“欺骗性或误导性内容……例如‘门径网页’等，将导致相关网页彻底移出 Google 搜索范围。”

通俗的说，封杀意味着用户在搜索栏内键入“BMW”或“BMWGermany”时，Google 搜索结果中将不再有宝马公司的德国网站地址 bmw.com.de。更糟的是，宝马德国网站的网页排名将被“归零”。这一数字显示了网页的受欢迎程度，网页排名越高，在 Google 搜索结果中出现的位置就越靠前。

■ Google 封杀 KinderStart

2006 今年 3 月，KinderStart 向美国圣何塞地方法院提起诉讼，指控 Google 对该公司进行了不当地“封杀”，损害了其互联网业务，这种行为违反了反垄断法和其它法律。据 KinderStart 称，该公司一年前在没有收到任何警告信息的情况下，被清除出了 Google 索引。KinderStart 的辩护律师格里戈尔·余表示：“Google 要做的就是阻碍竞争”。在对诉讼双方进行质询的过程中，反垄断问题引起了弗吉尔的共鸣，他表示能够理解 KinderStart 的立场。

■ 百度封杀 eczn.com

Eczn.com 为某 SEO 专家的站点，但由于涉嫌作弊被百度清除。除此之外，还大量站点与作弊站点链接或互相链接，也导致被清除。比如：电子商务指南 (<http://www.eczn.com>)、行业研究报告 (<http://www.cn006.net/>)、搜索引擎排名 (www.SEO.net.cn)、网站运营 (www.tongwang.net)、拓展训练 (www.03718.net)、都市网 (www.371.com) 等及大量与这些站点有相互链接的英文站点。



图 13-2-1 百度封杀 eczn.com 截图

第三节 搜索引擎惩罚和解除

一、搜索引擎惩罚的方式

一旦站点的搜索引擎作弊行为被搜索引擎发现，那么客户站点将会遭到降权或者直接被搜索引擎从索引中清除。而被清除的站点，几乎没有机会被重新收录，这就是片面追求搜索引擎排名效果的可悲之处。

根据网页的作弊情况，会受到不同程度的惩罚：

- 较轻微的惩罚，是系统在给网页排名时略为降低该网页的排名。比如 Google 曾经对 Donews 博客给予了轻微降权，Blog.donews.com 用户的排名受到影响；
- 较常见的惩罚，是作弊网站的某些关键词（或所有关键词）的排名降到其它任何网站之后。这种措施被 Yahoo 和 Google 使用较多，特别是 Google 的 Sandbox，也可以理解为这种惩罚方式；
- 较严厉的惩罚，对于作弊行为和后果特别恶劣的网站，将从搜索中消失；
- 特别补充，含恶意代码或恶意插件等用户极端反感内容的网页，虽然不是作弊，但只要用户投诉，一经确认也会被搜索引擎拒绝收录。

二、站点被惩罚后的影响

站点一旦被某一搜索引擎惩罚，一方面几乎很难再通过该搜索引擎获得巨大流量，另一方面对企业的品牌也会带来负面影响。在 2006 年初，Google 封杀宝马德国站后，宝马网页

作弊的丑闻，通过互联网迅速传播，使得宝马公司不得不出来公关，挽回因网站作弊事情导致的品牌损害。

该网站采用自动重定向的桥页作弊法让访问者和搜索引擎看到不同的网页，导致网站被 Google 删除。虽然 BMW.de 修正错误以后，Google 又重新收录了该网站。不过一定要牢记，宝马德国网站被迅速重新收录是因为它是宝马德国网站，并非意味着其它网站也有这样的特殊优待。实际上大部分网站被删除以后实难重获收录机会。或者在更改错误以后要经历至少好几个月的漫长等待时间。

在我们上述我们曾提到的 Google 封杀 KinderStart 案例中，被 Google “封杀”之后，该公司网络流量下滑了 70%，AdSense 广告营收也下滑了 80%。在顶峰时期，KinderStart 每月的访问人数超过 1000 万人。

三、如何投诉作弊

只有打击作弊网站网页，才能更好地保护正常网页的利益，百度、Yahoo、Google 等著名搜索引擎均设立了专门的作弊网站举报信箱或地址，接受用户的举报。

百度：

垃圾网页投诉中心：<http://www.baidu.com/search/spamreport.html>

中搜：

作弊网站举报信箱：anti-spam@zhongsou.com

Google：

作弊网站举报地址：<http://www.google.com/contact/spamreport.Html>

Yahoo：

作弊网站举报地址：<http://help.cn.yahoo.com/feedback.Html?product=websearch>

四、你的站点是否被惩罚？

要解除惩罚，首先要学会判断自己的站点是否被惩罚过。尤其是站点被清楚后的判断。

1、如何判断站点已被搜索引擎惩罚。

常见的几点参考因素如下：

- 网站的 PageRank 变为“0”，Google 工具栏上的 PageRank 指示为“0”；
- 网站在 Google 的外部链接数变为“0”；
- Google 工具栏上的 Backlinks，或在 Google 搜索框中键入“link:www.yourdomain.com”命令，显示没有外部链；
- 网站排名骤降。以极其冷门的几乎唯一相关的关键词搜索，搜索排名都明显下降或消失。不过这种情况有时候是算法更新的原因，与处罚无关。

2、网站在搜索引擎中消失

曾经收录的站点，直接输入网址，已经找不到该网站。此时，该网站极有可能已经被搜索引擎清除。

被处罚的网站一般情况下如果解除作弊，还是可以获得“新生”，情况特别严重者除外。

五、搜索引擎作弊惩罚解除流程

1、解除作弊惩罚

几乎所有的主要搜索引擎，比如 Google、百度、Yahoo、中搜、搜狗等，都在相关的网站管理员指南中注明：主要网站及时清除作弊内容，是可以申请搜索引擎解除站点的。

一般的申请解除的方式为电子邮件，电子邮件解除申请中一般要求注明以下几点：

- 需解除的网址；
- 作弊方式；
- 被惩罚的起始时间；
- 作弊行为是否已经清除。

尽管目前为止，通过这种方式解除惩罚的几率非常之小。大部分被解除的都是那些一直坚持高质量站点内容建设、真正受欢迎的站点、无意中涉及作弊的站点。也因为大部分站点，在清除作弊行为后，并没有被搜索引擎解除作弊。因而，搜索引擎也受到了很多网站管理员的指责。认为各站点公布的解除方式不过是虚设而已。

不管如何，在早期，与几乎没有搜索引擎声明只要解除作弊行为就有机会重新被搜索接受的情况相比，目前的搜索引擎已经人性化多了。

2、如何清除作弊行为

要清楚作弊行为，首先必须认真阅读搜索引擎的相关规则。如果无法深刻把握搜索引擎的判罚尺度，则最好是选择专业的机构，对站点进行诊断，找出所有可能涉嫌作弊的地方。

根据诊断结果，对于作弊行为或涉嫌作弊过渡针对搜索引擎而忽略用户体验的地方，进行改进。

但一定要注意，并非所有的作弊行为都是可以清除的。一般而言，如果是网站本身作弊，比如堆积关键词等是可以很容易清除的。但对于群发链接，链接工厂等方面是很难清除的。因为我们几乎没有办法控制其他站点的行为。

第四节 搜索引擎反作弊任重道远

一、搜索垃圾永不会消失

在与搜索引擎与垃圾网站的长期斗争中，搜索引擎总是处于守势。这不仅因为搜索引擎公司真正考虑搜索结果相关性和排序的只有几个人，而互联网上整天琢磨怎么对付搜索引擎的有几十万网站管理员。更重要的是，制造一个或几十个甚至几百个作弊网页，只要将一个目标网页稍作改动即可实现，并可以不断进行类似复制，而搜索引擎甄别、删除却需要相当的时间。这种“以一当百”，以一个网站对付数以万计作弊网站的工作，对搜索引擎而言，是一个巨大的人力资源投入与成本投入。

目前，垃圾网站的新攻击方式和攻击热点，不断变化和推陈出新，spam 如蝗虫般涌来，搜索引擎也不断需要改变自己的应对策略。

这一斗争过程，已完全类同于互联网的病毒与反病毒。所不同的是，制造病毒的成本有时会更些，它们往往需要很高的技术门槛。而制造垃圾网站的门槛却相当低，且制作方便。

因此，作为搜索引擎，只能以最严厉的方式来应对作弊站点，尽量提高作弊后的后果严重性来提升作弊成本和风险。

但在搜索中，永远会有垃圾网站的存在，因为它每时每刻都在诞生。由于垃圾网站的成本与收入间的巨大差异，SEO spam 们还会勇往直前，前赴后继。而到目前为止，政府以及相关部门还没有对这种互联网上的“假冒伪劣”制定相应的法律制度来制约。但相信这一天一定会到来。

二、搜索引擎反作弊的历程

乱世用重典，搜索引擎在与 spam 的长期战争中总是处于不利地位。当然，从某种角度来说，几乎是 spam 在推动着搜索引擎的进步。

搜索引擎诞生的早期，搜索引擎们为了搜索结果能更准确，为了能把最相关的搜索结果排在前面，发明了 meta tag。起初，meta tag 是很好用的，那时候，诚实的 webmaster 们如实地填写网站摘要，如实地列出跟他的网站最相关的关键词，搜索引擎们抓取、索引、检索这些诚实的 meta tag，并作为搜索结果排序的重要依据，然后，是一段相安无事的甜蜜时光。

岁月迅速流逝，甜蜜时光挡不住资本的力量。“一旦有适当的利润，资本就大胆起来……有 50% 的利润，他就铤而走险；有 100% 的利润，他就敢践踏一切人间法律；有 300% 的利润，他就敢犯任何罪行，甚至冒绞首的危险。”越来越多的网站开始追求收入追求 pageview，它们开始欺骗搜索引擎，它们在 meta tag 中塞进成堆的关键词，塞进跟自己的网站内容无关的但是流行的关键词。群众的力量是巨大的，搜索引擎的相关性被破坏了。

搜索引擎们对作弊者开始第一轮反击，但是无力的，它们的做法是：在排序时更多的依赖网页正文而不是 meta tag。

作弊者以进为退，开始第二轮攻击，他们在正文中大量使用与网页背景相同颜色的关键词，在 image tags 中塞进关键词，在网页代码加入“看不见的注释”。

搜索引擎开始第二轮反击，它们找到了有效的方法来过滤这些看不见的文字，取得了小小的胜利。

“善良是善良者的墓志铭，卑鄙是卑鄙者的通行证”，作弊者眼见常规武器无效，开始利用搜索引擎的真诚。搜索引擎的 spider 是互联网上的谦谦君子，它们访问任何网站都会自报身份，并且查阅网站的访问规定，完全按照各网站的规矩来办事。于是，作弊者精心制作了两个网站，一个专供网友访问，一个专供搜索引擎访问，每当发现访问请求来自搜索引擎时，就返回专供搜索引擎看的能令搜索引擎满意的网页，而正常访问者只能看到另一个截然不同的网站。这种方法叫做 cloaking。这第三轮攻击，作弊者大胜。

1998 年 3 月 1 日，分别在 AltaVista、Excite、HotBot、Infoseek、Lycos 搜索当时的热门词汇“Monica Lewinsky”，各搜索引擎首页居然分别有 50%、30%、30%、10%、50% 的搜索结果是 spam。由此可见当时作弊者们猖獗至何等程度！

搜索引擎们，因为它们道德上的弱点，因为有些手段不愿用，所以无法完全辨别 cloaking 网站，也就无法完全对付作弊者。搜索引擎能做的只是严厉的惩罚被发现的 spam，以警告作弊者，发展到今天的结果是，所有的搜索引擎对它们认定的 spam 行为，只有一种惩罚：永久屏蔽。

但是，严厉的惩罚从来只是无能者推卸责任的借口，而不能真正的改善状况。作弊者从来就不怕惩罚，当互联网狂潮来临，追求 pageview 的作弊者们更是肆无忌惮。

无可奈何的搜索引擎，只有去寻找新的相关性排序依据，只有去寻找作弊者不能控制的

排序依据，只有去网站以外寻找排序的依据，搜索引擎，真的别无选择。

于是，开始出现了超链分析，出现了 Pagerank。一个网页，说它自己是什么，它就是什么吗？物以类聚，如果它说的是真的，那么指向这个网页的链接应该来自具有同类内容的网页，所以指向这个网页的链接可以被用来验证这个网页的相关性。

谁，最不可能出于贪婪而作弊？谁，最可能如实提供网页内容？是政府网站，和教育网站。所以，在超链分析中，gov 和 edu 网站通常被赋予较高的排名，来自 gov 和 edu 的超链也有更高的价值。

但是 PageRank 也带来了一些不请自来的错误排序因素：新站不如旧站，小站不如大站，专业站不如通俗站。一个新网站，一个小网站，一个专业网站，它的内容再好，因为不可能迅速得到大量的超链，所以无法得到合适的排名，即使它的内容是无与伦比的。而一个历史悠久的老网站、大网站、通俗网站，即使它已久不更新，或者内容远不如某些小网站，却可以获得很高的排名。

当超链分析刚出来时，SEO 界一片悲观论调，认为这是无法作弊的，SEO 业完蛋了（就像近期各搜索引擎纷纷收费，SEO 业更觉得自己前景惨淡，其实虽然搜索引擎排名市场的大头要给搜索引擎们，但 SEO 还是有钱赚的，毕竟，只有劳动模范有下岗的可能，从来没听说过骗子会失业的），但是，道高一尺魔高一丈，很快 SEO 们发现了对付超链分析的方法：做大量的网站，准备大量的链接，哪个客户出钱，所有的网页都给它一个链接。这种 SEO 方法被称为 link firm。这种方法刚出现时还算有效，不过很快就不行了，因为一个网站是否有大量来自内容不相干网页的链接，或者一个不是搜索引擎的网站给出大量不相干的链接还不难判断。而每当发现一个 link firm，那些使用了该 link firm 的网站就全部死翘翘。

一种搜索结果的排序算法，如果更多的是基于概念、主题、基于网页内容的质量，而不是网络上互相链来链去的超链的重要性，那么搜索结果的相关性在本质上会更好。

只是，让计算机学会理解和评价网页内容吗？好像还是很遥远的事情；花大量的人力一个个网页评估么？这肯定是亏本的商业行为，只有靠志愿者。ODP 庶几近之，但也只能点击，用关键词搜索时还是无法跟超链分析的结果相比。

三、石头记录的 2005 年 SEO 作弊状况

2005 年已经过去了，对于许多 SEO 来说，注定不会是轻易被忘掉记忆。

我希望通过对我自己 2005 年工作经验的回顾来对中国的 SEO 进行一次回顾。如果读了之后，能对你的 SEO 事业有一点点提示作用的话，我那些用无数的痛苦和心血换来的经验，也算有有一点价值。

2005 年春节前后 那些疯狂的岁月

这是 SEO 疯狂和销帐的时刻，那时候，易赛诺，王者工作室，优构网络，联达网络，灿海网络等一连串如雷贯耳的名字，在网上无处不在；到处都是广告位出租的信息，到处搜索出来都是乱码……Google 的页面可以用惨不忍睹来形容。举个例子来说明，当时的 SEO 有多嚣张吧。那时候我打了一个电话到北京的一个 SEO，问他排队机要多少钱，他说八千，然后问他稳定不稳定，那个人嚣张地说：“我这个词在网上一年多就从来没有从第一名掉过！”。

总之，SEO 也在 Google 里也似乎无所不能，“led”，“电影”，“翻译”，“招聘”这些热门词前面照样充满着优化后的结果。那时候的优化手段就是作弊。疯狂地作弊，疯狂

地发留言，疯狂地堆积关键词。那时候最搞笑的是 wiki 里经常会看到这样的留言：“如果你删除我的链接，我就删除你的！”

2005 年 3 月 26 日凌晨四点 Google 不再沉默

Google 终于不再沉默。大刀砍向了 SEO，对一大批著名的 SEO 作弊站点进行了封杀。我记得那一天晚上我还象其他 SEO 一样在辛苦地工作过，我在查我自己的网站时，前一刻还被 Google 收录，突然就找不到，职业的敏感性是我意识到 Google 算法更改了。我测试了许多 SEO 站点，包括那时候最流行的优化资源 www.51.net, www.91i.net, 都已经被 Google 惩罚。那一刻，我心情复杂，用文章 Google 不再沉默记录了这一时刻。关于这次 Google 的算法更新，这是最早的一篇消息，第二天新华网才有了类似的新闻。

那一夜，我感悟很深，至今再读起那些文字时，我依然显得有些激动。近一年后，经历 Google 更多次算法变化后，我更坚信，那些原则依然适用。

51.net, 91i.net 到今天依然被 Google 惩罚，是 Google 对这次算法留给我们的见证。

此后虽然 Google 也进行了几次小规模删除行动，而且有认为干涉的嫌疑。但中国的 SEO 基本还是我行我素，作弊者依旧是前赴后继。

时间依旧流逝，街市依旧太平……

05 年 4 月--7 月 博客开始兴起

随着方兴东的摇旗呐喊，博客在 2005 年四月后开始急速成长，SEO 自然立刻掌握了这一先进生产力。开始学会利用博客，而不再是此前主要依靠免费空间来进行优化。以留言为主的优化方式开始过度到建立大量博客然后添加链接的新方法。当然，这一时间网页的跳转方式也出现了多样化。

到如今，还有不少的 SEO 们在博客垃圾的道路上越走越远，最常用来优化的博客包括 www.blogchina.com, www.blogsky.com, blog.donews.net, www.blogbus.com; blog.com 等。

博客与免费空间相比，具备申请快，申请站点多，发布内容简单，容易管理因而得到了最广泛的使用。免费空间随着国外著名站点如 www.freewebsite.com, www.bizhat.com 等站点的被封已越来越少用于优化。

05 年 7 月--10 月 博客疯狂，SEO 受质疑

Google 开始打算进入中国，SEO 们开始纷纷猜测，大家都相信 Google 进入中国后必然会对搜索结果的质量进行改进。然后 Google 选择了代理商，和这些代理商一起加紧了舆论上对 Google 排名的否定。不断地有报道来攻击 SEO，说 SEO 是非法的，甚至 Google 要起诉。

这一切，直接影响到了我们业务员的信心。我开始思考，SEO 到底有没有未来。我思考的结果自然就是搜索引擎优化必然存在而且合理。在一片质疑声中，我反而更坚定了我的信心。

这段时间也是博客疯狂的时候，最典型的的就是曾经输入“彩信”这种词，连续几十页都是博客的优化结果！（我当时有抓图→<http://soft.yesky.com/tools/169/2104669.shtml>）

05 年 10 月 24 日 Google 算法发变化，SEO 陷入痛苦中

不知道这一次算法对于其他 SEO 来手有什么影响，但对我来说，那绝对是一次刻骨铭心。我曾经写过《痛苦中煎熬，Google 算法调整后的 SEO 们》，那是我最真实心声的反应。

那一夜，中国大部分的 SEO 们站到了悬崖边，我也一样。

幸亏 Google 此次之前半年的时候就在英文版里进行了类似的算法调整，所以至少还有一点准备。连续几个晚上的辛苦后，20 多个关键词的排名终于恢复了。而我在公司里也遭遇了有史以来最严重的信誉危机，大家都开始动摇了 SEO 的信心，开始怀疑我的技术和操作手法。

05 年 11 月至 06 年初 塞翁失马，焉知非福? Google 拒绝排名

经历一次又一次的排名变化后，一次又一次地去维护，然后再下降，在维护的日子后，我彻底抛弃了以前所有的做法，从服务模式到实际优化方式。跟我最大的感触是，其实在中国按照正规的 SEO 方法去做，市场是照样接受的。而且最重要的，我现在客户总共 300 多个关键词全部在上面，达到要求，基本上都是客户的网站+我们自己的平台+免费空间。而且 Google 算法更新后，按照我的操作方法，即使我 10 天不理，这些排名也不会掉下来。

我能告诉你的就是，要想排名稳定，必须在优化的最开始就忘记外部链接!! 老实地调整你网站的结构，充实内容。一步一步做好才是根本! 不要以为这样就意味着优化需要很长时间，错了! 欲速则不达!

另外，我要告诉大家，Google 认同 SEO，但不认同 Google 排名，因为 SEO 不等于 Google 排名! 谁的网站敢标明是搞 Google 排名的，并且搜索相关关键词能靠前的话，最后的结局就是你的网站被 Google 干掉，不信? 马上搜索“Google 排名”，看那一天谁的网站排在前面，过段时间再去观察你就明白了!

四、曾经的作弊有多疯狂

为了让大家能回顾过去垃圾 SEO 那些疯狂作弊的日子，发一些图片，让大家见证以下当年的疯狂：

1、www.chunmeng.com



图 13-4-1 chunmeng.com 的 Google 收录页面

2、Bizhat.com 和 freewebpage.org

记不起什么时候开始流行使用 Bizhat 和 Freewebpage.org 来优化，如今唯一能看出的就是当年的火爆程度。



图 13-4-2 bizhat.com 的 Google 收录页面



图 13-4-3 freewebpage.org 的 Google 收录页面

3、优构网络 (YouGoogle.com)

这家自称深圳最早的 SEO 还声称拥有 Google 总部关系的 szhx.net, 如今似乎早已经消失在深圳的 SEO 届了。优构网络、华夏网渐渐成为记忆。



图 13-4-4 YouGoogle.com 的 Google 收录页面

4、Linkdata.cn 联达网络

联达网络的颠峰之作在于创造了网站不需要任何内容，只需要乱码+关键词，不断重复，制造大量页面和大量二级域名网站，然后互相连接在一起这一疯狂手法。辉煌的时候输入各种各样的热门词，总会有联达的声音，当时他们的这些网站 PR 达到了 5 以上。

五、SEO 的职业道德

网站排名所使用的优化技术可谓五花八门、形形色色。但并非所有的搜索引擎优化技术都是合乎道德规范的。这些不顾道德的作弊行为同样严重地危害了整个 SEO 行业的发展。因而 SEO 与搜索引擎有一个共同点，那就是反作弊。

但是由于许多 SEO 本身就参与了作弊，所以使得 SEO 的职业道德被提了出来。加强

SEO 的职业道德、对 SEO 进行积极的引导，也是搜索引擎应对作弊的重要手段。

1、SEO 职业道德的主观性

在探讨搜索引擎职业道德这个问题上，首先必须认识到这里面所涉及的问题是相当复杂的，而且带有很强的主观性。可能你认为一种优化技术很好，有人却觉得它拙劣不堪。所以，要想给搜索引擎优化技术用“优”、“良”和“差”进行划分的话，无异于是在瞄准一个飘乎不定的“移动靶”。最糟糕的是，由于我们对不同搜索引擎及各种不同的排名算法仍然缺乏全面深入的了解，因而要做这样一个性质上的准确划分也是几乎不可能的。但是我们可以给“白帽子”，“黑帽子”和“灰帽子”这三种技术下一个基本定义，这样读者就能够知道作者的态度：哪些技术合乎道德，哪些完全不符合道德，哪些又是介于两者之间的“灰色”技术。

对 SEO 职业道德所做的任何分析都要以网主对网站优化的真实意图为出发点。首先要考虑网主愿意承受的风险程度，即对被搜索引擎惩罚、甚至遭封所持的态度。其次要考虑相关关键词的竞争性。最后一点非常重要，那就是需要考虑网主是不是会为达到排名而不惜手段，无视一切职业道德。撇开道德这个因素不提，搜索引擎优化的最终目标无外乎让网站得到一个理想排名。让网站位居各大搜索引擎的显著位置实在是个巨大的诱惑，鲜有人面对这样的诱惑而无动于衷。

不管是为了增加网站的访问流量，提高销售额还是增加利润，或者仅仅只是为了满足一下自己的成就感，总之无论出于什么目的，网站管理员在优化策略上也是出尽百宝。这样就很可能面临这样一种状况，即你认为不错的优化技术却被别人所不耻。因而对一个网站所采用的优化技术是否合理（合乎道德），以及这些技术在搜索引擎优化中是否重要，就显得愈加重要了。

2、搜索引擎优化之白帽、黑帽、灰帽

搜索引擎优化专业人员大致可划分为三大阵营：“白帽子”SEO 人员，其对立阵营的“黑帽子”人员，及介于两者之间的“灰帽子”人员。

白帽子优化技术 白帽子优化其实就是已经得到大家广泛认可的优化技术，这些优化人员对 Google “网站管理指南”中所列出的种种被视为作弊的技术惟恐避之不及，同时杜绝使用一切带有任何可疑迹象的优化技术。标准的白帽子优化技术包括了：为网站添加更多主题相关的文字内容，定期更新网站的内容，为网站吸纳外部链接，并辅以合适的链接文字。保持网页标题的准确性和实时性。除此之外，将重点放在网站内容的良好建设上，例如增设 blog 来丰富充实网站的内容，同时注意关键词在网站上的合理分布，一切工作都围绕让网站成为一个权威（Authority Site）和中心网站（Hub Site）的目标来进行。

黑帽子优化技术 黑帽子技术一般意味着对网站使用在 Google 的“网站指南”中明令禁止的那些作弊优化方法，甚至包括被其它一些主流搜索引擎严禁的排名手段。例如内容隐藏技术、重定向技术、链接工厂机制、以及大量交叉互链做法，等等，这些均属于黑帽子技术。这其中花样最多的要属隐藏内容技术。一般是把字体和背景颜色设为相同或几近相同的颜色，也有通过在 CSS 中将字体设为与背景相同或不可见，或通过层（Layer）来达到隐藏内容的目的。

灰帽子优化技术 灰帽子技术介于这两者之间。这个概念比较难下定义，可以说亦正亦邪。一方面他们会使用一些白帽子的优化技术，但同时他们又会用链接工具和内容生成软件这一类带有黑帽性质的优化技术。

3、要求搜索引擎优化的职业道德是否有意义？

对于搜索引擎优化上的职业道德的探讨是否有意义？有些人会不假思索地说是，有些人的态度则比较犹疑，也有人认为只要客户的排名能够上去，这一点根本无关紧要。总而言之，对这个问题并没有一个明确积极的答案。

我们知道，一个搜索引擎优化人员要面对两个客户，这样他面对的就不仅是一种职业道德。首先，是他面对网主这个客户所需要具备的职业道德；其次，是面对搜索引擎需要具备的职业道德。但是，由于这两类客户在需求上存在必然的冲突和矛盾，必然会导致搜索引擎优化人员陷入两头为难的困境。让网站获得尽可能高的排名是搜索引擎优化专业人员责无旁贷的工作，但同时也要照章办事，考虑到搜索引擎对网站的各种要求。优化人员必须知道如何划分网站和搜索引擎之间的道德界限。而且重要的是，他应就以下两个问题与客户进行认真讨论：一个是客户对网站排名的真正目的，其次就是客户对因优化而带来的风险的承受底线。比较专业的优化人员往往不会让客户涉足黑帽子甚至灰帽子这样的领域。此外，对于搜索引擎对网站的要求和网主对网站的排名要求之间的界限的理解，对各位优化专家来说则是仁者见仁，智者见智了。

4、SEO 人员与搜索引擎的关系

优化人员对搜索引擎的态度是又一个重要的因素。有些人把搜索引擎视作自己的朋友般友好，但有些优化人员却将其视做自己的死敌一样充满敌意。还有一些专业的优化人员要么将搜索引擎视作一个虽然不好但却不可缺少的东西，要么就把它当作一个整体互联网营销项目中的一个工具。对搜索引擎为什么会有这么多不同态度，原因其实并不难理解。

在那些将搜索引擎视做朋友的人眼中，搜索引擎所提供的搜索结果就意味着免费的客户源。搜索引擎优化要比其它网络推广方式经济的多，而且通过合理充分的优化，网站完全可以获得完全免费而且源源不断的客户源。他们深信：搜索引擎实际上是在向公众提供免费服务。正是基于这个原因，他们才将搜索引擎视做自己的朋友，并乐意为其提供经过精心打造的“最好”的网站。

有些优化人员却把搜索引擎几乎视为敌人。他们坚持认为：搜索引擎并没有向用户提供与查询最相关的搜索结果，反而让那些使用不正当优化手段的网站大行其道，而那些真正与用户查询最相关的高质量的网站却被排挤到角落里，结果是把搜索结果弄到一团糟。同时，也正是由于搜索结果缺乏足够相关性以及作弊网站充斥其间，抱有这种想法的优化人员认为搜索引擎根本就没有履行他们自称的目标或服务条款。同时，也正是由于搜索引擎的算法系统为作弊网站提供了可乘之机，因而黑帽子一方觉得，无论是他们的优化工作还是被优化的网站，都没有法律或道义上的责任或义务来遵守搜索引擎制订的这些规矩。他们认为：如果搜索结果让一些不怀好意的网站得逞，而规规矩矩遵守服务条款的网站却受到惩罚，那么根本没有理由再去遵守搜索引擎所制订的所谓的“网站行为指南”。

此外，还有一些优化人员把搜索引擎视作一种商业模式，而且搜索引擎的排名算法充其量只是一种计算机程序，根本无所谓好和坏，它们所要执行的任务就是按网站对用户搜索条件的相关性来排列网站。而且正因为这是一种商业模式，如果要使用搜索引擎提供的业务或产品，用户需要遵守它们的商业规矩。在这种情况下，网站行为指南和服务条款就是搜索引擎为用户制订的商业规矩。就象饭店老板可以拒绝衣冠不整者入内就餐一样，搜索引擎也一样可以随时随地删除那些违反了它们的商业规矩的网站。

鉴于搜索引擎的职业道德及其排名算法上存在着太多争议，最好的办法就是把搜索引擎视做一种商业模式。把它们所制订的网站行为指南和服务条款视做商业规矩，对于不遵守这

些规矩的网站，搜索引擎安全有理由从其数据库中自行删除。此外，搜索引擎也并无索引网络上的每个网页或站点这样的义务和职责。

对于那些不遵守它们规则的网站，它们完全有权予以惩罚或删除。所以如果不想让网站遭到搜索引擎的惩罚甚至删除，就需要网站遵守这些规则。至于那些对搜索引擎提供的搜索结果不满意的人，搜索引擎都设置了检举投诉体系，你完全可以向它们反映问题。在接到这样的投诉后，搜索引擎的客服中心一经核实一般都会给予删除。不过删除的过程比较漫长，有可能要等上几周甚至几个月。

虽然有人怀疑搜索引擎是否有能力改观对作弊网站的监测、惩罚、以及删除，不过即使这个问题存在，它也并不意味着网站可以违反搜索引擎的服务条款而免受惩处。此外，一个作弊网站仍然占据搜索引擎结果的主要位置也并不意味着它就不会受到惩罚，这只能表明搜索引擎还没有发现它作弊而已。

六、搜索引擎反作弊的人性化

搜索引擎反作弊的手段，一直是 SEO 关注的焦点之一。通过对几大搜索引擎相关规则的分析发现，目前搜索引擎反作弊有人性化的趋势。

以 Yahoo 为例，在其站点上清楚地写到：“你可以重新整理清楚这些作弊行为并重新向雅虎提交你的网站，我们非常乐意重新收录你的网站。搜索引擎会在下一个索引周期（约 2 - 4 周左右）完成后重新检查是否有欺骗搜索引擎行为并决定收录”（来源：http://search.help.cn.Yahoo.com/h4_18.html）。

事实上，在实际案例观察中，发现 Yahoo 很少采用完全删除一个站点这样的行为，Yahoo 更多的是采用降权。而且一个被 Yahoo 删除的站点，有机会能被 Yahoo 在短时间内再次索引。相比百度，Yahoo 的反作弊策略更显人性化，避免了许多恶意的指责和猜测。

而以最近推出 Sogou Rank 的搜狗，一直以来对待删除站点的惩罚措施中都注明：

“被搜狗视为有作弊行为的网站，在清除作弊内容后，可以向我们写信申请解除处罚：anti_spam@sohu.com，经过搜狗监控人员的审查，如果确认网站已经不存在作弊行为，搜狗将恢复对该网站的正常收录。”（来源：<http://www.sogou.com/docs/help/webmasters.htm>）

对于 Google，事实上也是有跟上面这种差不多的条款的：“If your site has violated our webmaster guidelines, and you've made changes to it so that it meets our guidelines, you can request reinclusion and we'll evaluate your site.”中文意思即为：如果你的站点背离了我们的网络管理员须知，但是你已经作出了相应的改变，并符合了我们的要求，那么你可以重新要求我们评估并收录你的站点。

在实际操作中，能够得到 Google 重新评估和收录的站点并不多。如果站点本身就有一定的专业性和影响力，那么按照 Google 这样的指示，还是有机会被 Google 重新收录的。石头曾按照上面的提示，帮助了一个在某一领域有重要影响力的站点被 Google 重新收录。

遗憾的是，在百度的站点上目前还找不到类似意思的文字。但从上面这几大搜索引擎的表现来看，搜索引擎反作弊越来越考虑一些人性化的东西。比如 Google 的 Matt Cutts 的博客，也可以理解成 Google 反作弊人性化的一种理解。以前那种搜索引擎站在绝对统治，霸权地位的思维方式只会遭到更多人的质疑。

搜索引擎反作弊增加一点人性化的手段，能够迅速地缓解很多的误解和批评。特别是对于广大网站管理员，给予一个重新的机会能够获得他们的人心。而且最重要的一点是，人性化的存在，能够避免搜索引擎机器可能的出错。虽然机器能够做到更公正公平，但是谁能保

证机器不会犯错呢？

七、搜索引擎反作弊的法律支持

搜索引擎反作弊也引起了许多争议，很多站点因为不服搜索引擎对自己站点的惩罚行为而提起诉讼。尽管目前尚无任何法律对搜索引擎反作弊进行解释。但是基本的商业道德和搜索道德，还是赢得了法律的尊重。

Google 及其它搜索引擎已经多次因网页排名遭遇诉讼，但提起诉讼的一方从未获得过胜利，这是法律对于搜索引擎反作弊行为态度的最好解读。在国内方面，百度在 2006 年底也赢得了第一起搜索引擎作弊案。

2006 年初，枫叶之都使用百度搜索竞价排名时，发现自己的网站在许多与加拿大旅游相关关键词的搜索排序中发生了巨大变化——从前十位被调到倒数第一或第二页中。由此，枫叶之都认为，百度存在人为干涉搜索结果的可能，为配合其竞价排名销售推广业务，百度对该公司的搜索结果进行了人为技术降权和恶意锁定，构成不正当竞争，因此要求百度道歉并赔偿经济损失 10 万元。

百度就此解释称，枫叶之都网站在百度搜索排名中的位置变化，是百度搜索引擎算法策略自动计算所为。枫叶之都网站存在大量针对搜索引擎的垃圾链接，被百度搜索引擎算法规则系统识别，并调整了其网站权值所致。据百度提供给法院的证据显示，枫叶之都曾在多个压缩机、干燥机等与旅游业不相关的网站以及私人博客和留言板上进行了大量、重复的网站链接。而这种与该公司网站内容无任何关联性的网站超过百家。2006 年 3 月，枫叶之都的大量垃圾外链被百度的自动反垃圾外链系统发现，自动将其排名压后。

北京一中院经审理后判定，枫叶之都采用垃圾链接方式违反了搜索引擎业相关规则。

值得注意的是，在这次判决中，北京一中院为搜索引擎搜索结果的排名方式给出了法律界定：对于排名算法规则的设置尚没有明确的法律规定或行业规定，故各搜索引擎网站均自行设定排名算法规则。“这也表明了对于百度目前采用的‘竞价排名’和‘自然排名’相关规则在法律上获得的肯定。”

第十四章 附录

第一节 搜索引擎发展史

搜索引擎从一种技术、一种应用工具，已经演变成一种生活习惯、一种商业平台和一个巨大的产业了。我们不妨从了解搜索引擎的发展历史入手，来深入认识搜索引擎。

一、搜索引擎编年史：全球

搜索引擎是早于万维网出现的，但是在万维网出现之前，搜索引擎还仅仅是实验室的工具，并没有能够成为大众的工具。短短的十多年的时间，搜索引擎除了在技术上的进步有目共睹之外，这种技术的应用对互联网乃至大众生活产生的影响是任何一个搜索引擎发布之初都没有办法预见的。

1990 年

1990 年 12 月

蒙特利尔大学的学生 Alan Emtage 发明了 Archie。虽然当时 World Wide Web 还未出现，但网络中文件传输还是相当频繁的，而且由于大量的文件散布在各个分散的 FTP 主机中，查询起来非常不便，因此 Alan Emtage 想到了开发一个可以以文件名查找文件的系统，于是便有了 Archie。Archie 的工作原理与现在的搜索引擎已经很接近，它依靠脚本程序自动搜索网上的文件，然后对有关信息进行索引，供使用者以一定的表达式查询。

1991 年

1991 年 4 月

明尼苏达大学 (University of Minnesota) 的一个研究小组开发出了 Gopher，它最初的设计目的是在校园里建立信息共享系统，让每个人都可以简单地发布自己的研究成果，并能共享别人发布的信息。由于这个系统设计得界面简单，一致性好，很快就在 Internet 上流行起来。Gopher 最大的特点就是用户要作的只是从菜单中做出选择，而由 Gopher 完成所需的操作，它可能要与另外一台计算机建立连接，或者要下载一个文件，或者打开一个 Telnet 会话等，所有的过程用户都不必关心，也不必知道得到的数据是存放在哪台计算机中的。

广域信息服务 WAIS 出现，WAIS 是 Internet 提供的快速信息查询服务工具，将相关的资料组织成一个可按关键词检索的数据库资料，提供给公众检索。

1991 年 8 月

万维网诞生。万维网的发明者蒂姆·伯纳斯-李在 alt.hypertext 讨论组上贴出了一份关于其研究项目万维网的帖子。该项目旨在构建链接系统，“允许建立指向任意地点任意信息的链接”，万维网开始名副其实。

1992 年

1992 年 7 月

文字模式的网页浏览器 Lynx 出现，Lynx 是一种提供文字模式的浏览器，此程序支援

UNIX、LINUX、DOS 等多种操作系统。

1993 年

1993 年 2 月

Architext 发布。6 个斯坦福大学的学生想通过分析字词关系，以对互联网上的大量信息作更有效的检索。到 1993 年中，这已是一个完全投资项目 Architext，他们还发布了一个供 webmasters 在自己网站上使用的搜索软件版本，后来被叫做 Excite for Web Servers，也是后来的知名网站 Excite.com 的前身。

1993 年 3 月

美国总统 Bill Clinton 发出了全球第一封电子邮件。

1993 年 6 月

WWW Wanderer 诞生。世界上第一个用于监测互联网发展规模的“机器人”程序是 Matthew Gray 开发的 World wide Web Wanderer，刚开始它只用来统计互联网上的服务器数量，后来则发展为能够检索网站域名。

1993 年 11 月

ALIWEB 诞生。与 Wanderer 相对应，Martin Koster 于 1993 年 10 月创建了 ALIWEB，它是 Archie 的 HTTP 版本。ALIWEB 不使用“机器人”程序，而是靠网站主动提交信息来建立自己的链接索引。

1993 年 12 月

随着互联网的迅速发展，使得检索所有新出现的网页变得越来越困难，因此，在 Matthew Gray 的 Wanderer 基础上，一些编程者将传统的“蜘蛛”程序工作原理作了些改进。其设想是，既然所有网页都可能有连向其他网站的链接，那么从跟踪一个网站的链接开始，就有可能检索整个互联网。到 1993 年底，一些基于此原理的搜索引擎开始纷纷涌现，其中以 JumpStation、The World Wide Web Worm (Goto 的前身，也就是今天 Overture)，和 Repository-Based Software Engineering (RBSE) spider 最负盛名。然而 JumpStation 和 WWW Worm 只是以搜索工具在数据库中找到匹配信息的先后次序排列搜索结果，因此毫无信息关联度可言。而 RBSE 是第一个在搜索结果排列中引入关键词串匹配程度概念的引擎。

1994 年

1994 年 1 月

第一个既可搜索又可浏览的分类目录 EInet Galaxy (Tradewave Galaxy) 上线，除了网站搜索，它还支持 Gopher 和 Telnet 搜索。

1994 年 4 月

Yahoo 斯坦福大学的两名博士生 David Filo 和美籍华人杨致远 (Gerry Yang) 共同创办了超级目录索引 Yahoo，并成功地使搜索引擎的概念深入人心。

互联网上第一个支持搜索文件全部文字的全文搜索引擎 WebCrawler 是发布，在它之前，用户只能通过 URL 和摘要搜索，摘要一般来自人工评论或程序自动取正文的前 100 个字。

1994 年 7 月

Lycos，最早现代意义上的搜索引擎出现于 1994 年 7 月。当时 Michael Mauldin 将 John Leavitt 的蜘蛛程序接入到其索引程序中，创建了大家现在熟知的 Lycos，这是搜索引擎史上又一个重要的进步。当时 Lycos 数据量为 54,000，除了相关性排序外，还提供了前缀匹配和字符相近限制，第一个在搜索结果中使用了网页自动摘要。

1995 年

1995 年 2 月

Infoseek 的搜索引擎才与公众见面，并于 1995 年 12 月与 Netscape 签署战略性协议，使它成为一个强势搜索引擎（注：2001 年 2 月，Infoseek 停止了自己的搜索引擎，开始改用 Overture 的搜索结果）。

1995 年 4 月

Alexa 上线，并最终成为广受欢迎的网站排名系统。

1995 年 6 月

第一个元搜索引擎 Metacrawle 出现。

1995 年 9 月 26 日

Inktomi 创立。

1995 年 10 月

Excite 网站正式公布。

1995 年 10 月

LookSmart 上线，后来成为最有名的分类目录网站。

1995 年 12 月

DEC 的 AltaVista 登场亮相，大量的创新功能使它迅速到达当时搜索引擎的顶峰。AltaVista 是第一个支持自然语言搜索的搜索引擎，是第一个实现高级搜索语法的搜索引擎，AltaVista 也声称是第一个支持用户自己向网页索引库提交或删除 URL 的搜索引擎。

1995 年 6 月

Archive.org

1996 年

1996 年 5 月 20 日

Inktomi 公司成立，强大的 HotBot 出现在世人面前（Hotbot 曾是随后几年最受欢迎的搜索引擎之一，后被 Lycos 收购）。

1996 年

Mamma

Dogpile

1997 年

1997 年 4 月

Ask Jeeves，也就是现在的 ask.com 上线。

1997 年 7 月

FAST Search 上线

1997 年

GoTo 上线

1998 年

1998 年 1 月

Openfind 创立，其技术源自台湾中正大学吴升教授所领导的 GAIS 实验室。Openfind 起先只做中文搜索引擎，曾经是最好的中文搜索引擎，鼎盛时期同时为三大著名门户新浪、

奇摩、雅虎提供中文搜索引擎。

1998年2月

中国人自己的搜索引擎“搜狐”问世，“出门靠地图，上网找搜狐”开始了中国互联网的门户时代。

1998年5月

Yahoo 搜索采用 Inktomi 技术。

1998年6月

开发时目录 DMOZ 发布。

1998年9月27日

Google 诞生。

MSN Search 上线。

1998年10月

Lycos 收购 Wired/HotBot。

1999年

1999年1月

At Home 收购 Excite。

About.com 上线。

1999年2月

GO Network 上线。

1999年5月

Fast (Alltheweb) 公司发布了自己的搜索引擎 AllTheWeb。

1999年6月

CMGI 收购 AltaVista。

1999年7月

迪斯尼收购 Infoseek。

1999年9月

FindWhat 上线。

2000年

2000年5月

Terra 收购 Lycos。

2000年6月

Vivísimo 上线。

Yahoo 搜索采用 Google 技术。

2000年10月

Google 推出 AdWords 广告服务。

2001年

2001年4月

Teoma 登场。Teoma 支持类似自动分类和提供专业链接目录的功能，同年9月被提问式搜索引擎 Ask Jeeves 收购。

2001 年 8 月

LookSmart 整合 Zeal.com 并升级收费服务。非商业网站可通过 LookSmart 旗下的 Zeal.com 进行免费注册；商业网站则必须付费登录，还推出了“Express Modify”服务，允许付费快速修改网站信息。

2001 年 9 月

AskJeeves 收购 Teoma

WiseNut Wisenut 由韩裔 Yeogirl Yun 创立。2001 年春季发布 Beta 版，2001 年 9 月 5 日发布正式版，2002 年 4 月被分类目录提供商 looksmart 收购。

2001 年 9 月 21 日

搜狐公司在国内率先宣布对搜索引擎收费。但这项收费服务不是针对普通网民，而是对希望把自己的网站加入搜狐搜索引擎的企业提供分类搜索服务收费。

2001 年 10 月

GoTo 更名为 Overture。

2002 年

2002 年 1 月

Yahoo 收购 StarMedia Network 旗下的巴西搜索引擎 Cade，并将其与 Yahoo 搜索引擎整合。通过本次收购，Yahoo-Cade 将成为南美第三大网络媒体网站。

2002 年 2 月

Google 推出了 Adword Select 竞价排名广告业务，用户可针对目标关键词搜索结果页面中 Adword 广告链接排名进行竞价，最低竞价为 5 美分，Google 竞价排名在考虑竞价高低同时还参考链接的点击率。

2 月 28 日起，百度竞价排名前 3 位的网站开始出现在搜狐缺省搜索结果最前列。

2002 年 3 月 12 日

新浪网搜索引擎中的网页搜索服务被百度停止。

2002 年 5 月

微软拒绝继续与 RealNames 合作，导致后者宣布将在 6 月 28 日停止关键词搜索服务并宣布倒闭。

2002 年 12 月

Google 推出了商品购物搜索引擎 Froogle。

2003 年

2003 年 3 月

Yahoo 收购 Inktomi。

2003 年 4 月

FindWhat 收购 Espotting。

Google 收购 Applied Semantics。

Overture 收购 AltaVista 和 Fast/AlltheWeb。

4 月 15 日新浪正式宣布和中国搜索联盟结成战略合作伙伴，为企业提供搜索排名服务。

2003 年 6 月

Google 推出 AdSense 服务。

2003 年 10 月

Yahoo 收购 Overture。

2004 年

2004 年 2 月

Yahoo 整合其收购的多家搜索引擎技术，推出了 YST 全文搜索引擎技术，Yahoo 宣称其全文搜索引擎已索引了 43 亿个网页文件。Yahoo 推出新的搜索引擎，意味着与 Google 多年合作的终结。

Lycos 停止搜索服务。

2004 年 3 月

AltaVista、AlltheWeb 切换为 Yahoo 搜索。

AskJeeves 收购 Excite、iWon、My Way。

2004 年 4 月

Google 推出 Gmail 邮箱服务，全球率先推出 1G 免费邮箱。

2004 年 6 月

AOL 收购 Advertising.com。

Yahoo 在中国推出独立域名的搜索引擎一搜 Yisou.com。

Ask Jeeves 收购 Tukaroo Search。

百度 Baidu 获得新一轮投资，Google 成为其战略投资者。

2004 年 8 月

8 月 3 日，搜狐正式推出全新独立域名专业搜索网站“搜狗”（www.sogou.com），力推用户互动式搜索功能。

Yahoo 推出 Local Search Engine，提供本地搜索服务。

8 月 19 日

Google 首次公开发行股票 IPO，在美国 NASDAQ 成功上市，Google 的 IPO 采用荷兰式拍卖的形式，上市后市值达到 300 亿美元，超过众多老牌传统上市公司。

2004 年 9 月 15 日

全球最大的在线购物网站 Amazon 亚马逊发布了其 Web 搜索服务工具的新版本 A9.com，可从用户所选择的各种资源中组织有用信息，并能记住用户的个人搜索记录。

2004 年 11 月

MSN 发布新的搜索引擎。

Google 宣布收录网页超过 80 亿。

2005 年

2005 年 1 月

Gigablast 宣布收录网页超过 10 亿。

AOL 发布新版的搜索引擎，并采取多项措施加强搜索引擎。

2005 年 3 月

美国传媒大亨 Barry Diller 的 IAC 公司宣布收购 Ask Jeeves，涉及金额达 19.5 亿美元，IAC 拥有超过 40 家不同网站，其中包括在线旅行站点 Expedia、Hotels.com，社交网站 Match.com、Evite.com 及其它流行站点 Ticketmaster、LendingTree 等。

2005 年 4 月

4 月 12 日，搜狐以约 1000 万美元全资收购国内一家名为 Go2Map 的电子地图公司，搜

狐在完成对 Go2Map 的收购之后将把其资源整合进搜狐的搜索业务线当中。Go2Map 从 2000 年开始专注于网络地图搜索领域，Go2Map 在该细分市场内已经做到了接近 90% 的市场份额。

4 月 18 日，Yahoo 把旗下“Overture”更名为 Yahoo Search Marketing，美国市场将率先实施更名计划，海外市场也将陆续更名，但日本和韩国将继续保持“Overture”品牌。调整之后，Overture 的搜索广告产品将以“Sponsored Search（赞助搜索）”为名，而本地付费搜索则称为“Local Sponsored Search（本地赞助搜索）”

2006 年

1 月 27 日

Ask.com 上线，取代了 askjeeves.com。

2006 年 9 月

微软正式发布全新的搜索技术 Live Search。

二、搜索引擎编年史：中国

1996 年

1996 年 8 月

ITC 爱特信电子技术公司（北京）有限公司正式注册，也就是后来的搜狐。

1997 年

1997 年 2 月

爱特信公司正式推出 ITC 中国工商网络。

1997 年 10 月 29 日

由北大网络实验室研制开发的天网中英文搜索引擎在 CERNET 上向广大 Internet 用户提供 Web 信息导航服务。

1998 年

1998 年 2 月

搜狐推出。

1998 年 12 月

四通利方宣布并购海外最大的华人网站公司“华渊资讯”，成立全球最大的华人网站“新浪网”，引起海内外极大关注。

1999 年

1999 年 2 月 2 日

新浪网推出新一代中文搜索引擎“新浪搜索”测试版。

1999 年 3 月

在分类搜索的基础上，搜狐发展成为综合性网络门户，推出丰富的特色频道，提供多种网络服务。

2000 年

2000 年 4 月 13 日

新浪网在美国纳斯达克上市。

2000 年 6 月

百度正式推出中文搜索引擎，并且宣布全面进入中国互联网技术领域。

2000 年 7 月 12 日

搜狐公司在美国纳斯达克上市。

2000 年 9 月

网易正式推出了全中文搜索引擎服务，并拥有国内唯一的互动性开放式目录管理系统。

2000 年 10 月

百度始为新浪提供服务。

2000 年 11 月 1 日

新浪搜索正式推出国内第一家综合搜索引擎。

2001 年

2001 年 2 月

百度为 TOM.COM 提供全面搜索服务。

2001 年 10 月

百度推出全新商业模式“竞价排名”。

2002 年

2002 年 3 月

新浪网搜索引擎推出搜索引擎 new 指数。

2002 年 9 月

中搜与中国互联网新闻中心（中国网）联合发起成立“中国搜索联盟”。

2002 年 10 月

百度开始为雅虎中文提供搜索技术，雅虎同时加入百度竞价排名阵营。

2002 年 11 月

百度开始为网易提供搜索技术，百度发布 MP3 搜索。

2003 年

2003 年 4 月

新浪与中国搜索联盟结成战略合作关系共拓搜索引擎服务。

2003 年 7 月

百度推出图片、新闻两大技术化搜索引擎。

2003 年 8 月

中搜正式推出第三代智能搜索引擎，新浪、搜狐、TOM、网易等七大门户网站同时宣布采用其搜索引擎技术。

2003 年 12 月

百度陆续推出地区搜索、“贴吧”等划时代功能，搜索引擎步入社区化时代。

2003 年 12 月

“中国搜索”告成立，并推出搜索门户 www.zhongsou.com 及新闻中心。

2004 年

2004 年 2 月 26 日

中搜推出全球首款桌面搜索产品“网络猪”。

2004 年 3 月

首届搜索力经济高峰论坛在厦门召开。

2004 年 6 月

网易采用 Google 网页搜索技术。

2004 年 8 月 3 日

搜狐推出第三代互动式搜索引擎“搜狗”。

2005 年

2005 年 2 月

百度发布全球首款支持中英文的硬盘搜索工具。

2005 年 4 月

搜狐收购中国领先的在线地图服务公司 Go2Map。

2005 年 5 月 18 日

百度正式宣布推出影视搜索服务，提供影视资源在线观看及下载服务。

2005 年 6 月 23 日

百度推出名为“百度知道”的互动式知识问答分享平台，进军“知识搜索”领域。

2005 年 6 月 30 日

新浪正式推出自主研发搜索引擎“爱问”。

2005 年 8 月 5 日

百度在美国纳斯达克上市，首日股价涨幅达 354%。

2005 年 9 月 23 日

百度发布地图搜索服务，正式提供本地化地图搜索服务。

2005 年 11 月 7 日

搜狐公司正式成为 2008 年北京奥运会赞助商，成为百年奥运史上第一个互联网类别赞助商。

2006 年

2006 年 3 月 17 日

搜索引擎战略大会第一次中国会议在江苏南京举办。

2006 年 4 月

“网络猪”正式更名 IG，并定义为新一代个人门户，推出搜索市场。

2006 年 4 月

百度推出第三个社区类搜索产品——“百度百科”（baike.baidu.com）。

Google 中国公司正式成立并发布中文名称“谷歌”。

2006 年 7 月

百度发布新产品“百度空间”。

2006 年 9 月

百度竞价排名全面推出智能排名功能，以“综合排名指数”作为排名的标准。

2006 年 11 月

微软 Live Search 和中国电信合作推出 114 搜索。

2006 年 12 月

网易悄然推出自主开发的搜索引擎“有道”。

第二节 搜索引擎行业八卦

一、中国 SEO 市场概况

中国 SEO 市场开始萌芽于 2003 年初，2004 年底，达到了一个高潮。但那时候 SEO 引起关注，更多是因为大量作弊手段的流行，严重影响了搜索引擎的正常使用。

这一期间所谓的 SEO 服务公司，大都是以兜售 Google 左侧排名为生。由于排名的不稳定性，导致在 2005 年 5 月份后，一大批所谓的排名公司倒闭。这一期间，标准的遵循搜索引擎搜索道德的 SEO 服务开始诞生。

2006 年，是中国 SEO 迅速发展的一年，行业出现了出现了一批引导正确操作规范的个人，开始向中国的企业传播合理的 SEO 理念。SEO 也得到了广大媒体的正面报道。而且像“好耶”这种大品牌搜索引擎营销公司的介入，SEO 的形象得到极大的改善，也得到了 Google, Yahoo、百度的承认。这一年内，大量的行业领军站点开始尝试 SEO 的魅力，这个行业也得到了前所未有的关注。中国的 SEO 服务，也有了信心向它在国外一样，成为网站开展搜索引擎营销的核心战略之一。

经过几年的积累和沉淀，国内目前 SEO 服务公司的基本现状为：

- 大量以兜售关键词排名为服务核心模式的小作坊式企业，仍然大量存在。比如搜索“Google 排名”无处不在；
- 一些过去的 SEO 标志企业或品牌由于在发展中跟不上商业模式和技术的变化消失或者衰落或转型。比如上海启洋等。早期从事 SEO 的王通，更多的开始网络营销策划，和销售营销书籍为核心业务；
- 大量的传统网络产品销售渠道，为了吸引客户也打着 SEO 的旗帜提供服务。这些公司从过去的唱衰 SEO 到目前纷纷标榜自己可以提供 SEO 服务；
- 一些国际 SEO/sem 企业，尝试进入中国，但水土不服。只要表现在国外的 SEO/sem 企业天生只懂得英文推广，而且局限于 Google, Msn, Yahoo 国际。而对于国内最重要，占绝对优势的百度无能为力。比如尚奇，进军中国一年，反应平平；
- 一些新的 SEO 标志和品牌开始诞生。这些企业由于长期观察和从事 SEO 业务，但由于其没有受到早期错误作弊行为的干扰。赢得了搜索引擎的尊重，成为行业新的风向标；
- 一些从事传统网络营销研究的企业和以行业研究和观察为主的营销专家，为了生存，也提供简单的 SEO 咨询服务。比如新竞争力等；
- 一些传统的广告公司和管理咨询公司开始进入 SEO。但目前仍然缺乏专业的 SEO/sem 人才和服务经验。比如新华信和电通广告。他们的主要方式仍然是外包给其他 SEO 服务者。

二、杨致远为 Yahoo 做优化

Yahoo 是最早使用 SEO 技术的公司之一，那个时候，他们仍然没有估计到搜索产业巨大增长潜力在世纪之交即将来临。Yahoo 的联合创始人 大卫·费罗 和 杨致远 那时候为了增加站点的曝光度，想办法让他们的网站能被互联网上其他搜索引擎发现。通过对站点结构的优化和一些“狡猾”的代码编写技巧，他们的网站变得很容易被那些新用户检索到。当时他们也没有考虑将这种服务商业化操作，因为当时没有人清楚这样做到底是符合商业道德的还是非法的，也没有可以参考的简单标准。

三、SEO 大赛

以组织或者是赞助比赛的形式推广产品和服务，是近年来比较流行的。搜索引擎领域经常也会有一些赛事，从搜索引擎之间各种技术层面的较量，到广大用户搜索技巧的比拼，每次都搞得红红火火。最近，一场以搜索引擎优化为主题的比赛，在没有任何媒体宣传的情况下火爆了起来。

印象中比较早的一次 SEO 大赛发生在 2004 年 5 月，当时澳洲的 DarkBlue 和 SearchGuild 推出了一项名为“挑战 SEO”的 Google 左侧排名竞赛活动。竞赛以“nigritude ultramarine”为竞赛关键词，吸引了全球各地不少搜索引擎优化专家的眼球。主办方声称举办这个活动的目的是希望探讨一下 Google 究竟是如何进行排名的，设立了“竞赛奖”和“保持奖”两个奖项，竞赛日期是 2004 年 5 月 7 日-7 月 7 日，6 月 7 日晨 9 时，在 Google 中搜索竞赛关键词，排名第一的网站将获得“竞赛奖”，7 月 7 日晨 9 时，在 Google 中搜索竞赛关键词，排名第一的网站将获得“保持奖”。

但是由于在比赛规则上没有很好地设定，这次比赛最终沦为搜索引擎作弊者的温床，各种欺骗搜索引擎的手段层出不穷。虽然比赛产生了很大的影响，但是也让很多人对 SEO 产生了误解，把 SEO 当成了作弊、滥用关键词的代名词。

Carcasher.com 可以说是最精明的 SEO 大赛发起人。作为一个从事二手汽车在线交易的公司，carcasher.com 可以说和 SEO 毫无关系。但 carcasher.com 却想到了 SEO 大赛的方式，并设置了奖金。虽然 800 美元的奖金不算丰厚，但获得名次的意义可不仅仅在于奖金。于是 SEO 们纷纷加入，经过数月的竞争，大赛告一段落，真正受益的当然是 carcasher.com 了。这次 SEO 大赛可以说是一次经典的商业炒作案例，Carcasher 给出了丰厚的奖品，包括 14,000 美元、42"Plasma 彩电、索尼 PSP 游戏机和 iPod 这些奖品之外，还有一个标的为 12000 美元的 carcasher.com 的 2007 年的 SEO 服务合同，这些让 SEO 专家们趋之若鹜。当然，更能吸引高手的是奖品背后的荣誉。

国内比较有影响力的一次 SEO 大赛，应该是刘韧和老白之间的一次玩票性质的较量。二人以“毛主席语录”做 Google 排名竞赛关键词。虽然他们都不是真正意义上的 SEO，但是却有着非常广泛的人脉资源和网站资源。国内的 SEO 高手纷纷支招，网友们的主动链接也让这次二人之间的比赛迅速扩展成为业界的事情。

毛主席语录这样的关键词，虽然不算是热门词，但是长期积累下来的文章和网页多如牛毛。选择这样的词作为 SEO 比赛的结果就是很难找到。好在刘韧和老白都有不少朋友，否则不知道要优化到什么时候才能让自己的网页浮出水面。

最近一次由浩维传媒组织的 SEO 大赛就更有点传奇特征了。大赛发起前，有人就竞赛关键词是什么做出猜测和建议，比如有人就建议用月饼这个词。但是考虑到不利于完整观察

以及有可能因为竞赛影响原来的搜索结果，大赛组织方还是决定制造一个新的词汇。于是，一个绝对新词，在 Google、百度等搜索引擎搜索结果均为零的词语诞生了：“渡虎谷”。

渡虎谷 SEO 大赛有一些特点：

- 发起方是全球最大规模的搜索引擎营销和优化专业会议搜索引擎战略大会（Search Engine Strategies）的中国主办方浩维传媒，以引导 SEO 服务良性发展为目的，公允性和专业性都比较高；
- 竞赛关键词“渡虎谷”在赛前的没有搜索结果，能从零开始观察这个赛况；
- 选择了三个搜索引擎百度、雅虎和 Google 同时开始，强调了整体水平；
- 有严格的禁止作弊的要求；
- 强调了内容的重要性，并引导参赛者积极撰写相关的文字内容，估计赛后能出一本渡虎谷传说的八卦书籍了；
- 有很强的评委团队，将对大赛过程进行监督和指导。

香港、台湾、新加坡等地的 SEO 也加入了这一赛事，渡虎谷 SEO 大赛有点全球大赛的意思了。比赛不是目的，倡导正确的 SEO 理念、推动行业良性发展才是渡虎谷 SEO 大赛的意义所在。

第三节 SEO 工具介绍

一、SEO 工具软件种类

随着 SEO 的发展，整个行业经历不停的洗牌和调整，面对的是更多的工作量。为了减轻工作量，各种工具软件应用而生，这是一件好事情。SEO 软件类别有：

1、内容管理

特点是：生成静态页面系统，对各个频道进行全面更新维护，大大减轻服务器压力。

大部分网站管理的核心还是内容管理，因此，一套好的内容管理系统非常重要。在内容发布方面来看，国内有两套系统非常好，排名第一的因该是 TRS，排名第二的是 TURBOCMS。TurboCMS 和 TRS WCM，他们的客户群都是成百上千家了，产品经过了 4 到 5 年的发展，各方面考虑得已经非常周到。可以从如下几个方面进行比较：

灵活性 能够管理整个网站的所有频道。TurboCMS 可以针对每个频道设置扩展字段，TRS WCM 可以针对站点设置扩展字段。模板处理引擎技术方面，TurboCMS 使用脚本技术，TRS WCM 使用 TAGLIB 技术，然后结合可视化模板制作技术，都还不错，能从易用性与灵活性两个层次满足需求。

可扩展性 TurboCMS 提供了一个数据库部署工具，可以将内容复制到外部数据库，供其他应用使用，比如复制到 Web 上供做高级查询。但 TRS 不行，只能生成静态页面，用户要查询数据库，就只好把 CMS 跟 WEB 装到一起，很不安全。这个特性，感觉 TurboCMS 是跟国外的 Interwoven 学的，Interwoven 有个 DataDeploy。

自动内容采集，从外部数据源采集内容回来进行发布，减少人工干预 TurboCMS 内置了三种自动采集爬虫，数据库采集，Web 采集，文件采集。TRS WCM 提供一个单独的产品“网络雷达”来实现 Web 采集，另外一个产品叫“数据库网关”来实现从外部数据库取内容，然后可以导入 TRS 自己的文件。都做得不错。

站内全文检索 TRS 提供了一个独立的产品做站内检索，与 TRS WCM 可以很好地集成。TurboCMS 没有自己的全文检索系统，需要与第三方的系统做集成。

这两个产品都很贵，都要几十万，不是个人网站可以使用的，这里提出来，做为内容管理领域的优秀软件简单分析，供参考。两者都是生成静态页面，同时允许.shtml 来包含动态查询，调用外部数据库。且两者的很多功能已经具备了页面优化的特征，在此软件基础上做优化工作会很轻松容易。

适合个人网站的也有不错的系统，还有电子商城类的，可以探讨挖掘。

2、即时通信

特点是：和潜在客户即时沟通。这类软件有两种。

被动沟通：MSN、QQ 等 可以加入很多好友进行交流沟通，洽谈生意。现在很多电子商务网站都有这两类联系工具。大家都很熟悉。

第二种是主动沟通：“网眼”类交流工具 它们有一句经典的广告语是：“网眼，让访客变顾客，流量变销量！”安置在网站上，这类工具能够让你主动联系上正在浏览你网站的每一个客户，查看到对方身份信息（如来自地区、登陆次数、使用搜索引擎和所查询的关键词等等），分析出最有价值的潜在客户，同时获取对方的采购意向、联系方式等重要的商业信息，从而将你网站上的浏览者都转变成为真正的客户，不再像过去那样，因为要等着客户打电话过来而流失大量宝贵的商业机会，让你的网站变成真正意义上的销售门户。但是这类产品中有些产品由于其技术原理不合理，会导致整个被框在 IFRAME 中，对搜索引擎极不友好。具体采用时，请注意它们的技术原理。如果选对了产品，对销售量的转化和用户行为分析还是很有效的。

3、邮件群发

特点是：直接投放到用户面前，定位比较精确。分垃圾群发和许可营销。

好的邮件群发系统能对大批量的邮件群发进行有效的跟踪和分析，让邮件发送者全面掌握邮件群发的状态和效果，结合消费者数据仓库进行客户关注分析，分析数据成为后续电子商务营销的决策依据。垃圾群发的效果越来越差，要进行细分定位才会改良效果。邮件列表是基于用户许可的原则，用户自愿加入、自由退出，稍微不同的是，Email 营销直接向用户发送促销信息，而邮件列表是通过为用户提供有价值的信息，在邮件内容中加入适量促销信息，从而实现营销的目的。

4、留言板群发

特点是：制造大量垃圾链接。

目前大家都在使用它做外部连接，很多 SEO 公司和 SEO 个人在客户急功近利的压力下，天天启动它群发，似乎都没有别的事情可以做了。如果把留言板略微修改一下，它的效果就大打折扣。现在这样的链接群发很容易被搜索引擎封杀。

5、论坛群发

特点是：做品牌推广。

论坛以软文性质出现的最多，重点在于品牌产品的宣传推广。

《暴力营销》是其中的代表作，内置式 OCR 解码模块，可绕过论坛“验证码”。

论坛群发对机器性能要求很高，一般机器发 10 万个论坛需要好几天甚至 10 来天。使用

虽然简单，但对结果无法跟踪。发送成功率在 65%左右，但是最后真正保存下来的帖子估计能在 10%就不错了。因为论坛是一个有斑竹管理的地方，一般的文不对题就很容易被删掉，略微有点经验的就能看出来是机器所发，那就更加是删你没商量，因此实际效果不怎么样。要想多保存帖子存活量，还得在文章内容方面下工夫。它无法选择论坛，这个是最大的弊端。

该软件由于天天骚扰论坛管理，引起很多斑竹的痛恨，已经掀起反《暴力营销》狂潮，下面是一个斑竹反《暴力营销》的一个例子：《垃圾网站》论坛群发软件的推出，标志着中国网络营销正式进入“暴力时代”。《暴力营销》被自动替换成《垃圾网站》了。

由于主站点访问人数过多，速度较慢，建议访问镜像站点：

[美国]http://病毒网站已被屏蔽/

[美国]http://病毒网站已被屏蔽/

上面的网址也被替换成“病毒网站已被屏蔽”字样。

6、博客群发

特点是：发垃圾链接。

国内已经出来了专门往博客的留言评论里发垃圾链接的软件，这个跟留言板群发类似。但还没有看到能够自动创建博客、自动登陆发帖的国产博客发帖软件，要是有的话，估计博客网站就要大大遭殃了。

7、在线分析诊断

特点是：分析网页的各种页面元素关系等。

这些工具对于 SEO 专职人员来说，用得较多。主要是分析网页中存在的问题，减轻人工工作量。这一类工具大家都很熟悉，这里列举一些如下：

关键词搜索引擎排名速查

网页坏链接在线检测

网页在线 SEO 体检

Google Dance 查询

中文域名删除组合查询

■ Google 相关

Google PR 查询

Google PR 预测

Google Dance 查询

计算别人网站贡献的 PR

Google Adsence 预览

■ SEO 查询

关键词密度查询工具

页面相似度检测

搜索引擎抓取内容模拟器

网页信噪比查询

搜索引擎收录数量查询

网页在线 SEO 体检
网页 Meta 信息检测
Dmoz 收录查询

■ 链接相关

链接广泛度查询
网页坏链接检测
判断对方是否给你做正确的链接

■ 域名相关

国际域名过期查询
国内域名过期查询
域名/ip 专业定位
whois 查询
同一主机的网站（域名）查询
域名拼写错误分析

■ 其他杂类

Alexa 世界排名查询
网页属性查看器
搜索引擎关键词排名查询
网站历史纪录查看
网站上线时间查看

8、文字编码

繁简体转换
Html/jS 互转
HTML encode/decode 互转
URL encode/decode 互转
Escape/unescape 互转

9、流量统计系统

特点是：详细分析用户行为动作，统计各种数据。

根据统计系统，我们可以从中发现很多有价值的信息，对网站进行持续改进指导很有意义。一般可以统计到的基础数据包括：

访问量 包括网站的独立访客数（IP）及页面访问量（Pageview）

访问时段 即一天 24 小时内的访问量的分布

访问者来自地区 对国内访问的分析可以精确到省

来自搜索引擎的访问 可以统计出搜索引擎种类及关键词分布

客户访问时所使用的浏览器及操作系统

客户访问时所使用的分辨率

访问来源 可以统计出来自其他网站的链接所导入的访问量

页面热点统计 可以统计出网站上最受欢迎的页面的排名
同时在线统计 可以统计出 15 分钟网站上的同时在线人数
回头率分析 可以统计出访客回访网站的频率

比较熟悉的大概有如下几个：

Google 网站访问统计 (Google Analytics) : <http://www.Google.com/analytics/>

ITSUN 网站流量统计: <http://www.itsun.com>

51yes 网站流量统计: <http://count.51yes.com>

阿江统计: <http://www.ajiang.net>

武林榜: <http://www.50bang.com>

二、Google Sitemaps

Google Sitemaps 是 Google 的一个和网站管理员相关的工具，有点象 Blog 的 RSS 功能，是一个方便自己的服务，如果大家（互联网上所有的信息源）都采用了这种方式提交自己的更新的话，Google 就再也不用派出那么多爬虫辛辛苦苦的到处乱窜了。任何一个站点，只要有更新，便会自动“通知”Google，方便 Google 进行索引。

Google Sitemaps 的用法非常简单。

首先在自己网站根目录上建立一个名为 Google11e5844324b7354e.Html，文件本身可为空。然后在 Google Sitemaps 主页上单击验证按钮，Google 会验证这一文件。如果验证通过，Google 会显示一个已验证状态，并提供一个指向用户网站统计信息的链接。经过验证后，用户随时可以通过单击“网站概述”页上的链接来查看自己网站的统计信息。

接下来，建立 Sitemaps 文件。Sitemaps 文件是 XML 格式的，Google 建议使用 Sitemaps 协议格式的文件。不过使用 BLOG 的 RSS 甚至纯文本文件也都可以。XML Sitemaps 格式其实也很简单，建议多花几分钟转换为 XML Sitemaps 格式后再提交 Google。Sitemaps 协议是 XML 的变种，用于概括与 Web 抓取工具相关的 Sitemaps 信息。对于每个网址，都可以加入抓取“提示”（如最后修改日期和近似更改频率）。Sitemaps 协议以 <urlset> 开始标记开始，以 </urlset> 结束标记结束。每个网址包含一个作为父标记的 <url> 条目。每一个 <url> 父标记包括一个 <loc> 子标记条目。<loc>是网页文件的网址。此网址应以 http 开始并少于 2048 个字符。<lastmod>表示该文件上次修改的日期。此日期允许删除时间部分，例如 YYYY-MM-DD。其他的可选项可以不加。

最后，在 Google Sitemaps 的提交页面填写你生成的 Sitemap 地址，并提交即可。

过了几个小时后，再登录 Google Sitemaps，可以看到功能菜单。

三、通过 sitemaps 分析死链

Google Sitemaps 还提供了 Google Spider 对用户网站访问情况的说明报告，如：查询统计信息、抓取统计信息、网页分析、索引统计信息等信息。“查询统计信息”中，有列出几个进入你网站的热门搜索关键词。在“抓取统计信息”中，可以看到 Google Spider 抓取网站的概要统计和抓取错误的地址。实现了网站主和 Google Spider 的信息交互。通过该表可以详细得知站点的死链情况。

第四节 Google 的几个著名现象

一、Google Trust Rank

1、TrustRank 的提出背景

搜索引擎在计算网页排名的时候，非常依赖链接，而且链接的质量越来越显得重要。这种情况下，需要对链接的来源站点质量进行判断。更重要的是，以前依靠链接和相关性来决定排名的方式，已遭到了各种各样作弊行为的挑衅，Spam 的横行，直接导致了 Google 必须找到一种新的反作弊机制，以确保高质量的站点来获得搜索引擎的亲睐。因而 Sandbox 和 TrustRank 被提了出来。意图确保好的站点能获得更高的搜索表现，并加强对站点的审核。Google 自己关于 TrustRank 的最初论述也提到了这些：

垃圾页面经常使用各种各样的作弊方法来获取一个好的搜索引擎排名。虽然人工可以很容易识别这些垃圾，但问题是这么多的页面，靠人工是无法完成的。所以我们就提出了我们最新的半自动的一种技术方案。我们首先让我们的专家选择一些好的样板页面，然后分析他们的链接结构并以此结构来比较其他页面，进而发现那些没有作弊可能性较大的页面。我们讨论了好几种样板选择和如果发现优质页面的方案。我们利用 AltaVista 在因特网上进行了实验，并评估了我们这项技术的效果。结果证明我们能够很好的通过网上一小部分优质的网站来辨别垃圾页面，当然这些优质的样板网站数量不得少于 200 个。

2、什么是 Google Trustrank?

Google 诚信指数 (TrustRank) 是指网站对于 Google 而言值得信赖的程度。

网站 TrustRank 的计算采用人工和机器链接分析相结合的方式。通过 Google 或其他一些检索机构的专家，可以先确定一批站点的 TR 值，在通过机器的链接结构分析来确定互联网上其他站点 TrustRank 值，然后以 TR 值的高低来做为网页排名的一个重要依据。跟 PR 值原理类似，如果其他站点获得了来自高 TR 值站点的链接也将获得更高的 TR 值。Google TrustRank 应该是以站点而不是页面为单位的。

3、Google TrustRank 的影响

Google TrustRank 的影响现在已经越来越明显，大量的 SEO 通过实践都证明了 TrustRank 是真实存在的。而 Google 也早就注册了 TrustRank 的商标。

Google TrustRank 对于网站排名有种非常重要的影响：

- 站点内的页面在其他情况参数接近的情况下，高 TR 值的站点内页面将获得比其他站点页面更高的排名；
- 高 TR 值站点的页面收录速度加快，因为 Google 对它认为重要的站点会频繁访问；
- 获得足够的 TR 值的站点可以避免 Sandbox；
- 如果一个站点的信任指数太低，Google 将可能会将其进行惩罚，包括进入 Sandbox 等；
- 如果一个站点的信任指数太低，即使其他参数非常理想，在较热门关键词上，也很难获得好的排名表现。

二、Google Sandbox

1、什么是 Sandbox

所谓沙盒效应指的是，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链接，网站既搜索引擎友好，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。可以说，Google 的 Sandbox 有点像给予新网站的一个试用期。在这段试用期内，新网站几乎无法在竞争比较激烈的关键词下得到好的排名。

这个现象最早是在 2004 年 3 月开始被注意到。沙盒效应更多的是发生在以竞争比较激烈的关键词为目标的网站上。那些不太商业的，竞争比较少的关键词，发生沙盒现象的机会就比较小。

虽然沙盒效应是这一两年最热门的话题，但到目前为止，没有人能够确认确实存在沙盒效应，Google 从来没有肯定或否定过。有人认为，沙盒效应本身就是一个独立的过滤算法，把所有新网站的排名暂时都排到后面。

也有人认为沙盒效应只是一种效应或现象，它本身并不是一个过滤器，而是很多其他排名因素所造成的一个现象而已。比如说，Google 确实在近一两年开始在把链接的年龄以及链接页的历史情况都考虑在内。还有网站获得链接的速度，如果一个网站在短时间内获得了大量的链接，Google 会怀疑这些链接是买来的，或者是通过作弊手段得到的。这些与时间有关的因素组合起来，就可能对新网站产生这种沙盒现象。

虽然对 Sandbox 的本质是什么众说纷纭，但一般都认为这种现象是存在的，因为是可以观察到的。

大部分人认为 Google 之所以会制造这种 Sandbox 效应，是为了清除那些垃圾网站。通常这些垃圾网站都会快速买大量链接，得到好的排名，赚一笔钱后，这些作弊手段被发现了，网站被删除或被惩罚。但是这些人也不在乎，这个域名也就被放弃了，转而开始做另外一个新的网站。

在 Google 开始出现 Sandbox 以后，有可能大部分的这些垃圾网站制造者都没什么耐心，要等他的网站从沙盒里面出来，从而也就不建这些垃圾网站了。

通常沙盒效应会维持六个月，有一些针对竞争性不高的关键词的网站，可能在沙盒里会短一些。行业竞争越高，沙盒效应会越长。

Sandbox 仅是发生在 Google 排名中的一种现象。一般认为，Sandbox 效应是发生在距今最近的 Google 算法更新后（即业界称为“Florida”与“Austin”更新）。具体是这样的：Google 会给新网站额外加上一定的分值，使得新网站的排名迅速上升（某几种目标关键词）。随后的一段时间里（大概是一到三、四个月），网站的排名就会逐渐下降，有的主要关键词排名甚至会消失（尽管你的 PR 很高，页面内容与原来一样）。这就如沙漠里的沙流将周围的物体慢慢吞噬，所以业界给这种现象取名为 Sandbox 现象。

2、怎样去正确看待 Sandbox 效应？

当你的网站进入了 Sandbox 效应时期，你并没有受到 Google 的处罚，你也很确定 Google 依然收录你的网站。你应该保持平静，即使你的网站是在一开始就有很多重要的关键词都排得很好，现在突然间消失了或者下降了很多，我们仍然需要保持心理平衡。

首先要放宽心，因为从根本上说，你没办法逃避沙盒现象。随着时间的推移，过大概六

个月以后，你的新网站或新域名自然会从沙盒里出来。

同时，当你的网站还在被 Sandbox 沙盒效应影响的时候，你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。

实际上有了沙盒效应，对很多网站运行者来说，可能是一件好事。因为在大概半年的时间里，你只能把精力放在网站内容上面。而不会去考虑排名。

从长远来看，一旦沙盒效应结束，你的导入链接的年龄也足够长了，会被记入算法当中。你的网站也有了足够的内容，那么你的网站排名会有一个质的飞跃。

3、Sandbox 效应时期采取的措施？

在 Sandbox 效应时期，我们应该采取有效的措施，尽快从 Sandbox 效应中跳出来，重新获得排名。以下是几种常用的措施：

- 花更多的时间去更新你网站的内容；
- 利用 Sandbox 效应这段时间加入更多高质量的分目录，交换更多的高质量链接；
- 采取更多的营销手段保持网站正常的流量。

4、避免陷入 Sandbox 效应的方法

据业界的很多专家观察、实践，他们总结了几个较为有效的方法。但这毕竟是没有一个标准或者正确的答案，只是仁者见仁，智者见智，供大家一个参考而已。或者你还有可能发现更多有效的方法。

- 在你的网站完成之前就把所申请的域名挂上去，等过了 Sandbox 效应时期再上传网页（这个手法不提倡）；
- 对于一般零售业的网站来说，最重要的是尽量避免在购物的旺季前一到两个月上传网站；
- 新网站被收录后，坚持每天或者至多三天就要更新一至两页的内容；
- 向更多的高质量目录提交，交换更多相关主题高质量网站链接；
- 如果前期排名理想的话，千万不要大动网站的整体结构。

对于第一种手法，之所以不认同、不提倡是由于以下原因：

- Google 不喜欢那些只有域名而没有任何内容的网站；
- 既然网站完成没有内容又何来排名？Sandbox 效应更是无从谈起了；
- 如果真的存在着 Sandbox 效应的话，我们何妨不如“怎样去正确看待 Sandbox 效应”中所说，利用这个时机把优势一直保持下去呢？

5、如何摆脱 Sandbox

当我们的网站被“放入”了 Sandbox 后应该怎么办？

耐心等待！

这是无奈但又最有效的答案。但我们必须明白：Google 的排名规则本身并没有大的变化，网站的排名取决于内容的质量、反向链接的质量等。我们要做的便是在这一段时期进一步丰富内容、建立链接等，这样，在离开 Sandbox 时才能保证页面在 SERP 中能有较理想的位置。

一般说来，在这段时间应该：

- 正常地更新内容，特别是高质量的原创内容；
- 尽可能稳定地、逐步地得到一些高质量的外部链接，这也可以成为访问量的另一个来源；
- 通过其他的搜索引擎给网站带来访问量。

6、针对非竞争性关键词优化

被置入 Sandbox，并不意味着被 Google 屏蔽。相比较而言，Sandbox 效应在竞争性强的关键词上表现异常明显，而在竞争性较弱的关键词上则放开一面。通过对非竞争性的关键词进行优化，依然可以让我们出现在 SERP 的前面，虽然这样带来的访问量不大。另一方面，有关研究表明，选用竞争性强关键词的网站其在 Sandbox 中呆的时间要长于选用非竞争性关键词的网站，这样，通过选用非竞争性关键词可以让我们尽快摆脱 Sandbox。

相信没人愿意自己的新网站被置入 Sandbox，但是，我们对此没有选择，只能尽量减小其带来的影响。

7、Yahoo 也可能有 Sandbox

经过观察发现，与 Google 类似，Yahoo 的算法中也可能包含 Sandbox 机制。

Yahoo Sandbox 的表现跟 Google 类似。一方面新站可能列入黑盒，其次，对于老站点，如果有不合理的状况或作弊行为，也可能导致进入 Sandbox。另外，也跟 Google 一样，Sandbox 的反面 Trustrank 也存在，Yahoo 对于站点也开始给予不同的信任指数。

但 Yahoo Sandbox 的诱发因素、影响的时间和范围，与 Google 是有所不同的，具体的情况还需要进一步的探讨。上述的结论是基于 Yahoo 目前对一些作弊站点的惩罚措施，新站在 Yahoo 表现情况监测后做出的猜测。目前还无法通过官方的渠道对此进行证实。相信不同的人对上述的看法也会持不同的见解。

三、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），你可以发现 Google 的搜索结果会有大幅度的波动。事实上，Google 的更新过程可以很容易地通过搜索结果的显著变化来识别，通过 Google 对搜索页面的网页快照的更新也能够看出来。在 Google Dance 过程中，几乎每一分钟都会有变化。

在 Google 的更新过程中，其搜索始终在变化，所以它的搜索结果也会有很大的变动，页面等级（PageRank）和返回链接的更新也是在这个时期完成的。

如何使用 Google Dance 工具？

众所周知，Google 有三个主要服务器，分别是：

www.google.com：主服务器，Google 主页所在位置；

www2.Google.com：测试服务器；

www3.Google.com：测试服务器。

除 Google 自己外，没人知道其 www2 与 www3 服务器的真正用途，大多数人推测，其可能派作测试用场。在 Google 更新期间，这三个服务器的返回结果有相当程度的差异，正因为这种搜索结果的大幅度波动，Google 的索引更新才被称为“Google Dance”。

除 www2 和 www3 外，使用本工具还可对 Google 其他数据中心同时进行查询。通过

相应的返回结果，你可对确知 Google 是否正在进行更新，及 Google 更新的规律与趋势。

四、Sitelinks

1、什么是 Sitelinks

经常使用 Google 的人，一定对下面这情况并不陌生：可能搜索某一关键词的时候，会发现排在第一名的结果除了正常的显示标题、描述和网址之外，下面还可能会以链接的方式列出几个该网站其他页面的标题。这种情况下，第 1 名在 Google 显示的位置范围差不多是第 2 名的二倍，更加吸引注意和彰显行业至尊地位。



图 14-4-1 搜索“新浪”，新浪网的表现



图 14-4-2 搜索“太平洋电脑网”，太平洋电脑网的表现

这种情况在 Google 以前出现了很久，但似乎最近才从 Google 的网站管理员帮助中心得到了比较详细的解释：

The links shown below some sites in our search results, called Sitelinks, are meant to help users navigate your site. Our systems analyze the link structure of your site to find shortcuts that will save users time and allow them to quickly find the information they're looking for.

We only show Sitelinks for results when we think they'll be useful to the user. If the structure of your site doesn't allow our algorithms to find good Sitelinks, or we don't think that the Sitelinks for your site are relevant for the user's query, we won't show them.

At the moment, Sitelinks are completely automated. We're always working to improve our itelinks algorithm, and we may incorporate webmaster input in the future.

显示在有些网站的搜索结果下面的链接，叫做站内链接（Sitelinks），这些链接是为了帮助用户更好地访问你的站点。我们的系统通过分析网站的链接结构，来发现这种快捷方式，以帮助用户接省时间、更迅速地找到他们想找的东西的。

只有当我们觉得这些站内链接对我们的用户非常有价值的时候我们才会把他显示出来。如果你网站的结构不能让我们的算法找出高质量的站内链接，或者我们认为这些链接跟我们的用户搜索请求并不相关的时候，我们也不会把它们显示出来。

目前，站内链接完全是机器自动生成的，我们也一直在致力于提高我们的站内链接算法，将来我们可能会让我们的广大网站管理员参与进来。

根据 Google 的说法，Sitelinks 的主要目的在于帮助用户更快速、以更简单的方式找到所需的信息。当然，Google 只是对部分关键词排在首位的网站才提供 Sitelinks，一般说来，这类关键词多为网站名称、品牌、商标。

2、如何获得 Google Sitelinks?

既然 Google Sitelinks 是通过算法自动得出的，那么肯定有些较重要的因素能够帮助网站获得 Sitelinks，也许，这是下一步网站建设者努力的一个方向，毕竟，与单纯的所谓排名优化相比，让更多的用户点击访问网站才是实现 SEO 目标的真正保障。

首先要在特定的搜索项中排名第一，这是最根本的，因为 Sitelinks 只向在 SERP 中排名第一的网站提供。

使用网站名称或名称中的一部分作为主关键词，将是获得 Sitelinks 的决定性因素。仔细撰写网页的描述与内容部分，对被 DMOZ 收录的网站，应考虑是否使用 NOODP 标签。据大多数人观察，Google Sitelinks 与网页在 SERP 中的点击量有关，也即是说，单纯排名是不够的，必须能在 SERP 页中引起用户的注意并点击。

网站的内部链接也相当重要，并要保证 Googlebot 对网站深层页面的访问。此外，通过对大多数 Sitelink 页面的观察，内部链接与导航出现在网页 HTML 代码靠前位置具有一定优势。

很多人推测 Google Sitelinks 可能使用 Google toolbar（工具条）数据来采集数据，包括用户对特定网页的访问情况、收藏情况等等，这就又回到了 SEO 的老问题上，即让用户感到有用，感到有价值而收藏并多次访问。

3、Sitelinks 的其他推测

当然，在现阶段，仍然没有太肯定的关于 Sitelinks 实现机制的结论，但一般说来，如果某个网站在 Google 的 SERP 中拥有 Sitelinks，可以认为 Google 将该网站判定为该搜索项方面的权威网站。

此外，在 Sitelinks 中，链接采用的方式使得玩味：SERP 中的其他链接均是直接指向目标地址，而 4 个 Sitelinks 则采取类似 JavaScript 实现的 302 跳转，联系到 Google 说明中的“and we may incorporate webmaster input in the future”，说明了什么？

五、补充材料

1、什么是 Google 补充材料

在 Google 搜索结果中除了列出网页的网页快照，类似网页以外，有时候 Google 还会把网页标志为补充材料。

到目前为止，Google 并没有太多信息告诉大家补充材料是怎么定义的。在 Google 自己的网站管理员指南中所给的官方信息是这样写的：

补充网站是 Google 辅助索引的一部分。对辅助索引进行抓取的限制少于主要索引。例如，一个网址的参量可能使该网站无法被抓取到主要索引中，但仍可能被抓取到我们的补充索引里。

将网站包括其中的索引是完全自动的，没有任何方式可以选择或更改你网站的索引。请放心，包含网站的索引不会影响该网站的 PageRank。

实际的情况是，被标为补充材料的网页在 Google 搜索结果中一般都会被排在普通网页的后面，所以通常是找不到的。除非关键词是比较长，并且比较具体的时候，补充材料才会出现。

被标为补充材料的网页是在单独的数据库（补充索引）里，不是主索引数据库，在补充索引里的网页被爬行的次数会减少，网页快照的日期也通常都是很旧的。

如果你的网站有很多网页被标为补充材料，在绝大部分情况下都不是什么好事。因为只有普通结果不够的时候，被标为补充材料的网页才会出现在搜索结果中，自然对网页的排名有很大影响。

2、补充材料产生的原因

被标为补充材料的网页，通常有下面几种情况：

第一，有网址规范化问题。也就是带有 www 和没有 www 的两个网址版本都在数据库中，那么其中一个就有可能被标为补充材料。

第二，有时候已经被删除的网页，也就是实际上应该返回 404 错误的 URL，会被列为补充材料。或者域名已经过期了，也有时会被标成补充材料。

第三，有时候网站管理员在改动网页之后，新旧内容的两个版本的网页都会出现在 Google 索引中。而糟糕的是，其中一个会被标为补充材料。

第四，复制内容网页。很多转载或抄袭的内容会被标为补充材料。因为 Google 也没办法判断哪一个是原创的，所以有时候原创者也会遭殃。

第五，网页上相同或相似的内容太多。比如导航系统占网页内容比例比较大，而正文部

分比较小。

从上面这几种情况来看，补充材料在大部分情况下和网页内容重复或类似有关。

如果你的网页被标志成补充材料，有什么办法解决呢？基本上没有什么太好的办法。你只能改正你能控制的部分，比如说网址规范化问题，正文内容太少等。然后就等着 Google 解决他们自己的算法问题吧。

3、补充材料对排名的影响

这个问题就比较复杂，要具体情况具体分析。

像前面讲的，同一个 URL 可能出现在不同的数据库中，新版本的 URL 可能会出现在正常的主要关键词的搜索中。但是在其他关键词搜索中，它也可能出现在补充材料中。

如果你的网站有很多补充材料，有可能影响排名。但是如果这些补充材料都有一个相应的新的 URL+日期+网页快照版本，对排名的影响就不大。如果这些补充材料没有新的版本，影响就比较大。

补充材料 URL 会不会被更新呢？ 会被更新，但是被重新抓取的频率要比较低。当 Google 蜘蛛来抓取这个 URL 的时候，如果以前使它成为补充材料的那个原因已经不存在了的话，这个 URL 可能就被重新归入主数据库。

补充材料会存在多长时间？ 一般来说补充材料会继续存在长达一年或更长的时间。不过出现补充材料并不是最大的问题，如果你的网站新版本已经都被重新抓取，就没关系了。