

# 搜索引擎优化基础教程

本教程由站长百科 ([www.zzbaike.com](http://www.zzbaike.com)) 教程制作组整理



作者: 月飞春秋@站长百科  
制作者: 站长百科教程制作组  
制作时间: 2009年3月21日  
制作网站: 站长百科 [www.zzbaike.com](http://www.zzbaike.com)  
版权: 站长百科

# 使用说明

本电子书的内容来源于网上搜集的资料以及一些 SEO 相关的教程。制作组人员把搜集的相关的资料进行了整理“去其糟粕，取其精华”，把 SEO 最基本的内容、最基本的知识，以最简便的方式展现给大家。

既然是基础教程，讲的当然都是 SEO 最基本的内容了，例如：佩奇等级、robots.txt 等相关知识，对于刚刚涉足于搜索引擎的朋友来说，可能对这些比较迷茫，但，通过本书能很快的了解这些内容。

本电子书为 PDF 格式，为了方便您浏览，我们为电子书制作了书签。可以用两种方式来导航：

- 1.点击左侧的书签标志，打开书签，点击书签的内容就可以转到对应页面

- 2.使用本电子书的目录，点击可进入相应页面。

您也可以利用 PDF 的搜索功能来快速查找您需要的内容。

本书是我们推出的第一本电子教程，搜索引擎更高级的优化方法，我们将在第二本的电子书中进行介绍，届时我们讲介绍最常用的优化方法以及优化技巧，热爱搜索引擎优化的朋友敬请期待吧！

本电子书由站长百科教程组制作，虽然是对收集的资料进行整理，但是，教程制作组人员为此也付出了的辛勤的汗水，所以，请勿私自将全部或部分用于商业用途。转载书中内容请注明出处。

第一章:搜索引擎基础.....	3
第一节: 什么是搜索引擎.....	3
1.1.1.搜索引擎的作用.....	3
1.1.2.搜索引擎的定义.....	3
第二节: 搜索引擎的工作原理.....	4
1.2.1 探寻蜘蛛和机器人.....	4
1.2.2. 如何辨别搜索引擎机器人身份.....	5
第三节: SiteMap 介绍.....	6
1.3.1.什么是 SiteMap.....	6
1.3.2. Sitemaps 文件的位置.....	7
1.3.3.XML Sitemaps 格式.....	8
1.3.4.创建 SiteMap.....	11
1.3.5. 使用 Sitemaps 索引文件.....	12
1.3.6. 验证您的 Sitemaps.....	13
1.3.7.提交 SiteMap.....	14
第四节: Robots.txt 介绍.....	15
1.4.1.什么是 Robots.txt.....	15
1.4.2. robots.txt 的语法格式.....	16
1.4.3. robots.txt 使用技巧.....	18
1.4.4. robots.txt 用法举例:.....	19
1.4.5.robots.txt 举例.....	20
1.4.6.常见 robots.txt 错误.....	20
第二章: 主要搜索引擎介绍.....	22
第一节: Google.....	22
2.1.1 关于 Google.....	22
2.1.2. Google 的 PageRank 技术.....	23
2.1.3. Google 的 SiteMap 技术.....	24
2.1.4. 网站管理员工具.....	27
第二节: 百度.....	28
2.2.1.关于百度.....	28
2.2.2.百度的 SiteMap.....	29
2.2.3.百度论坛的开发协议.....	31
第三节: YaHoo (雅虎).....	34
2.3.1. 关于雅虎.....	34
2.3.2.YaHoo 中国.....	35
2.3.3.中文搜索引擎列表.....	35
第三章: 网站的基本查询方式.....	35
第一节: 域名查询.....	35
第二节: 收录查询.....	36
3.2.1. 网站是否被收录.....	36
3.2.2. 网页收录数量.....	37
第三节: 反向链接查询.....	38
第四节: 网站 PR 查询.....	38
第五节: 关键词排名查询.....	38

第六节: ALEXA 查询.....	39
3.6.1. 什么是 Alexa? .....	39
3.6.2.Alexa 的主要数据 .....	39
3.6.3.Alexa 的作用 .....	40
3.6.4. 提高 Alexa 排名 .....	40
第四章: 搜索引擎优化基础.....	42
第一节: 搜索引擎优化的定义和价值.....	42
4.1.1 搜索引擎优化的定义.....	42
4.1.2.搜索引擎优化的价值.....	43
第二节: 搜索引擎优化究竟为了什么 .....	44
4.2.1.SEO 体现网站效果 .....	44
4.2.2.SEO 的终极目标和实现目标 .....	44
第三节: 白帽 SEO 和黑帽 SEO.....	44
第四节: 关于搜索引擎优化的一些观点 .....	47
4.4.1. 搜索引擎优化合法吗? .....	47
4.4.2.会不会和搜索引擎对着干? .....	47
4.4.3.我的网站有大量的优质原创内容, 所以我不需要 SEO 了。.....	47
4.4.4.SEO 没有技术含量或 SEO 需要很高深的技术.....	48
4.4.5.大网站那样做了, 我也向他们学习 .....	48
4.4.6.SEO 企业或 SEO 工具的保证排名宣传 .....	48
4.4.7.稳定的排名.....	49
4.4.8.竞争对手无法破坏我的排名 .....	49
第五章: 网站设计的优化操作.....	49
第一节: 域名对 SEO 的影响.....	49
5.1.1.域名的基本常识.....	49
5.1.2 不同后缀的域名在搜索引擎中的权重 .....	51
第二节: 主机对 SEO 的影响.....	51
5.2.1. 共享主机、独立主机的区别? .....	51
5.2.2. 主机服务商选择.....	52
5.2.3.主机服务商转移: .....	53
第三节: 对搜索引擎友好的设计.....	53
5.3.1. 目录结构.....	53
5.3.2.URL.....	54
5.3.3.导航结构.....	55
5.3.4. 框架结构.....	56
5.3.5. 图片优化.....	57
5.3.6. FLASH 优化 .....	57
5.3.7.404 错误页设置.....	57
第四节: 网站设计思维.....	58
5.4.1.重视网站结构一致性.....	58
5.4.2.W3C 浏览器兼容验证 .....	58
附录 A: 搜索引擎优化常用术语 (中英对照) .....	59
附录 B:网站管理员指南(Google).....	70
附录 C: 百度的建站指导 .....	73

## 第一章:搜索引擎基础

### 第一节:什么是搜索引擎

#### 1.1.1.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。现在的网站用户大部分是通过搜索引擎进入的,如果用户要购买一件商品,他不知道去那家网上的商店去购买,他就会去搜索,我们的搜索引擎就是满足用户的这些需求。

我们都知道,现在的互联网上有大量的垃圾信息,而我们的搜索引擎的作用就是将比较好的网站,和质量高的网站排在页面的前面,这样用户就能很快通过引擎找到他们的东西了,这应该就是引擎的作用。

#### 1.1.2.搜索引擎的定义

给搜索引擎下定义是困难的,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。

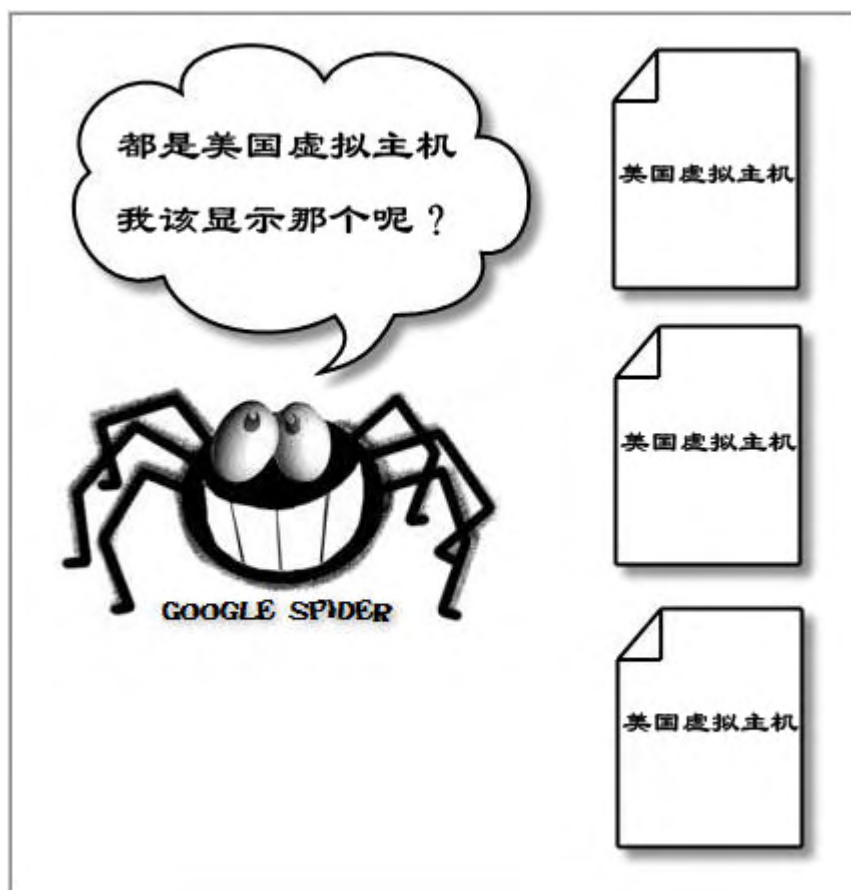
所谓的搜索引擎(Search Engines)是一些能够主动搜索信息(搜索网页上的单词和简短的特定的内容描述)并将其自动索引的Web网站,其索引内容储存在可供检索的大型数据库中,建立索引和目录服务。它是一个为你提供信息“检索”服务的网站,它使用某些程序把互联网上的所有信息归类,以帮助人们在茫茫网海中搜寻到所需要的信息。

## 第二节：搜索引擎的工作原理

### 1.2.1 探寻蜘蛛和机器人

搜索引擎要知道网上的新生事物，就得派人出去收集，每天都有新的网站产生，每天都有网站内容的更新，而且这些产生的网站数量、更新的网站内容是爆炸式的，靠人工是不可能完成这个任务的，所以搜索引擎的发明者就设计了计算机程序，派它们来执行这个任务。

探测器有很多叫法，也叫 Crawler(爬行器)、spider(蜘蛛)、robot (机器人)。这些形象的叫法是描绘搜索引擎派出的蜘蛛机器人爬行在互联网上探测新的信息，Google 把它的探测器叫做 Googlebot，百度就叫 Baiduspider,Yahoo 称为 Slurp，无论它们叫什么，它们都是人们编制的计算机程序，由它们不分昼夜的访问各个网站，取回网站的内容、标签、图片等，然后依照搜索引擎的算法给它们定制索引。



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。

当抓取到这些网页后，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。

建立网页索引数据库后，当用户输入关键词搜索，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。

### 1.2.2. 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, 你可以通过服务器上的设置来阻止这个非法机器人。

## 第三节: SiteMap 介绍

### 1.3.1.什么是 SiteMap

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 在用户无法通过可浏览界面访问网站的所有区域时作用尤其明显。(通常, 指用户无法通过追踪链接访问网站的特定页面或区域。)例如, 那些只能通过搜索表单才能访问其中某些页面的网站都会从创建 Sitemaps 并将其提交到搜索引擎中获益。此文件说明 Sitemaps 文件的格式, 并解释您张贴 Sitemaps 文件的位置以便搜索引擎能够检索到。

请注意 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您的网站。

就算已经被 Google 和 Yahoo 收录, 也还是有必要用 SiteMap 的, 这是为什么呢? 有以下两大原因:

- a) 用 SiteMap 有利于搜索引擎更加友好的对网站进行收录, 不要让网站的收录有些漏洞或者收录不全。有了 SiteMap 可以很好的在这方面发挥作用。
- b) 有了 SiteMap 并提交给搜索引擎之后, 方便你在以后进行下一步的工作, 比如对网



站的外部链接和内链错误进行更好的调整, 这些都要用到 SiteMap 的提交功能。

因此, SiteMap 的整个作用是不容忽视的。如果你的网站属于下列情况, 那么, 使用 SiteMap 会特别实用:

- A. 网站含动态内容。
- B. 网站有不容易被发现的页面, 如有大量富 AJAX 或 Flash 内容的页面。
- C. 网站为新网站且指向网站的链接不多。
- D. 网站有大量内容页存档, 这些内容页相互没有很好地链接, 或根本就没有链接。
- E. 网站上网页的更改频率。
- F. 各网页上次修改的日期。
- G. 您网站上各网页的相对重要性。

如果你的网站出现上面介绍的情况, 那么, 你就需要创建 SiteMap 了。

### 1.3.2. 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。

如果您有更改 `http://example.org/path/sitemap.gz` 的许可, 您应该也可以提供以 `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, 您需要针对不同的路径创建对应的 Sitemaps, 例如, 如果在您的组织中, 安全许可对应不同目录划分上传权限。

1.3.3.XML Sitemaps 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。

XML 标记定义

以下对可用 XML 标记进行说明。

<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。如果您的 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 <a href="#">W3C Datetime</a> 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	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： <ul style="list-style-type: none"><li>• always</li><li>• hourly</li><li>• daily</li><li>• weekly</li><li>• monthly</li><li>• yearly</li></ul>

		<ul style="list-style-type: none"> <li>• never</li> </ul> <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<b>提示</b>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	<p>此网址的优先级与您网站上其他网址的优先级相关。有效值范围从 0.0 到 1.0。此值不会影响您的网页与其他网站上网页的比较结果，只是告诉搜索引擎您认为您的那个网页最重要，从而它们对您页面的抓取可以按照您最喜欢的方式进行排序。</p> <p>一个网页的默认优先级为 0.5。</p> <p>请注意，您为页面指定的优先级 <b>不影响</b>您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，您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，请注意为您网站中的所有网址都指定高优先级对您没什么好处。由于优先级是相互关联的，只用于在您自己网站的网页之间进行选择；您网页的优先级不会与其他网站的页面优先级进行比较。</p>

## 实体转义

我们要求您的 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。

字符		转义码
& 符号	&	&amp;
单引号	'	&apos;
双引号	"	&quot;
大于	>	&gt;
小于	<	&lt;

此外，所有的网址（包括您的 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，如果您使用任何脚本、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），通常这部分已经替您完成了。如果您提交了 Sitemaps 却收到 Google 无法找到某些网址的错误消息，请查看并确保您的网址遵循 [RFC-3986](#) URI 标准、[RFC-3987](#) IRI 标准以及 [XML 标准](#)。

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

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

以下是 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`

以下是 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`

以下是实体转义过的同一网址:

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

### 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>
  </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&desc=vacation_usa</loc>
  <lastmod>2004-11-23</lastmod>
</url>
</urlset>

```

您可以使用 gzip 压缩您的 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。

### 1.3.4.创建 SiteMap

如果您有多个网站，可以创建一个或多个 Sitemap 将所有经过验证网站的网址包含在其中，然后将这些 Sitemap 保存在一个位置，以此来简化 Sitemap 的创建和提交过程。您可以通过以下多种方式创建 Sitemap：

#### 1.3.4.1 根据 Sitemap 协议创建 Sitemap：

1. 创建一个文本文件并将其带 .xml 扩展名保存。
2. 将以下内容添加到文件顶部：

```

<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">

```

3. 将以下内容添加到文件底部：      </urlset>
4. 为各网址创建一个条目。 <loc> 为必需，其余标记为可选。

```

<url>
  <loc>http://www.example.com/</loc>
  <lastmod>2005-01-01</lastmod>
  <changefreq>monthly</changefreq>
  <priority>0.8</priority>
</url>

```

5. 将 Sitemap 上传到您的网站。 然后使用网站管理员工具将其提交给 Google。

#### 1.3.4.2.Sitemap 生成器

如果您可以访问网络服务器并且服务器装有 Python，则可以利用我们的脚本来创建采用 Sitemap 协议的 Sitemap。

Google Sitemap 生成器（后面的章节介绍）是一个 Python 脚本，可以使用 Sitemap 协

议为您的网站创建 Sitemap。此脚本可以通过网址列表、网络服务器目录或通过访问日志创建 Sitemap。要使用此脚本：

1. 您需要连接到您的网络服务器，并在上面运行脚本。
2. 您的网络服务器需安装 Python 2.2 或更高版本。
3. 您需要知道启动 Python 的命令。（该命令通常为 python，但可能会因安装不同而异。）例如，如果网络服务器安装了两种版本的 Python，命令 python 可调用较早的版本，而命令 python2 则可调用较晚的版本。
4. 您需要知道指向贵网站的目录路径。如果您的网络服务器托管一个网站，则路径可能是 var/www/html 等。如果您拥有一个托管多个网站的虚拟服务器，路径可能是 home/virtual/site1/fst/var/www/html 等。
5. 您需要能够将文件上传到自己的网络服务器（例如，使用 FTP）。
6. 如果您要根据访问日志生成一个网址列表，则需要了解用于这些日志的编码，以及指向它们的完整路径。

Sitemap 生成器可以用于生成常规 XML Sitemap。不过，一些特殊种类的 Sitemap（如代码搜索和视频 Sitemap）有非常独特的要求。因此，您最好不要使用 Sitemap 生成器来创建这些 Sitemap。

#### 1.3.4.3 使用第三方工具。

很多第三方提供了可用于创建有效的 Sitemap 的工具。

#### 1.3.5. 使用 Sitemaps 索引文件

您可以提供多个 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，您需要创建多个 Sitemaps 文件。如果预计您的 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。如果您的确提供了多个 Sitemaps，您可以将其列在 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](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, 您的 Sitemaps 索引文件应为 UTF-8 编码。

### 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。

### 1.3.6. 验证您的 Sitemaps

Google 使用 XML 架构定义可以出现在 Sitemaps 文件中的元素和属性。您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>

有多种工具可帮助您根据此架构来验证您的 Sitemaps 结构。在下面的每一个位置您都可以找到 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您的 Sitemaps 或 Sitemaps 索引文件, XML 文件需要有附加的标头。如果您使用的是 Sitemaps 生成器, 这些标头已经包含其中。如果您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>
```

### 1.3.7.提交 SiteMap

SiteMap 制作完成以后直接传到空间上(只要传到网络上, 能访问就行), 然后在各个搜索网站提交。

向 Google 提交网站地图 Sitemap: 通过 <http://www.google.com/webmasters> 管理提交;

向 Yahoo!提交网站地图 Sitemap: 通过 <http://siteexplorer.search.yahoo.com> 管理提交;

向百度 Baidu 提交网站地图 Sitemap: 没办法, 现在百度不支持 Sitemap。但可通过 [http://www.baidu.com/search/url\\_submit.html](http://www.baidu.com/search/url_submit.html) 来提交你的网址。百度自行搜索, 更新速度很快。

## 第四节: Robots.txt 介绍

### 1.4.1.什么是 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。举例来说, 当 spider 访问一个网站 (比如 <http://www.abc.com>) 时, 首先会检查该网站中是否存在 <http://www.abc.com/robots.txt> 这个文件, 如果 Spider 找到这个文件, 它会根据这个文件的内容, 来确定它访问权限的范围。

就算已经被 Google 和 Yahoo 收录, 也还是有必要用 SiteMap 的, 这是为什么呢? 有以下两大原因:

- c) 用 SiteMap 有利于搜索引擎更加友好的对网站进行收录, 不要让网站的收录有些漏洞或者收录不全。有了 SiteMap 可以很好的在这方面发挥作用。
- d) 有了 SiteMap 并提交给搜索引擎之后, 方便你在以后进行下一步的工作, 比如对网站的外部链接和内链错误进行更好的调整, 这些都要用到 SiteMap 的提交功能。

因此, SiteMap 的整个作用是不容忽视的。如果你的网站属于下列情况, 那么, 使用 SiteMap 会特别实用:

- A. 网站含动态内容。
- B. 网站有不容易被发现的页面, 如有大量富 AJAX 或 Flash 内容的页面。
- C. 网站为新网站且指向网站的链接不多。

- D. 网站有大量内容页存档, 这些内容页相互没有很好地链接, 或根本就没有链接。
- E. 网站上网页的更改频率。
- F. 各网页上次修改的日期。
- G. 您网站上各网页的相对重要性。

如果你的网站出现上面介绍的情况, 那么, 你就需要创建 SiteMap 了。

### 1.4.2. robots.txt 的语法格式

“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:

在"robots.txt"文件中, 如果有多条 User-agent 记录说明有多个 robot 会受到"robots.txt"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\*, 则对任何 robot 均有效, 在"robots.txt"文件中, "User-agent:\*"这样的记录只能有一条。

如果在"robots.txt"文件中, 加入"User-agent:SomeBot"和若干 Disallow、Allow 行, 那么名为"SomeBot"只受到"User-agent:SomeBot"后面的 Disallow 和 Allow 行的限制。

#### Disallow:

这个值可以是一条完整的路径, 也可以是路径的非空前缀, 以 Disallow 项的值开头的 URL 不会被 robot 访问。例如"Disallow:/help"禁止 robot 访问/help.html、/helpabc.html、/help/index.html, 而"Disallow:/help/"则允许 robot 访问/help.html、/helpabc.html, 不能访问/help/index.html。

"Disallow:"说明允许 robot 访问该网站的所有 url, 在"/robots.txt"文件中, 至少要有一条 Disallow 记录。如果"/robots.txt"不存在或者为空文件, 则对于所有的搜索引擎 robot, 该网站都是开放的。

- ◆ 要拦截整个网站, 请使用正斜杠。

Disallow:/

- ◆ 要拦截目录及其中的所有内容, 请在目录名后添加正斜杠。

Disallow:/private\_directory/

- ◆ 要拦截网页，请列出该网页。

Disallow:/private\_file.html

#### **Allow:**

与 Disallow 项相似，这个值可以是一条完整的路径，也可以是路径的前缀，以 Allow 项的值开头的 URL 是允许 robot 访问的。例如"Allow:/hibaidu"允许 robot 访问/hibaidu.htm、/hibaiducom.html、/hibaidu /com.html。一个网站的所有 URL 默认是 Allow 的，所以 Allow 通常与 Disallow 搭配使用，实现允许访问一部分网页同时禁止访问其它所有 URL 的功能。

注意：Disallow 与 Allow 行的顺序是有意义的，robot 会根据第一个匹配成功的 Allow 或 Disallow 行确定是否访问某个 URL。

#### **使用 \* 匹配字符序列**

您可使用星号 (\*) 来匹配字符序列。例如，要拦截对所有以"专用"开头的子目录的访问，可使用下列条目：

User-Agent:Googlebot

Disallow:/专用\*/

要拦截对所有包含问号 (?) 的网址的访问，可使用下列条目：

User-agent:\*

Disallow:/\*?

#### **使用 \$ 匹配网址的结束字符**

您可使用 \$ 字符指定与网址的结束字符进行匹配。例如，要拦截以 .asp 结尾的网址，可使用下列条目：

User-Agent:Googlebot

Disallow:/\*.asp\$

您可将此模式匹配与 Allow 指令配合使用。例如，如果 ? 表示一个会话 ID，您可排除所有包含该 ID 的网址，确保搜索引擎蜘蛛不会抓取重复的网页。但是，以 ? 结尾的网址可能是您要包含的网页版本。在此情况下，可对 robots.txt 文件进行如下设置：

User-agent:\*

Allow:/\*?\$

Disallow:/\*?

Disallow:/ \*? 一行将拦截包含 ? 的网址（具体而言，它将拦截所有以您的域名开头、

后接任意字符串，然后是问号 (?)，而后又是任意字符串的网址)。

`Allow: /*?$` 一行将允许包含任何以 ? 结尾的网址（具体而言，它将允许包含所有以您的域名开头、后接任意字符串，然后是问号 (?)，问号之后没有任何字符的网址)。

### 1.4.3. robots.txt 使用技巧

1. 每当用户试图访问某个不存在的 URL 时，服务器都会在日志中记录 404 错误（无法找到文件）。每当搜索蜘蛛来寻找并不存在的 robots.txt 文件时，服务器也将在日志中记录一条 404 错误，所以你应该在网站中添加一个 robots.txt。

2. 网站管理员必须使蜘蛛程序远离某些服务器上的目录——保证服务器性能。比如：大多数网站服务器都有程序储存在“cgi-bin”目录下，因此在 robots.txt 文件中加入“`Disallow: /cgi-bin`”是个好主意，这样能够避免将所有程序文件被蜘蛛索引，可以节省服务器资源。一般网站中不需要蜘蛛抓取的文件有：后台管理文件、程序脚本、附件、数据库文件、编码文件、样式表文件、模板文件、导航图片和背景图片等等。

下面是 VeryCMS 里的 robots.txt 文件：

```
User-agent: *  
  
Disallow: /admin/ 后台管理文件  
  
Disallow: /require/ 程序文件  
  
Disallow: /attachment/ 附件  
  
Disallow: /images/ 图片  
  
Disallow: /data/ 数据库文件  
  
Disallow: /template/ 模板文件  
  
Disallow: /css/ 样式表文件  
  
Disallow: /lang/ 编码文件  
  
Disallow: /script/ 脚本文件
```

3. 如果你的网站是动态网页，并且你为这些动态网页创建了静态副本，以供搜索蜘蛛更容易抓取。那么你需要在 robots.txt 文件里设置避免动态网页被蜘蛛索引，以保证这些网页不会被视作含重复内容。

4. robots.txt 文件里还可以直接包括在 sitemap 文件的链接。就像这样：

```
Sitemap: http://www.\*\*\*.com/sitemap.xml
```

目前对此表示支持的搜索引擎公司有 Google, Yahoo, Ask and MSN。而中文搜索引擎公司, 显然不在这个圈子内。这样做的好处就是, 站长不用到每个搜索引擎的站长工具或者相似的站长部分, 去提交自己的 sitemap 文件, 搜索引擎的蜘蛛自己就会抓取 robots.txt 文件, 读取其中的 sitemap 路径, 接着抓取其中相链接的网页。

**5. 合理使用 robots.txt 文件还能避免访问时出错。**比如, 不能让搜索者直接进入购物车页面。因为没有理由使购物车被收录, 所以你可以在 robots.txt 文件里设置来阻止搜索者直接进入购物车页面。

#### 1.4.4. robots.txt 用法举例:

**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 Baiduspider)**

```
User-agent: Baiduspider  
Disallow: /
```

**E. 只允许某个搜索引擎的访问 (下例中的 Baiduspider)**

```
User-agent: Baiduspider  
Disallow:  
User-agent: *  
Disallow: /
```

**F. 允许访问特定目录中的部分 url**

```
User-agent: *  
Allow: /cgi-bin/see  
Allow: /tmp/hi  
Allow: /~joe/look  
Disallow: /cgi-bin/  
Disallow: /tmp/  
Disallow: /~joe/
```

**G. 使用 "\*" 限制访问 url**

```
ser-agent: *
```

Disallow: /cgi-bin/\*.htm

禁止访问/cgi-bin/目录下的所有以".htm"为后缀的 URL(包含子目录)。

**H. 使用"\$"限制访问 url**

User-agent: \*

Allow: .htm\$

Disallow: /

仅允许访问以".htm"为后缀的 URL。

**I. 禁止访问网站中所有的动态页面**

User-agent: \*

Disallow: /\*?\*

**J. 禁止 Baiduspider 抓取网站上所有图片**

User-agent: Baiduspider

Disallow: .jpg\$

Disallow: .jpeg\$

Disallow: .gif\$

Disallow: .png\$

Disallow: .bmp\$

仅允许抓取网页，禁止抓取任何图片。

**K. 仅允许 Baiduspider 抓取网页和.gif 格式图片**

User-agent: Baiduspider

Allow: .gif\$

Disallow: .jpg\$

Disallow: .jpeg\$

Disallow: .png\$

Disallow: .bmp\$

### 1.4.5.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>

### 1.4.6.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**

**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**

**错误写成:**

```
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。**

**错误写成:**

```
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。

**正确的应该是:**

```
user-agent:GoogleBot
```

```
disallow:
```

**F. 语法中只有 Disallow, 没有 Allow**

**错误写成:**

```
User-agent: Baiduspider
```

```
Disallow: /john/
```

```
allow: /jane/
```

**G. 忘记了斜杠/**

**错误写成:**

```
User-agent: Baiduspider
```

```
Disallow: css
```

**正确的应该是:**

```
User-agent: Baiduspider
```

```
Disallow: /css/
```

## 第二章：主要搜索引擎介绍

### 第一节：Google

#### 2.1.1 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](http://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搜索获取更多信息请访问站长百科[www.zzbaike.com](http://www.zzbaike.com)

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。

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, 全球受众: 超过百分之五十的点击量来自美国境外。

### 2.1.2. Google 的 PageRank 技术

PageRank(佩奇等级)是 Google 能够超越其它搜索引擎的一个法宝。“佩奇等级”是 Google 算法的重要内容, 2001 年 9 月 Google 被授予美国专利, 专利人就是 Google 的创始人之一拉里.佩奇, 因此, PageRank 里面的 Pages 不是网页, 而是指佩奇, 即这个等级方法是以佩奇来命名的。

“佩奇等级”是着重考察网站的权威性, 即越有权威的网站越容易被其它网站主动链接, 被链接的越多, 就意味着其它网站投票越多, 这个就是所谓的“链接流行度”-----衡量多少人愿意将他们的网站和你的网站挂钩。

“佩奇等级”的初衷是非常理想化的, 他认为, 一个网站的好坏不是有网站自己吹出来的, 也不是网站自己作假做出来的, 而是大家评选出来的, 你链接一个网站, 那个网站就获得了一个投票。这样的评比结果很难被操纵, 网站的排名也就很难有作弊获得。

但是, 这个逻辑现在面临着一个挑战, 搜索引擎优化者为了提升一个网站的“佩奇等级”价值, 总是在试图获取最多的导入链接, 因为, 一个导入链接就相当于一份投票, 更多的网站主靠虚拟的链接来积累这个“投票”的作弊, 因此, Google 现在对于搜索引擎的排名并不像以前那样重视“佩奇等级”。

PageRank 有一个比较复杂的计算公式, 这里我就部列出来了, 这个公式对于我们来说, 一般情况下是用不到的, 只要我们了解这个概念就 OK 了。我们不鼓励大家刻意的去追求 PR, 因为决定排名的因素刻意有上百种, 但是, 我们建议网站设计者要充分认识佩奇等级在 Google 判断网站质量中的重要作用, 从设计前的考虑到这些方面, 如果在设计中考虑这些问题, 我们将在后面的章节中讲述。

### 2.1.3. Google 的 SiteMap 技术

前面我们也简单的介绍了 SiteMap 的相关知识，这里我们就来具体的讲讲 Google 的 SiteMap 技术。

#### 2.1.3.1 Google 的 SiteMap 技术是什么？

Google Sitemap 可以让网站建设者通过使用特定格式的 sitemap 文件，通知 Google 并指引 Google spider 收录相应网页。正确地使用 Google Sitemap，可以确保让 Google spider 不遗漏网站内的任何页面，及时地、连续地收录进 Google 的索引数据库。无论从哪个角度，网页被收录进索引数据库是网页出现在 SERP 的最基本条件，而 Google Sitemap 则提供了网页收录的捷径，这在某种意义上可以视之为 Google 的后门。

既然说是 Google 某种意义上的后门，那么，Google 一定会抓取 Sitemap 中的所有网址并为其编制索引吗？这个问题 Google 的官方也做出了相关的回答，内容如下：

我们不保证一定会抓取所有网址并为其编制索引。例如，我们不会抓取您的 Sitemap 中所含的图片网址或为其编制索引。但是，我们会使用 Sitemap 中的数据了解网站的结构，这样可以让我们改进抓取工具的日程，并在日后能更好地对网站进行抓取。大多数情况下，网站管理员会从 Sitemap 提交中受益，您决不会因此而受到处罚。

从 Google 的官方回答来看，Google 一定会抓取 Sitemap 中的所有网址并为其编制索引，而只是有利于其进行抓取，某种意义上来说，Google 的 SiteMap 对于提供网站搜索的位置有很大的帮助，这个帮助基本上有两个用途：

**A：Google 站点地图给你提供了一个途径来告诉 Google 你的网站有价值的信息。**

**B：你可以通过 Google 站点地图来了解 Google 是怎么看待你的网站的。**

#### 2.1.3.2 怎样设立 Google 站点地图

设立 Google 的站点地图，第一件事情就是要创建一个 Google 账户，这个账户可以是你的 Gmail 信箱的账户，或者是 Adwords, AdSense 的账户就可以了。


一般来说，有两种类型的 sitemap。第一种类型的 sitemap 是通常按部分列出您网站网页的 HTML 网页，用于帮助用户查找所需的信息。虽然此类 sitemap 能够帮助访问者和搜索引擎浏览网站，但是它不同于 XML Sitemap，XML Sitemap 可通过网站管理员工具提交给 Google。

#### A: 使用 RSS / Atom 供稿作为 Sitemap

Google 接受 RSS（真正简单的联合供稿）2.0 供稿和 Atom 1.0 供稿。如果您的博客有

RSS 或 Atom 供稿, 您可以将该供稿的网址作为 Sitemap 提交。大多数博客软件会为您创建供稿。如果您的网站没有供稿, 您可以使用 Feedburner 创建并发布供稿。

### **B:创建基于文本文件的 Sitemap**

一个 Sitemap 应包含一个您网站的网址(最多 50,000 个)列表。如果您的网站很大, 包含的网址超过 50,000 个, 则应创建多个 Sitemap, 并为它们提交一个 Sitemap 索引文件 。

您可以向 Google 提供纯文本文件, 其中每行包含一个网址。例如:

`http://www.example.com/file1.html`

`http://www.example.com/file2.html`

要获得最佳效果, 请按照下列指南操作:

1. 由于 Google 完全按照所提供的网址进行抓取, 您需要完整指定网址。
2. 每个文本文件最多可以包含 50,000 个网址。如果网站所包含的网址超过 50,000 个, 则可以将列表分割成多个文本文件, 然后分别添加每个文件。
3. 文本文件需使用 UTF-8 编码。在保存文件时您可指明此项(例如, 在记事本中, 此项会在"另存为"对话框中的"编码"菜单中列出)。
4. 文本文件不应包含网址列表以外的任何信息。
5. 文本文件不应包含任何页眉或页脚信息。
6. 您可以对文本文件进行任意命名。Google 建议将文件的扩展名定为 .txt, 以表明其为文本文件(例如 sitemap.txt)。

您应将文本文件上传到服务器, 通常传至希望搜索引擎抓取的最高级目录。创建此文件后, 您即可将其作为 Sitemap 提交。如果您对编制网络服务器的脚本或管理网络服务器不熟悉, 手动进行此过程是最简单的方法, 也不失为一种最好的方法。

如果您有多个网站, 可以创建一个或多个 Sitemap 将所有经过验证网站的网址包含在其中, 然后将这些 Sitemap 保存在一个位置, 以此来简化 Sitemap 的创建和提交过程。所有网站都必须使用网站管理员工具进行验证。

### **C:根据 Sitemap 协议创建 Sitemap**

Google 可以接受多种格式的 Sitemap, 但是我们建议您根据 Sitemap 协议创建 Sitemap, 因为这种格式的文件也可以提交至作为 sitemaps.org 成员的其他搜索引擎, 如 MSN 和 Yahoo!。

您可以通过以下多种方式创建 Sitemap:

1. 根据 Sitemap 协议[手动创建 Sitemap](#)
2. [使用 Sitemap 生成器](#)。如果您可以访问网络服务器并且服务器装有 Python，则可以利用我们的脚本来创建采用 Sitemap 协议的 Sitemap。[Google Sitemap 生成器](#)是一个 Python 脚本，可以利用 Sitemap 协议为您的网站创建 Sitemap。此脚本可以通过网址列表、网络服务器目录或通过访问日志创建 Sitemap。
3. 使用第三方工具。很多第三方提供了可用于创建有效的 Sitemap 的工具。

### 2.1.3.3. 向 Google 添加网站

加入 Google 搜索结果是不需要付费的，并且非常容易执行；您甚至不需要将网站提交给 Google。Google 是一个完全自动化的搜索引擎，它使用称为“信息采集软件”的软件定期抓取网络并找到可添加到 Google 索引的网站。事实上，列入搜索结果中的绝大多数网站都不是人工提交的，而是信息采集软件抓取网络时找到并自动添加的。

但是，如果您的网站提供专门的产品、内容或服务，则您可以把它们投放在[Google 网络搜索](#)和其他 Google 服务上，以便在全世界范围内进行传播。有关更多信息，请访问[Google 内容中心](#)。

要判断您的网站当前是否包含在 Google 索引中，请对您的网址执行"site:"搜索。例如，[索 \[site:google.com\]](#) 会返回以下结果：<http://www.google.com/search?q=site%3Agoogle.com>。

虽然 Google 可抓取超过数十亿个网页，但遗漏少数网站也在所难免。造成 Spider 遗漏网站的常见原因如下：

- \* 网站未通过多个链接牢靠地连接到网络上的其他网站。
- \* 网站在 Google 最近一次抓取后才启动。
- \* [网站的设计](#)让 Google 很难有效抓取其中的内容。
- \* 网站在我们抓取时暂时不可用，或我们在抓取时收到错误消息。您可以使用 Google 网站管理员工具来查看当我们[试图抓取您的网站](#)时是否收到了错误。

公平准确地展示互联网内容是我们的宗旨。为实现此目标，我们提供了有关构建便于抓取的网站的[指南](#)和[提示](#)。虽然我们无法保证信息采集软件能找到某个特定网站，但遵循以下指南应该可以增加网站在我们的搜索结果中显示的机会。

考虑[创建并提交关于您的网页的详细 Sitemap](#)。[Sitemap](#) 是您向 Google 索引提交所有网址并详细了解您的网页在 Google 显示情况的一条捷径。通过 Sitemap，我们可以随时自动获得有关您当前网页及其更新的所有信息。请注意，提交 Sitemap 并不能保证您网站上的所有网页都会被抓取或包含在我们的搜索结果中。



## 2.1.4. 网站管理员工具

### 2.1.4.1. 网站管理员工具介绍

Google 网站管理员工具提供免费的简易方法,让您的网站更容易被 Google 抓取。该工具可显示 Google 对您网站的意见,帮助您诊断问题并让您与我们共享信息来提高您网站的展示率。

使用 Google 网站管理员工具,具有以下优点:

#### I. 获取 Google 对您网站的意见,并诊断出可能的问题

提高您网站在 Google 上展示率的第一步,就是先了解我们的漫游器如何抓取您的网站并编制索引。

1. [抓取信息](#): 您可以确定我们是否已经访问过您的网站,并查看 Googlebot 上次访问的时间。您还可以查看我们在抓取时遇到问题的网址,并了解无法抓取的原因。这样您便可以修正所有问题,让我们能够为您的所有网页编制索引。
2. [Robots.txt 文件验证](#): 查看我们在处理您的文件时是否遇到问题,并先对文件的修改进行测试,然后在服务器上修改该文件。
3. [网站内容](#): 查看您网站的热门内容以及其他网站用来链接到您网站的字词。

#### II. 查看您网站的运作效果

第二步是了解如何提高网站点击量。

1. [热门查询](#): 查找可提高您网站点击量的热门查询,以及您的网站可出现在热门搜索结果中的哪个部分。这可让您了解用户如何找到您的网站。
2. [索引信息](#): 了解编制网站索引的方式以及您的哪些网页已编制到索引中。如果我们发现您的网站上有违规行为,会给您机会修正问题,并要求您的网站重新加入。

#### III. 与 Google 共享关于您网站的信息

因为没人比您更了解您的网站,因此您也可以与 Google 共享此信息并提高您网站的易抓取程度。

1. [提交 sitemap 文件](#): 提交 Sitemap 文件来告诉我们关于您网页的一切信息,帮助我们了解哪些网页对您最重要,以及那些网页更改的频率。



2. [指定首选域](#)：请告诉我们编制您网站索引时要使用的网址，我们会尽量为您首选的网址编制索引。

## 第二节：百度

### 2.2.1.关于百度



百度在国内，也可谓妇孺皆知，只要稍微了解点电脑的，就一定会知道百度。下面我们来看看百度的成长历程：

百度，2000 年 1 月创立于北京中关村，是全球最大的中文网站、最大的中文搜索引擎。

2000 年 1 月 1 日，公司创始人李彦宏、徐勇从美国硅谷回国，创建了百度。创立之初，百度便将“让人们更便捷地获取信息”作为自己的使命，并为此始终如一、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，从后台技术提供者转为面向公众独立提供搜索服务，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。

2003 年 12 月，开创性地推出贴吧，开搜索社区化之先河。

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。

2008 年 1 月 23 日，百度日本公司正式运营，百度全面启动国际化战略。

九年来，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。

### 2.2.2.百度的 SiteMap

准确的说，叫它 baidu sitemap 不太准确，而且会有朋友问，百度有类似于 google 的 xml 格式 sitemap 吗？答案是有，但是它又不完全等同于 sitemap。根据百度官方的描述，我们应该管它叫做《互联网新闻开放协议》。但是我还是喜欢叫它 baidu sitemap，我觉得这个名称对站长来说更亲切一些：)

其实按照百度官方的解释，我们应该称之为《互联网新闻开放协议》。但 Google 的 Sitemap 叫习惯了，所以一时不改不了口，暂且就叫它百度 Sitemap 吧。

从官方的公布的文档名称，我们清楚的知道这个 XML 文档只针对的是站内的新闻，似乎对我们没有什么价值。这点就反映在我们对“新闻”一词的理解上，我认为百度的“新闻”一词应该包括：新闻、文章、文档资料等，而这些都是很多网站所必备的东西。而且蜘蛛是跟据链接来爬行的，所以当我们的网站通过互联网开放协议的审查后，百度就会来抓取这个 xml 文档里的信息，而且还会顺着链接找到其它的内容，百度也会连这些内容一并抓取了。

有这些好处，为什么有很多站长到现在还不知道有这个东西的存在呢，可能主要的原因是百度的 Sitemap 在提交会由百度审核，目前国内许多的站长的抄袭严重，所以提交审核自然有许多不会被通过了，自然打消人们的积极性。

下面来说说百度 Sitemap 中 XML 标签所代表的含义的：

\*< document>——标记整个 XML 文件内容的开始和结束。

\*< website>——站点地址。

\*< webmaster>——负责人员的 Email。当有必要时，我们通过这个地址与您联系。

\*< updateperi>——更新周期，以分钟为单位。搜索引擎将遵照此周期访问该页面，使页面上的新闻更及时地出现在百度新闻中。

\*< item>——标记每篇新闻信息的开始和结束。标记内为单篇新闻信息，不包括新闻专题。

\*< title>——新闻标题。

\*< link>——新闻 url 地址，与单篇新闻一一对应；若分页的新闻存在多个 URL，相当于多篇新闻。

< description>——新闻内容简介。

\*< text>——完整的新闻正文（仅包含正文文字，不包含 HTML 语言等其它字符）。此项的目的是使该篇新闻更多、更准地出现在搜索结果中。

\*< image>——新闻正文内相关图片，采用绝对地址。若该篇新闻无相关图片，可以为空；若含有多张图片，请重复使用该标签。此项的目的是使该篇新闻的相关图片展现在搜索结果中。

< headlineimg>——为有可能成为头条的新闻制作的头条图，采用绝对地址。

< keywords>——反映新闻主题内容的一个或多个关键词，关键词之间以空格隔开。此项仅作参考，检索结果不完全依赖于此标签中的内容。

< category>——新闻分类，可以遵循网站自身的分类体系，最好采用一级分类。

< author>——新闻作者，可以为机构或个人。

< source>——新闻来源，即原创媒体或其它机构。

\*< pubdate>——新闻发布时间，与该篇新闻 HTML 页面上的发布时间保持一致。请精确到分钟；若您网站的发布时间未记录小时分钟，提供年月日即可。

不知道大家注意到没有，在百度的 Sitemap 中所有的新闻正文内容全都放在 Text 标签内，这样做虽然能让百度有效的收录所有的新闻正文内容，但同时也存在风险，如果 XML 文件地址被泄露的话，那网站内容就会被轻意的读取、复制。

是否应该使用 baidu sitemap，相信大家心里应该有数，每个网站都有自己的算盘。如果觉得自己的网站不错，以我建议在做搜索引擎优化的时候还是做一做 baidu sitemap。

### 2.2.3. 百度论坛的开发协议

开放协议概述:

使用此开放协议, 将会为您的网站带去更多流量!

《互联网论坛收录开放协议》是百度网页搜索制定的论坛内容收录标准, 论坛网站可将论坛内发布的帖子制作成遵循此开放协议的 XML 格式的网页供搜索引擎索引, 将论坛发布的帖子主动、及时地告知百度搜索引擎。

采用了《互联网论坛收录开放协议》, 就相当于论坛中的帖子被搜索引擎订阅, 通过百度--全球最大的中文搜索引擎这个平台, 网民将有可能在更大范围内更高频率地访问到您网站论坛中的帖子, 进而为您的网站带来潜在的流量。

开放协议内容

下图为遵照《互联网论坛收录开放协议》制作的 XML 格式的网页, 网页上按照标准格式列出了网站论坛发布的帖子的相关信息。

```
<?xml version="1.0" encoding="GB2312" ?>
- <document xmlns:bbs="http://www.baidu.com/search/bbs_sitemap.xsd">
  <webSite>post.baidu.com</webSite>
  <webMaster>webmaster@baidu.com</webMaster>
  <updatePeri>12</updatePeri>
  <updatetime>2007-05-16 19:00:00</updatetime>
  <version>百度贴吧</version>
  - <item>
    <link>http://post.baidu.com/f?kz=176459170</link>
    <title>百度有奖调查活动, 快来参加~</title>
    <pubDate>2007-02-26 16:36:00</pubDate>
    <bbs:lastDate>2007-05-16 14:31:00</bbs:lastDate>
    <bbs:reply>48</bbs:reply>
    <bbs:hit>6446</bbs:hit>
    <bbs:mainLen>447</bbs:mainLen>
    <bbs:boardid>百度</bbs:boardid>
    <bbs:pick>1</bbs:pick>
  </item>
  <item>.....</item>
  .....
</document>
```

**XML 标签说明:** 其中带星号标记的为必选项, 未带星号标记为可选项。

\***<document>**——标记整个 XML 文件内容的开始和结束。后面的" **xmlns:bbs = "http://www.baidu.com/search/bbs\_sitemap.xsd"**是该 xml 文件遵守百度《互联网论坛收录开放协议》的声明, 如没有该段声明, xml 文件不支持下面的**<bbs:reply>**等语法。

\***<webSite>**——站点地址。

\***<webMaster>**——负责人员的 Email。当有必要时, 我们通过这个地址与您联系。

\***<updatePeri>**——sitemap 更新周期, 以小时为单位。搜索引擎将遵照此周期访问该页面, 使页面上的内容更及时地被百度 spider 发现。

\***<updatetime>**——该 xml 文件最近一次的更新时间。

\***<version>**——论坛所使用程序的版本。

\***<item>**——标记每个帖子信息的开始和结束，标记内为单个帖子信息。

\***<link>**——帖子的 url 地址，即帖子第一页的地址，请尽量去除 url 中不必要的参数。

**<title>**——帖子标题。

**<pubDate>**——帖子主贴的发布时间。

**<bbs:lastDate>**——帖子最后一条回复的发布时间。

**<bbs:reply>**——帖子总的回复数，不包括主贴。

**<bbs:hit>**——帖子总的查看次数。

**<bbs:mainLen>**——帖子主贴的长度，单位字节。

**<bbs:boardid>**——帖子所在版面，提供版面的 id。

**<bbs:pick>**——是否设为精华，精华标为 1，非精华标为 0。

推荐时间格式：年月日小时分钟秒

如：2005-11-09 10:37 | 2005/11/09 10:37:00 | 2005.11.09 10:37:00 |

2005 年 11 月 09 日 10 时 37 分 00 秒 | Fri, 09 Nov 2005 10:37:00 GMT

## 开放协议使用

使用之前，您需要了解以下几点：

- 无论您的网站已经被百度 spider 抓取，还是尚未被百度 spider 抓取，都可以使用此开放协议。
- 您使用开放协议提供的内容应全部符合下面的《论坛内容收录标准》。
- 《互联网论坛收录开放协议》只是对百度 spider 原有收录方式的一种协助和有益补充，而不是取代。

## 论坛内容收录标准：

百度希望论坛内容丰富，帖子质量较高，版面氛围和谐。只要是具有大量有价值的内容并能及时更新的正规、合法的网站，且网站服务器稳定、高速，就符合了百度论坛内容收录的基本原则。

百度 spider 收录使用任何程序架设的论坛，不过您应对提供的全部内容承担一切法律责任，保证您提供内容的真实性、合法性，并不得侵犯任何第三方的权益。

XML 文件的大小最大为 10M，大于 10M 的文件百度 spider 将不能保证正确获取。

对于通过遵守《互联网论坛收录开放协议》的 xml 文件发现的内容，百度 spider 将在进行质量判断后展现给用户，即被发现的内容并不一定最终会展现在百度网页搜索的结果中。

下面就开始吧！

### 第一步：创建 XML 文件

请按照上面公布的[开放协议内容](#)制作 XML 文件。

**其它说明:**

- 支持的编码格式有 GB18030、UTF-8、GB2312、BIG5 等, 推荐使用 GB2312 和 UTF-8。注意: XML 文件头部的 encoding 值必须与 XML 文件的实际编码类型一致。
- 请保持每个 XML 文件处于按更新周期持续自动更新的状态。
- XML 标签内容中的一些特殊字符需要转换为 XML 定义的转义字符, 否则将发生错误导致搜索引擎不能获取页面上的帖子, 具体参见下表。

字符		转义后的字符	
		HTML 字符	字符编码
和(and)	&	&amp;	&#38;
单引号	'	&apos;	&#39;
双引号	"	&quot;	&#34;
大于号	>	&gt;	&#62;
小于号	<	&lt;	&#60;

- ?? 转义字符中的"&"无需再转。

**第二步: 验证 XML 文件**

下面的地址提供了帮助您验证 XML 文件结构的多种工具:

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

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

通过验证的 XML 文件能够使您提供的信息更加标准, 确保您发布的视频信息不被搜索引擎遗漏。

**第三步: 使用 XML 文件**

请将 XML 文件命名为 `sitemap_baidu.xml` 上传到您网站根目录下, 并保证文件所在的 url 地址能够被百度 spider 正常访问。例如, 网站为 `post.baidu.com`, 则将 xml 文件上传至 `post.baidu.com/sitemap_baidu.xml`

当百度 spider 发现了 xml 文件后, 会根据上面提供的参数自动对 xml 文件进行更新, 并抓取高质量的内容。

**注意:** 百度 spider 不保证一定能收录您提交的全部内容。



## 第三节：YaHoo（雅虎）

### 2.3.1. 关于雅虎

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对关键词查询的需要，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。



### 2.3.2.YaHoo 中国

雅虎([www.yahoo.com](http://www.yahoo.com))是全球第一门户搜索网站,业务遍及 24 个国家和地区,为全球超过 5 亿的独立用户提供多元化的网络服务。

1999 年 9 月,雅虎中国网站开通。2005 年 8 月,中国雅虎由阿里巴巴集团全资收购。中国雅虎([www.yahoo.com.cn](http://www.yahoo.com.cn))开创性地将全球领先的互联网技术与中国本地运营相结合,并一直致力于以创新、人性、全面的网络应用,为亿万中文用户带来最大价值的生活体验,成为中国互联网的“生活引擎”。

口碑网([www.koubei.com](http://www.koubei.com))是中国最大的生活搜索引擎,涵盖餐饮娱乐、租房、买房、工作、旅游等生活消费信息,信息遍布所有大中城市。自 2004 年 6 月成立以来,一直致力于做百姓的生活好向导,2006 年 10 月,阿里巴巴集团正式战略投资口碑网。本地生活搜索是口碑网提供给用户的重要服务功能,口碑网两大行业频道餐饮休闲、房产交易都在同类网站中居首位。

2008 年 6 月,中国雅虎和口碑网整合,成立雅虎口碑网,正式进军生活服务领域。以全网搜索为基础,为生活服务消费者打造出一个海量、方便、可信的生活服务平台-雅虎口碑网。网站一经推出,就确立了在同行业的领先地位。

### 2.3.3.中文搜索引擎列表

名称	网址	所用技术
谷歌	<a href="http://www.google.cn">http://www.google.cn</a>	Google
百度	<a href="http://www.baidu.com">http://www.baidu.com</a>	百度
雅虎中文	<a href="http://www.yahoo.cn">http://www.yahoo.cn</a>	YaHoo
易搜	<a href="http://yisou.com">http://yisou.com</a>	YaHoo
有道	<a href="http://www.youdao.com">http://www.youdao.com</a>	网易
SOSO	<a href="http://www.soso.com">http://www.soso.com</a>	Google
爱问	<a href="http://iask.com">http://iask.com</a>	Google
搜狗	<a href="http://www.sogou.com">http://www.sogou.com</a>	搜狗

## 第三章：网站的基本查询方式

### 第一节：域名查询

#### 3.1.1.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](http://www.whois.com)

中文域名信息查询：<http://whois.zzbaike.com/>

一般域名注册商网站也提供此类查询功能。

### 3.1.2. 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>

### 3.1.3. 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.zzbaike.com`，返回以下内容：

```
Pinging www.zzbaike.com [74.86.118.181] with 32 bytes of data:

Reply from 74.86.118.181: bytes=32 time=234ms TTL=46
Reply from 74.86.118.181: bytes=32 time=240ms TTL=46
Reply from 74.86.118.181: bytes=32 time=236ms TTL=46
Reply from 74.86.118.181: bytes=32 time=238ms TTL=46

Ping statistics for 74.86.118.181:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
    Approximate round trip times in milli-seconds:
        Minimum = 234ms, Maximum = 240ms, Average = 237ms
```

其中 74.86.118.181 就是网站的IP地址，time=234ms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，受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。

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。

## 第二节：收录查询

### 3.2.1. 网站是否被收录

直接在搜索引擎搜索域名。比如在百度搜索 [www.zzbaike.com](http://www.zzbaike.com)

[站长百科-首页](#)

设为首页 添加收藏 主页 站长百科 站长论坛 站长下载 主机侦探 网站导航 Alexa中文专题站WordPress中文专题站WordPress文档翻译计划站长网址 百科导航  
WordPressDiscuz!DedeCMS帝国CMSPHPWind 论坛板块 Blog程序交流区CMS程序交流区论坛程序交流区...

[www.zzbaike.com/](http://www.zzbaike.com/) 22K 2009-3-12 - [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。

您可以直接访问: [zbaike.com](http://zbaike.com)

**百度建议您:**

- 如果网址正确,请点击上面的链接
- 如果您在寻找包含"zbaike.com"的所有网页,请点击[这里查询](#)

在这里需要注意的是,在解析域名时,使用的那个域名,就查询解析的那个,例如:你解析的域名是 [zzbaike.com](http://zzbaike.com),而不是 [www.zzbaike.com](http://www.zzbaike.com) 时,那么,你查询 [www.zzbaike.com](http://www.zzbaike.com),当然就不会有内容显示了,记住要查询解析的那个域名。

### 3.2.2. 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:www.zzbaike.com** ,找到相关网页约2,820篇。

查看网址中包含[www.zzbaike.com](http://www.zzbaike.com)的网页收录数,例如[bbs.zzbaike.com](http://bbs.zzbaike.com)的网页都会被统计在内。

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

网站的反向链接比较重要，前面我们也说过，反向链接对于Google的佩奇等级也是有一定的影响的。

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.zzbaike.com**

如果你想要在多个搜索引擎里查询这些信息，每次都得重新打开一个搜索引擎窗口，重新输入一次命令，比较的麻烦，你可以使用下面介绍的工具：

<http://linksincount.zzbaike.com/>

### 第四节：网站 PR 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，如果该网页已经被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 Explore 浏览器，也支持 Firefox 浏览器了。

当然，如果感觉下载、安装这些插件比较麻烦，你可以使用在线的 PageRank 查询工具，也比较的方便：<http://pr.zzbaike.com/>

### 第五节：关键词排名查询

如果你想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：

<http://keywordsrank.zzbaike.com/>

## 第六节：ALEXA 查询

### 3.6.1. 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。

### 3.6.2.Alexa 的主要数据

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：访问用户数 = （访问人数/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。

### 3.6.3.Alexa 的作用

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，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。

**对于非商业的网站来说，排名的作用为：**

- a、满足虚荣心
- b、如有人要在你的网站投放广告，alexa排名是收费的重要依据。（通常排名比较高，才会有人找你投入小量广告，就当挣些小钱吧。）

**对于商业网站而言，排名比较重要：**

- a、是广告收费的依据之一。
- b、评估网站价值的依据之一（如果以后要转卖域名或网站）
- c、炒作新闻的一个报道点。
- d、严密监测竞争对手的重要方法。（对于自己的竞争对手网站，你是否每天去看他的流量和排名变化。除alexa之外别无他法，换言之：只有alexa才能告诉你答案。当某竞争对手的流量出现突变时，必是其投入了广告或在某处加了搜索，自己留心去找，肯定能找到。）

### 3.6.4. 提高 Alexa 排名

当知道 Alexa 是什么，也知道 Alexa 的作用之后，接下来的重点，就是谈谈如何提高 Alexa 排名。

这里所讲的绝不会是某种作弊的方法。比如你用搜索来搜一下“Alexa 刷”，你就会看到许多通过刷页面来提高 Alexa，或者是互刷联盟，或者是工具来刷，再或者通过模拟 IP 来提高。这些都无法反映你的真实流量。这里所说的方法，都是合理合法的，既不会吹出大泡泡，也不会受到搜索引擎的惩罚。

1) 安装 Alexa 工具条, 或者安装 Firefox 下的 SearchStatus 插件。并且设置自己的 Blog 作为默认首页, 这样每次你打开浏览器的时候, Alexa 统计就开始工作了。

2) 把 Alexa 排名统计的 widget 放在 Blog 上, 当它被点击的时候, 也相当于 Alexa 统计到一次 PV, 即使那个用户没有安装 Alexa 工具条。

3) 如果可以的话, 在你的公司或者办公室里的电脑里都安装上 Alexa 工具条。

4) 在站长论坛里发贴子的时候留下你的链接, 一般站长安装 Alexa 的较多, 当他们访问你的 Blog 的时候, 当然会有帮助。

5) 给你的 Blog 添加一个 Alexa 分类, 把有关 Alexa 的文章都放在里面, 这样站长有更大的机会来到你的 Blog, 你的排名当然就容易上升。

6) 找到你的 Blog 上最受欢迎的文章, 把 Alexa 的链接添加上, 或者使用第五点中所说的重定向链接到你自己的文章。

7) 提供一些站长工具, 这样站长就会经常来造访, 并且提高你的 Alexa 排名, 比如 SEO 在线工具中的 Google PageRank Checker 就是一个很好的例子。

想要在网站上即时显示你的 Alexa 排名吗? Alexa 排名动态图标帮你实现! 生成的图标能够即时更新, 及时反应贵站最新的 Alexa 排名, 要生成你的 Alexa 排名动态图标, 只需以下两步:

1. 输入域名, 选择喜欢的图标背景
2. 点击“提交”, 然后将代码复制到你的网站上, 就 OK 了

一切就是这么简单!

网址: <http://alexaicon.zzbaike.com/>



## 第四章：搜索引擎优化基础

### 第一节：搜索引擎优化的定义和价值

#### 4.1.1 搜索引擎优化的定义

##### WIKI 定义:

(Search Engine Optimization, 简称 SEO)

搜索引擎优化即 Search Engine Optimization, 用英文描述是 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。与之相关的搜索知识还有 Search Engine Marketing (搜索引擎营销), Search Engine Positioning (搜索引擎定位)、Search Engine Ranking (搜索引擎排名) 等。

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。

##### 搜索引擎优化基本定义:

**搜索引擎优化**是针对搜索引擎对网页的检索特点, 让网站建设各项基本要素适合搜索引擎的检索原则, 从而获得搜索引擎收录尽可能多的网页, 并在搜索引擎自然检索结果中排名靠前, 最终达到网站推广的目的。

搜索引擎优化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, 来对网页内容进行相关的优化, 使其符合用户浏览习惯, 在不损害用户体验的情况下提高搜索引擎排名, 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。所谓“针对搜索引擎优化处理”, 是为了要让网站更容易被搜索引擎接受。搜索引擎会将网站彼此间的内容做一些相关性的资料比对, 然后再由浏览器将这些内容以最快速且接近最完整的方式, 呈现给搜寻者。由于不少研究发现, 搜索引擎的用户往往只会留意搜索结果最开 首的几项条目, 所以不少商业网站都希望透过各种形式来干扰搜索引擎的排序。当中尤以各种依靠广告维生的网站为甚。目前 SEO 技术被很多目光短浅的人, 用一些 SEO 作弊的不正当的手段, 牺牲用户体验, 一味迎合搜索引擎的缺陷, 来提高排名, 这种 SEO 方法是不可取的。

在国外,SEO 开展较早,那些专门从事 SEO 的技术人员被 Google 称之为“Search Engine Optimizers”,简称 SEOs。由于 Google 是目前世界最大搜索引擎提供商,所以 Google 也成为了全世界 SEOs 的主要研究对象,为此 Google 官方网站专门有一页介绍 SEO,并表明 Google 对 SEO 的态度。

搜索引擎优化对于任何一家网站来说,要想在网站推广中取得成功,搜索引擎优化都是至为关键的一项任务。同时,随着搜索引擎不断变换它们的排名算法规则,每次算法上的改变都会让一些排名很好的网站在一夜之间名落孙山,而失去排名的直接后果就是失去了网站固有的可观访问量。所以每次搜索引擎算法的改变都会在网站之中引起不小的骚动和焦虑。我们可以说,搜索引擎优化也成了一个愈来愈复杂的任务。

#### 4.1.2.搜索引擎优化的价值

##### 1、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。

##### 2、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。

##### 3、自然排名结果,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。

## 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。

## 第二节：搜索引擎优化究竟为了什么

### 4.2.1.SEO 体现网站效果

搜索引擎的目标是获得网页在搜索引擎结果中的前十名，然而，我们必须明白，获得这个排名并不等于搜索引擎优化取得了成功，获得排名只是检验优化效果的显示，而优化效果的体现则是这些排名是否能真正地带来我们期盼的访客。排名在高，带来的访问者不能成为你的客户，那么，这就是“竹篮打水一场空”，因此，检验搜索引擎优化是否成功的标准不是看网页在某些关键词的搜索结果中是否排到首页。

### 4.2.2.SEO 的终极目标和实现目标

搜索引擎优化（SEO）的终极目标是从根本上改善网站的结构和内容，从而提供给搜索引擎一个非常宝贵的信息源，对于经过 SEO 获得的收益，则是我们的现实目标，现实目标的实现是取得成功的初期目标，而终极目标就是将这个成功长久保持下去的永久目标。

SEO 对网站的处理，基本目的是优化网站的内部结构、增加信息的承载量、扩充网站的数据流通、提升网站的表现，将一个网站应该发挥的作用最大化，并且随之将网站收益和投资之比大幅提高。

只有无限接近永久目标，SEO 才能完成使命，因此，SEO 工程是一定是长期的工程，而不是寄希望于短期的效益，因为，远期的目标一旦实现，我们的网站就会真正成为企业市场推广中举足轻重的力量。

## 第三节：白帽 SEO 和黑帽 SEO

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，到其他留言本或博客留一些言（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）。

## 第四节：关于搜索引擎优化的一些观点

### 4.4.1. 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。

### 4.4.2. 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。

再者，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。

### 4.4.3. 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。



#### 4.4.4.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。

#### 4.4.5.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；

再者，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。

对这些大型网站不要盲目的攀比、也不要盲从，要结合自身情况，实事求是的使用有利于自己网站的优化方式。

#### 4.4.6.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。如果你要的是“保证排名”的效果，也许你应该考虑竞价排名，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。

#### 4.4.7.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。

#### 4.4.8.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。

## 第五章：网站设计的优化操作

### 第一节：域名对 SEO 的影响

#### 5.1.1.域名的基本常识

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。

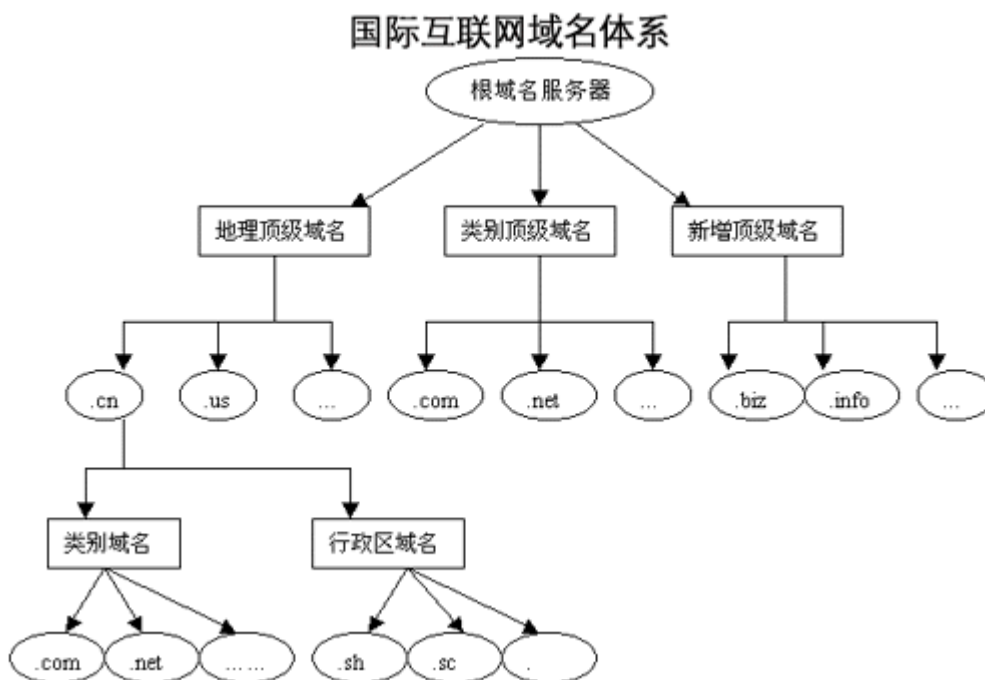
为了便于大家进一步了解域名的实质,有必要在这里谈谈域名的体系结构。从[www.zzbaike.com](http://www.zzbaike.com)这个域名来看,它是由几个不同的部分组成的,这几个部分彼此之间具有层次关系。其中最后的.com是域名的第一层,.zzbaike是真正的域名,处在第二层,当然还

可以有第三层，如：[bbs.zzbaike.com](http://bbs.zzbaike.com)，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.com 叫做类别顶级域名。

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。



### 5.1.2 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。

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。

## 第二节: 主机对 SEO 的影响

### 5.2.1. 共享主机、独立主机的区别?

有些朋友对于主机的概念,不是很清楚,“共享主机、独立主机”一个名词让你感觉有点迷茫,所以,为了解决这个问题,看看这篇文章吧!

**共享主机(Share Hosting)**就是把一台主机服务器划分成多个“虚拟”主机,并与其它数百甚至数千个网站共享一台主机服务器(web server)的资源,如CPU时间、硬盘空间等等。也因为如此,共享主机是市场上最便宜的主机选择了,由于许多网站共享一台主机服务

器，所以，当超过某个数量时，网站的速度就会急剧下降。

有信誉的主机商都会经常观察并确保一台主机服务器，不会过量负荷，造成网站的访问速度变慢的情况出现，由于主机低廉和执行速度不错，所以，对大多数网站来说，共享主机是最佳的首选。

**独立主机 (Dedicated Server)** 刚好与共享自己相反，独立主机完全有你一个人使用，所以速度快，那是肯定的，而且，你也有很大的灵活性来安装任何网络软件或架设多个网站，但是，独立主机的费用不便宜，月费都比较高，所以，通常只有非常高流量的网站，才需要独立主机。

**VPS 主机 (Virtual Private Server/VPS)** 是一种通过技术手段，把一台实体主机分割为数个单位的虚拟化专属主机，从形式上，你可以把它看成，介于独立主机和共享主机之间，但在执行上，它犹如独立主机一样具有灵活性，所以执行速度方面肯定比共享主机好的多，价格也要比独立主机便宜。

在购买主机时，要选择适合自己的主机？建议站长先从共享主机 (share hosting) 开始，等网站的流量增加了，出现速度缓慢的现象后，在考虑升级虚拟主机。

### 5.2.2. 主机服务商选择

主机对网站的搜索引擎排名影响很大。在寻找主机提供商的时候，千万不要只考虑价格因素。如果你多次向 Google 提交网站，但 Google 却一直没有对其进行索引，则有可能是网站所在的服务器出了问题。因此选择服务器时要特别注意以下几点：

- A. **避免使用免费主机：**由于免费主机里面经常会出现 Spammers、镜像网站、桥页等“搜索引擎垃圾”，很多搜索引擎都不愿意索引免费主机上的网站；同时搜索引擎目录如 Yahoo 也很难收录来自免费主机的商业站点。此外，免费主机的服务很难保证，常常服务器超载，速度奇慢，宕机频繁，甚至关闭服务，这都会直接影响网站排名。
- B. **选择有信誉的主机提供商：**搜索引擎的索引程序定期或不定期地来访问收录的网站。对经常更新的网站，Google 的 Googlebot 漫游器——俗称“蜘蛛”，一般每天都会小幅度地“爬”一下，一个月进行一次大的索引。如果蜘蛛在索引时出现网页打不开或下载速度缓慢，则放弃索引。所以你的网站在任何时候都要正常工作，以博取蜘蛛的欢心。这就要求网站存放的主机系统提供快速、稳定、安全的

服务, 最好有 24 小时客服支持和维护, 保证你的网 站快速、可靠、稳定地运行。

虚拟主机的选择还是非常重要的, 国内的虚拟主机商, 像[万网](#)、[新网](#), 不过感觉有点贵, 国外的也有不错的虚拟主机, 笔者在建站以来, 使用的都是国外主机, 对国外主机也算有点见解, 这里就部详述了, 如果对国外主机有兴趣, 你可以到这里看看 ([点击进入](#))

### 5.2.3.主机服务商转移:

如果你想转移网站的主机服务, 要在取消原来的服务提供商之前, 先找好新的供应商。先在新的服务商处开个帐号, 确保所有的文件都正确无误地传到新服务器上。网站在新服务器上运行无碍后再联系你的原供应商, 告知对方将终止合作, 但确保他们不会马上关你的网站。再联系你的域名注册商 (可能是原主机提供商), 变更域名服务器(DNS)。

域名服务器的功能就是把你的域名解析成计算机能够识别的 IP 地址。把旧的域名服务器改成新的, 这个过程到生效时间大概 24-48 小时之内, 少数情况下可能要 72 小时。这期间, 如果一个搜索引擎刚好来抓你的网页, 要么抓到旧的 IP 地址上的网站, 要么抓到新的域名解析生效后的网站, 总之都不会出现网站打不开的空档。

## 第三节：对搜索引擎友好的设计

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。

### 5.3.1. 目录结构

清晰简短的目录结构和规范的命名不仅有利于用户体验和网址传播, 更是搜索引擎友好的体现。

#### 1.目录访问层次:

即通过几层能够访问到最终页面, Google 最好为三层。

以 <http://www.zzbaike.com/> 为例。

<http://www.zzbaike.com/> 首页 为 第一层, <http://www.zzbaike.com/wiki> 为 第二层,

<http://www.zzbaike.com/wik/wordpress> 为第三层。

这样的结果便于搜索引擎索引, <http://www.zzbaike.com/wiki> 这个页面在百度、 google 里面都有收录。当网站目录层次大于三层, 这个时候就要使用二级域名的方式扩大级数, 因为二级域名算独立网站, 目录层次从当前二级域名算起。

例如: 站长百科论坛, [bbs.zzbaike.com](http://bbs.zzbaike.com)

## 2.目录和文件命名:

根据关键字无所不在的原则, 可以在目录名称和文件名称中使用到关键词。但如果是关键词 组, 则需要用分隔符分开。我们常用连字符 “-” 和下划线 “\_” 进行分隔, URL 中还经常出现空格码 “%20”。因此, 如果以 “中国制造” 作文件名, 就可能出现以下三种分隔形式:

`made-in-china.htm`

`made_in_china.htm`

`made%20in%20china.htm`

连在一起之后, 关键词就失去了意义。但事实上, 至少在目前 Google 并不认同 “\_” 为分隔符。对 Google 来说, `made-in-china` 和 `made%20in%20china` 都等于 `made in china`, 但 `made_in_china` 就被读成了 `madeinchina`,

## 5.3.2.URL

**URL** 是统一资源定位, 即每个网页的网址、路径。网站文件的目录结构直接体现于 URL。

### 1.绝对 URL 和相对 URL:

绝对路径 URL: 使用完整的 URL 指向指定网页。

相对路径 URL: 使用自动的 URL 指向指定网页。

例如:对指定网页 [www.yoursite.com/page1/index.html](http://www.yoursite.com/page1/index.html), [www.yoursite.com/page1/index.html](http://www.yoursite.com/page1/index.html) 为绝对路径, `/page1/index.html` 则为相对路径。对于相对路径来说, 浏览器会自动在该链接前加上 “[www.yoursite.com](http://www.yoursite.com)”

### 1. 动态 URL:

目前很多网站都有数据库驱动生成的 URL, 即动态 URL, 往往表现为在 URL 中出现 “?”、“=”、“%”, 以及 “&”、“\$” 等字符。动态 URL 极不利于搜索引擎抓取网页, 严重影响网站排名, 通常是通过技术解决方案将动态 URL 转化成静态的 URL 形式。

例如:

<http://www.domain.com/messages.php?id=2&type=5> 转化为



<http://www.domain.com/messages/2/5/>

总的来说, URL 应该越短越好, 有人为了单纯增加关键字而额外建多一个带有关键字的子目录, 改变目录结构。由于 URL 中含有关键字本身对排名提高帮助并不大, 因而这种做法多此一举, 也是搜索引擎反感的。

### 5.3.3.导航结构

#### 1.主导航醒目清晰

主导航一般体现为一级目录, 通过它们用户和蜘蛛程序都可以层层深入访问到网站所有内容。因此主栏目必须在网站首页第一屏的醒目位置体现, 并最好采用文本链接而不是图片。

例如: 非常代码网, 导航栏为一级目录。首 页 程序代码(C) 共享软件(S)

#### 2. “面包屑型(Breadcrumbs)” 路径 :

所谓“面包屑”是比喻用户通过主导航到目标网页的访问过程中的路径提示, 使用户了解所处网站中的位置而不至于迷失“方向”, 并方便回到上级页面和起点。路径中的每个栏目最好添加链接。如下:

例如: 外国驻华使馆-天天收藏夹目录

首页 >> 政法军事(27631) >> 国际政治与外交(591) >> 外国驻华使馆(88)

即使没有详细的路径来源, 也至少应该在每个子页面提示回首页的链接, 包括页面的 LOGO 作链接。良好的目录导航, 可以提高网站的流量。

#### 3.首页突出重要内容 :

除了主栏目, 还应该将次级目录中的重要内容以链接的方式在首页或其它子页中多次呈现, 以突出重点。搜索引擎会对这种一站内多次出现的链接给予充分重视, 对网页级别 (PageRank) 提高有很大帮助, 这也是每个网站首页的网页级别一般高于其它页面级别的重要因素, 因为每个子页都对首页进行了链接。

例如: 以前很多人使用博客做排名, 原因就在于博客首页 PR 值大, 而刚刚更新的博客会出现在首页, 如果此时 搜索引擎来索引, 收录, 对个人博客作用很大, 通常能够获得不错的排名。

#### 4.使用网站地图:

网站地图 (Site Map) 是辅助导航的手段, 多采用文本链接, 以加快页面加载速度。尤其对于那些采用图片导航和动态技术生成的网页, 通过在网站地图中进行文本链接, 可在



一定程度上弥补蜘蛛程序无法识别图片和动态网页造成的页面不可见的风险。

需要注意，网站地图也要突出重点，尽量给出主干性内容及链接，而不是所有细枝末节。一页内不适宜放太多链接。Google 明确提出“如果网站地图上的链接超过大约 100 个，则最好将网站地图拆成多个网页”。

若页面太多，可以考虑使用前面讲到的二级域名扩展。

#### 5.3.4. 框架结构

框架结构，即帧结构（Frame），包括 IFrame,Frame。

例如：框架结构示例，登录好后即框架结构。

```
<frameset rows="30,1%" frameborder="NO" border="0" framespacing="0" >
  <frame name="mainFrame" src="top.jsp" scrolling="NO">
  <frameset cols="168,1%" frameborder="NO" border="0" framespacing="0">
    <frame name="avdNsVDmJQC7" scrolling="AUTO" src="left.jsp">
    <frame name="rightFrame" scrolling="AUTO" noresize src="right.jsp">
  </frameset>
</frameset>
<noframes>
  <body bgcolor="#FFFFFF" text="#000000">
  </body>
</noframes>
```

框架型网站的优越性体现在页面的整体一致性和更新方便上。尤其对于那些大型网站而言，框架结构的使用可以使网站的维护变得相对容易。但框架对搜索引擎来说是一个很大的问题，这是由于大多数搜索引擎都无法识别框架，也没有什么兴趣去抓取框架中的内容。此外，某些浏览器也不支持框架页面。

如果网页已经使用了框架，或出于某种原因一定要使用框架结构，则必须在代码中使用“Noframes”标签进行优化，把 Noframe 标签看做是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，搜索引擎才能够正确索引到框架内的信息。

### 5.3.5. 图片优化

一般而言,搜索引擎只识读文本内容,对图片文件是视而不见的。同时,图像文件直接延缓页面加载时间,如果超过 20 秒网站还不能加载,用户极有可能离开你的网站。因此,除非你的网站内容是图片为主,比如游戏站点或者图片至关重要,否则尽量避免使用大图片,更不要采用纯图像制作网页(SPLASH PAGE)。

网站图片优化的有三点:

1. 在保持图像质量的情况下尽量压缩图像的文件大小。
2. Alt 属性: 每个图像<IMG>标签中都有 ALT 属性,搜索引擎会读取该属性以了解图像的信息。因此,最好在所有插图的 ALT 属性中都有文字描述,并带上该页关键字在其中。``
- 3.在图片上方或下方加上包含关键词的描述文本;
- 4.使用链接链接到这个图片

### 5.3.6. FLASH 优化

Flash 会使页面很好看,不过 FLASH 网页有一个非常致命的问题,即大部分搜索引擎无法识别 FLASH 中的信息。

FLASH 优化可以从以下三个方面来考虑:

#### 1、做一个辅助 HTML 版本:

保留原有 FLASH 版本的同时,还可以设计一个 HTML 格式的版本,这样既可以保持动态美观 效果,也可以让搜索引擎通过 HTML 版本的网页来发现网站。

#### 2、将 Flash 内嵌 HTML 文件:

还可以通过改变网页结构进行弥补,即不要将整个网页都设计成 Flash 动画,而是将 Flash 内容嵌入到 HTML 文件中,这样对于用户浏览并不会削弱视觉效果,搜索引擎也可以从 HTML 代码中发现一些必要的信息,尤其是进入内容页面的链接。

### 5.3.7.404 错误页设置

网页设计中,出现错误是常见的,但是错误页我们也需要将它制定一个缺省页,使它成为信息页,导引访问者继续访问网站,而不要让它成为空白的“Sorry,你访问的网站不存在”。

方法如下:

首先设计一张网页, 建议和主页的格式保持一致, 将这一页命名为 404.html。

然后上传到网站的根目录, 如: [www.yourdomain.com/404.htm](http://www.yourdomain.com/404.htm)

修改你的.htaccess 文档, 写入:

ErrorDocument 404 <http://www.yourdomain.com/404.htm>

上传这个文档到根目录, 如果没有这个.htaccess, 可以用写字板来写成.htaccess.txt 文件, 上传, 然后, 在服务器的文件存放处将.txt 这个后缀删除。

## 第四节: 网站设计思维

### 5.4.1. 重视网站结构一致性

网站的每个页面遵循一个主题, 所以在格式设计上力求保持一致或者大体相仿, 突然的一个页面改变会打乱访客的思维影响他们对网站内容的吸引。这个是从可用度方面来观察的。另一方面, 不同的格式也会影响搜索引擎优化的努力。搜索引擎优化是个整体的、连续的工作, 但是从开始就要围绕网站的主题设计。不稳定的网站结构会增加搜索引擎优化的工作量, 也影响对优化结果的评估。

微软的网站 [www.microsoft.com](http://www.microsoft.com) 虽然产品众多, 但是, 每个产品的网站都基本上遵循一个风格, 在导航设计上也力求简单明了, 不参杂无关的信息。这种设计于绝大多数的网络使用者贴近, 对维护客户对微软的亲密度具有非常重要的作用。

### 5.4.2. W3C 浏览器兼容验证

网页是由 HTML 或 XML 语言写成的, 就如任何语言, 它们都包含一些法则或者标准, 针对这些标准所进行的测试称为校验。HTML 的标准是 World Wide Web Consortium([www.w3c.org](http://www.w3c.org))来制定的。这个组织帮助网站编辑者编写标准的网页, 以便通过不同的浏览器都能阅读。

验证同时也要保证搜索引擎免于遭遇网页中的编写句法的错误而不能理解网页内容, 或者不明白网页中什么是重要的, 什么不是, 遵循 W3C 标准, 也便于不同的网页设计者使用不同的网页设计工具来编写一个网站的不同网页, 使得各个网页的结构得到统一。

虽然许多网站没有遵循 W3C 标准也获得很好的排名, 但是经过验证后, 网站能保证你

的网的式样不会被浏览器改变，使得你网站的访问者看到的网页与你设计出来的完全一致。你可以在网页的最首段加入这一行来告诉你的访问者、浏览器、验证机制和搜索引擎的蜘蛛机器是遵循 W3C 标准的：

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
```

```
"http://www.w3.org/TR/html14/strict.dtd">
```

验证的方法是到达 <http://validator.w3.org> 输入网址，一些不匹配的错误就将得到反馈。

## 附录 A：搜索引擎优化常用术语（中英对照）

为了便于读者熟悉搜索引擎优化中的常用术语，更是为了方便读者了解这些术语的英文起源，从而可以参考英文有关搜索引擎的文章，这里将这些术语汇总，并以英文字母作为顺序如下。

### A

#### **Alexa**

Alexa 是亚马逊公司([amazon.com](http://amazon.com))下属的一个网站，用来估算世界各国网站的大体流量以确定网站的受欢迎程度。它给网站统一排名，越小数目的 Alexa 值说明排名越高，网站的流量越大。但是，这个数据仅能作为一个参考。因为它是第三方的估算，不能直接得到流量数据，而且与其依靠它的 Alexa 工具条的使用者作为基础，因此给出的流量数据都是有相当的夸大或者压缩，因此不具备真正的衡量一个网站表现的功能。网站的实际表现必须看网站自己的独立统计系统的数据，而这个数据 Alexa 是得不到的。Alexa 数据对搜索引擎优化没有参考价值。

#### **Algorithm 算法**

算法是解决一个问题所形成的公式。在搜索引擎中，算法是程序员编写的用于自动分析网站内容、结构和相关性的一组程序，来确定网站的质量和值，为关键词搜索的结果的排列提供依据。

#### **Anchor Text 链接源头文字**

链接源头文字是一张网页中被划线强调出的一段文字，用来指明连向别的网页的说明。

点击这段文字，浏览器就调出这段文字后的目标，就是另外一张网页。链接源头文字的写作需要和指向页的内容相关，为访问者言简意赅地引见指向页。链接源头文字的编写对网站 PR 值的提高具有重要作用。

## B

### **Backlinks (Inbound Links) 导入链接**

导入链接是由别的网站指向你的网站的一个链接，和导出链接(outbound links)相反。导入链接的多少是搜索引擎，尤其是 Google 来衡量一个网站是否受欢迎的若干因素中重要的一个。追求导入链接，需要注意链接方与你的网站的相关性。

### **Bad Neighborhood 坏邻居**

坏邻居是指你的网站所链接的网站中的质量低下的网站。坏邻居网站很可能包括我们称为“链接养殖场”(link farm)这类专靠链接来骗取 PR 值的网站。坏邻居也包括过分优化的网站。这类网站最容易被搜索引擎惩罚。我们在建立链接的时候要避免和这类网站挂钩。

### **Black-Hat-SEO 黑帽优化手段**

靠捕捉搜索引擎的漏洞和缺陷，将本身不怎么样的网站或者垃圾网站塞给搜索引擎，以次充好，骗取搜索引擎的排名，这样的所谓 SEO 做法是“黑帽手段”。“黑帽”优化往往是过时的优化做法，在三、四年前搜索引擎反垃圾、反滥用的技术不够完美的时期“曾经有效”，而现在使用这些做法就导致网站被搜索引擎拒绝收录。与“黑帽”相反的自然就是“白帽”，即通过搜索引擎普遍接受的方法，比如增加文章的相关性来优化网站。

## C

### **Cache 页存**

一些搜索引擎，如 Google，百度不定时地将一个网站的内容储存在它们的数据库里面，方便搜索者能在某些网页因故暂时无法出现的时候能够参考。因为是不定时地更新，页存出现的网页内容很可能与网站当今真正的内容不同。一般在搜索结果下面有个页存的链接。

### **Cloaking 障眼法**

障眼法是采用伪装网页，先判断来访者是普通浏览者还是搜索引擎，从而展示出不同的网页。这是一种典型的欺骗搜索引擎的障眼法。搜索引擎看到的网页是个优化非常严重的一篇内容，而一般浏览者看到的则非常不同或者根本就不一样。搜索引擎对使用这种做法的网站都给予剔除。

### **CPC 竞价广告**

也叫 PPC。广告商与其他广告商依照关键词来争夺在搜索引擎右侧的赞助商排列(sponsors list)。广告商允许使用简单的文字链接和概述来制作小型广告。当这个广告被点击一次,广告商就支付一次的单价。因为广告商之间在竞争这个付费排列的最佳位置,点击的单价就会逐渐上升。支付最高价格的广告商就获得首位。虽然这种竞价方式的广告节省了许多盲目的传统广告方式,但是因为竞争的原因,广告费依然会有一定的规模。

### **Conversion 转化**

在搜索引擎营销中,成功地引来大量的访问者不是最终目的。这些访问者必须成为一个网站的顾客才有意义。这个顾客和访客的比例的大小说明网站是否达到设计目的。这个从访客变为顾客的过程,就是转化。

### **Crawl 爬行**

这是对搜索引擎来访问网站的一个形象说法。搜索引擎派出机器人(程序)顺着导入链接到达一个网站并且顺着内部链接访问网站内页,再顺着导出链接访问别的网站。

### **CSS 网页样式表**

全称为 Cascading Style Sheets。网页样式表用形象的话来描述就是将网页的规范“一气呵成”。Cascading 在英文计算机语言中是指对命令的一环扣一环的统一设定。通过 CSS,网页可以遵循统一的文字格式,使得各个网页做到完整一致,特别对大型网站的多人编辑非常有用。在搜索引擎优化中,可以通过 CSS 将难看、冗长的 JavaScript 等置于网页之外而以链接方式引入使用,还可以将网页开头的一些不重要的编码移到网页末尾,以便搜索引擎快速阅读中心内容。

## **D**

### **Description Tag 叙述标签**

叙述标签是存在于网页的 HTML 编码之内,只有搜索引擎才会去阅读。这个标签的使用是为了弥补标题标签(title tag)的不足,通过较长的一段叙述,把网页的大意让搜索引擎了解。因此,这个标签的重要性仅次于标题标签。在撰写这段叙述的时候,如何在其中适当地融入关键词是个重要的写作技术。

### **Doorway Pages 门页**

门页,也叫桥页、跳页,是纯粹为了某个特别的关键词获得搜索排名而设计的网页。这些网页一般不在网站的导航中出现,但是被用来导引访客更深入地进入网站的别的页面。这些网页的内容很不讲究,但是在网页的底部有个链接,指导访客进入真正的有实质性内容的

网站部分。搜索引擎对这个做法很反感。

### **DMOZ 开放目录**

它又叫 Open Directory (开放目录), 是最重要的名录导航站。这个网站属于互联网最早的公司 Netscape (网景公司), 历史悠久, 来自全球的志愿者来作为编辑来维护、管理这个网站。由于搜索引擎将这个网站视为最有信用的目录站, 因此能将网站登录在这个名录站的适当分类中, 非常符合搜索引擎最看中的相关性原则。

### **Dynamic URL 动态网址**

动态网址是以 ASP, PHP, Perl 或者 JSP 等编程语言来制作的网页出现的地址。这类网页往往使用数据库来依照访问者的需要提供个性化的网页或者进行适应性的改编。由于网页的生成是采用数据库的内容, 因此网页的永恒性不存在。这样造成了搜索引擎的阅读困难, 即搜索引擎抓不住一个永久的主题, 因此不能输入到搜索引擎中的缓冲 (cache)内。同时, 由于编程中使用循环语句, 搜索引擎在访问中会陷入死循环, 而且使搜索引擎不知道究竟有多少网页需要阅读。搜索引擎优化的重要方面就是将动态网页静态化, 以便搜索引擎收录相关网页。

## **E**

### **Entry Page 入口页**

入口页就是访客进入你网站的那一页。网站的任何一页都有可能是入口页。入口页越多, 访客的来源越广。入口页的数量多少, 也可以用来衡量搜索引擎优化的效果。搜索引擎可能认为你的任何一张网页与搜索者的需求相符, 而把此页的链接提供给搜索者。

### **Exit Page 出口页**

出口页就是访客离开你网站的地方, 是他们停留的最后一页。分析出口页的情况可以明白网站什么原因让访客跑掉, 以便进行改进。

## **F**

### **Florida Update 佛罗里达更新**

2003 年 11 月上旬 Google 开始的对排名算法的剧烈更新。这个更新, 犹如飓风, 席卷了许多网站原先的排名, 使很多网站在即将到来的圣诞节的购物黄金季节失去大量的客源。这个被称为 Florida Update 的大规模调整, 是 Google 反击黑帽 SEO 制作垃圾网站来获取不合法的排名的历史性战役。持续若干月的这个更新, 使许多 SEO 从业者反思如何正确地使用技术来减少这种灾难似的打击。



## **Freshbot 和 Deepbot 刷新探测器和深度探测器**

Google 使用两个探测器来抓取网站上的内容。深度探测器每月一次地出击, 受访内容在 Google 的主要索引之中。刷新探测器是持续不断地发现新的内容, 例如新的网站、论坛、博客等。看起来, Google 是发现了一个新的网页, 之后频繁地再访来看看是否还有什么新的更新。如果有, 这个新网站就被加入到刷新探测器的名单中进行访问。

# **G**

## **Google API 谷歌应用界面**

Google 的网络服务 API 能让程序员通过自己编制的程序来索询超过 80 亿的网页。这个服务是建立在标准的 SOAP 和 WSDL 界面上, 因此能够在很广泛的技术范围内进入。这样, 有不少的搜索引擎优化者就可以建立一个程序来查询网站在 Google 搜索中的排名。

# **H**

## **Hidden Text 隐藏文字**

隐蔽文字(hidden text)是许多人喜欢使用的作弊方式。由于并不想多余的用来引诱搜索引擎阅读的文字来影响网页的面目和感觉, 那么就将这些多余的文字隐蔽在 HTML 页码之中只让搜索引擎看见或者使之透明而让浏览者看不见。搜索引擎现在已经能够侦察出这作弊手法。

# **I**

## **Index 索引**

作为名词, 也是动词。作为名词, 是指搜索引擎存储有价值的网站信息的收录数据库。作为动词, 是指搜索引擎将网站访问后获取有价值的信息, 给以存储, 以便提供给搜索者。

## **Internal Links 内部链接**

内部链接将一个网站中的各个文件串联起来组成一个网站。广泛、准确地使用链接源头文字来优化网站的内部链接, 对一个网站顺利获得比较理想的 PR 值有重要的意义。好的内部链接结构也能让有较高 PR 值的网页把 PR 值传递到其他网页。

## **Inbound Links 导入链接**

导入链接, 见 Backlinks

# **J**

## **Jump Pages 跳页**

跳页, 见 Doorway Pages

## K

### **Keywords 关键字词**

关键字词是人们在搜索引擎中输入的所寻找的信息的核心代表词或者词组。网站优化要在网页中使用人们会经常在搜索引擎中输入的关键字词, 才能有机会在搜索结果中被人们发现。关键字词研究是搜索引擎优化的最重要的一步。科学选择关键字词, 不仅能使网页获得较高的排名机率, 而且能在 Google 的算法不断更新中保持排名。

### **Keyword Density 关键字词密度**

关键字词密度是指一张网页的所有文字中关键字词出现的次数的百分比。它是若干年前搜索引擎优化中非常注重的一个策略。关键字词密度需要掌握得恰到好处, 一般是在 15-25% 之间。过少的密度不足以说明一张网页说明的信息的主题, 过多的密度会被搜索引擎以为是恶意操作。因为容易操作, 现在这个做法已经不具有当初的重要性。

### **Keyword Stacking 关键词叠加**

关键词叠加是指在网页中过分重复关键词。最基本的叠加方式是在网页中访客看不见的 HTML 文件中的一些地方, 如标题签、描述标签、图片的替代文字中, 等等。

### **Keyword Stuffing 关键词堆积**

关键词堆积(keyword stuffing)和关键词叠加常常指的是同样的情况。一些搜索营销的人士将两者分开。关键词叠加一般指写些垃圾句子。关键词堆积一般指将这些垃圾句子放在图片之中, 比如使用<alt 标签。

### **Keyword Tag 关键词标签**

关键词标签是网页的 HTML 用来告诉搜索引擎网页内容的一个标签。早期搜索引擎希望通过这个标签能很快了解网页的主体内容。但是由于搜索优化者滥用这个标签, 堆积往往和网页内容不相干的词语, 现在已经基本上不被搜索引擎所使用。

## L

### **Landing Page 着陆页**

一个访问者顺着链接到达一个网站的那一页。这一页可以是首页, 也可以使广告对应页。从搜索引擎优化的观点看, 网站的所有页面都是可能的着陆页。网站的着陆页越多, 表明网站的可获得的流量资源越丰富。

### **Links 链接**

链接是将网站内部内容串联起来和与外面网站挂钩的方式之一。链接是通过超文字来实现的。超文字(hypertext)这个词是由泰德·奈尔逊(Ted Nelson)在上世纪六十年代设计第一个使用 hypertext 的工程 Xanadu 时候铸造的。奈尔逊受到 1945 年由凡丽娃·布什(Vanneva Bush)写过的一篇题目叫《我们可以试想》(As We May Think)的论文的启发。布什在文章中设想有种机器能将靠阅读许多文件中由词或者词组,并根据词义和来表达的概念,通过一种非线性的途径将众多的文件首尾相接。链接的好坏与否直接关系到网站的可用性和在搜索引擎中的价值。链接分内部链接和外部链接两种。

### **Link Farm 链接养殖场**

“链接养殖场”是指一个全无有价值信息的网页,这个网页站除了人为罗列的一个个指向其他网站的链接外,没有其他内容或者极少的内容。“链接养殖场”的任务就是交换链接,寄希望通过链接互相传递 PageRank 来提升这个网页的 PR 值。这个网页,或者数个网页的内容不一定和你的网站内容相关,甚至没有一点相关。同样,内容也不一定和它链接的其他网站的内容相关,甚至没有一点相关。这样的纯粹用于交换链接,添加链接数量的网页就成了一个“链接养殖场”。与“链接养殖场”相连的网站有被搜索引擎拒绝收录的危险。

### **Link Popularity 链接广泛度**

也有人称呼为“链接流行度”。这个是影响搜索引擎,尤其是 Google 决定一个网站排名的标准之一。搜索引擎检查一个网站有多少人愿意链接他,这些链接的网站的质量是否优秀,来判断这个网站的可信度和权威性。

## **N**

### **Natural Search 自然搜索**

自然搜索是搜索引擎业者为了区分于 PPC(搜索竞价排名广告)而引入的一个术语,也叫有机搜索(organic search)。它是指搜索引擎根据自己的算法给予所有在它们索引数据库中的网站针对搜索关键词而伺服给用户的搜索结果。这类结果一般显示在网页的左侧,因此才会有人对在自然搜索中的排名叫做左侧排名。这种搜索不由广告所控制,完全由算法程序给予自动排列。搜索引擎优化的现实目标是追求在自然搜索中的优先排列位置,以吸引最大程度的来自搜索引擎的流量。

## **O**

### **ODP 开放目录**

又叫 DMOZ Open Directory ([www.dmoz.org](http://www.dmoz.org)) 是最重要的名录导航站。这个网站属于互

联网最早的公司 Netscape (网景公司), 历史悠久, 来自全球的志愿者来作为编辑来维护、管理这个网站。由于搜索引擎将这个网站视为最有信用的目录站, 因此能将网站登录在这个名录站的适当分类中, 非常符合搜索引擎最看中的相关性原则。

### **Outbound Links 导出链接**

导出链接是从你的网站连到另外一个网站的链接。导出链接可以在不增加你自己的网站额外内容的情况下给访问者提供更多的信息源。这个导出链接是访问者跳向另外一个网站的关节。同时, 它也给指向的网站传递 PageRank。搜索引擎优化者一般不情愿在网站中放置过多的导出链接以免造成 PageRank 的流失。但是, 随着搜索引擎的算法更加成熟, 导出链接不一定有不利的影 响, 除非是指向“坏邻居”垃圾网站。

## **P**

### **PageRank 网页等级**

“网页等级”是 Google 能在九十年代后期超越其他搜索引擎的一个重要概念和法宝。2001 年 9 月 Google 被授予美国专利。“网页等级”着重考察网站的权威性, 即越有权威的网站越容易被其他网站主动链接。被链接越多, 就意味着被别的网站投票越多。这个概念引自学术中一篇论文的被引述的频度, 即被引述的次数越多, 这篇论文的权威性越高。“网页等级”评价一个网页用 1-10 的数字来显示在 Google Toolbar 中。搜索引擎优化者为了提升一个网站的“网页等级”价值总是在试图获取最多的导入链接, 因为一个导入链接就相当于一份投票。

### **PageView 受访页面量**

一个受访页面是指从网站服务器里面成功调出一个完整文件, 也就是成功显示一个网页。你可以通过统计系统来读一天、一周和一个月的数量。并在一起, 受访页面量就是一个衡量网站受欢迎程度的指标之一。

### **PPC 竞价广告**

全称是 Pay Per Click。见 CPC。

### **Page Jacking 网页劫持**

网页劫持是将别人的网站内容或者整个网站全面复制下来, 偷梁换柱放在自己的网站上。这个方法是对网页内容极其匮乏的站主颇有吸引力。但是, 这个做法是相当冒险的, 更是不齿的。搜索引擎的专利技术能从多个因素上来判断这个被复制的网页或者网站不是原创, 而不予以收录。

## **PODCAST 文件广播机制**

PODCAST 起源于为苹果公司的 iPOD 来提供可以下载的音乐文件。现在泛指可以通过互联网下载的多媒体文件的一个机制,如同在网上“广播”这些多媒体节目一样,欢迎人们来索取这些文件。这些多媒体文件包括语音和录像,也可以包括图片、文字、PDF 或者任何种文件。SEO 专家可以对文件的内容进行关键词优化,也可以添加链接等,使这些文件在搜索引擎的搜索中获得优先的推荐,从而达到推广的目的。

## **Reciprocal Linking 互换链接**

互换链接是介于两个网站之间通常是 1:1 的链接交换。从搜索引擎优化的观点来看,这个互换链接必须是由两个话题类似的网站之间来进行的,搜索引擎才会认可这两个网站获得的链接广泛度。

## **Referrer 来路**

一个来路就是一个访客来到你的网站的途径。这个途径可以是搜索引擎自然搜索结果、搜索引擎广告、网站链接、email 里面的链接,直接敲入网站地址或者是书签。这个来路可以被分为两半:来路的域名和来自的网页。这个分析的统计对搜索引擎优化的成功相当重要。

## **Robot 探测器**

探测器有多种叫法,也叫 crawler (爬行器), spider (蜘蛛)。这个形象的叫法是描绘搜索引擎派出的蜘蛛机器人爬行在互联网上探测新的讯息。这个探测器实际上是人们编制的计算机程序,由它不分昼夜地进入访问各个网站,取回网站内容、标签、图片等,然后依照搜索引擎的算法给它们制定成索引。但是,探测器对许多网站是不能完整取回信息的,这个大多是由于网站的设计没有按照搜索引擎探测器的思路来进行优化。

# **S**

## **Sandbox 沙盒**

“沙盒”是 Google 反击垃圾网站的重要措施,出现在 2004 年 3 月,也就是举世瞩目的 Google 佛罗里达风暴和奥斯丁风暴之后的半年。有了沙盒,Google 仍然象过去一样迅速收录网站,从互联网上从新的网站里努力获取新鲜的信息,但是不再象以前那样信任这些新网站。所有新收录的网站都要经过“适用期”或者“观察期”。经过时间的考验,如果这些网站能顺利通过 Google 多次的反“泛滥”过滤器(spam filter)的考验,最终这些网站就被放回搜索的“大海”之中,参与正常的排名竞赛之中。这个过程少则两、三个月,长则要一年。

新网站可以采取一些积极的措施，争取早日走出沙盒。

### **SEM 搜索引擎营销**

搜索引擎营销，全称为 Search Engine Marketing，简单缩写为 SEM。搜索引擎营销，顾名思义是试图从搜索引擎上获得新鲜的访问者。从广义上说，所有依靠搜索引擎来实施营销方案的网站推广方式，都是 SEM。目前搜索引擎营销主要包括搜索引擎优化、搜索竞价广告两种常见形式。

### **SEO 搜索引擎优化**

搜索引擎优化，全称为 Search Engine Optimization，是针对搜索引擎的网站收录和评价规律，对网站的结构做合理的调整，对网页的内容和页面因素包括标题、描述等做出科学性的处理，使每个网页能够最有效地反映网页需要表述的信息，而总体上使网站集中反映一个主题。这个处理的方式是一种技术，需要对搜索引擎的网站评价方式有个系统的了解，对搜索者的行为有个深切的感受，这样对网站进行一种最优化的组建或者改造，使搜索者在使用搜索引擎的时候，在搜索结果页面上排列的一系列相应网站的大致介绍的排列中出现在前列，比如最前面的 10 位之内。

### **SERP 搜索结果页面**

搜索结果页面，是 Search Engine Results Page 的缩写。这个是搜索引擎针对用户搜索的关键词而提供的搜索引擎认为最相关的网站链接。搜索结果页面有几十甚至上百页，而只有前面一、二页会被用户所重视。因此，搜索引擎优化的目的之一，就是要让网站的某页出现在最前面一、二页的搜索结果中。

### **Sitemap 站点地图**

站点地图，是搜索引擎优化中的一个重要步骤。其作用远超过给网站访客做向导。更主要的作用是给搜索引擎的探索机器人或者蜘蛛来了解一个网站，指导它们按照地图上指向各个网页的链接顺利到达各个网页。Google SiteMap 服务则允许站主用途 XML 文件，罗列网站中所有网页送交 Google，确保 Google 能访问网站中的每一页。

### **Social Bookmarking 社交书签**

“社交书签”(social bookmarking)是美国 Web 2.0 的另一个新生的应用。这个书签的意义在于一个网上社区的会员来共同推荐一个网站，给这个网站列为大家共同喜欢的名单中。能被大家共同推荐的网站，自然地就获得了来自许多有相似爱好的访问者的丰富流量。“社交书签”的网站代表者是 Del.icio.us 和 Digg。

### **Spam 滥用**



许多搜索引擎优化者通过制作无用、无价值的网页来试图影响搜索引擎的判断，投机取巧地获得在搜索引擎中的优先排名或者影响这个排名的结果。这种行为就是被称为“滥用”，而制造出来的网页或者网站被称为“垃圾”。泛滥的种类还包括域名泛滥等试图愚弄搜索引擎。搜索引擎从诞生起就没有停止对泛滥垃圾的斗争。

### **Static Page 静态网页**

静态网页是以对访问者任何需要都给予同样内容的网页。这类网站需要制作各个不同的网页来应对访问者寻求的信息。静态网页的最典型的例子就是老式的 HTML 网页。由于搜索引擎的技术的限制，它们目前对静态网页的阅读和抓取比较好，因此动态网页往往需要首先变成静态网页方能很好地被搜索引擎吸收。

## **T**

### **Title Tag 标题标签**

标题标签是网页 HTML 编码中的一段概述网页内容的短句。它显示在浏览器窗口的最上方的蓝色横条里，一般不太引起用户的注意。但是，搜索引擎非常看重它的写作，以它描述的文字来决定网页的主题和内容。因为搜索引擎只阅读这个标签一定长度的文字，比如 Google 限制在英文字母总长度 75 个之内，所以如何撰写这个标签变得至关重要。这个标签内一般含有最能说明网页内容的关键词。

## **U**

### **Unique Visitor 独一访客**

独一访客是网站统计中的一个观察因素。独一访客数量扣除了一定时间段里面，比如一天里回头又来访的访客次数。这个数据的统计是用 cookie 或者 IP 地址来跟踪的。来自同一 IP 地址的人只算一个。

## **V**

### **Validation 效验**

网页是由 HTML 或 XML 语言写成的。就如任何语言，它们都包含一些法则或者标准。针对这些标准所进行的测试称做效验。HTML 的标准是 World Wide Web Consortium ([www.w3c.org](http://www.w3c.org))来制定的。这个组织帮助网站编辑者编写标准的网页，以便通过不同的浏览器用户都能完美阅读。验证的方法是到达 <http://validator.w3.org> 输入网址，一些不匹配的错误就将得到反馈。



## W

### W3C 万维网联盟

全称是 World Wide Web Consortium, 即世界万维网联盟。网站是 <http://www.w3.org/>, 中文网站在 <http://www.chinaw3c.org/>。这个联盟以开发互通技术(规格、指引、软件和工具), 来引发网络的全部潜能作为一个讨论对于信息、商贸、通信及集体理解的论坛。

## X

### XML Sitemap XML 站点图

Google 首先采用接受 XML 为模式的文档, 便于 Google 方便地知道一个网站的全貌和所有网页的存在, 为爬行这个网站提供一个路线图。

## 附录 B:网站管理员指南(Google)

遵循这些指南有助于 Google 查找您的网站并将其编入索引和进行排名。即使您选择不采纳这些建议, 我们也强烈建议您对"质量指南"多加留意, 其中简要说明了一些可能造成网站从 Google 索引中彻底删除或受到其他处罚的违禁行为。如果网站受到处罚, 则可能不会再出现在 Google.cn 或任何 Google 合作伙伴网站的搜索结果中。

### 网站准备就绪后:

- 通过 <http://www.google.com/addurl.html> 将其提交给 Google。
- 通过 [Google 网站管理员工具](#)提交 [Sitemap](#)。Google 使用您的 Sitemap 来了解您网站的结构, 并提高对您网页的抓取率。
- 确保应了解您网页的所有网站都知道您的网站已处于在线状态。

### 设计与内容指南

- 网站应具有清晰的层次结构和文本链接。每个网页应至少可以通过一个静态文本链接打开。
- 为用户提供网站地图, 列出指向网站重要部分的链接。如果网站地图上的链接超过或大约为 100 个, 则需要将网站地图拆分为多个网页。
- 网站应实用且信息丰富, 网页文字应清晰、准确地表述要传达的内容。
- 要考虑到用户会使用哪些字词来查找您的网页, 确保网站上确实包含了这

些文字。

- 尽量使用文字而不是图片来显示重要的名称、内容或链接。Google 抓取工具无法识别图片中所含的文字。
- 确保 <标题> 元素和 ALT 属性具有描述性且准确无误。
- 检查链接是否损坏，并确保 HTML 格式正确。
- 如果要使用动态网页（即网址中包含"?" 字符），请注意并非每一个搜索引擎抓取工具都能抓取动态和静态网页。缩短参数长度和减少参数数目都会对这种情况有帮助。
- 将特定网页上的链接限制在合理的数量内（少于 100 个）。

## 技术指南

- 使用诸如 [Lynx](#) 的文本浏览器来检查您的网站，因为大多数搜索引擎抓取工具查看您网站的方式与 Lynx 几乎一样。如果诸如 Javascript、Cookie、会话 ID、框架、DHTML 或 Flash 等复杂功能造成您在文本浏览器中无法看到整个网站，则搜索引擎抓取工具在抓取您的网站时可能会遇到问题。
- 允许搜索漫游器抓取您网站时不抓取会话 ID 和跟踪其访问网站路径的参数。这些技术对跟踪单个用户的行为非常有用，但漫游器的访问模式却完全不同。如果采用这些技术，可能会导致对网站的索引编入不完整，因为漫游器可能无法排除那些看上去不同，但实际却指向同一个网页的网址。
- 确保您的网络服务器支持 If-Modified-Since HTTP 标头。通过该功能，您的网络服务器可以告诉 Google 自上次抓取您的网站以来，内容是否已发生变化。该功能可以节省您的带宽和开销。
- 利用网络服务器上的 robots.txt 文件。该文件会告诉抓取工具哪些目录能够抓取，哪些不能。确保该文件是网站的最新版本，以免意外拦截 Googlebot 抓取工具。请访问 <http://www.robotstxt.org/wc/faq.html>，了解如何在漫游器访问您的网站时对其进行指示。您可以使用 Google 网站管理员工具中的 [robots.txt 分析工具](#) 来测试 robots.txt 文件，以确保正确使用此文件。
- 如果您的公司购买了内容管理系统，请确保该系统能导出您的内容，以便

搜索引擎抓取工具可以抓取您的网站。

- 使用 robots.txt 可避免抓取搜索结果页或其他自动生成的页面，这些页面来自搜索引擎的用户没有附加价值。
- 对网站进行测试，[以确保其在不同的浏览器中均能正确显示](#)。

## 质量指南

这些质量指南涵盖了最常见的作弊形式或操纵行为，对于此处未列出的其他误导行为（例如，通过注册知名网站的错拼形式来欺骗用户），Google 仍可予以拒绝。不要抱有侥幸心理，认为某种欺骗手段未在本页中列出，Google 就会认可该手段。作为网站管理员，与其花费大量时间寻找漏洞加以利用，不如尽其所能维护基本原则，以便为用户带来更好的体验，从而使网站获得更高的排名。

如果您认为其他网站正在滥用 Google 质量指南，请通过 <https://www.google.com/webmasters/tools/spamreport?hl=zh-CN> 举报该网站。Google 希望能开发出灵活的自动解决方案来解决上述问题，因而尽量避免一一处理违禁行为。我们收到的滥用行为报告将用于创建灵活的规则系统，以便于识别并阻止以后的滥用企图。

## 质量指南 - 基本原则

- 设计网页时该考虑的是用户，而不是搜索引擎。不要欺骗用户，或提交给搜索引擎一种内容，而显示给用户另一种，这种做法通常称为“隐藏”。
- 请不要为了提高搜索引擎排名而弄虚作假。一个简单分辨是非的方法是：您是否可以坦然地跟竞争对手网站解释您对网站所做的事情。另一个有用的测试手段即扪心自问：“这能否给我的用户带来帮助？如果不存在搜索引擎，我是否还会这样做？”
- 请不要[参与旨在提高您的网站排名或 PageRank 的连接方案](#)。尤其要避免链接到违禁的网站或“恶邻”，因为您自身的排名会受到这些链接的负面影响。
- 请不要使用未经授权的计算机程序来提交网页和检查排名等，此类程序会耗用计算资源并违反我们的[服务条款](#)。Google 不建议使用 WebPosition Gold™ 这类产品向 Google 发送自动查询或用程序编写的查询。

## 质量指南 - 具体指南

- 请不要使用[隐藏文本或隐藏链接](#)。

- 请不要使用[隐藏真实内容或欺骗性的重定向手段](#)。
- 请不要向 Google 发送[自动查询](#)。
- 请不要[加载使用无关关键字的网页](#)。
- 请不要创建包含大量[重复内容](#)的多个网页、子域或域。
- 请不要制作欺诈性或安装有病毒、特洛伊木马或其他[有害](#)软件的网页。
- 请不要[专门针对搜索引擎创建“桥页”](#)，或使用其他如联属计划这类[原创内容很少或几乎没有原创内容](#)的俗套 (cookie cutter) 方法。
- 如果您的网站参与联属计划，请确保您的网站可为其增添价值。请[提供独特且相关的内容](#)，使用户有理由首先访问您的网站。

如果断定您的网站不符合这些指南，可进行修改使其满足要求，然后[提交网站以便重新审核](#)。

## 附录 C：百度的建站指导

### A:网页收录问题

#### 1. 如何让我的网站（独立网站或者 blog）被百度收录？如何查看我的网站是否被百度收录？

- 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 更快的发现您的站点，您也可以向我们提交一下您的网站的入口网址。提交地址是：  
[http://www.baidu.com/search/url\\_submit.html](http://www.baidu.com/search/url_submit.html)。您只需提交首页即可，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。
- 百度是否已经收录您的网站可以通过执行 site 语法查看，直接在百度搜索中输入 site:您的域名，如 [site:www.baidu.com](http://site:www.baidu.com)，如果 site 语法查询可以查询到结果，那您的网站就已经被百度收录。
- site 语法得到的搜索结果数，只是一个估算的数值，仅供参考。

#### 2. 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/>）。
- 您可以设置一个 Robots 文件以限制您的网站全部网页或者部分目录下网页不被百度收录。具体写法，参见：[如何撰写 Robots 文件](#)。
- 如果您的网站在被百度收录之后才设置 Robots 文件禁止抓取，那么新的 Robots 文件通常会在 48 小时内生效，生效以后的新网页，将不再建入索引。需要注意的是，robots.txt 禁止收录以前百度已收录的内容，从搜索结果中去除可能需要数月的时间。
- 如果您的拒绝被收录需求非常急迫，也可以发邮件给 [webmaster@baidu.com](mailto:webmaster@baidu.com) 请求处理。

### 3. 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，某些网页即便通过您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：
  - A. 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
  - B. 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
- 如果您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](mailto:webmaster@baidu.com) 进行解决。

### 4. 为什么我的网站收录数量越来越少？

- 您的网站所在的服务器不稳定，Spider 在检查更新时抓取不到网页而被暂时去除。
- 您的网站不符合用户的搜索体验。

#### 5. 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。
- 如果您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：
  - A. 您的网页不符合用户的搜索体验
  - B. 您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
  - C. 您的网页内容有不符合国家法律和法规规定的地方
  - D. 其他技术性问题
- 以下的说法是错误的和毫无根据的：
  - A. 参与了百度竞价排名但未续费，会从百度搜索结果中消失
  - B. 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
  - C. 与百度旗下网站产生了竞争，会从百度搜索结果中消失
  - D. 从百度获得的流量太大，会从百度搜索结果中消失

#### 6. 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？

- 百度只收录对用户有价值的网页。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：
    - A. 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。
- 如果您的网站中有较多这种页面，那么这可能会使您的整个网站的页面收录和排序受到影响。

- B. 网页是复制自互联网上的高度重复性的内容。
- C. 网页中有不符合中国法律和法规的内容。

7. 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，都有机会被百度重新收录。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：
  - A. 我成为百度的广告客户或者联盟网站，就可以重新被收录
  - B. 我给百度若干钞票，就可以重新被收录
  - C. 我认识百度的某某人，就可以重新被收录

8. 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（包括去除死链接，更新域名变化，更新内容变化）。因此请耐心等待一段时间，您的网站上的变化就会被百度察觉并修正。

9. 为什么我的网站在百度收录的数量和其他搜索引擎相比相差很多？

- 通常情况下，这是正常的现象，不同的搜索引擎判断网页价值的算法不同。

**B. 网页排序问题**

1. 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？

答：排序算法非常复杂。我们的目标，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。我们会非常欢迎您把您遇到的困惑和问题，反馈给我们。我们的工程师，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，您可以在那里提交您的问题，或者将您的问题发送到 [webmaster@baidu.com](mailto:webmaster@baidu.com)，以协助我们改进。

我们一直在改进搜索算法，以使得百度的搜索结果更加符合用户的搜索需求。



2. 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？

答：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：

- A. 特定关键词所涉及的您的网页发生了变化
- B. 特定关键词所涉及的其他网页发生了变化
- C. 百度的排序算法发生了变化

3. 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？

答：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。

a) 我请一些“SEO”来为我的网站或者网页做优化，会有什么后果？

答：合理的搜索引擎优化，参见百度的“给站长的建站建议”。

外界很多打着 SEO 旗号的公司或者个人，也许能为您的网站带来短期的排序收益，但是，这会使您将面临更大损失的风险。在您把网站资源交托给别人之后，很多 SEO 会使用作弊的手法来提高排名，甚至会利用您的资源进行他们个人的运营项目，最终导致您的利益受损。

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：

- A. 我和百度的人很熟，想怎么干就怎么干，没风险
- B. 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
- C. 我把 xxx、yyy、zzz 这些关键词都搞到第一了，所以我是牛人啊

您也可以向百度[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。

C. 商业客户相关的问题

1. 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？

答：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。

如果您收到任何类似威胁的说辞,请您直接发邮件至 [ask@baidu.com](mailto:ask@baidu.com) 举报。

2. **我是百度的竞价排名客户,为什么停止续费后网站就从百度消失了?**

答:网站能否被百度收录只与您网站的质量有关,与竞价排名没有任何关系。网页搜索结果中的竞价排名不代表您的网站被百度收录。如果您的网站从百度消失了,请参考[网页收录问题](#)的说明。

3. **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?**

答:不可以。我们对网站的收录,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 7 中的叙述。

4. **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?**

答:不可能。

**D.给站长的建站建议**

1. 为每个网页添加合适的标题,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,这可以让您的潜在用户通过搜索引擎结果中的标题快速访问到您的页面。
2. 充分利用网站首页或者频道首页的 description 标签,提供此网页内容的概括说明,形式为<meta name="description" content="此网页内容的概括说明" />,这将帮助用户和搜索引擎加强对你的网站和网页的理解。
3. 网站应该有明晰的导航和层次结构,网站上重要的网页,应该能从网站比较浅层的位置找到,确保每个页面都可以通过至少一个文本链接到达。
4. 尽量使用文字而不是 flash、Javascript 等来显示重要的内容或链接,百度暂时无法识别 Flash、Javascript 中的内容,这部分内容可能无法在百

度搜索得到；仅在 flash、Javascript 中包含链接指向的网页，百度可能无法收录。

5. 尽量少使用 frame 和 iframe 框架结构，通过 iframe 显示的内容可能会被百度丢弃。
6. 如果网站采用动态网页，减少参数的数量和控制参数的长度将有利于收录。
7. 网站改版或者网站内重要页面链接发生变动时，应该将改版前的页面 301 永久重定向到改版后的页面。
8. 网站更换域名，应该将旧域名的所有页面 301 永久重定向到新域名上对应的页面。

只有当搜索引擎、站长、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，只会使您与用户、与搜索引擎越来越远。搜索引擎与站长之间，宜和谐发展，共同拥抱美好的愿景。

以下是我们给出的一些网站质量方面的建议：

1. 网站的内容应该是面向用户的，搜索引擎也只是网站的一个普通访客，放置任何用户不可见、或者欺骗用户的内容，都可能被搜索引擎当做作弊行为，这些行为包括但 不限于：在网页中加入隐藏文字或隐藏链接；在网页中加入与网页内容不相关的关键词；具有欺骗性跳转或重定向；专门针对搜索引擎制作桥页；针对搜索引擎利用程序生成的内容；具有大量重复无价值内容；充斥大量恶意广告或恶意代码等。
2. 百度更喜欢独特的原创内容，如果您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。
3. 谨慎设置您的友情链接，如果您网站上的友情链接，多是指向一些垃圾站点，那么您的站点可能会受到一些负面影响。

4. 谨慎加入频道共建、内容联盟等不能产生或很少产生原创内容的计划，除非您能为内容联盟创造原创的内容。
5. 百度会尽量收录提供不同信息的网页，如果您网站上相同的内容可以通过不同形式展现（如论坛的简版页面、打印页），可以使用 robots.txt 禁止 spider 抓取您不想向用户展现的形式，这也有助于节省您的带宽。

#### E. 互联网论坛收录开放协议

《互联网论坛收录开放协议》是百度网页搜索制定的论坛内容收录标准，论坛网站可将论坛内发布的帖子制作成遵循此开放协议的 XML 格式的网页供搜索引擎索引，将论坛发布的帖子主动、及时地告知百度搜索引擎。采用了《互联网论坛收录开放协议》，就相当于论坛中的帖子被搜索引擎订阅，通过百度——全球最大的中文搜索引擎这个平台，网民将有可能在更大范围内更高频率地访问到您网站论坛中的帖子，进而为您的网站带来潜在的流量。

#### F. 其他

**我发给百度的线上反馈，或者发给 [webmaster@baidu.com](mailto:webmaster@baidu.com) 邮件，是否会得到及时的回复？**

答：百度负责网页搜索质量的工作人员，虽然无法对反馈和邮件进行回复，但对于每一个线上反馈和邮件，都会认真仔细的阅读和归类，并及时的转给相应的负责部门处理。