

第 3 章

站内优化（一）

【学习目标】

- 掌握搜索引擎的排名机制
- 了解影响关键词排名的因素
- 掌握网站导航优化方法
- 掌握网站 URL 设计方法
- 掌握 301 重定向的设置方法
- 学会如何设置网站的站内链接
- 学会撰写 robots.txt 文件
- 掌握网站地图的构建方法

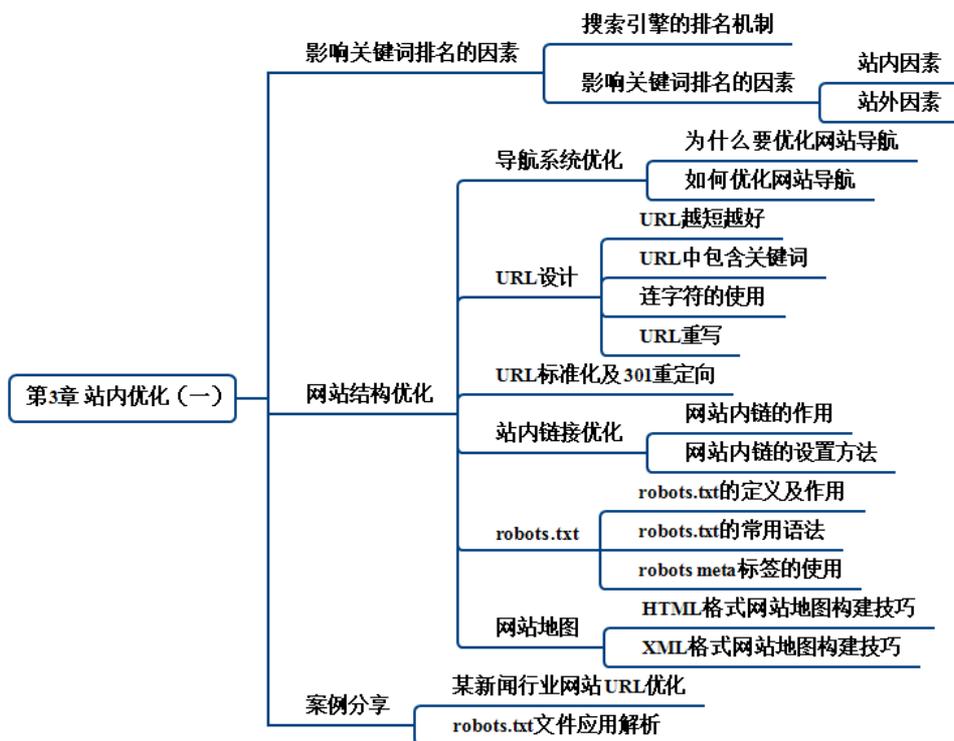
【导读】

一名优秀的 SEO 人员从事网站优化工作不会从单一的角度对一个网站进行优化，而是全方位的进行，这样可以使网站有一个综合性的提升。

网站优化一般分为站内优化和站外优化。站内优化，顾名思义就是网站内部的优化。随着百度算法的不断更新，站内优化对网站关键词排名越来越重要了。站外优化，即网站外部的优化，也就是增加外部链接。

本章会介绍站内优化的具体内容。首先介绍搜索引擎的排名机制，然后根据排名机制拓展出影响关键词排名的因素，同时重点分析站内优化中网站结构影响关键词排名的原理及优化方法。





3.1 影响关键词排名的因素

对于很多 SEO 新手来说，最棘手的问题就是接手一个网站自己不知道如何去优化。实际上，做 SEO 很重要的一点是要了解网站优化各个因素的原理。而想要了解影响关键词排名的因素，首先要清楚搜索引擎的排名机制，知道搜索引擎重视哪些因素。

3.1.1 搜索引擎的排名机制

将有价值的内容展现给用户是搜索引擎的服务宗旨。对于用户来说，有价值的内容可以从两个方面去理解：一是用户想要看到的内容；二是对用户来说相对重要的内容。而“用户想要看到的”和“重要的”这两个方面在搜索引擎中可以理解为“相关性”和“页面权重”，这就是搜索引擎的排名机制。

相关性，是指页面与关键词的关联性。关联性越高，页面越容易获得排名。从另一个角度理解，用户在搜索引擎上搜索某个关键词后，搜索引擎所展示的内容一定是与用户搜索的内容相关的。如图 3.1 所示（本图为参考图，以实际查询结果为准），在百度搜索引擎中搜索“旅游”，排名靠前的都是与“旅游”相关的搜索结果，而且相关性越高，排名越靠前。

页面权重，搜索引擎会将互联网中的页面按照重要程度进行划分，将其认为重要的页面展示给用户。权重的详细解释，请参照“第 1 章 SEO 基础”中常用名词含义部分。在其他因素一定的前提下，权重越高页面排名越靠前。



图 3.1 百度中搜索“旅游”

3.1.2 影响关键词排名的站内因素和站外因素

既然搜索引擎根据相关性和权重来计算页面排名。那么根据相关性和权重就可以扩展出影响关键词排名的因素包括站内因素和站外因素。

1. 站内因素

影响关键词排名的站内因素可以概括为：网站结构、网站代码、关键词布局及密度、网站内容。

网站结构：涉及的内容比较多，包括导航系统优化、URL 设计、URL 标准化及 301 重定向、站内链接优化、robots.txt 与网站地图。

网站代码：指的是网站基本标签优化对排名的影响，包括<h>、、<a>、、<iframe>等基础标签。

关键词布局及密度：包括关键词布局优化、关键词密度优化。

网站内容：网站内容更新及网站收录优化。

影响关键词排名的站内因素比较多，但是每一个因素都很重要。SEO 人员在优化网站时，应该重视每一个因素的优化。

2. 站外因素

站外因素主要包括单向外部链接和友情链接的优化，虽然站外因素相对于站内因素来说比较少，但是站外因素对于关键词排名的作用非常大，因此也需要重视。

3.2 网站结构优化

网站结构指的是网站中页面之间的层次关系。站在用户的角度考虑，能够使得用户在网站中浏览时可以方便地获取信息，不至于迷失。从 SEO 的角度来看，优化网站结构要达到以下几个目的。

1. 建立良好的用户体验

用户体验是检测一个网站质量好坏的标准，从根本上说，用户体验好的网站也是搜索引

喜欢的网站，而用户体验会影响网站的关键词排名。当用户在访问一个网站时，希望能够自如点击网站中的链接，找到自己想要的信息的，要做到这一点，有赖于良好的导航系统、适时出现的内部链接、准确的锚文本指示。

2. 提升网站收录

网站收录是 SEO 人员比较关注的问题，网站页面收录的多少在很大程度上依靠网站结构布局的好坏。从理论上讲，网站结构清晰，更容易吸引搜索引擎蜘蛛的优先关注，更有利于搜索引擎蜘蛛顺利的爬行抓取。

3. 便于页面权重的合理分配

除了外部链接能给网站页面带来权重外，网站本身的结构及链接关系是内部页面权重分配的重要因素。哪些页面具备比较高的排名能力，取决于页面得到的权重高低。网站的页面权重高低会影响到网站页面的排名。因此，好的网站结构对权重分配影响很大。

4. 锚文本的建设

锚文本是指把关键词做一个链接，指向别的网页。例如，图 3.2 中标注的就是一个锚文本，点击“出境长线”，即可跳转到“出境长线”页面。通过锚文本链接，可以增加页面与该关键词之间的相关性，因此锚文本也是影响网站排名的一个重要的因素。在实际优化网站过程中，网站内部的锚文本是 SEO 人员能控制的。

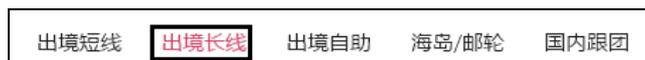


图 3.2 锚文本

网站结构优化的目的主要是督促 SEO 人员完成整个网站内部的优化，最终达到的目的是网站有好的排名，有更多的人浏览查看，有更高的转化率。

通常来讲，网站结构优化可以从导航系统优化、URL 设计、URL 标准化及 301 重定向、站内链接优化、robots.txt 与网站地图几个方面进行分析。

3.2.1 导航系统优化

网站导航，是为用户提供全面的引导，让用户快速进入相关栏目，从而找到相关内容的版块。当然网站导航中的各种链接实际上也为百度蜘蛛提供了抓取的入口。

1. 为什么要优化网站导航

- 网站导航决定用户在网站中穿梭浏览的体验，促使用户能够在网站中继续浏览下去。
- 网站导航设计合理，可以将网站的内容和产品尽可能多地展现在用户面前。
- 合理的网站导航可以增加用户粘性，提高网站的浏览深度，增加浏览量，降低网站的跳出率。
- 清晰大方的网站导航更容易获得用户的好感，提高网站的回访率。
- 网站导航简洁，搜索引擎蜘蛛能最大限度的抓取网站，从而增加网站的收录量。

2. 如何优化网站导航

(1) 主导航栏目结构和名称要清晰

网站主导航就是放在网站最上部、网站的栏目或主要内容的导入链接。一般情况下，导航上的栏目或内容，是这个网站最主要的内容，也是除了首页外，站长最希望用户进入的页面，

图 3.3 所示是某旅游网站的主导航。



图 3.3 某旅游网站的主导航

要优化网站的主导航，首先，要减少导航链接的层次，做到链接结构扁平化。例如，途牛旅游网导航中“跟团游”栏目的链接为 www.tuniu.com/tours/，只有两层，结构非常清晰。

另外，主导航名称要清晰、醒目。主导航一般体现为一级栏目，通过它用户和搜索引擎蜘蛛都可以进入到下级栏目页。因此，主导航名称要能够高度概括下级栏目的内容，让用户和搜索引擎蜘蛛通过该名称就能够知道栏目介绍的是什么内容，而且尽量使用目标关键词。如图 3.3 所示，某旅游网导航中的“跟团游”栏目，介绍的是出境跟团、国内跟团、周边跟团等与跟团游相关的内容，导航名称直接使用跟团游，清晰、明了。

（2）使用文字导航

网站导航尽量使用文字形式不要使用图片形式，因为文字导航比图片导航能更好地传递权重。如果导航中的文字可以复制就是文字导航，如图 3.4 所示。也不要使用 flash、JavaScript 生成的导航系统，因为搜索引擎不能分析 flash、JavaScript 内的链接。



图 3.4 文字导航

一般情况下，网站导航的形式需要配合技术人员进行调整。SEO 人员在优化网站时发现导航存在问题，应及时与技术人员协商解决。

（3）网站导航包含关键词

导航系统中的链接通常是分类页面获得内部链接的最主要来源，数量很大。所以在选择导航链接关键词时要注意分类名称，尽量使用目标关键词。图 3.5 是某旅游网站导航，“旅游攻略”“自由行商城”等都是网站优化的关键词。在实际工作中，SEO 人员在确定了页面要优化的关键词之后，首先规划好网站导航栏目，然后在导航栏目中设置合适的关键词。



图 3.5 某旅游网站导航

需要注意的是，导航中不要堆积关键词，很多 SEO 人员喜欢在导航堆积关键词，来提升关键词排名。实际上这既不利于用户体验，搜索引擎也会察觉。

（4）面包屑导航的设置

面包屑导航是网站导航的一种，位于主导航以下，页面主要内容以上，用于展示当前页面在网站中的位置和层级，帮助用户返回到上级页面。图 3.6 所示为途牛网站的面包屑导航。



图 3.6 途牛网站的面包屑导航

面包屑导航的作用非常大，如通过面包屑导航，可以让用户了解当前页面的位置，清楚网站的架构层级，并且可以根据导航快速返回各个层级。另外，面包屑导航中的各个层级以锚文本的形式展示，这也增加了网站的站内链接结构，有利于搜索引擎蜘蛛对网站的抓取。

因此，在优化网站过程中，发现网站没有面包屑导航，SEO 人员可以配合技术人员为网站添加上面包屑导航。面包屑导航的层级关系，可以根据所优化网站的层级来设置。

(5) 尾部导航避免堆积关键词

在网站底部添加一个网站导航是很多网站采用的一种优化方式，图 3.7 是某旅游网站尾部导航展示。

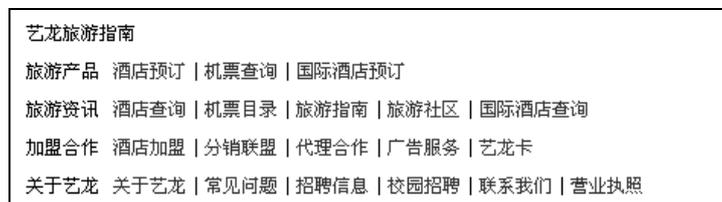


图 3.7 某旅游网站尾部导航展示

网站首页设计尾部导航有两方面的好处：

- 有利于提高首页以及其他页面导航关键词密度，更多链接指向导航中添加的页面，有利于提高这些页面的权重。
- 有利于用户体验，让用户随时都能找到重要导航的链接。

当然，尾部导航不需要把主导航的链接全部放上，只放一些最重要的即可，可以再放一些其他内容。如图 3.7 所示的艺龙旅游网尾部导航，添加的是网站一些重要栏目的入口。

因此 SEO 人员在优化网站时，可以为网站添加尾部导航，并适当植入关键词，但是要避免关键词堆积。

3.2.2 URL 设计

设计网站 URL 结构时需要对 URL 层级及文件命名系统事先做好规划。总的原则是首先从用户体验出发，URL 应该清晰、友好、方便记忆，然后才考虑对排名的影响。在设计 URL 时通常考虑以下几个方面。

1. URL 越短越好

URL 层级越短越有利于搜索引擎抓取收录。页面被收录是排名的前提，URL 层级越短对应的页面权重就越高，有利于提升搜索结果排名。

通常情况下，网站 URL 层级建议最多保留 3 个层级，以下面 URL 为例来分析。

优化前：<http://www.sina.com.cn/mobile/c/2017/08/29/07121212.html>，URL 层级是 7 层。

优化后：<http://www.sina.com.cn/mobile/c-2017-08-29-07121212.html>，URL 层级为 3 层。

对于网站 URL 层级过多的情况，SEO 人员需要在保证网站 URL 能够清晰表达网站结构的前提下，减少不必要的目录层级。URL 层级缩短操作方法请参照本章视频讲解。

2. URL 中包含关键词

URL 中出现关键词有利于提升关键词与页面的相关性，对于提升页面排名有一定的作用。关键词可以出现在域名中，也可以出现在路径中。

域名中包含关键词，可以是关键词本身，也可以是关键词的拼音或拼音首字母。如 www.toberp.com，深圳拓步软件公司，关键词是“erp”，在域名中出现了关键词。

路径中包含关键词，指的是路径中的目录名、文件名可以包含关键词的英文、拼音、拼音首字母等，图 3.8 所示是房天下家居装修网站的五金栏目，目录名称为关键词“五金”的拼音。

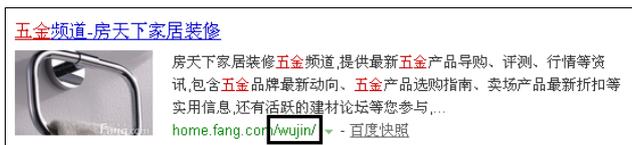


图 3.8 房天下家居装修网站的五金栏目

3. 连字符的使用

目录名或文件名中单词之间一般建议使用连字符“-”分隔，不要使用下划线或者其他符号。搜索引擎把 URL 中的连字符当做空格处理，下划线则被忽略。例如，文件名 `seo-tools.html` 会被读取为 `seo` 与 `tools` 两个单词，而文件名 `seotools.com` 就不能被分解出两个单词。

在优化网站时，发现网站路径名称中有关键词，就可以利用连字符将多个单词进行分割。需要注意的是，域名中尽量不要使用连字符。

4. URL 重写

URL 重写即把动态页面的 URL 转化成伪静态 URL 的过程。

动态页面指的是通过执行 `asp`、`php`、`jsp`、`asp.net` 等程序生成客户端页面代码的网页，后缀对应为 `.asp`、`.php`、`.jsp`、`.aspx`。图 3.9 所示为某网站的动态页面 URL。



图 3.9 某网站的动态页面 URL

静态页面一般是以 `.html` 为后缀的页面，不需要经过网站服务器的编译，比动态页面易收

录。图 3.10 所示为某网站的静态页面 URL。



图 3.10 某网站的静态页面 URL

伪静态 URL 是模拟静态页面的 URL，伪静态的页面后缀可以是.html、.htm 或者目录格式。例如，URL 重写前是 <http://www.xxx.com/c/userfile.aspx?id=1321>，URL 重写后是 <http://www.xxx.com/userfile-1321.html>。

需要注意，URL 重写只是修改了网址的格式和后缀，修改后的页面并没有变为纯静态页面。伪静态只是改变了 URL 的表现形式，实际上还是动态页面。

为什么要进行 URL 重写呢？动态 URL 不太容易被百度蜘蛛所识别，因此优化后的 URL 更易于百度蜘蛛的爬行和抓取，对于网站收录的提升有较大帮助。

一般情况下，有专门的技术人员负责将动态页面的 URL 重写成伪静态的 URL，SEO 人员需要做到的是能够判断出网站中是否存在动态页面 URL，并将其整理出来交由开发人员去处理。

3.2.3 URL 标准化及 301 重定向

在日常浏览网页过程中，你是否见过多个 URL 打开同一个页面的情况？例如：

<http://tour.eyuyao.com/>

<http://tour.eyuyao.com/index.php>

这两个 URL 打开的是同一个页面，图 3.11 是它们对应的网站页面。



图 3.11 两个 URL 对应的网站页面

当多个 URL 打开同一个页面时，会分散页面目标页面的权重，这时搜索引擎会对页面进行 URL 标准化处理。

1. URL 标准化

（1）定义

当多个 URL 对应的内容一样，搜索引擎要从多个 URL 中挑选一个具有代表性的，而忽略其余的，这个选择的过程称为 URL 标准化。

这个被搜索引擎选择了的 URL，就是一个标准化的 URL。一般情况下，搜索引擎根据被指向链接数量来判断标准化的 URL，被指向链接多的则为标准化的 URL。

（2）URL 未标准化会出现的问题及导致的后果

URL 未标准化可能会出现的问题有：从技术上讲，不同的 URL，服务器可以返回不同的内容，但是这种情况下返回的是相同的内容；从搜索引擎的角度来讲，几个不同网址对应的内容相同，搜索引擎认为是复制的内容。

URL 未标准化可能会导致的后果：外部链接指向不同的 URL，分散页面权重，不利于排名；搜索引擎判断的标准化网址不是我们想要的那个；复制内容过多，搜索引擎可能会认为有作弊的嫌疑。

要解决以上问题有很多方法，301 重定向就是其中一个。

2. 301 重定向

301 重定向的一个重要作用是集中权重，使首页权重不分散。另外，页面做 301 重定向之后，选中的目标页面是我们制定的页面，同时排除了重复内容的风险。

怎样做 301 重定向呢？

这里以 LAMP（Linux+Apache+MySQL+PHP）主机为例讲解 301 重定向的操作方法。

LAMP 主机使用 .htaccess 文件作 301 转向，.htaccess 是一个普通文字文件，存放在网站根目录下。这里以从 myweb.com 跳转到 www.myweb.com 为例进行讲解。

第一步：登录主机独立控制面板，在“常用管理功能导航”下选择“301 转向”，如图 3.12 所示。

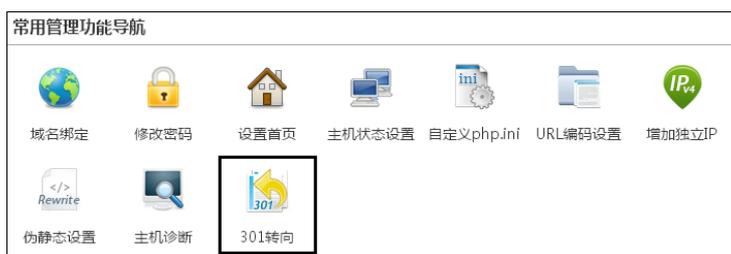


图 3.12 “301 转向”功能选项

第二步：确定原域名及目标域名（目标域名只能填写一个），会自动生成一段指令，如图 3.13 所示。之后将自动生成的一段指令复制到新建的文本文档中。

第三步：将文本文档上传到空间 wwwroot 文件夹下，并将该文本文档重命名为 .htaccess，如图 3.14 所示。

301 重定向的具体操作步骤，请参照本章视频解析。需要注意的是，尽量在网站上线前由技术人员来完成 301 重定向，设置完成后若没有特殊情况，则不需要改动。

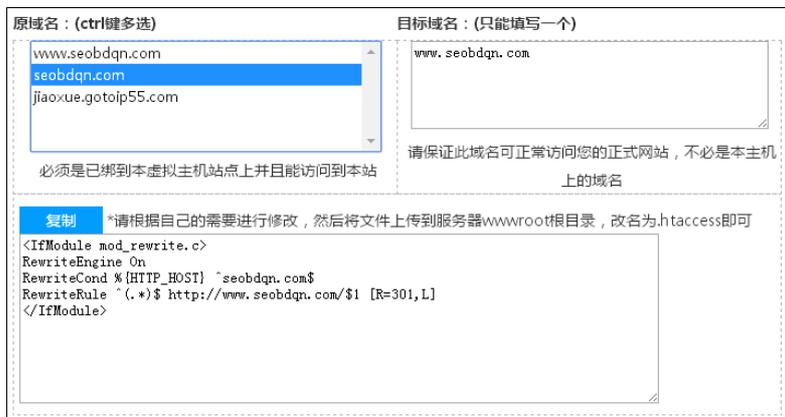


图 3.13 确定原域名及目标域名

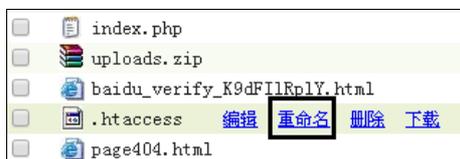


图 3.14 重命名“.htaccess”文档

除了网址规范化问题，还有哪些情况需要做 301 重定向呢？

需要用到 301 重定向的情况有很多，比如，网站更换域名，需要利用 301 重定向将旧域名跳转到新域名。网站改版也经常需要用到 301 重定向，如页面删除、改变地址、URL 命名系统改变等。

SEO 人员需要能够判断出何时要做 301 重定向，并配合技术人员来完成。

3.2.4 站内链接优化

在网站结构优化中，站内链接是很重要的一部分内容。因为站内链接设置不合理，对网站关键词排名影响很大。

1. 站内链接的作用

站内链接的作用可以概括为以下 4 点：

(1) 提升用户体验

如果能在相关页面设置对应页面的链接，浏览会变得非常顺畅，用户体验也会比较好。如果页面中链接设置不合理，有的页面甚至没有链接，只能重新输入网址才能访问其他页面，这种体验会非常差。

这也就告诉 SEO 人员，在优化网站时，想要让你的网站用户体验比较好，设置好网站的站内链接是其中一个方面。

(2) 增加网站收录

蜘蛛爬行网站时，并不是每次都爬行全站，而是有选择地爬行。同等情况下，站内链接设置合理的网站更容易被搜索引擎蜘蛛抓取到新内容。

因此站内链接是影响网站收录的一个重要因素，在优化网站过程中，如果网站内容收录

比较差，可以分析站内链接设置这一方面。

（3）提升页面权重

权重是靠链接进行传递的，站内页面之间的链接也可以传递权重。因此在优化网站时，对于重点优化的页面，可以合理地设置内部链接。

（4）提高页面关键词排名

权重是影响关键词排名的一个重要因素，站内链接在提升页面权重的同时，也有助于提升页面关键词的排名。

总体来看，站内链接对于网站优化的作用还是很大的，SEO 人员需要重视站内链接的优化，学会站内链接的优化。

2. 站内链接的优化方法

搜索引擎对互联网上的内容进行索引时，是按照树状结构抓取的。离“树干”越近，从“树根”获得的营养越多；离“树干”越远，获得营养的机会和数量越少。

搜索引擎就是这个道理，离根目录越近的页面，得到的权重越多，反之越少。所以网站内链应该设置成扁平化的结构，可以让网站的内页获得更高的权重。也就是说，网站内链在设置时可以参照扁平化的结构。

那么网站内链具体应该如何设置呢？

页面链接要做到通过每一个页面都可以到达网站内任意页面。虽然不同网站在设置站内链接时链接的位置不甚相同，但思路是类似的。

- 首页链接：网站首页要有链接指向重要栏目页和内容页。

设置网站首页指向重要栏目页的链接，例如可以在首页的导航、版块标题、尾部导航中添加重要栏目页的入口，这样就可以从首页进入到重要栏目页，如图 3.15 所示。



图 3.15 首页指向重要栏目页

设置网站首页指向内容页的链接，例如，可以将重要内容页推荐到网站首页展示，很多网站会有“推荐到首页”的功能，可以直接将重要的内容页推荐到首页，如图 3.16 所示。

- 栏目页链接：网站栏目页要有链接指向首页，指向其他栏目页，指向栏目内的内容页。

网站栏目页指向首页，可以设置的入口比较多，例如，在网站主导航中添加首页的链接，通常网站主导航全站通用，这样在网站任意页面都可以跳转到首页。另外，面包屑导航中的链接也可以返回到首页，如果网站中没有添加面包屑导航，可以添加上。



图 3.16 首页指向重要内容页

由栏目页指向其他栏目页，实现起来也比较简单，例如在主导航中添加其他重要栏目，就可以实现从栏目页指向其他重要栏目页。

因为网站栏目页一般是文章列表页，因此从栏目页指向栏目内的内容页，只需在栏目中更新内容即可。图 3.17 所示是栏目页站内链接的设置。



图 3.17 栏目页站内链接的设置

- 内容页链接：网站内容页要有链接指向首页，指向上级栏目页，指向栏目内相邻和重要内容页，指向其他栏目重要内容页。例如：

由内容页中设置链接指向首页，可以在主导航中添加首页链接入口，设置面包屑导航也可以实现，还可以在内容中添加首页的锚文本链接，在内容页面中首次出现关键词时，设置锚文本链接跳转到首页。

由内容页指向上级栏目页的链接设置方法与指向首页的设置方法类似。只不过在内容页中设置锚文本链接时，是跳转到上级栏目页。

由内容页指向栏目相邻和重要内容页以及指向其他栏目重要内容页时，可以在内容页中添加几篇推荐的相关文章，设置成锚文本的形式。如图 3.18 所示，网站内容页中添加相关推荐的文章以及上一篇下一篇功能。也可以与技术人员协商，添加相关推荐的栏目。

在站内链接设置过程中，还需要注意一些特殊情况。

(1) 重要内页

一般来说，网站首页获得的内外部链接最多，权重最高。首页链接到栏目页，这些栏目页权重仅次于首页。大部分网站有多层分类，权重依次下降，权重最低的是最终产品页面。



图 3.18 网站内容页中添加相关推荐的文章以及上一篇下一篇功能

有时某些具体产品页面需要比较高的权重，例如某些转化率最高、利润率最高或者新推出的重点产品、搜索次数很多的产品，还有为特定节日或促销活动制作的专题页面。这些页面按照经典树形结构安排，离首页通常有一定的距离，权重不会太高。要想使这类重要内页获得高权重，最简单的方法就是在首页上直接加上几个重要内页的链接。距离首页近的网页更重要，无论这个页面层级目录有多深，只要在首页有入口搜索引擎就会认为这个页面是重要的。SEO 人员需要明白，首页可以放上自己想要重点优化排名的产品页面，使这些页面权重提高。

一般情况下，将产品内页的链接放在首页上，哪怕没有其他外部链接的支持，这些内页的排名也会有显著提高。

(2) 非重要页面

每个网站都有一些在功能及用户体验方面很有必要，但是在 SEO 角度没有必要的页面，比如“关于我们”“添加收藏”“用户注册”等，这类页面通常不会去重点优化让其获得较高排名。

一般情况下，这类页面在整个网站每个页面上都会有链接，造成权重传递的浪费。因此，对于这类页面可以使其不能被跟踪或传递权重。例如使用 `nofollow` 属性可以避免搜索引擎蜘蛛抓取该链接，权重就不会传递。如图 3.19 所示，某网站“帮助中心”页面对关键词排名无作用，因此链接中添加了 `nofollow` 属性，避免权重传递的浪费。当然也可以使这类页面只在首页出现，在其他页面删除掉。

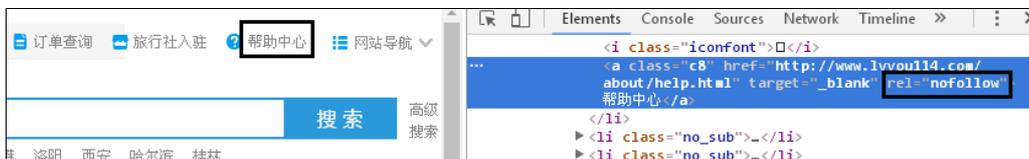


图 3.19 某网站“帮助中心”页面

实际工作中，SEO 人员面对的网站多种多样，如果遇到不确定的网站内部链接结构问题，抓住站内链接设置的核心：使搜索引擎蜘蛛以最短的距离爬行到最多的页面，最好结构清晰、简单地将权重传递到重要页面。

对于 SEO 人员来讲，需要对整个网站非常熟悉，知道网站中哪些页面是重点优化的，链接设置是否合理等。

3.2.5 robots.txt

在优化网站过程中，有些时候，网站中有重要及私密的内容，站长并不希望某些页面被蜘蛛抓取，比如后台的数据、测试阶段的网站。还有一种很常见的情况，搜索引擎抓取了大量没有意义的页面，如电子商务网站按照各种条件过滤、筛选的页面等。而要确保页面不被抓取，需要使用 robots.txt 文件。

robots.txt 是一个纯文本文件，用于声明该网站中不想被蜘蛛访问的部分，或者指定蜘蛛抓取的部分。当蜘蛛访问一个站点时，它会首先检查该站点是否存在 robots.txt 文件，如果找到，蜘蛛就会按照该文件中的内容来确定抓取的范围；如果该文件不存在，那么蜘蛛就会沿着链接直接抓取。即，只有在需要禁止抓取某些内容时，写 robots.txt 才有意义。

robots.txt 文件放置在一个站点的根目录下，而且文件名必须全部小写，正确的写法是 robots.txt。要查看某网站的 robots.txt 文件，在浏览器中输入的网址是 <http://www.xxx.com/robots.txt>，其中 www.xxx.com 是要查询网站的域名。

1. robots.txt 的语法

robots.txt 文件的常用语法有 3 种，分别是 User-agent、Allow、Disallow，下面讲解 3 种语法的具体用法。

(1) User-agent

指定 robots.txt 中的规则针对哪个搜索引擎蜘蛛。

针对所有搜索引擎蜘蛛的写法是 User-agent:*，通配符*代表所有搜索引擎。只适用于百度蜘蛛的正确写法是 User-agent:Baiduspider。不同的搜索引擎，其蜘蛛名称也不相同，表 3-1 所示是不同搜索引擎的蜘蛛名称。

表 3-1 不同搜索引擎的蜘蛛名称

| 蜘蛛名称 | 搜索引擎 |
|--------------|-------------------|
| Baiduspider | 百度 www.baidu.com |
| Googlebot | 谷歌 www.google.com |
| Sogou spider | 搜狗 www.sogou.com |
| Bingbot | 必应 cn.bing.com |
| Youdao Bot | 有道 www.youdao.com |

(2) Allow

允许搜索引擎蜘蛛抓取某些文件。例如允许蜘蛛访问网站中的/a/目录，正确写法是 Allow:/a/。

\$: 表示匹配 URL 结尾的字符。例如允许蜘蛛抓取以.htm 为后缀的 URL，写法是 Allow:.htm\$。

(3) Disallow

告诉搜索引擎蜘蛛不要抓取某些文件或目录。例如禁止蜘蛛抓取/admin/目录的写法是 Disallow:/admin/。

禁止的目录或文件必须分开写，每个一行，例如禁止所有的搜索引擎蜘蛛抓取/a/、/b/、/c/目录，正确的写法是：

```
User-agent:*
Disallow:/a/
Disallow:/b/
Disallow:/c/
```

在网站优化中，SEO 人员需要熟练掌握 robots.txt 的基本语法。下面分享一些常见的 robots.txt 语法使用案例，如表 3-2 所示。

表 3-2 robots.txt 语法使用案例

| 案例 | robots 代码 |
|---------------------------------|--|
| 禁止所有搜索引擎蜘蛛访问网站的任何部分 | User-agent: * Disallow: / |
| 允许所有搜索引擎蜘蛛访问网站的任何部分 | User-agent: * Allow: / |
| 仅禁止 Baiduspider 访问您的网站 | User-agent: Baiduspider Disallow: / |
| 仅允许 Baiduspider 以及 Googlebot 访问 | User-agent: Baiduspider Allow: / User-agent: Googlebot Allow: / User-agent: * Disallow: / |

2. robots.txt 应用

robots.txt 文件的一个用法是在 robots.txt 文件中指定 sitemap 的位置。具体的用法是 sitemap:http://www.seobdqn.com/sitemap.xml，这样就告诉搜索引擎蜘蛛这个页面是网站地图。

robots.txt 文件是搜索引擎蜘蛛进入网站后访问的第一个文件，在编写时确实有很多需要注意的地方，如果日常使用中不注意语法的正确使用，有些语句可能就发挥不了应有的作用，会影响搜索引擎对网站的访问，因此要正确编写。robots.txt 文件的编写方法请参照本章视频。

对于 SEO 人员来讲，在优化网站过程中，当发现网站中有隐私文件需要屏蔽搜索引擎蜘蛛抓取时，可以设置 robots.txt 屏蔽搜索引擎蜘蛛抓取这些隐私文件。而网站中没有要屏蔽的内容时，可以不添加 robots.txt。

3. robots meta 标签

如果搜索引擎已经收录网页，而这个网页是我们不想让搜索引擎收录的，robots.txt 文件解决不了这个问题，而 robots meta 标签却可以解决。

robots.txt 文件主要是限制整个站点或者目录的蜘蛛访问情况，而 robots meta 标签则主要是针对某个具体的页面。robots meta 标签放在页面中，专门用来告诉搜索引擎蜘蛛如何抓取该页的内容。

robots meta 标签的基本写法是<meta name="robots" content="index,follow">，其中有几项需要特别注意的内容。

(1) 在 robots meta 标签中, name="robots"表示所有的搜索引擎,也可以针对某个具体的搜索引擎,如针对百度搜索引擎可以写为 name="Baiduspider"。

(2) content 部分有 4 个指令选项,以英文逗号“,”隔开,分别是: index、follow、noindex、nofollow。

- index 指令告诉搜索引擎蜘蛛可以抓取该页面。
- noindex 指令与 index 指令相反,表示搜索引擎蜘蛛不可以抓取该页面。
- follow 指令表示搜索引擎蜘蛛可以爬行该页面上的链接。
- nofollow 指令与 follow 指令相反,表示搜索引擎蜘蛛不可以爬行该页面上的其他链接。

综上所述, robots meta 标签有以下 4 种组合:

```
<meta name="robots" content="index,follow">
<meta name="robots" content="noindex,follow">
<meta name="robots" content="index,nofollow">
<meta name="robots" content="noindex,nofollow">
```

当 robots meta 标签的 content 值为"index,follow"时,表示该页面可以被抓取,该页面上的链接也可以被继续爬行下去, robots meta 标签可以简写为<meta name="robots" content="all">。

当 robots meta 标签的 content 值为"noindex,nofollow"时,表示该页面不可以被抓取,该页面上的链接也不可以被继续爬行, robots meta 标签可以简写为<meta name="robots" content="none">。

robots meta 标签的具体使用方法请参照本章视频。

robots meta 标签是限制某个具体页面的蜘蛛访问情况,因此当发现网站中某个页面需要屏蔽蜘蛛抓取时,可以在该页面的源代码中添加 robots meta 标签, robots meta 标签添加在页面的<head>和</head>之间。

3.2.6 网站地图

网站无论大小,单独的网站地图页面都是必需的。通过网站地图,不仅用户可以对网站的结构和内容有清晰的认识,搜索引擎也可以跟踪链接爬行到网站所有主要部分。

网站地图分为两种形式,一种是 HTML 格式,一种是 XML 格式,二者有一定的区别,但是作用类似。

1. HTML 格式网站地图构建技巧

(1) 地图需要包含网站主要的链接

对于稍具规模的网站来说,一个 HTML 格式的网站地图不可能罗列出所有的页面链接,在设置网站地图页面时可以采取两种办法:一种方法是网站地图只列出网站最主要部分的链接;另一种方法是将网站地图分成多个文件,主网站地图列出通往次级网站地图的链接,次级网站地图再列出部分网页的链接。多个 HTML 网站地图页面加在一起,就可以列出大部分的页面链接了。图 3.20 所示是新浪网 HTML 格式网站地图,导航中添加了网站中重要内容的链接。



图 3.20 新浪网 HTML 格式网站地图

(2) 地图排版美观简洁

HTML 地图是给用户看的，HTML 地图虽然处在网站中不起眼的位置，但是万万不可忽视这样的细节，如果网站地图分类模糊不清，用户找不到想看的東西，有可能会跳出。如果分类太杂太多，也会影响用户的体验。一般来说，把主要的版块以及一些目录层次较深但有价值的页面展示出来就可以了。例如图 3.20 中所示的新浪网网站地图，页面排版比较简洁，类别划分清晰，方便用户浏览。

HTML 格式网站地图可以用程序生成，也可以手工制作和编辑。以 DedeCMS 网站为例，生成网站地图的步骤是：进入网站后台，单击左侧“生成”标签，选择“更新网站地图”选项，选中“普通地图”单选按钮，单击“开始更新”即可。DedeCMS 生成网站地图的操作界面如图 3.21 所示。

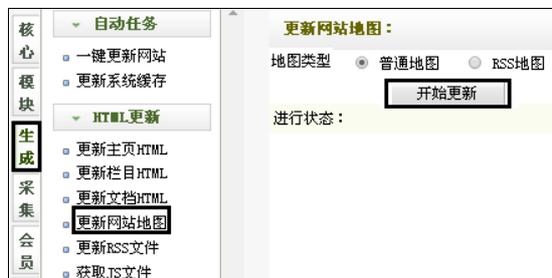


图 3.21 DedeCMS 生成网站地图的操作界面

2. XML 格式网站地图构建技巧

XML 格式网站地图由 XML 标签组成，文件本身必须是 utf-8 编码。最简单的网站地图可以是一个纯文本文档，文件只要列出页面 URL，搜索引擎就能抓取并理解文件内容。

XML 网站地图文件如图 3.22 中代码所示。

```
<?xml version="1.0" encoding="utf-8">
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>http://www.seobdqn.com</loc>
    <lastmod>2017-05-10</lastmod>
    <changefreq>daily</changefreq>
    <priority>1.0</priority>
  </url>
</urlset>
```

图 3.22 XML 网站地图文件

XML 网站地图一般都是由软件生成，比较程序化，可变化性较低。制作好网站地图文件之后，可以有两种方式通知搜索引擎网站地图的位置，一是在站长工具后台提交网站地图文件。另一种方式是在 robots.txt 文件中通知搜索引擎网站地图文件的位置，图 3.23 所示是在 robots.txt 文件中添加网站地图。

中大型网站结构相对来说比较复杂，蜘蛛在爬行时会有一定压力，因此中大型网站提交网站地图通常有比较好的效果，能使收录增长不少。HTML 网站地图不仅对于用户的网站体验是友好的，对于网站的优化也非常有利。对于小型网站来说可以没有 XML 网站地图，但不能没有 HTML 网站地图。

两种格式的网站地图生成方法，请参照本章视频。

```
User-agent: *
Disallow: /plus/ad_js.php
Disallow: /plus/advancedsearch.php
Disallow: /plus/car.php
Disallow: /plus/carbuyaction.php
Disallow: /plus/shops_buyaction.php
Disallow: /plus/erraddsave.php
Disallow: /plus/posttocar.php
Disallow: /plus/disdls.php
Disallow: /plus/feedback_js.php
Disallow: /plus/mytag_js.php
Disallow: /plus/rss.php
Disallow: /plus/search.php
Disallow: /plus/recommend.php
Disallow: /plus/stow.php
Disallow: /plus/count.php
Disallow: /include
Disallow: /templates
Sitemap:http://www.seobdq.com/sitemap.xml
```

图 3.23 在 robots.txt 文件中添加网站地图

3.3 案例分享

案例 1：某新闻行业网站 URL 优化

【案例描述】

北京搬家信息网，网址：<http://www.92bjw.com/>。

请分析该网站 URL 设计中存在的问题，并给出解决方案。

【案例分析】

对网站 URL 进行优化时，可以从四个方面进行分析优化：URL 中是否包含关键词、URL 层级是否合理、URL 中连字符的使用是否合理、是否包含动态 URL。根据这四个方面进行分析，可以发现该网站在 URL 优化方面主要存在两大问题。

1. 包含动态 URL

例如网站中“联系我们”的 URL 为：<http://www.92bjw.com/about.php?part=aboutus&id=5>，包含动态参数，属于动态页面。

优化方法：将动态页面重写成伪静态页面，例如将上述链接改写为：<http://www.92bjw.com/about-5.html>。

2. 部分栏目 URL 中不包含关键词

例如网站“北京搬家/货运”栏目路径为：<http://www.92bjw.com/info.php?catid=155>，不包含关键词。

优化方法：可以将 URL 的路径名称改为关键词的拼音、英文、首字母缩写等。例如可以将上述链接改写为 <http://www.92bjw.com/bjhy-155/>，其他栏目页面的 URL 可以参照该方法进行修改。

该网站中动态页面以及 URL 的路径中不包含关键词的页面有很多，在实际优化过程中，还需要 SEO 人员将所有存在问题的页面的 URL 查找出来，然后进行统一的修改。

案例 2: robots.txt 文件应用解析

【案例描述】

很多站长在编写网站 robots.txt 文件时，会将不希望被搜索引擎抓取的目录或内容添加到 robots.txt 文件中。这种做法本身无可厚非，但是如果编写时语法使用不当，得到的结果就会大相径庭。

【案例分析】

我们重点分析 Allow 和 Disallow 的顺序问题。

某个网站的 robots.txt 文件如下：

```
User-agent:*
```

```
Allow:/
```

```
Disallow:/abcd/
```

从这样一段代码中可以理解到，肯定是想屏蔽“abcd”文件夹，不让搜索引擎蜘蛛爬取这个文件夹中的文件。但是经过分析却发现，这段语句最终的结果和预期的效果却是大相径庭的，搜索引擎蜘蛛仍然会爬取“abcd”文件夹。

原因就在于搜索引擎蜘蛛从 robots.txt 文件中读取内容的规则是从上而下的，下面写的规则不能打败上面所写的规则。如果按照影响范围来解释上面的情况，那就是因为已经设置了“Allow:/”，那么对蜘蛛的影响就是全局性的，允许蜘蛛访问这个网站的所有文件，而在第三行中，设置“Disallow:/abcd/”之后，就是在前一个影响范围之中进行的，本身“abcd”文件夹就是在网站目录中，也受到了第二条的影响，所以，第三条是无效的，因此这段 robots.txt 文件的写法中，蜘蛛仍然可以抓取“abcd”文件夹。

再如：

```
User-agent:*
```

```
Disallow:/abcd/
```

```
Allow:/
```

这种写法仅仅是将顺序颠倒了一番，蜘蛛就不能访问“abcd”文件夹了，如果看懂了上一个例子的解释，那么这个例子就比较容易理解了。这个例子中，“Disallow:/abcd/”在前面，由于限制的出现，它所影响的是“abcd”文件夹，在第三行中“Allow:/”影响的范围本来是全局的，但是这条语句对蜘蛛的影响不能妨碍前一条语句。蜘蛛本来想通过第三条语句访问全部的目录，但是第三条语句的权限打败不了第二条语句，所以就只能访问除了“abcd”之外的文件夹。

章节总结

本章主要介绍了以下内容：

- 搜索引擎的排名机制：相关性和权重。
- 影响关键词排名的站内因素：网站结构、网站代码、关键词布局及密度、网站内容。
- 影响关键词排名的站外因素：单向外链、友情链接。
- 网站结构优化的目的：建立良好的用户体验、提升网站收录、便于页面权重的合理分配、锚文本的建设。

- 网站结构优化的要点：导航系统优化、URL 设计、站内链接优化、robots.txt 与网站地图。
- 如何优化网站导航：主导航栏目结构和名称要清晰、网站导航包含关键词、网站导航关键词的位置、面包屑导航和子导航的设置、避免页脚堆积。
- URL 的组成：传输协议、服务器地址、端口号、路径、询问。
- URL 的设计规则：URL 越短越好、URL 中包含关键词、连字符的使用、URL 重写。
- URL 标准化：当多个 URL 对应的内容一样，搜索引擎要从多个 URL 中挑选一个具有代表性的，而忽略其余的，这个选择的过程称为 URL 标准化。
- 301 重定向：301 重定向也叫 301 转向，301 跳转，是用户或蜘蛛向网站服务器发出访问请求时，服务器返回 http 数据流中头信息（header）部分状态码的一种，表示网址永久性转移到另一个地址。
- 站内链接：站内链接也称为内链，是指同一域名下页面之间的链接。
- 站内链接的作用：提升用户体验、增加网站收录、提升页面权重、提高页面关键词排名。
- 站内链接的优化方法：页面链接要做到通过每一个页面可以到达网站内任意页面。
- robots.txt：robots.txt 是一个纯文本文件，用于声明该网站中不想被蜘蛛访问的部分，或者指定蜘蛛抓取的部分。
- robots.txt 的作用：防止私密或重要内容被搜索引擎抓取、节省服务器资源，从而提高服务质量、减少重复抓取，提高网站质量、指定 sitemap 文件位置。
- robots.txt 的语法：User-agent、Allow、Disallow。
- robots meta 标签：robots meta 标签主要是针对某个具体的页面。robots meta 标签放在页面中，专门用来告诉搜索引擎蜘蛛如何抓取该页的内容。
- 网站地图：网站地图又称 sitemap，是一个网页，里面放置了网站上需要蜘蛛抓取的重要页面的链接。

作业

一、简答题

1. 简述站内链接的作用。
2. 简述 301 重定向的作用及原理。

二、实战题

1. 为网站域名设置 301 跳转，将不带 www 的域名跳转到带有 www 的域名。
2. 将网站的 URL 层级缩短至 3 层。
3. 为网站编写 robots.txt 文件，并将网站地图地址添加到 robots.txt 文件中。