**腾讯SOSO/搜狗搜索优化指南**

**前言**

　　搜索引擎优化，即SEO(Search Engine Optimization)，指为了提升网站/网页在搜索引擎搜索结果中的收录数量和排序位置，为了从搜索引擎中获取更多免费流量、高质量用户，针对搜索引擎的检索特点、排序规律，合理调整优化网站设计和建设方法，使其符合搜索引擎的检索规则的网站建设、网站运营行为。除此之外，SEO工作还有利于搜索引擎快速收录新页面，提高收录覆盖率的效果。

　　SEO是网络营销的重要手段，通过了解搜索引擎工作原理和用户需求，在尊重搜索引擎用户价值的前提下，对网站进行合理优化，可以获取更多的用户、流量和品牌。SEO具有成本低、易操作、有益于用户体验等特点，是网站自我营销时必须重视的一个环节。

　**搜索引擎简介和SEO基础**

**搜索引擎及其用户价值**

　　搜索引擎，是对互联网的数据信息进行收集、整理，并提供检索查询服务的系统。搜索引擎擅长理解用户搜索需求、对海量数据抓取、信息甄别和打分、快速查询等工作。

　　搜索引擎根据用户的检索查询要求，对数据信息进行打分排序，并按照顺序输出呈现，帮助用户方便、快速、准确、全面地找到所需信息，这是搜索引擎最重要的用户价值。

　　[提示]一切SEO工作，都应以遵循搜索引擎工作原理和用户价值为前提。

　**白帽SEO和黑帽SEO**

　　白帽SEO，指通过合理优化网站，提升用户体验，与搜索引擎良好交互，真实反映网站质量和权威性的搜索引擎优化方法。

　　黑帽SEO，以搜索引擎禁止的方法影响搜索引擎排序(作弊)，与用户价值和搜索体验相背离，以短期利益为主要目标的搜索引擎优化方法。搜索引擎会维护互联网网站的公正排序，给用户提供正确合理的搜索结果，而黑帽SEO的做法与此相违背。

　　搜索引擎在不断通过技术革新、策略优化来识别作弊行为。任何利用搜索引擎的策略缺陷漏洞，利用恶意手段获取与网页质量不符的排名，引起用搜索结果质量和用户搜索体验下降的行为都可能被搜索引擎识别为作弊行为。对待作弊行为，轻则被搜索引擎剥夺作弊权重，重则可能整个网站被搜索引擎拒绝收录。

　**搜索引擎工作流程图**



　　[提示]对外，面向用户和整个互联网内容;对内，分为抓取、索引、检索几个重要模块。

**搜索引擎爬虫**

　　搜索引擎的爬虫程序(也叫Spider、蜘蛛、爬虫机器人)，是搜索引擎收集信息、数据的工具。爬虫程序不断发现和下载互联网网页，这个过程是互联网网页进入搜索引擎内部必须经历的一关。

　　爬虫程序擅长：分配下载资源，海量并发下载，读取文字(特别是网页文本)，分析站点地图并依此指引抓取页面。相反爬虫程序不擅长：读取图片、flash、JS、Ajax、frame，深层次、持续性抓取，读取有权限控制的页面，发现孤岛页面。

　　针对爬虫程序的SEO工作应该根据其特点针对性进行。



　　[提示]搜索引擎爬行轨迹示意图

　　**搜索结果排序**

　　搜索引擎通过理解用户搜索词、对数据进行检索，可以得到诸多可能的搜索结果。对这些结果的排序要靠成百上千的相关性策略实现。

　　相关性排序的目的是将结果按照对用户价值的高低排列，整个排序过程会参考网站、页面的相关度、质量、时新性、权威性等诸多因素综合考察。

　　SEO工作很大一部分跟搜索引擎相关性策略有关，简单来讲即通过SEO来告诉搜索引擎：哪些页面更有价值，哪些页面更搜索词更相关，而应该被排序到更靠前的位置。

　**网站建立**

**域名注册**

　　域名是网站的重要品牌印象。域名注册时应选择简单、易记、方便输入、贴近网站名称、能反映网站内容的域名。如fanli.qq.com、www.paipai.com。

　　域名一旦选定，不应随意变更;同一网站拥有多个域名应确定主要域名，并设置其他域名向主要域名跳转。

　　**服务部署和网站上线**

　　网站需要部署上线才能被用户访问。

　　租用/购买安全可靠、高速稳定的服务器部署网站，是用户访问体验的有效保障，同时会影响搜索引擎爬虫的爬取。

　　网站部署上线并进行域名解析后，即面向用户开发访问，同时也意味着向搜索引擎爬虫开放。

　**让搜索引擎发现网站**

　　新建立的网站是互联网中的孤岛，搜索引擎也很难发现和收录它。在网站建立之初，可以将网站推荐给搜索引擎或者引导爬取收录。

　　向搜索引擎提交网站域名或入口。如向SOSO/搜狗的提交新网站，http://help.sogou.com/addurl.html;

　　建立外链，如创建友情链接、向导航网站申请收录等;

　　网站建立之初可重点经营内容列表页，在不影响用户体验的前提下，方便搜索引擎爬虫一次发现更多链接。

　　**内容建设**

**树型扁平网状结构**

　　层级分明的树型扁平网状结构是较理想的网站部署结构，每个页面与其父子页面有链接关联，可以帮助用户快速定位到感兴趣的频道、正文，也有利于搜索引擎理解网站结构层次和更好的爬取内容。



　　明确区分出，网站首页、列表页/索引页/频道页、内容页，根据各个页面所在层级用链接相连，无孤岛页面;

　　层级、各层级分支内容不要过多，便于搜索引擎爬虫爬取;

　　内容和质量相似的页面组织在同一层级/目录中，便于搜索引擎进行聚合挖掘;

　　越重要的内容，越应放置于高层级位置，或在高层级页面中有链接直接可达;



　　[提示]腾讯网新闻频道，头条新闻放置于醒目位置

　　在页面中建立导航链接、面包屑，便于用户和爬虫理解网站结构和在不同层级内容间跳转;

　　站点中新的内容应该尽量在深度较浅的列表页或首页发布入口，便于被搜索引擎爬虫快速发现。

　**页面标题**

　　页面标题是页面内容的重要标识，准确简洁的标题有助于用户快速识别页面内容。页面标题会被搜索引擎展现在搜索结果列表中，是用户通过搜索进入网站页面的重要入口。

　　标题内容在页面代码中被标签包围，会被展现在搜索结果页、浏览器标题、页面标签、任务栏等位置，是用户识别/区分页面的重要依据;



　　标题内容是页面的最佳概括，应当主题明确，突出重点，内容不宜过长;

　　每个页面都应该拥有标题内容，避免不写标题，或无意义标题(如“无标题文档”、“网站建设中”);

　　避免不同页面使用相同标题(甚至全站公用相同标题);

　　不在标题内容中堆砌关键词，避免影响可读性和被判作弊;

　　“标题党”、“文不对题”对用户搜索体验和搜索引擎收录排序均有副作用，应坚决杜绝。

　　**Meta信息**

　　Meta信息中，keywords和description分别以关键词和概括摘要两种方式描述页面内容，会被搜索引擎特别关注。

　　keywords部分应填入最能表现页面内容的几个关键词，词与词之间以半角逗号间隔开;

　　description部分应填入一段介绍页面内容的文字，这段文字有可能被搜索引擎选取作为结果摘要。网站首页、索引页、频道入口等无大篇幅页面正文的页面，应该尽可能提供description信息。

　　keywords和description信息缺失不会对排序权重产生直接影响;在keywords和description信息中堆砌关键词作弊可能受到搜索引擎惩罚。

　　腾讯网的description信息：



　　腾讯网在网页搜索结果中的摘要展现：



　　**链接**

　　制定规范统一、简洁、可读的url，有利于搜索引擎的收录和排序，也会方便用户识别页面内容，记录和输入链接。

　　每个页面都应拥有其独一无二的正常态url：

　　网站内统一使用正常url，不出现异常url;

　　指向同一页面异常url，应跳统一跳转到正常url上;

　　必须带有参数的url需要精简参数个数和内容，对页面功能/展现无作用(如统计参数)的应尽量去除。

　　链接结构设计建议：

　　url结构的设计应同时考虑用户和搜索引擎的需求，方便用户识别、记忆，并利于搜索引擎爬取和展现;

　　url应尽量短。较短的url会在搜索结果页完全展现，方便用户通过读取其以了解页面内容，也易于记忆和传播。搜索引擎本身不会对url进行限制，排序跟url长度无关;

　　url目录层级越少越好。层级过多的url设计可能对搜索引擎的爬取形成障碍，使得收录变少或不收录;

　　url中目录名、文件名应具有描述性，不使用无意义的名称;

　　url中尽量只使用小写字母或数字，特别注意不要使用特殊字符，如“.”“,”“;”等;

　　去除url中不必要参数和静态化。无论动态或静态url ，搜索引擎均可正常爬取和检索，但建议尽量精简去除url中不必要的参数，特别是与页面功能/展现无关的参数。

　**锚文本设计建议**

　　链接应选择文字作为链接描述，即锚文本，尽量少使用图片作为链接内容，必须使用图片时，应同时使用alt标记补充文字描述，不要用flash、js替代文本链接(特别是重要入口);

　　锚文本应简洁精炼，明确描述页面内容;不要在锚文本里编写与页面内容无关的文字，更不要堆砌关键词或者“标题党”;

　　非必要的情况下，不要使用url本身作为锚文本。

　**链接设计GoodCase**

　　http://finance.qq.com/stock，http://t.qq.com/liuxiang，子域名、目录描述了页面主题或内容;

　　http://finance.qq.com/a/20110317/005789.htm，新闻页面目录标识了发布日期;

　**链接设计BadCase**

　　链接url为js生成，搜索引擎无法爬取到链接指向页面。





　　http://7.sixjoy.com/webplat/info/news\_version3/87/97/98/99/m51/201103/2763.shtml，链接陷阱/黑洞，这类url目录构造对用户和搜索引擎均无意义，搜索引擎可能不会爬取或者不索引这类url。

　　锚文本内容无意义。类似“更多”“返回”的文本对页面内容的描述没有贡献，不建议使用。



　　首页链接全部在flash内部，无法获取更多站内链接，同时结果页摘要显相关较差。



　　全站页面标题全部相同，无法体现页面内容，在搜索结果和浏览器中均不易让用户识别，同时搜索引擎无法通过标题了解页面内容;



　　**页面内容和服务**

　　提供高质量的网站内容和服务是网站的立足之本，同时也是影响搜索引擎对网站内容评价的重要因素和指标。

　　提供围绕网站主题或页面主题的内容或数据，自然、美观的书写页面内容，满足用户的阅读需求，这是对用户最有价值的服务：

　　给用户提供便于阅读的页面。页面布局合理，突出展现重点内容，内容间有主次、段落，避免一大堆文字的粗暴罗列;

　　在页面中放置面包屑导航。面包屑对用户辨识页面作为位置、层级有帮助，还能快速在不同频道间切换。SOSO音乐/SOSO问问/腾讯网新闻频道的面包屑导航：

　　提供与众不同的原创内容或服务，避免抄袭内容或重复发布相似、相同内容;

　　与主题不相关或弱关性的内容，如广告、站内推广链接等，应放置于页面次要内容，并且不占用过多篇幅;

　　不要有拼写错误或错别字。

　　同时，页面内容建设也照顾搜索引擎读取页面的“感受”：

　　页面正文用文本书写，不要用图片、flash代替，或用ajax生成;

　　站内导航和面包屑易于搜索引擎爬取页面、获取锚文本和分析链接权重;

　　使用图片作为内容时，为图片添加alt信息，图片文件使用简洁直白的命名(类似给页面的命名);

　　不要使用iframe构造页面。

　　**robots.txt和Sitemap**

　　robots.txt和sitemap是站长和搜索引擎交流的工具，为搜索引擎的爬取、收录、排序行为提供参考。robots.txt和sitemap均不涉及对用户体验的直接影响。

　　robots.txt：

　　robots.txt文件，需要准确命名为”robots.txt”(文件名小写，扩展名为txt，)，放置于网站域名的根目录下，能被正常访问;

　　可以使用robots.txt限制搜索引擎爬虫访问某些页面、目录或者整合网站。如供网站管理员使用的在线管理系统，或用户隐私信息页面，均可以此方法控制;

　　可以在robots.txt指定sitemap地址。

　　User-agent: \*

　　Disallow: /admin/

　　Disallow: /ads/

　　Sitemap: http://www.soso.com/sitemap.xml

　　[提示]以上是一段robots.txt文件的示例。标识了拒绝搜索引擎访问的路径和sitemap地址

　　Sitemap：

　　Sitemap中可以指定url的更新时间、更新频率、相对权重等信息;

　　重要的url应该出现在Sitemap中，给予较高权重;

　　搜索引擎会参考Sitemap中的信息进行爬取、索引和排序，但不会完全遵守;

　**其他**

　　搜索引擎在响应网站内容变化需要时间，特别是针对网站内容建设的SEO是逐渐生效的，改变网站几个页网页内容就能迅速提高网站排名的观点是错误的;

　　搜索引擎对于不符合规范的网站建设方式或者过度SEO有一定容忍能力，并不会因为网站建设不当而直接拒绝爬取、收录、索引某个网站;

　　面向SEO的网站建设策略和用户体验应兼顾，甚至更多考虑用户体验，避免为了SEO而SEO。

　　网站运营和推广

　　面向用户的网站运营和推广工作，可以起到引入和沉淀用户、建设网站口碑、增强用户黏性的作用。而然，网站运营和推广同样可以在SEO工作中发挥作用，下面从这个维度来简单介绍。

　**内容运营**

**原创内容建设**

　　网站建设以内容为王，在内容质量、原创、全面、权威等方面有优势的网站，更容易赢得用户好感，同时也会赢得搜索引擎收录和排序的倾向。

　　搜索引擎非常喜欢原创内容(特别是对于新站来说)在网站上主动发布或引导用户来发布原创内容，增加原创内容比例，会让搜索引擎更快收录，同时在内容排重、排序上占得优势。

　　**规律的内容维护**

　　规律的内容更新时间和更新规模，有助于搜索引擎合理调度爬虫进行抓取。更新较快较多的网站，也容易赢得搜索引擎爬虫的高频率光顾。

　**广告内容运营**

　　广告能为网站带来收入，是网站运营不能回避的运营工作。从SEO角度看来，也有些应注意的优化点。

　　网站内容和广告内容应区分出主次，突出页面核心内容，利于搜索引擎理解页面内容，而不应满屏广告;

　　广告内容应配合页面内容，与页面内容、主题相关(搜索引擎爬虫可能采集广告内容作为页面正文);

　　不要将广告内容伪装成页面正文或链接，诱导点击。

　　**增加外链和积累有效锚文本**

　　站外链接和锚文本是网站的重要SEO资源，承担着入口、页面内容识别、权重引入等诸多参与搜索引擎爬取、排重、排序工作的信息。对外链和锚文本的建设，是SEO工作中极重要的一环。

　　以下是一些外链、锚文本建设的建议：

　　互换友情链接、申请导航站收录;

　　建设外链资源时，应合理设置锚文本，尽量不要以url做锚文本;

　　不止网站首页需要外链，二级列表页、专题页、频道首页、特色页面等，同样需要外链建设;

　　可以发布网站功能特色介绍的文章，期间包含网站外链;

　　博客博文、微博、论坛帖子中出现的网站链接，对外链和锚文本建设有益，可将网站介绍文章转发于此，但不要泛滥建设;

　　博客评论区、留言板、bbs签名档等非主要区域的外链价值极低，可能被搜索引擎忽略;

　　外链、锚文本应避免泛滥建设，否则可能被判为作弊。

　**用户推荐和评价**

　　用户有鉴别内容质量的能力，通过用户对网站内容的推荐和评价，可以帮助网站筛选好的内容，甚至制造外链。

　　建立用户对内容投票机制，获取用户评价;

　　提供方便的分享工具，可将内容推荐到其他产品平台，形成有价值外链;



　　[提示]一个常见的分享功能设计

　　培养用户黏性，并适当鼓励用户对朋友推荐网站内容(包括在IM、博客、论坛等平台上推荐)。