



# 网页质量白皮书

360站长平台 / 360搜索技术团队

联合发布

2017.10.26

# 目录 / Catalogue

前言 .....	3
网站价值判断标准	
1. 网站主题	
1.1 体现网站价值的主题 .....	4
1.2 符合法律法规及政策导向的原则 .....	4
2. 网站的稳定与安全	
2.1 网站访问的稳定性 .....	4
2.2 网站安全 .....	5
网页质量评估维度	
3. 内容质量	
3.1 不同类型网页的质量关注维度 .....	6
3.2 网页内容质量现状与收录的权衡 .....	7
3.3 内容质量好的网页 .....	8
3.4 内容质量一般的网页 .....	8
3.5 内容质量差的网页 .....	9
3.6 作弊网页 .....	9
4. 用户体验	
4.1 广告对主体内容的干扰度 .....	11
4.2 内容整洁度 .....	11
5. 网页资源可访问性	
5.1 具有打开及获取权限的网页 .....	11
5.2 死链及失效页面 .....	12
总结与建议 .....	14

# 前言

在过去的一段时间，搜索引擎的策略保守且封闭。站长们往往通过试错的办法，试探搜索引擎对网站、网页的质量判定标准，但这种尝试可能只会取得短期的正向效果。而各种针对搜索引擎刻意作弊的手段，对用户不友好的行为也开始大量出现。

网页质量，是网页在设计、版式、内容等维度上对用户需求是否满足的一个衡量标准。基于搜索引擎生态及用户体验的大背景，360搜索制定了一套判断网页质量的标准。在这个标准框架下，360搜索对高质量、原创并且符合用户需求的网页内容，会给予收录和排序的机会。同时，对一些存在安全风险，针对搜索引擎作弊，影响用户体验及需求满足的低质量内容，在收录与排序上进行控制！

此次推出的《360搜索网页质量白皮书》（1.0版本），旨在公开360搜索在网页质量方面的判断标准，为站长们提供建议指南，指导网站的内容建设，进而促进更多高质量原创内容的产生，构建良性的互联网内容生态。我们愿意与站长们携手，在满足搜索用户需求，为网站带来流量的共赢基础上，共同为广大用户服务，让优质网站更好更长远地发展！

## ● 网站价值判断标准

### ● 网站主题

#### 1.1 体现网站价值的主题

网站在建立伊始，一般都会确定后续的内容主题方向，比如图片站、企业站等。然后再根据这个主题方向，扩充相关的内容。

360搜索建议：网站应紧扣自身所在的行业领域及网站主题框架，产出符合用户需求的优质内容，形成自己的特色，从而体现出自身的价值。我们非常不建议网站通过大规模批量采集的方式，摘录其他网站的内容来扩充自己的网站，对于这种情况，我们会视情况降低其收录量和排序，甚至直接拒绝收录！

但是，目前也存在着一个不可否认的现状：随着网站数量的增长，网站同质化问题越来越严重，例如：新闻资讯站、小说站、商情站的数量非常多。一些内容在互联网中重复性太高，可替代性很强。建议站长朋友们在选择网站内容主题方向时，可以优先考虑小众，并且可深入挖掘内容的门类，在内容上与其他网站形成差异化，做成更加精细化和专业化的网站！

#### 1.2 符合法律法规及政策导向的原则

网站的内容需和国家现行法律法规、政策导向、社会正向思潮保持一致。站长朋友们应坚决避免网站中出现各类政治有害、诈骗、博彩、虚假、色情敏感及灰色产业信息。涉违法信息的网站会被主管部门问责整改或处以行政、刑事处罚，违法类有害信息也是360搜索坚决抵制的一种类型！

## ● 网站的稳定与安全

### 2.1 网站访问的稳定性

网站存在访问异常，一般分为两种情况：

类 型	说 明
访问不稳定	是指因设置问题或故障，导致网站时而可访问，时而又无法访问的情况。
域名失效 / 无法访问	是指域名设置错误、停止解析、服务到期、故障等原因，而导致网站无法访问的状态。

360搜索会对存在访问异常的网站进行处理，具体来讲：

- 如果网站经常处在访问不稳定的状态（特别是首页都经常无法打开的），360搜索会降低网站的收录量及排序；
- 如果域名停靠或者无法访问，我们将直接删除网站在360搜索中的收录。

因此，我们建议：

- 建站初期：在遴选域名及空间服务器时，选择品质优秀、服务稳定的产品，并建议配备相应的容灾机制和应急方案，避免因服务故障导致网站访问异常；
- 运营过程中：不建议频繁变更网站设置，特别是与网站访问直接相关的服务器设置；做好网站可持续访问的准备，提前做好域名、空间续费等事宜，防止他人恶意抢注域名，给您造成损失；
- 网站关闭：做好网站关停准备及数据处理，并及时通过站长平台申请删除网站死链。

此外，我们在这里也建议各位站长，将[360搜索的蜘蛛IP](#)添加到网站的访问白名单中，保证360搜索能够正常抓取网站的内容。

## 2.2 网站安全

网站不安全，用户在访问网站过程中可能会遭遇中毒、受骗。

网站被黑，一般是网站对安全问题不受重视，从而导致大规模被挂马和被篡改。在搜索引擎结果里，被黑的网站往往会有以下体现：

- 在搜索引擎中，网站标题摘要被篡改成博彩、色情、违法信息，如下图所示：



图1 网站被黑导致标题摘要被篡改

- 使用 site 语法，可发现域名下出现大量垃圾子域名或垃圾网页；
- 通过搜索引擎的访问请求，会被强制跳转到博彩、色情等非法内容站点；
- 用户去点击这些网站时，可能还会出现被某些安全软件提示报毒等。



图2 访问被黑站点时被提示风险

360搜索对于网站是否安全的态度：

- 对于网站内容优质、访问安全的网站，我们会优先考虑收录；
- 对于不安全网站，我们会采取降低排序的办法，或直接拒绝收录。

建议网站完善风控机制，规范账号密码及信息管理的流程。及时备份网站数据。防范潜在的挂马、被黑的风险，防止让他人有机可乘。保障网站的稳定运行！

## ● 网页质量评估维度

网站是多个网页共同组成的集合。细化到单个网页的质量时，360搜索在一般会按以下几个维度进行综合衡量：

- 内容质量；
- 用户体验；
- 网页资源可访问性。

下面，我们向大家具体介绍这些维度：

## ● 内容质量

在评价网页的内容质量时，360搜索首先会考虑这个网页内容的优劣，以及对用户需求是否可以被满足。

### 3.1 不同类型网页的质量关注维度

网页的形态不尽相同，针对不同类型的页面，360搜索在质量上也会有不同的关注维度。我们在此举一些常见的例子，供大家参考：

- **网站首页**：标签是否完善、清晰；网页中的内容及链接是否可访问。
- **列表页、搜索结果页**：内容与网页主题、搜索词是否有相关性，并且内容清晰，符合用

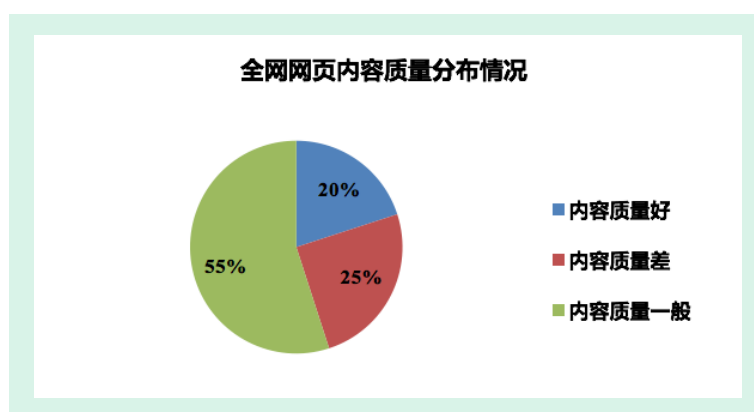
户的习惯。

- **文章页**：是否提供完整、真实的内容，图文并茂更佳。
- **新闻页**：具备新闻属性，对新近发生的事实有完整准确的报道，图文并茂更佳。
- **小说页**：能否提供清晰可阅读的小说内容，无错章，章节之间连贯，更新及时。
- **文档页**：用户是否可以正常阅读，是否有权限限制。
- **视频页**：是否清晰且能正常播放，是否有权限限制。
- **图片页**：图片是否清晰且能浏览，有文字信息介绍该图片中的内容更佳。
- **百科及问答知识页**：是否对某一事物、词语、现象进行了准确的解释，或具有参考价值的答案。带有图片或参考资料说明更佳。
- **以UGC为主要特征的网页（如：博客、自媒体、贴吧、社区帖子）**：是否能产出独具自身特色或观点的内容，内容完整丰富，具有一定的阅读价值。
- **商品页**：是否提供了完整真实的商品信息，是否提供了可成功购买入口。如有产品评价的内容会更好。
- **分类信息页**：是否提供了完整有效准确的信息（如黄页中的联系方式，地址，联系人，公司，产品信息是否准确有效），并且没有过期。
- **下载页**：是否提供了下载的入口，是否有下载权限限制，下载的资源安全且真实。

### 3.2 网页内容质量现状与收录的权衡

根据2017年7月发布的第40次《中国互联网络发展状况统计报告》：截止2017年月，中国网站的总数已达 506 万个，半年增长率为 4.8%。中国网民规模已达到 7.51亿。

目前，全网网页数量已接近万亿级别。网页内容的质量情况大体分布如下：



数据来自：360站长平台

通过数据我们可以知道：尽管网站、网页数量目前已经达到了一个惊人的数量级，但优质内容仍然较为稀缺。互联网中大部分内容都属于质量一般的内容，还有相当一部分内容是属于内容质量差的。因此，基于网页资源质量的实际情况，从用户需求和内容索引择优的原则出发，并非所有网页都会被360搜索收录。

### 3.3 内容质量好的网页

一个内容质量好的网页，往往是网站投入了一定精力来制作的。具备原创属性。资源有效、权威，能充分满足用户的需求，能彰显出网页及网站的价值。个别内容甚至是全网独家拥有的稀缺资源。

对于此类网页，360搜索会在一定程度上给予优先收录和排序的机会。我们列举了其中一小部分网页作为例子，供大家来感受一下：

示 例	说 明
<a href="#">case1</a>	政府网站一个权威且完整有效的通知。
<a href="#">case2</a>	新闻网站的原创内容，倾注了精力运营的新闻报道或专题。
<a href="#">case3</a>	医疗网站发布的内容丰富且权威的页面。针对感冒,提供了问答、文章、症状、诊断和鉴别等内容，可满足用户的需求。
<a href="#">case4</a>	经过网友用心编辑，内容丰富的 UGC 内容。
<a href="#">case5</a>	房产网站的一个丰富且内容有效的列表页。
<a href="#">case6</a>	一个 IT 技术类博客文章，内容完整且能解决专业技术性问题。
<a href="#">case7</a>	购物网站一个包含商品信息、评价的完整页面。
<a href="#">case8</a>	汽车网站提供的车系聚合信息页,包含了汽车价格,配置参数,经销商,评价等全面且多样的信息。

### 3.4 内容质量一般的网页

相较于内容质量好的网页，这种类型的网页往往可能不是网站自身原创的，也能在一定程度上满足用户的需求，但本身网页中的内容可替代性较强。对于此类网页，360搜索会综合其资源可替代性、需求满足程度、用户体验等维度考虑其是否收录及排序高低。



### 3.5 内容质量差的网页

这种类型的网页，往往主体内容不完整，缺失有效信息，无法从实质性上解决用户的需求。请站长们尤其需要注意：如果在一个网站中此类网页占比很大，不仅影响用户的体验，也会影响到360搜索对这个网站或网页的评价。

提供一些 badcase 例子，供站长们参考：

示 例	说 明
<a href="#">case1</a>	网页主体内容缺失。
<a href="#">case2</a>	交易、团购、活动结束，商品下架，招聘失效等，不能满足用户当前时效需求。
<a href="#">case3</a>	问答类网页存在有问无答、回答无用、回答推广等特征。
<a href="#">case4</a>	差标签页、差搜索结果页，列表页的标题和内容相关性差。
<a href="#">case5</a>	视频页标题文不对题，还带有推广的性质。
<a href="#">case6</a>	网页中的文字、链接、图片、视频或必要的下载链接失效。
<a href="#">case7</a>	论坛水贴。
<a href="#">case8</a>	通过新闻网站、博客等渠道发布的医疗推广内容，以及差医疗站点。
<a href="#">case9</a>	网页或文章的标题与内容相关性差，或者毫无关系。

### 3.6 作弊网页

此类网页往往有作弊行为，内容粗制滥造，完全不能为用户提供任何帮助，甚至还夹带病毒，没有任何价值。360搜索会依据实际情况，做降低排序和直接拒绝收录的处理！

在此，我们列举一部分例子，请各位站长知悉：

示 例	说 明
<a href="#">case1</a>	在正常文章的句子或段落之前穿插其他无关联的关键词或锚文本。

示例	说明
<a href="#">case2</a>	在网页或标题中刻意的大量重复关键词或锚文本，进行恶劣优化，提高关键词密度，从而欺骗搜索引擎。
<a href="#">case3</a>	网页打开跳转到其他带有明显推广盈利性质的页面，如博彩、色情等，或跳转到其他明显作弊特征的页面。
<a href="#">case4</a>	把转载的文章进行再加工，欺骗搜索引擎，让其认为这是一片原创文章。
<a href="#">case5</a>	页面有虚假电话、虚假下载、钓鱼网站等危害用户安全的行为。
<a href="#">case6</a>	网站内容采集于其他网站，通常有规模大、范围广、板式差、不完整等特征，并可能兼有其他黑帽 SEO 或作弊行为，达到被搜索引擎收录的目的。
<a href="#">case7</a>	网站将目录或频道页租售出去，或自行有意而为之。在全站或其中几个目录下，发布大量与网站主题完全无关的内容，包括蹭热点事件、编造八卦、发广告软文、甚至发布低俗色情及非法的内容，以此来达到吸引用户点击的目的。
<a href="#">case8</a>	在首页或子页面刻意设置垃圾外链，形成恶劣链接关系，达到欺骗搜索引擎或恶意推广的目的。部分网站甚至利用这一手段，形成了链接买卖和链接黑市交易的行为。
<a href="#">case9</a>	刻意生成大量二级或多级子域名，制造作弊垃圾内容，获取搜索引擎的广泛收录达到推广目的。
<a href="#">case10</a>	利用模板粗制滥造，生成大量相同但质量很低网站，达到欺骗搜索引擎排序目的。

## ● 用户体验

对于用户体验，我们主要从以下两个维度进行判定：

- 广告对主体内容的干扰度；

- 内容整洁度。

#### 4.1 广告对主体内容的干扰度

广告是网站收入的一个来源，我们理解站长朋友们在网站运营过程中耗费的心血。在一个网页中插入广告，如果数量合适、排版合理、内容优质、不影响用户对网页主体内容的阅读，可视为对页面内容的一种补充。相反，网页上如存在大片广告，干扰用户阅读，恶意让用户点击、下载、收藏、转发，甚至连续跳转、弹窗霸屏，都会伤害到用户的体验！

#### 4.2 内容整洁度

用户访问一个网页，他们首先感知到的，就是网页内容及版式是否利于阅读。一个整洁清爽的网页，用户能很方便地阅读到其中的内容。相反，一个内容及版式很差的网页，不仅影响用户的体验和评价，也增加了用户获取内容的时间和精力。

提供一些 badcase 例子，供站长们参考：

示 例	说 明
<a href="#">case1</a>	网页有主体内容，但是广告占据了大量篇幅，在低分辨率的显示器或移动设备上阅读较为吃力，不利于用户获取信息。
<a href="#">case2</a>	满屏都是广告，访问后完全看不到网页的内容。
<a href="#">case3</a>	页面广告很多，板式很差，正文几乎不能阅读。
<a href="#">case4</a>	编辑失误导致页面段落杂揉在一起，没有分段，阅读非常吃力。
<a href="#">case5</a>	页面内容不分段，部分语句前言不搭后语，用户完全不能正常阅读。

对于用户体验好的网页，360搜索会给予其收录和排序，而对于用户体验差的网页，360搜索会降低其排序，甚至拒绝收录！

## ● 网页资源可访问性

### 5.1 具有打开及获取权限的网页

一些网站会通过权限，来限制用户访问页面。限制的类型往往有：要求用户安装插件或软件，以及要求用户付费。

我们并不主张这种提高内容获取门槛的行为。对于一些非原创，非优质，来自于转载并且可替代性很强的内容，用户在访问时还被权限限制的，360搜索会根据实际情况决定其收录及排序。

但一些提供高质量、原创、正版内容的网站，在内容建设上投入了很大精力和成本。用户在获取页面或资源时，会被要求安装软件和插件、甚至付费。此类情况多出现在一些具有自制节目或购买了版权的视频站，以及一些提供专业文献检索的网站中。对于此类页面，360搜索不会将其视为对用户不友好的行为。

注册和登录页面，其本身目的就是要求用户进行注册和登录，因此这种类型的页面不会被归为“具有打开及获取权限的网页”。

以下是一些举例，供站长朋友们参考：

示 例	等级程度	说 明
<a href="#">case1</a>	可接受	某视频网站上的一部电影，普通用户只能看 5 分钟，完整观看需要购买VIP。网站在获得电影时需要支付很高的成本，这类情况360搜索会认为是可接受的。
<a href="#">case2</a>	不建议	内容并非来自于网站自身的原创，互联网上存在大量可替代的资源，但仍然要求付费才能下载。
<a href="#">case3</a>	不建议	页面的内容不完整，并且内容替代性很强，但仍然需要权限才能阅读。
<a href="#">case4</a>	不建议	视频播放和下载时，被要求安装已经无法使用或非常冷门的软件。

## 5.2 死链及失效页面

网站在运营的过程中，往往会出现死链和失效页面。我们在这里为大家罗列几种常见的类型：

类 型	说 明	示 例
协议死链	指通过 http 协议状态码，明确表示网页已经无法访问的页面类型。常见的协议死链状态码包括：404、403、503、504 等状态。	<a href="#">case1</a>
跳转死链	跳转死链，是指一个页面内容失效之后，以 301、302 跳转的形式，跳转到特定的死链特征页面上。例如这个跳转到了： <a href="https://store.taobao.com/shop/noshop.htm">https://store.taobao.com/shop/noshop.htm</a>	<a href="#">case2</a>
内容死链	内容死链的产生，主要是由于网站自身内容变化引起。网页可以正常打开，http 协议状态码为 200。常见的类型包括：帖子被删除、内容已转移、空间被关闭、信息已过期、交易已关闭等。	<a href="#">case3</a>
视频死链	视频页面中的视频主体不能播放，提示已经删除、已经下线等。	<a href="#">case4</a>

死链及失效页面对用户没有任何价值。网站中如果这一类型页面数量占比很大，会影响到整个网站的收录与排序，请站长们务必重视！站长们可通过360站长平台，及时提交死链进行删除。

站长们在设置页面死链的类型时，请严格按照规范进行设置。建议大家将死链优先设置为“协议死链”的状态（即表格中第一种类型）。360搜索蜘蛛也会在重新抓取后识别出这些死链，进而推送删除。

## 总结与建议

作为国内第二大搜索引擎，360搜索致力于为广大网民提供信息检索服务，并通过一系列策略，努力将更多优质内容提供给用户，因此，我们希望能和广大站长朋友们携手，一起更好地为用户服务。

总结《360搜索网页质量白皮书》(1.0版本)的内容，我们向广大站长们提供以下建议，即“一个避免、五项关注”：

- 避免针对搜索引擎的作弊和黑帽 SEO 手段。我们非常不提倡此类行为，轻者导致排序降低，重者收录被拒绝；
- 关注用户需求的满足与用户体验。网站应该优先考虑用户的利益与需求，并同时提升用户体验，而不是针对搜索引擎的单方面迎合优化；
- 关注高质量原创内容的产出，减少低质量内容在网站中的数量。网站的价值不是通过采集等手段获得，而是通过自身的高质量原创内容，以合理的方式体现网站的特色，考虑站点的短期收益，更重视站点的长远发展；
- 关注失效内容与死链的清理。及时处理网站中的失效数据、过期信息与死链，并及时提交给搜索引擎进行删除；
- 关注网站安全与访问的稳定性。防范潜在的网站安全风险，保证网站的可用性和访问稳定；
- 关注360搜索的公告和动向，并积极自查是否存在违规行为。积极使用360站长平台提供的工具，进行数据查询和监控。

360站长平台地址为：<http://zhanzhang.so.com/>，欢迎大家注册使用！