

基于搜索引擎优化的出版社网站设计研究

撰文：靳帆，邓三鸿，江莹

Research on the Design of Publishing House Website Based on Search Engine Optimization

[摘要] 出版社网站是出版社数字化转型的关键，但低迷的流量限制了其功能的发挥，搜索引擎优化（SEO）工作可以提高网站访问量。文章选取10家少儿类图书出版社网站，从网站内容、title和meta标签（keywords、description）、网站结构、图片alt属性、flash、内外链接等影响搜索引擎排名因素的角度，对出版社网站的设计进行分析，提出五条出版社网站优化设计的建议：重视网页内容、合理设置网页的keywords和description、编制网站地图、优化图片、优化导出链接。

[关键词] 搜索引擎优化；出版社网站；少儿类图书

[Abstract] Publishing house website plays an important role in communicating and releasing information, but the slump flow limits its function. This article selected 10 websites of children's books publishing house as research object, thought analysis the factors that affect search engine ranking include website content, title and meta tag (keywords, description), website structure, image alt attribute, flash, internal and external links, etc. Finally, five suggestions are put forward: attaching importance to the content of web pages, setting up the keywords and description of webpage reasonably, compiling website map, optimizing pictures and optimizing export links.

[Key words] SEO; Publishing house website; Children's books

[基金项目] 本文系南京大学优质课程建设项目示范性精品课程“信息检索”（项目编号：IM2017A004）阶段性成果。

从1995年电子工业出版社凭借自己的信息技术优势率先建立起网站开始，我国出版社网站建设已经走过了二十多年的发展历程。截至2014年4月，全国580家出版社中，已经有438家建立了属于自己的官网。^[1]出版社网站是出版社数字化转型的关键，它的主要功能包括沟通和服务读者、作者、分销商、求职者，以及发布一般性的信息。此外，出版社网站还有一个重要功能就是宣传推广书籍和出版社的品牌形象，这项功能的实现需要建立在提高网站访问量的基础之上。出版社网站51%的访问量是来自搜索引擎^[2]，另有一项调查显示，有75%的网民会点击进入搜索引擎提供的检索结果的第一页，检索结果排名越靠后，被点击的几率就越小。因此，网站的建设者需要根据搜索引擎的排名规则来设计网站，这项工作便是搜索引擎优化（SEO，Search Engine Optimization）。

搜索引擎优化是和搜索引擎同步发展起来的。国外对于SEO的研究最早可追溯到20世纪90年代中期^[3]，其目的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，来对网站内容进行相关优化，使其在不损害用户体验的情况下提高网站的搜索结果排名，提高网站的访问量。笔者试从影响搜索引擎排名因素的角度切入，结合实证研究，探讨提高出版社网站访问量的设计方案。

一、影响搜索引擎排名的因素

搜索引擎的工作过程可分为网页抓取、信息提取、关键字搜索、建立关键字索引和网页排序五个部分^[4]。目前主流的搜索引擎大多采用爬虫程序来完成网页的抓取。这些程序像虫子一样以极快的速度从一个链接爬到另一个链接，不间断地发现并下载新的网页，然后对网页中的链接、title、keywords和description以及文字等内容进行分析，再对网页进行编目、索引、排序。当用户在搜索引擎中输入关键字时，搜索引擎便将匹配上的网页按排序的先后呈现给用户。影响搜索引擎对网站排序的因素主要包括网站内容、title和meta标签（keywords、description）、网站结构、图片alt属性、flash、内外链接等。

（一）网站内容

内容是网站建设最重要的方面，也是实施SEO策略的基础。这个观点自SEO出现以来，被业内反复地强调。网站要保证提供

鸟与兽的通俗生活 (豆瓣) title

keywords

图书鸟与兽的通俗生活 介绍、书评、论坛及推荐... 多识鸟兽草木之名《鸟与兽的通俗生活》记得是在博物学者刘华杰教授的书中看到,他说为什么要了解那些植物的名字...

https://book.douban.com/subjec... - 百度快照

description

图1《鸟与兽的通俗生活》在百度中的一条检索结果

丰富的原创性内容，同时保持固定的更新频率。在更新内容时，要注意内容与网站（网页）的主题和关键词相关。每个页面的内容应该主题一致，在格式上保持统一的设计，突然变化的网页会影响搜索引擎对内容的吸收，降低搜索引擎对网站的评价。

（二）title 和 meta 标签（keywords、description）

搜索引擎非常重视 title 标签和 meta 标签中的 description、keywords 属性，合理利用这三个关键内容可以极大地提高搜索引擎友好性。

title 应简短，包含网页关键词。譬如国内常用的搜索引擎——百度对标题最多显示 30 个中文字符，过长的标题不利于突出目标关键词，降低相关性。一个网页中的 keywords 不宜超过 20 个，密度最好为 2%~8%，每页重点使用的关键字词不要超过 3~4 个。关键词的选取上，编辑可围绕主关键词制定一些“长尾关键词”。长尾关键词派生于长尾理论，指的是与页面内容相关，但较为冷门的词汇。在搜索引擎中检索时，包含该关键词的网页数量较少。一个经过长尾关键词优化的网站，长尾关键词带来的流量汇聚起来将会达到甚至超过网站主关键词带来的流量。^[5] description 帮助搜索引擎和用户快速了解网页的大致内容。百度 description 最多可显示 70 个左右汉字，其中一定要包含关键词，如图 1 所示。

（三）网站结构

网站设计应尽量避免使用框架结构，因为搜索引擎不能迅速、完整地带走全部网页信息，通过 iframe 显示的内容可能被搜索引擎抛弃。如果网页已经使用了框架，则最好在代码中使用 noframes 标签进行优化，noframes 中包含指向 frame 页的链接以及带有关键词的描述文本，这样搜索引擎才能够正确索引到框架内的信息。控制好 url 层级也很重要。探测器在遍历页面的时候是按照 url 层级进行的，当层级超过三层的时候

就很难被探测器搜索。层级主要是为了区分页面的重要性，因此重要性区别不大的项目尽量放在一个目录里。

（四）图片 alt 属性、flash

大部分搜索引擎不能识别图片主题内容，为了不损失图片搜索带来的流量，网站设计者应为图片设置包含关键词的 alt 属性，让搜索引擎了解图片的主要信息。alt 属性是在图像无法显示时的替代文本，即使图片不能正常显示也能够帮助用户了解图片内容。flash 的搜索引擎友好性较差，网页关键词不能放在 flash 中，如果必须使用 flash，则应该同时制作一个文字版供搜索引擎收录，并在首页使用文本链接指向文字版。

（五）内外链接

爬虫程序收集页面的轨迹是顺着一条链接到另一个链接，如果网站内部存在死链、断链，网站信息就不能被完整地收集。长尾关键词的链接资源几乎全部来自内链。为帮助搜索引擎全面了解网站的结构，可以建立网站地图，引导爬虫程序按照地图上的链接到达每个网页。同时，网站地图也可以方便用户使用，提升用户体验。

建立优质的外部链接同样重要。外部链接包括导入链接和导出链接。导入链接也称为反链，假设有两个页面 A 和 B，B 页面通过一个链接指向 A 页面，那么，B→A 这个链接就是 A 的一个反链。Google 评价网站质量的一个重要指标是 PageRank（简称 PR 值，分为 10 个等级），指的就是一个网站有多少反链，以及这些反链网站的质量如何。导出链接即友情链接，就是页面中指向别的网站的链接。某网站的 PR 值 = $(1 - 0.85) + 0.85 \times (\text{友情链接网站的 PR 值} / \text{友情链接网站的外链数})$ ^[6]，当友情链接网站的 PR 值一定时，其外链数越少越好。nofollow 标签（即“反垃圾链接”标签）的作用是指示搜索引擎不要抓取网页上的带有 nofollow 属性的任何出站链接，以避免垃圾链接分散网站权重。

二、出版社网站搜索引擎优化效果调查

出版社每年都会出版发行一定数量的图书，可提供给网站更新的内容非常多。图书内容涵盖的范围广，可以产生大量的关键词和长尾关键词。因此图书网站在搜索引擎优化方面有天然的优势。

（一）调查对象及方法

根据上文总结的影响搜索引擎排名的因素，本文制定了表 1 所示的 SEO 状况评价指标。衡量 SEO 效果的指标，国内外学者主要是从网站流量数据方面提出的。^{[3][13]} 在样本的选取上，全出版业综合实力排名前 10 位的出版单位网站，其流量数据差别较小，不利于 SEO 效果研究。本文选取的样本为新闻出版广电总局《2015 年新闻出版产业分析报告》中总体经济规模排名前 10 位的少儿类图书出版单位网站。少儿出版是 21 世纪以来中国出版业增长最快、连续增长时间最长的一个细分市场，既可以避免网站流量低迷，失去 SEO 效果研究的意义，样本之间的差距也更为合理。SEO 综合查询工具可以检测出一个网站的搜索引擎流量、权重、友情链接情况等数据，TOOL 站长工具和爱站网站长工具是公认较好的 SEO 综合查询工具，笔者利用它们检测了网站链接、权重、百度日均 IP 数和百度收录、谷歌收录几项数据，然后将数据结果依据表 1 的评价指标分析出结果。

（二）数据收集与分析

数据收集时间为 2018 年 1 月 5 日，主要选取各网站最重要的首页进行分析。从表 2 的数据可以看出，对最重要的原创性内容给予重视的出版社有 4 家。以中国少年儿童新闻出版总社为例，在图书介绍页面中，列出的项目有：书籍基本信息（包括作者、出版社、出版时间、版次、开本等）——这一项内容几乎所有参与调查的出版社都具备；3 至 5 分钟的书籍内容介绍视频；500 字左右的内容简介；作者简介；媒体评论。后

四项内容其余 6 家出版社都没有涉及。湖南少年儿童出版社的每本图书都只给出了书名，配有一张图片，作者信息还需要将鼠标放在图片上才能显示。中国少年儿童新闻出版总社列出的上述五项内容可以帮助读者快速了解一本书的方方面面，且内容均为原创。如百度词条“盘中餐（于虹呈的绘本）”的内容就来自该页面，该词条下面列出的参考资料链接指向该页面，这就是一条高质量的反链。

网站首页源代码中设置了 title、keywords 和 description 的只有 5 家出版社。其中，安徽少年儿童出版社的 keywords 和 description 都是“安徽少年儿童出版社”一个词，海燕出版社的仅仅是“首页”两个字。新疆少年儿童出版社虽然也有 keywords 和 description，但其 keywords 的内容是“精品童书包括阿凡提故事系列、贝贝熊系列、亚瑟小子双语阅读全 18 册系列……”，关键词过长，并且词之间没有用英文逗号隔开，因此也起不到应有的作用。接力出版社和长江少年儿童出版社设置的 keywords 和 description 符合 SEO 策略，后者就其社名设置了“长江少年儿童出版社有限公司”“湖北少年儿童出版社”等 5 个同义词，保证读者即使不知道准确的社名，也能搜索到该网站。

虽然表 2 的数据显示，中国少年儿童新闻出版总社的图书介绍内容很丰富，但其宣传某一本书的内页源代码中并未设置 keywords 和 description，如在百度搜索引擎中输入“盘中餐图书”，内容不如该页面丰富的东莞图书馆网站排在了它前面，东莞图书馆介绍《盘中餐》这本书的页面设置了 title、keywords 和 description，且三项内容都符合 SEO 策略，均经过精心编写。

在图片 alt 属性和 flash 设置上表现最好的是明天出版社，10 家出版社都没有编制网站地图。

表 3 中数据来自 TOOL 站长工具^[7]和爱

评价指标	说明
网站内容	内容是否原创、丰富。这项指标比较主观，主要是查看书籍内容介绍、书评等内容，分为丰富、一般和无三个层次；网站内容是否有固定的更新频率，可通过新闻、新书发布时间查看。
title 和 meta 标签 (keywords、description)	页面是否设置了 title，是否和内容一致；meta 标签中是否设置了关键词和相关描述。
网站结构	主要查看 url 层级数，查看正文页在第几层。
图片 alt 属性、flash	。
站内链接	不同页面之间是否连通，是否有死链。
网站地图	是否设有网站地图。
反链数	。
友情链接	友情链接的质量。

表 1 SEO 状况评价指标

网站名称	原创性 内容	更新	标题	关键词	描述	url 层级	图片 alt 属 性	flash	网站 地图
中国少年儿童新闻出版总社	丰富	√	√	×	×	3	√	√	×
二十一世纪出版社	丰富	√	√	×	×	3	√	√	×
明天出版社	丰富	√	√	×	×	4	√	×	×
安徽少年儿童出版社	一般	√	√	√	√	4	√	√	×
接力出版社	无	√	√	√	√	4	√	√	×
浙江少年儿童出版社	一般	√	√	×	×	4	×	√	×
长江少年儿童出版社	丰富	√	√	√	√	4	√	√	×
海燕出版社	一般	√	√	√	√	4	×	√	×
湖南少年儿童出版社	无	√	√	×	×	3	×	√	×
新疆少年儿童出版社	无	√	√	√	×	4	√	√	×

表 2 10 家少儿类图书出版单位网站内容、title 和 meta 标签等状况

站站长工具^[8]。从表 3 可以看出，有的出版社对导出链接不重视，如海燕出版社没有设置导出链接；有的出版社对导出链接没有进行适当选择，二十一世纪出版社、长江少年儿童出版社和湖南少年儿童出版社的导出链接的平均权重较低。中国少年儿童出版社、明天出版社和接力出版社在链接方面的优化工作做的较好。

综上，我们可以看出，上述 10 家少儿类图书出版单位网站中，SEO 工作做的最好的是中国少年儿童新闻出版总社、明天出版社和接力出版社。表 4 是使用站长工具对上述 10 家少儿类图书出版单位网站进行测试后获得的数据，与表 1 的评价指标分析出

的结果有部分出入，但基本一致。

三、出版社网站设计建议

从上文的分析中可以看出，作为本文研究对象的 10 家少儿类图书出版单位网站每天来自搜索引擎的访问量都很低，大部分出版社网站的 SEO 工作都很不到位。出版社需要增强 SEO 意识，依靠自身在内容方面的资源优势，在网站设计上实施全面的 SEO 策略，使出版社网站发挥更大的作用。据此，对出版社网站设计提出以下几点建议：

（一）重视原创内容

新书宣传页面是出版社网站最常见也是最重要的页面。围绕一本新书，网站设计者

应尽可能详细地编制内容介绍、作者介绍、媒体评论等栏目，提高读者的阅读体验。丰富的原创内容才是搜索引擎欢迎的。这一点要改进很容易，因为每本书本身都有充足的信息，只要编辑在发布网站内容时选取合适的信息即可。

（二）合理设置网页 keywords 和 description

新书宣传页不应该只选取书名作为关键词，为了方便不知道该书全名的人进行查找，还应该围绕该词选取一些长尾关键词。选取长尾关键词的方法，一是从搜索者的需求出发，用尽可能多的词汇和主关键词组合；二是留意百度搜索页面底部“相关搜索”列出的词汇，然后将关键词自然地融入到网页描述中。

（三）将 url 层级控制在 3 层以内，编制网站地图

可以将书籍分类栏目设置在首页，层级越少就越方便用户查找和记忆。有些网站尽管可以从首页看出网站大体布局，但编制一个清晰的网站地图不仅方便爬虫程序抓取网站页面，而且可以极大地提高用户体验。网站地图入口可以放置在网站首页顶部或底部。

（四）优化站内图片

出版社网站需要使用大量图片，如果图片不能被检索、加载速度慢，将会影响网站的搜索引擎排名。网站设计者应为有价值的图片增加 alt 属性，并在不影响显示效果的前提下尽量压缩图片大小。flash 等对搜索引擎不友好的代码也要尽量避免使用。

（五）对导出链接进行优化

导出链接的选择要慎重。上述 10 家少儿图书出版单位网站中，中国少年儿童新闻出版总社、浙江少年儿童出版社等几家出版社网站的导出链接都是自己所属集团的兄弟网站。抱团推广的初衷可以理解，但如果兄弟网站质量很差，反而会稀释自己网站的权重。表现最好的接力出版社网站导出链接就选取的是相对优质的网站。

结语

随着信息技术的快速发展，我国出版行业正在从传统出版向数字出版转型。出版社网站作为数字出版的窗口，在出版社的日常运作和经营管理中大有可为。出版社应重视网站建设，充分发挥自身的资源优势，做好网站 SEO 工作，提高网站访问量，为出版社的未来发展提供更广阔的平台，为用户提供更便捷的服务，为国家书香工程的建设和全民阅读助力。

参考文献

[1] 苗倩楠.我国出版社官网用户体验研究[D].南京:南京大学,2014.
[2] 扈欣悦.我国图书出版网站建设研究[D].北京:北京印刷学院,2013.
[3] 周鑫.基于谷歌搜索引擎的企业 SEO 方法及其效果的实证研究[D].宁波:宁波大学,2014.
[4] 徐炜.浅析搜索引擎优化技术及其应用[J].科技风,2017(4):53-54.
[5] 邓孜.长尾关键词理论在提高出版网站访问量中的应用[J].中国出版,2009(Z3):89-91.
[6] PR 值 [EB/OL].[2018-01-05].[https://baike.baidu.com/item/PR 值](https://baike.baidu.com/item/PR值).
[7] 站长工具 [EB/OL].[2018-01-05].<http://tool.chinaz.com>.
[8] 爱站网 [EB/OL].[2018-01-05].<https://www.aizhan.com>.

网站名称	站内死链接数	反链数	导出链接数	导出链接平均百度权重	导出链接平均 PR 值
中国少年儿童新闻出版总社。	2。	5。	26。	2.96。	5.58。
二十一世纪出版社。	4。	19。	18。	0.89。	2.39。
明天出版社。	2。	10。	5。	2.2。	2.2。
安徽少年儿童出版社。	1。	1。	7。	1.14。	1.71。
接力出版社。	5。	21。	18。	4.17。	5.22。
浙江少年儿童出版社。	4。	8。	19。	1.53。	2.63。
长江少年儿童出版社。	5。	4。	9。	0.56。	3.11。
海燕出版社。	2。	10。	0。	0。	0。
湖南少年儿童出版社。	4。	13。	8。	0.88。	2.88。
新疆少年儿童出版社。	2。	11。	8。	2。	4.75。

表 3 10 家少儿类图书出版单位网站链接状况

网站名及网址	通过百度访问的日均 IP 数	百度权重	Google 权重	百度收录	Google 收录
中国少年儿童新闻出版总社。	70。	1。	0。	1030。	307。
二十一世纪出版社。	204。	1。	6。	5570。	8950。
明天出版社。	48。	0。	5。	1820。	2790。
安徽少年儿童出版社。	0。	0。	4。	465。	-。
接力出版社。	11。	0。	7。	993。	-。
浙江少年儿童出版社。	2。	0。	0。	517。	-。
长江少年儿童出版社。	10。	0。	0。	723。	-。
海燕出版社。	0。	0。	4。	374。	-。
湖南少年儿童出版社。	166。	1。	4。	360。	-。
新疆少年儿童出版社。	13。	0。	0。	470。	-。

表 4 站长工具数据结果

靳帆 邓三鸿 江莹

南京大学