

搜狗搜索页面质量白皮书

搜狗搜索页面质量白皮书.....	1
1. 必备概念.....	2
1.1 页面主体内容/content.....	2
1.2 页面标题/title.....	2
2. 页面质量评估维度.....	2
2.1 可访问性.....	3
2.2 内容质量.....	3
2.3 浏览体验.....	4
2.4 特殊情况说明.....	5
3. 给各位站长的建议.....	7
附录.....	8
4. 一般质量.....	8
4.1 一般质量例 1.....	8
4.2 一般质量例 2.....	9
4.3 一般质量例 3.....	10
5. 低质量.....	11
5.1 低质量例 1.....	11
5.2 低质量例 2.....	12
5.3 低质量例 3.....	13
5.4 低质量例 4.....	13
5.5 低质量例 5.....	14
5.6 低质量例 6.....	14
6. 无质量可言.....	15
6.1 无质量可言例 1.....	15
6.2 无质量可言例 2.....	16
6.3 无质量可言例 3.....	17
7. 浏览体验差.....	18
7.1 浏览体验差例 3.....	18
7.2 浏览体验差例 4.....	19
7.3 浏览体验差例 5.....	20
7.4 浏览体验差例 6.....	21

搜狗搜索页面质量白皮书以达到用户与站点的双赢为目的，积极为用户营造良好的网络环境。我们希望能与站长共同努力、一起进步，为用户提供更优质、更有价值、更加满足用户需求的信息。

以下是页面质量的评估标准：

1. 必备概念

1.1 页面主体内容/content

页面主体内容指网页中最关键、最能引起用户注意的地方。一般来说，除去网页的导航条、边框和版权信息部位外，网页的中间偏左上的位置是网页的主体内容。对于 bbs 和博客来说，文章是网页的主体内容。对于搜索结果页来说，搜索结果列表就是主体内容。对于一句话新闻来说，那句新闻和标题就是主体内容。

1.2 页面标题/title

页面标题，一般称 title，指的是页面的标题。一般在浏览器顶端显示，如下图红框内所示：



一般情况下，页面标题包括页面内容标题和站点信息，以上图为例：

- ① “搜狗搜索”是站点信息——这部分可以从网页左上角得到印证；
- ② “页面质量标准”才是真正的内容标题，content title。

2. 页面质量评估维度

页面质量评估的思考维度主要包含页面可访问性、页面内容质量以及浏览体验三个方面。分别从这三个维度出发，页面在每个维度上又可以细分为以下类别。

可访问性	内容质量	浏览体验
正常页面	优质内容页面	浏览体验好
需要权限，不可直接获取主体内容	一般内容页面	浏览体验一般
无法访问或资源失效	低质内容页面	浏览体验差

2.1 可访问性

可访问性主要指网页可访问性、主体内容可获取性以及主要资源的有效性。从可访问性的角度出发，主要有以下几种情况：

■ 正常页面

页面可以正常打开，并且用户可**直接浏览全部**页面主体内容，或**直接**获取主要资源（比如：下载软件、播放视频等）。

■ 需要权限，不可直接获取主体内容

页面可以正常打开，但是用户如果不登陆或不注册，就**不能**浏览页面**全部**主要内容，或**不能**获取页面主要资源，而是有一定的权限限制。

搜狗 Tips:单纯的登录页面的主要目的就是供用户登录，用户查看登陆页的目的也是想要登录，因此登录页面不算权限页面。

■ 无法访问或资源失效

页面无法打开，或页面可以正常打开，但是没有用户所需内容或资源。

2.2 内容质量

内容质量主要看页面**主体内容**的好坏，主体内容的真实性、完整性、丰富性、权威性、原创性等都是影响内容质量的因素。

■ 优质内容页面

页面内容经过精心编辑，内容丰富完整、清晰有效或资源优质可用，能够提供给用户真实安全的信息，不存在作弊或其他不当意图。

示例	内容质量	说明
例 1	优质	医疗专题页，内容丰富，能够为用户提供大量有价值的信息
例 2	优质	视频专辑页面，有版权资源，视频资源丰富优质，更新进度快，另外还提供了剧情介绍、更新时间、演员等信息
例 3	优质	权威来源的文章，内容丰富，完整
例 4	优质	认真编辑，内容丰富的百科词条
例 5	优质	专题页，页面内容围绕圣诞橱窗给出了丰富的介绍，页面布局美观

■ 一般内容页面

页面内容完整但不丰富，资源可消费但质量欠佳，信息真实有效但权威性或原创性差，安全无毒，能够满足用户需求。

示例	内容质量	说明
例 1	一般	普通有好答案的页面，页面的主体内容问题和答案都具备，但是答案并不丰富。
例 2	一般	视频可以正常播放，但是视频质量一般，画面不清晰
例 3	一般	页面主体内容是一篇完整的新闻报道，非权威来源，互联网上存在大量可替换的资源

■ 低质内容页面

低质内容页面往往存在主体内容不完整、有效信息少、信息失效或过期、资源不可消费等问题，无法为用户提供实质性信息或帮助。

示例	内容质量	说明
例 1	低质	问答页面的主体内容是问题、答案，本页面虽然有答案，但是答案与问题完全不相关，不能给用户有效信息
例 2	低质	有问无答
例 3	低质	已经失效的租房页面
例 4	低质	已过期的团购页面
例 5	低质	已下架的商品信息页
例 6	低质	信息量少的用户主页

■ 无质量可言

无质量可言的网页是指那些内容粗制滥造、有病毒、含有作弊行为或不当意图、页面标题与主体内容不符、主体内容不能支撑页面主题，完全不能为用户提供任何价值的页面。

示例	内容质量	说明
例 1	无质量可言	页面内容与页面标题“穿僧衣男子数分钟内对十多名女子袭胸：僧衣猥琐男”不相关
例 2	无质量可言	页面主体内容中跟婴儿奶粉链接，点击之后都没有有价值的信息，只是不停的又点进该站内的另外一些页面，这种情况属于主体内容不能支撑页面主题，对用户毫无价值并且欺骗用户点击
例 3	无质量可言	页面标题显示“一起作业网_17 作业网_一起作业_17 作业”，而实际页面主体内容仅仅是一个截图，并不是一起作业网，与页面标题不符，不能给用户任何有价值信息

2.3 浏览体验

影响用户浏览体验的因素主要是页面排版和广告。

■ 浏览体验好

页面布局合理，用户获取主体内容成本低，一般具有以下特征：

- ① 排版合理，版式美观，易于阅读和浏览；
- ② 用户需要的内容占据网页最重要位置；
- ③ 能够通过页面标签或页面布局十分清楚地区分出哪些是广告；
- ④ 广告不抢占主体内容位置，不阻碍用户对主要内容的获取。

■ 浏览体验差

页面布局和广告极大干扰用户对主体内容的获取，用户很难直接获取到有效信息，令用户反感。包括但不限于以下情况：

- ① 页面排版混乱，几乎无法获取有效信息；
- ② 广告与主体内容混淆或广告遮挡主体内容，用户无法区分；
- ③ 页面布局不合理，在通用分辨率下，首屏都是广告，用户不能直接看到主体内容；
- ④ 大量的色情和赌博广告，令人厌恶；
- ⑤ 弹窗广告过多，令用户反感或点击链接时，出现预期之外的弹窗；
- ⑥ 多次点击跳到有效内容；
- ⑦ 页面为两个到多个页面拼接而成。

示例	浏览体验	说明
例 1	好	页面布局合理，主体内容新闻占据醒目主要位置，上方和右侧有少量广告，与主体内容区分明显，对用户阅读新闻无任何影响
例 2	一般	页面布局差，图片占据了绝大部分位置，其他主要信息不能在首屏显示
例 3	差	首屏都是广告，用户不能直接看到主体内容
例 4	差	页面排版混乱，内容互相遮挡，用户无法阅读
例 5	差	页面排版混乱，首屏有大量留白，且首屏看不到主体内容
例 6	差	大量的赌博色情广告，令人厌恶
例 7	一般	wap 页面，布局更合适触屏和手机使用，PC 端浏览体验较差

2.4 特殊情况说明

2.4.1 违法内容页面

页面主体内容为违法信息的均为垃圾页面，主要有以下几类：

■ 淫秽色情

页面主体内容是色情内容的，直接判为垃圾，具体内容主要包括视频、图片和小说等；

■ 非法博彩

大陆合法的彩票、双色球、时时彩等为正常页面，其余博彩页面为垃圾；

■ 网络诈骗

网络诈骗，主要有假发票、假币、出售个人信息、假证等几类。

- ① 假发票：涉及代开、虚开、伪造、变造、擅自制造和买卖各类发票的，一律违法
- ② 假币：有关明确或隐晦（如含有暗示性词语的如“一比一”、“高仿”、“不要拿去银行”等）销售假币的信息违法
- ③ 出售个人信息：涉及销售公民个人信息的帖文，如明确标注包括姓名、地址和联系方式等内容且有出卖人联系方式的信息违法

④ 假证：有关买卖或者使用伪造、变造的国家机关、人民团体、企业、事业单位或者其他组织的公文、证件、证明文件和身份证的信息违法

■ 涉枪涉爆

涉枪涉爆，主要有买卖枪支弹药、弓弩、涉毒、涉爆、管制刀具等几类。

- ① 枪支弹药：买卖各类枪支及其配件、弹药，以及枪支制作方法的信息一律违法
- ② 弓弩：单独售弩、弩和毒箭混合销售的信息，求购弩信息均为违法
- ③ 涉毒：利用互联网走私、贩卖毒品。
- ④ 涉爆：买卖、供应、出售、制造炸弹等
- ⑤ 管制刀具： 买卖、出售、订购贩售管制刀具信息均为违法

■ 网络招嫖

网络招嫖，主要有小姐、缓交妹、妓女等提供服务或技师上门的信息。

■ 政治敏感

政治敏感，主要分敏感人物和敏感事件两类。

- ① 人物类包含：人物的人事任免、被调查、贪污、腐败、大老虎、情人、暗杀、密谋、政变、内幕等为敏感信息。
- ② 事件类包含：六四事件、藏独、疆独、法轮功等。

2.4.2 跳转页面

跳转页面分为以下两种情况：

- ③ 页面无任何提示直接跳转时，以跳转后的页面为准进行评估；
- ④ 页面提示网站变更，并自动跳转到最新网址时，该页面为正常页面。

3. 给各位站长的建议

- **网页的设计是为了用户而不是搜索引擎**
不要企图用任何方式作弊方式来提升排名，作弊带来的利益是短暂的，但是负面影响会很久远。
- **永远将用户体验放在首位**
- **根据用户需求制作内容**
- **多考虑如何让自己的网站具有独特价值**
- **将目光放长远，滥放广告弊大于利**
- **及时删除低质内容**



附录

4. 一般质量

4.1 一般质量例 1

hey不黑 | 分类: 历史话题 2008-09-20
民族英雄代代传资料 5

非常急!资料!有那些英雄?什么英雄事迹

我来回答 ▾

问问APP 答题更简单
问问手机版拥有语音输入和图片答疑功能哦! **立即下载**

回答 (2)

冷笏醉影 ❤️❤️❤️ 2008-09-20
林则徐(1785-1850), 字元抚, 又字少穆, 晚号俟村老人、谥村退叟、七十二峰退叟。福建侯官(今福建福州)人。 主要功绩是从英国手里收缴全部鸦片近2万箱, 约237万余斤。于四月二十二日(6月3日)在虎门海滩上当众销毁。

评论(0) 0 0

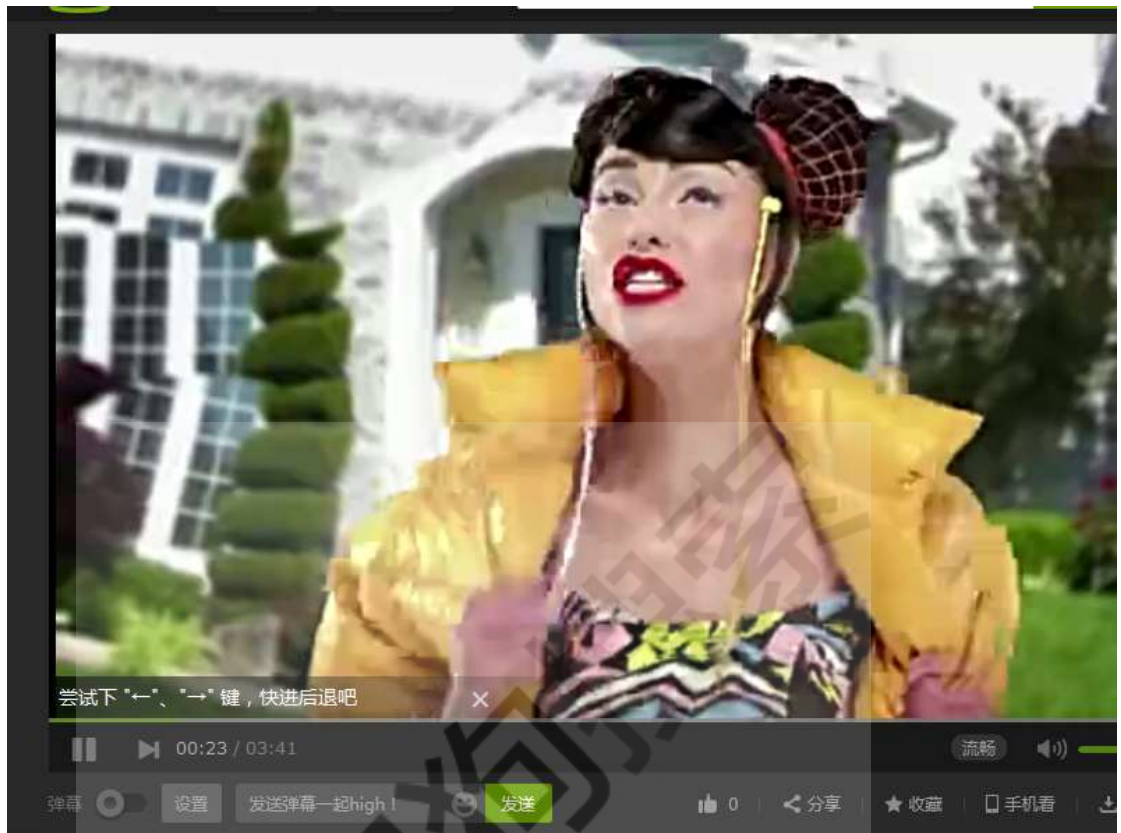
刘小松 2008-09-20
郑成功, 明清时期的, 1662年率领两万将士从福建出发, 收复了被荷兰占领的殖民地台湾, 成为民族英雄

评论(0) 0 0

普通有好答案的页面，页面的主体内容问题和答案都具备，但是答案并不丰富。

[返回一般质量说明列表页](#)

4.2 一般质量例 2



视频可以正常播放，但是视频质量一般，画面不清晰

[返回一般质量说明列表页](#)

4.3 一般质量例 3

合肥房叔方广云：番禺的房叔蔡彬


2014年12月15日 10:36 来源：三军网



纹身一般多少钱 ▶

洗纹身价格表 ▶

怎么去掉纹身 ▶

蜜蜂养殖视频 

合肥房叔方广云：番禺的房叔蔡彬方广云在职期间，利用职权，伪造、假借身份证等手段，以亲属及他人名义非法套取侵占十多套拆迁安置房。

合肥房叔方广云：番禺的房叔蔡彬方广云，男，安徽省合肥市新站区站北社区原党委书记、管委会主任。在职期间，利用职权，伪造、假借身份证等手段，以亲属及他人名义非法套取侵占十多套拆迁安置房。2014年9月29日上午，方广云被指控涉嫌贪污、受贿和滥用职权。

页面主体内容是一篇完整的新闻报道，非权威来源，互联网上存在大量可替换的资源

[返回一般质量说明列表页](#)



5. 低质量

5.1 低质量例 1



问答页面的主体内容是问题、答案，本页面虽然有答案，但是答案与问题完全不相关，不能给用户提供有效信息

[返回低质量说明列表页](#)

5.2 低质量例 2

你是礁的不二依赖 | 分类：音乐 2013-07-01

歌词：“小时候有个梦，长大要去你身边...”是谁唱的？ 5

我来回答

提交回答

问问APP 答题更简单
问问手机版拥有语音输入和图片答疑功能哦！

立即下载

[帮助](#) | [提意见](#)

©2015 SOGOU - 京ICP证050897号

有问无答

[返回低质量说明列表页](#)



5.3 低质量例 3



已经失效的租房页面

[返回低质量说明列表页](#)

5.4 低质量例 4

南宁团购【团宝网南宁】仅需¥58元, 金大叔胡子鸡团购! 原价68元, 活鸡毛重约3斤, 屠宰后净重2斤, 赠送全部内脏! 炖、蒸、炒, 样样清香可口! 南宁市快环内免费送货到家, 如需屠宰请预约时说明。



已过期的团购页面

[返回低质量说明列表页](#)

5.5 低质量例 5



已下架的商品信息页

[返回低质量说明列表页](#)

5.6 低质量例 6



信息量少的用户主页

[返回低质量说明列表页](#)

6. 无质量可言

6.1 无质量可言例 1

穿僧衣男子数分钟内对十多名女子袭胸：僧衣猥琐男

更新时间：2015/5/9 15:20:56 编辑：中华军事网 来源：互联网

美女白领被农民工公交车摸胸：含泪忍气吞声

有网友在论坛上发帖称自己是北京的一个小白领，一个土生土长的北京人，不是有车族，所以该女士只能乘坐公交上下班，在公交车上遭遇农民工让她感到了耻辱，这耻辱就是被农民工给摸了。

今天早上搭公交车上班，遭到几个外地民工性的骚扰，现在想起来还像吞苍蝇一样的恶心。因为早上上班高峰，所以人多的就别提了。还赶上呼啦呼啦上来五六



页面内容与页面标题“穿僧衣男子数分钟内对十多名女子袭胸：僧衣猥琐男”不相关

[返回无质量说明列表页](#)



6.2 无质量可言例 2



页面主体内容中跟婴儿奶粉链接，点击之后都没有有价值的信息，只是不停的又点进该站内的另外一些页面，这种情况属于主体内容不能支撑页面主题，对用户毫无价值并且欺骗用户点击

[返回无质量说明列表页](#)

6.3 无质量可言例 3



页面标题显示“一起作业网_17 作业网_一起作业_17 作业”，而实际页面主体内容仅仅是一个截图，并不是一起作业网，与页面标题不符，不能给用户提供任何有价值信息

[返回无质量说明列表页](#)



7. 浏览体验差

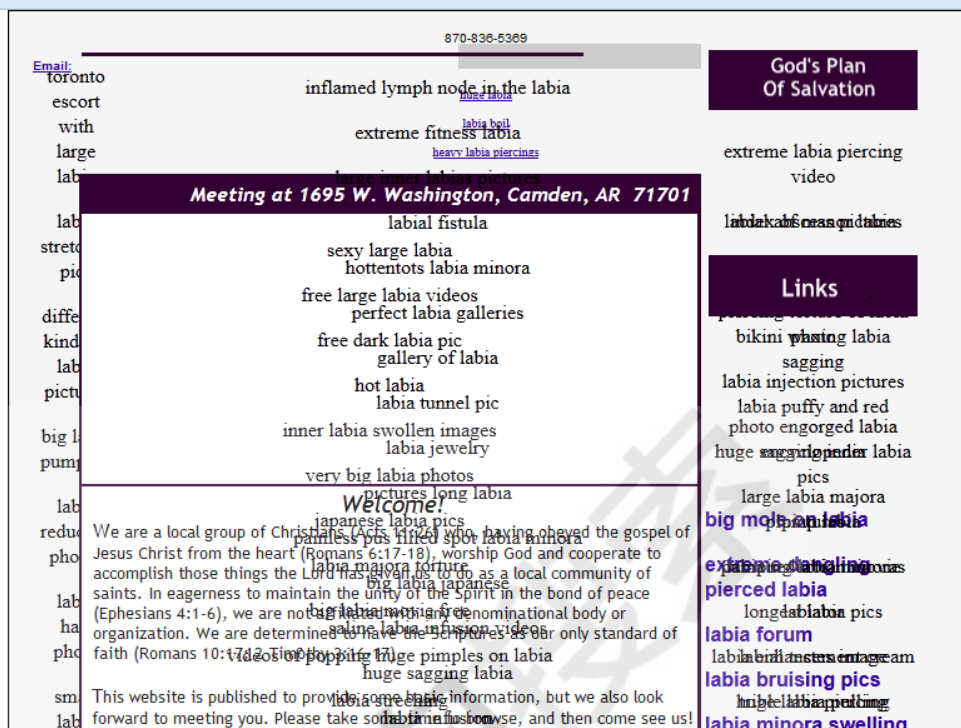
7.1 浏览体验差例 3

The screenshot shows the homepage of 163 Guizhou. The top navigation bar includes links for '163贵州网首页', '综合信息', '贵州游乐', '163贵州招聘吧', '163贵州人事考试信息网', '贵州房产', '贵州三农', '贵州网址', '网络世界', and '贵州商城'. Below the navigation bar, there are several large, colorful advertisements for '政图教育' (Zhengtu Education) and '农村信用社' (Rural Credit Co-operative) interviews. The main content area is dominated by these ads, with a list of recruitment notices below. The list includes various job openings such as '贵州物产五张名片传播中心', '贵州工商职业学院2015年招聘信息', and '贵州新西南教育发展有限公司2015年春季招聘'. The page is cluttered with text and graphics, making it difficult to find the main content.

首屏都是广告，用户不能直接看到主体内容

[返回浏览体验说明列表页](#)

7.2 浏览体验差例 4




页面排版混乱，内容互相遮挡，用户无法阅读

[返回浏览体验说明列表页](#)



7.3 浏览体验差例 5

亚马逊  10周年 手机访问Z.cn, 直达亚马逊

[Z秒杀](#) |  | [我的心愿单](#)

搜索

页面排版混乱，首屏有大量留白，且首屏看不到主体内容

[返回浏览体验说明列表页](#)

7.4 浏览体验差例 6



大量的赌博色情广告，令人厌恶

[返回浏览体验说明列表页](#)

