

原创系统

作者: [88250](#)

原文链接: <https://ld246.com/article/1544941500253>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

原创的重要性

这里是一个注重原创内容分享和讨论的地方，这里不欢迎全文转载，即使是授权转载。

为了让“注重原创”这件事落地，我们上线了“原创系统”来对帖子进行原创性评分，分值越低原创也越低：

1. **10-50**：基本是转载，灰色标识。这类内容在这里不受欢迎，并且有很大概率会被管理员删帖，40分以下的帖子不会出现在全局列表上
2. **60**：大部分是转载，浅灰色标识。这类内容大部分都是摘抄或是没有规范使用“引用”
3. **70**：少部分摘抄转载，绿色标识。这类内容可能是没有规范使用“引用”导致，需要认真排版
4. **80**：极少部分摘抄，蓝色标识。这类内容可能是没有规范使用“引用”导致，需要认真排版，不是的原创内容
5. **90**：个别术语或短语摘抄，紫色标识。这类内容是很受欢迎的原创内容，会在社区内部以及外部转发
6. **100**：完全原创，橙色标识。这类内容是最受欢迎的原创内容，排版优秀且具有很大价值

目前只有在超过 1024 字符（代码和引用不计入）的情况下才会进入原创系统评分，短文或者简单的答不会进行评分。

实现细节

原创性我们是通过搜索引擎后验实现评分判断的：

1. 根据内容 HTML 提取若干段纯文本短句
2. 将这些短句通过搜索引擎查询，结果排除站内索引，每个短句根据重复搜索引擎重复次数进行评分其中搜索引擎收录时间和发帖时间的差值会作为一个重要的权重项
3. 根据每个短句的评分得到该帖的平均原创评分

我们会不断调整原创系统的评分算法，大家可随时跟帖提建议。

这里绝不会成为一个内容农场（Content Farm），这里以后会有越来越多有价值的讨论。