

US$9.90 成功

【哥飞SEO教程】如何制作出SEO友好的网页？先从学习谷歌是如何理解我们网页开始

原创 我是哥飞 哥飞

大家好，我是哥飞。
上一次发文，还是上次，这几天在忙，新项目起量了，耽误写文章了。









今天终于有空，所以来回答下大家最近比较关心的问题。
时不时有朋友问哥飞一些问题：
1. 做工具站要不要做内容？
2. 为什么不要前端渲染？
3. 为什么多语言不能用js切换？
其实这些问题的答案，都跟谷歌抓取方式有关。
所以今天哥飞跟大家聊聊谷歌是如何抓取我们的网页，并且如何理解网页的。
注意，下文哥飞会在不同的句子里用“网址”、“网页”、“链接”等不同描述，有时候是同一个东西。

一、爬取

不管是什么网页，文章也好、新闻也好、图片页面、视频页面、音乐页面， 谷歌理解网页，依然靠的是从网页获取到的文字信息。
那么是怎么获取的呢?
用的是谷歌的爬虫。
爬虫是什么?
爬虫是形象化的说法，其实是一段运行在谷歌服务器里的程序代码。
拿到一个网址后，程序发出一个 GET 请求，获取网址对应的 html 代码，这就是爬虫最主要做的事情。
注意，这里只是获取到了 html 代码，并没有去执行 html 代码。
我们 可以想到，如果你的网页是前端渲染的，那么拿到的就只是 html 模板，没有任何内容，那自然就无法参与排名。

二、解析

爬到 html 代码之后，还需要用程序去解析 html 代码。
解析 html 代码目的是提取网页的正文部分，当然页头、页尾、Meta信息等也会被识别和提取出来。
以及最重要的，当前 html 页面里的所有的链接都会被提取出来。

三、索引

之后，当前页面解析出来的信息就会入库，然后进行分词、抽取主题、语义理解、建立正向索引、反向索引等步骤，这里哥飞就不展开讲解了。

四、处理链接

对于拿到的每一个链接，都会判断是否已经爬取过，多久之前爬的，是否还需要再爬一遍。
对于需要再爬的，会被放入到待爬列表中。
所以之前哥飞在之前的文章《 新上线的网站，如何快速让谷歌收录? 》中提醒大家，要想快速被收录，除了需要提交到GSC后台外，还可以去别的爬虫来得特别频繁的网站发帖，吸引爬虫把我们的网址加入到待爬列表里。
每一个进入待爬列表的网址，都会走上面这些步骤，不算循环。
这样谷歌就能够通过链接，爬取完互联网上所有网站的所有网页。

五、内链和外链和反链

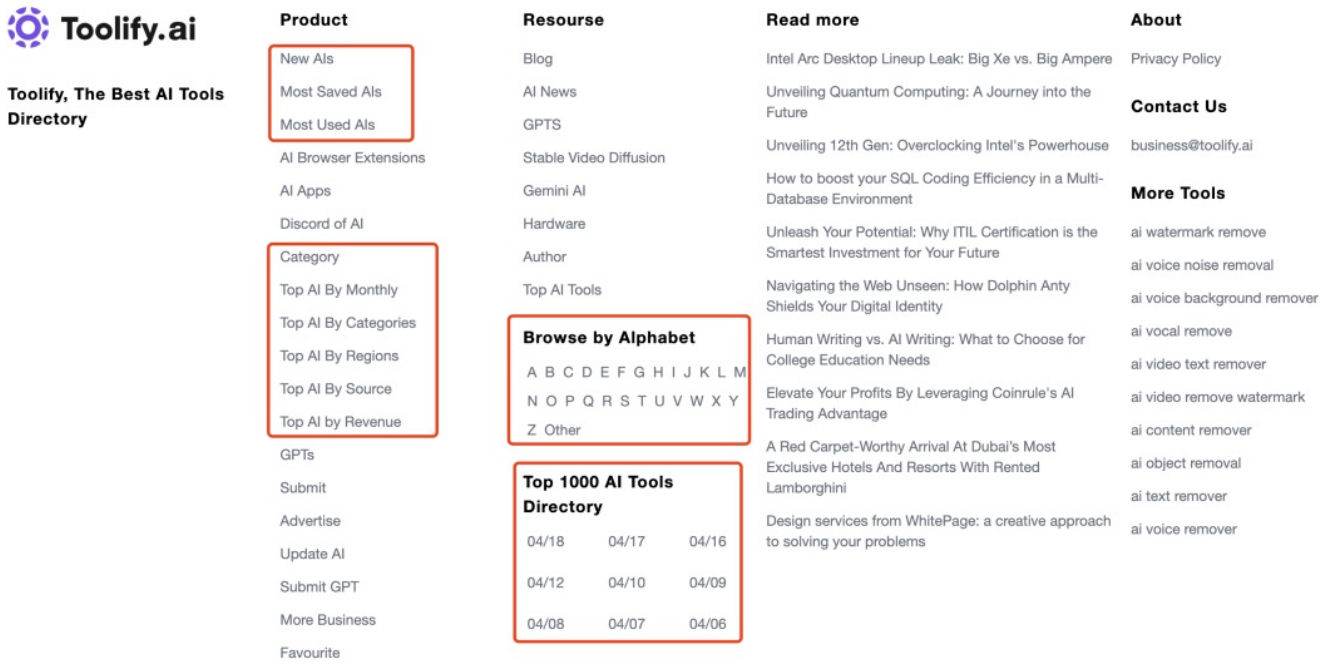
对于一个网站来说，链接分为内链 Internal Link 和外链 External Link 。
内链，是站内链接的简称，是 指向到当前网站其它页面的链接。
外链，是站外链接的简称，又名出站链接， 是指向到别的网站的链接。
假设A网站有一个指向到B网站的链接。
那么这个链接对于A网站来说，是一个 External Link， 也就是外部链接，可以简称为外链。
这个链接对于B网站来说，是一个 Back Link， 也就是反向链接，可以简称为反链。
有时候，我们会称反链为外链，其实就是同一个链接针对不同网站的说法。
如下面两句，其实是同一个意思:
“我要去搞一个 Reddit 的外链”;
“我要去 Reddit 给我网站搞一个反链”。

六、Sitemap

我们把网页在 s itemap.xml 文件里列出来， 目的也是为了把网页 加入待爬列表。
有人会问，有了 Sitemap 是不是就不用建立内链了呢?
哥飞告诉你，并不是这样的。
Sitemap 对于谷歌是一个建议，谷歌可以不听，或者部分听取。
谷歌更喜欢通过链接去不断爬行探索网站。
甚至哥飞目前最新的这个网站不到一个月收录十几万，压根就没有 Sitemap 。
方法就是在首页不断的列出最新的内容，同时有一个页面能够一页一页列出网站所有页面。
这样就可以吸引谷歌的爬虫不断来我们网站爬最新的网页。

七、页面深度

从首页开始，不断的点击链接，到达了某个页面的最短路径，就是这个页面的深度。
通常我们会说，页面深度不要超过5层，但哥飞建议，最好在3层内。
也就是你要保证你网站的任何一个页面，都可以点击三次就能够到达。
那就要把一些深度的页面改成广度的页面。
具体怎么做，大家可以参考 Toolify 的页脚部分。



如果看不懂，请加入哥飞的朋友们这个付费社群，哥飞会为大家讲解。

八、权重传递

权重会在站内传递，也会在网站之间传递。
我们可以把每一个网页都当作一个最小单元，这个网页有很多出去的链接，也有很多进入的链接。
出去的链接和进入的链接，都可能是站内链接，也可能是站外链接。
不管什么类型，每一个链接都在传递权重。
一般站内链接都是 do follow 的，也就是会完全传递权重，而外链则为了不传递太多权重出去，我们会设置为 no follow 。
谷歌有自己的一套复杂的算法，会综合各种情况去计算每一个链接能够传递的权重到底是多少。

九、关键词召回

当用户搜索一个关键词时，谷歌会找出所有内容匹配的网页。
怎么找?
根据反向索引来找。
这里哥飞就不深入讲解了，如果大家想了解，可以谷歌搜索了解更多关于搜索引擎的原理。
但是我们知道，要想网站被召回，就一定要内容足够匹配。
这就要求第一是后端渲染，能够让谷歌爬虫可以抓取到内容。
第二则要求我们即使是工具站，也要做内容，让谷歌能够根据内容知道我们的工具到底有什么作用。

十、搜索排名

被召回的结果有几千万个，怎么给网页排序呢?
谷歌最初使用 Page Rank 算法，根据不同网站的权重来进行排序。
但是经过二十多年的发展，排序算法已经复杂无比了，会根据各种因素去综合排序。
具体有哪些因素，可以看《 【哥飞解读】2024年谷歌算法排名因素变化 》。

哥飞建议大家好好看看上面这篇文章，然后在做网页时，把每一个关键因素都好好考虑到。
好了，以上介绍了怎么多，哥飞再来回答开头的三个问题。
1. 做工具站要不要做内容?
哥飞答：当然要做内容，你不做内容，谷歌怎么知道你的工具到底有什么用，谷歌不可能去运行你的工具来判断你的工具的用途，所以只能根据你写的内容来判断。
你可能会问，谷歌就这么相信我们写的?
哥飞答：谷歌一开始会把所有人都当作好人，去信任所有网页，然后根据搜索用户的表现来判断你的文字到底有没有如实的描述工具，如果发现言过其实，就会给你降低排名。

2. 为什么不要前端渲染?
哥飞答：因为谷歌无法获取到前端渲染网页的文字内容。
你可能会问，谷歌不会去运行我们的js代码吗?
哥飞答：运行代码需要消耗额外的算力，谷歌只会给一些大网站有这样的优待，因为谷歌很需要那些大网站的内容。对于我们做的小破网站，谷歌不算浪费算力的。所以我们就不要去麻烦谷歌了，我们自己做好后端渲染不是更好吗?

3. 为什么多语言不能用js切换?
哥飞答：因为js切换不会改变网址，这就导致谷歌爬虫来的时候，只能抓取到一种语言的内容，也就只能参与一种语言的搜索排名。那么这个网页的多语言就相当于白做了。
好了，以上就是今天的内容，总共2500字，够干的了。
如果你还想学习更多SEO知识，想要跟着哥飞做出海网站赚美元养老，请加入哥飞的朋友们社群。

关于社群的更多介绍，请看《 是时候给大家好好介绍一下哥飞的社群了，毕竟刚被二十年站长大佬夸过 》。

先给 大家看看已经在社群里的其他朋友怎么评价的吧。

那么这到底是一个什么样的社群呢?

一句话介绍，这是一个教你 出海做网站赚美元 的社群。

哥飞知道，肯定会有人说，教人赚钱最赚钱。

但是，哥飞这个社群，你学了，是真的能够赚美元的。

因为哥飞不是教你做一个类似的社群去赚别人的钱。

哥飞教你怎么挖掘需求，怎么搞SEO，怎么搞流量，怎么把流量变现。

或者说，哥飞其实是在给大家做微培训。

如果想更多了解社群，或者想加入社群，请加哥飞的微信 GeFeiSvip 。

当然，即使你现在不加社群，也可以先关注哥飞的公众号，多看几篇文章，你会有收获的。

修改于