

# 前端SEO优化

## 前言

以前开发独立站，涉及过一些SEO的内容，为了避免后面会遗忘，感觉有必要记录下相关经验，也算是一次总结吧！

## SEO概念

SEO 是 **Search Engine Optimizatio**（搜索引擎优化）的首字母缩写，利用搜索引擎的规则对网站进行内部及外部的调整优化，提高网页或网站在搜索引擎中关键词的自然排名，以求得获得更多的展现量和吸引免费的点击流量，从而达到互联网营销及品牌建设的目标。

## SEO的方式

### 1. Title标签

对页面点击率有直接影响，因为这是用户对网站页面的第一印象，而且也是爬虫重点的爬取对象，填写的文字要对网页内容有准确而简洁的描述，能够吸引用户点击，而且长度要适中。

### 2. Meta标签

meta标签共有两个属性，分别是http-equiv属性和name属性，不同的属性又有不同的参数值，另外还有主要作用于社交平台的OG标签。

#### (1) name属性

- **Meta Keywords:** `Meta Keywords` 是排名的重要组成部分，因为早期的搜索引擎机器人使用该值对网站进行分类，但在后来逐渐被放弃了。

- (1) [Google 在2009年正式宣布](#)，keyword标签不再是排名的一部分。
- (2) 在百度站长论坛上官方曾表示这个标签目前对SEO的影响可以忽略不计。

```
1 html
2 复制代码
3 <meta name ="keywords" content="这是keywords的内容"
```

- **description:** 告诉搜索引擎你的网站主要内容

```
1 html
2 复制代码
3 <meta name ="description" content="这是description的内容"
```

- **robots:** 告诉搜索机器人哪些页面需要索引，哪些页面不需要索引

```
1 html
2 复制代码
3 <meta name="robots" content="none"
```

- **author:** 标注网页的作者

```
1 html
2 复制代码
3 <meta name="author" content="广白,掘金"
```

## (2)http-equiv属性

- **Expires:** 可以用于设定网页的到期时间，一旦网页过期，必须到服务器上重新传输

```
1 html
2 复制代码
3 <!-- 必须使用GMT的时间格式 --><meta http-equiv="expires" content="Fri, 12 Jan
  2001 18:18:18 GMT"
```

- **Pragma:** 禁止浏览器从本地计算机的缓存中访问页面内容

```
1 html
2 复制代码
3 <meta http-equiv="Pragma" content="no-cache"
```

- **Refresh:** 自动刷新并指向新页面

```
1 html
2 复制代码
3 <!-- 其中的2是指停留2秒钟后自动刷新到URL网址 --><meta http-equiv="Refresh"
  content="2; URL=http://www.juejin.com"
```

- **Set-Cookie:** 如果网页过期，那么存盘的cookie将被删除

```
1 html
2 复制代码
3 <!-- 必须使用GMT的时间格式 --><meta http-equiv="Set-Cookie"
   content="cookievalue=xxx; expires=Friday, 10-Nov-2023 18:18:18 GMT; path=/"
```

- **Window-target:** 强制页面在当前窗口以独立页面显

```
1 html
2 复制代码
3 <!-- 用来防止别人在框架里调用自己的页面 --><meta http-equiv="Window-target"
   content="_top"
```

- **content-Type:** 设定页面使用的字符集

```
1 html
2 复制代码
3 <!-- 设定页面使用的字符集 --><meta http-equiv="content-Type" content="text/html;
   charset=gb2312"
```

### (3) OG标签

它是 Facebook 在 2010 年 F8 开发者大会公布的一种网页元信息（Meta Information）标记协议，属于 Meta Tag（Meta 标签）的范畴，是一种为社交分享而生的 Meta 标签。为了让信息内容加速流动和准确呈现，Facebook 早年极力推动这个协议，到目前几乎主流的社交媒体网站都支持 OG 协议。包括 Twitter、Pinterest、LinkedIn 和 Google+ 都可以识别 OG 协议。虽然 Twitter 也有自家的 Twitter Cards 协议，但是 Twitter 只要发现网页上没有使用自家的协议，就会用 OG 协议代替。**国内的百度、360 搜索、微博、微信、人人网等也支持该协议。**

- 使用方法：

```
1 ini
2 复制代码
3 <html prefix="og: https://ogp.me/ns#"><head><title>The Rock (1996)</title><meta
   property="og:title" content="The Rock" /><meta property="og:type"
   content="video.movie" /><meta property="og:url"
   content="https://www.imdb.com/title/tt0117500/" /><meta property="og:image"
   content="https://ia.media-imdb.com/images/rock.jpg" />...</head>...</html>
```

- 作用：

OG在社交媒体上具有丰富的内容展示，比如分享一个网址链接在社交平台后，这个链接会显示缩略图、标题和描述等，增加访客点进来的概率，左图加了OG标签，右图没有。



标签详细属性可以看官网：[ogp.me/](http://ogp.me/)

### 3. 语义化标签

#### H系列标签

从H1到H6（重要性从高到低）一共有6个，而且在页面中的作用性极高，所以不能乱用，得根据内容的重要性进行排列，最后不要有断层，比如从H1直接到H3，错过了H2；另外，H1标签最好只存在一个，更多细节可以网上搜下。

#### img标签

img的alt属性，为搜索引擎提供替代文本，图片使用alt标签优化，对搜索引擎排名产生积极影响；另外，网速不佳等原因造成无法加载图片文件时，将在图片的位置显示alt里的文字

#### 其他

还有HTML5新出的 Header, Nav, Aside, Article, Footer等语义化标签，这些都能帮助爬虫更好的获取页面内容

### 4. sitemap(站点地图)

站点地图一般是xml格式的文件，放在网站的根目录下，有些网站甚至可以通过 [网址/sitemap.xml](#) 直接访问到（手动狗头），文件里包含了每个网页的链接（loc），更新时间（lastmod），权重（priority）等信息，权重从0到1，依次递增，一般主页设为1，然后其他按重要性递减，拿掘金



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocat:
  <!-- created with Free Online Sitemap Generator www.xml-sitemaps.com -->
  ▼<url>
    <loc>https://juejin.cn/</loc>
    <lastmod>2023-11-10T13:57:44+00:00</lastmod>
    <priority>1.00</priority>
  </url>
  ▼<url>
    <loc>https://juejin.cn/pins</loc>
    <lastmod>2023-11-10T13:57:44+00:00</lastmod>
    <priority>0.80</priority>
  </url>
  ▼<url>
    <loc>https://juejin.cn/course</loc>
    <lastmod>2023-11-10T13:57:44+00:00</lastmod>
    <priority>0.80</priority>
  </url>
```

搜索引擎可以通过查看 [站点地图](#) 快速获取网站的整体结构，并将精力集中在重要页面的索引上，这对于提高搜索引擎的爬行效率和索引速度非常有帮助。

## 5. robots文件

robots.txt，是一个给爬虫下指令的文本文件，能让其合理地抓取网站内资源，而且可以将网站不重要的内容、模块等进行屏蔽，从而抓取更多有价值高质量的内容和网页，提高网站排名。大多数网站都可以通过 [网址/robots.txt](#) 进行获取，还是拿掘金来举例：



- **User-agent:** 后面填你要针对的搜索引擎，\*代表全部搜索引擎
- **Disallow:** 后面填你要禁止抓取的网站内容和文件夹，/做前缀
- **Allow:** 后面填你允许抓取的网站内容，文件夹和链接，/做前缀
- **Crawl-delay:** 后面填数字，意思是抓取延迟，小网站不建议使用
- **Sitemap:** 站点地图的存放位置

## 6. 内链和外链

### 内链

在自己的网站当中通过链接的方式在各个内容页面之间相互链接，从而提高爬虫以及一些搜索引擎对于网站的爬行索引效率；另外，为了避免内链起到反作用，在进行网站待更新的时候，一定要注意定期清理死链和断链，方便爬虫可以顺着链接进行收录爬行，达到更好的收录效果。

### 外链

通过在其他一些高流量的网站放置自己的链接，相较于内链的内部操作，外链可以达到网络之间的信息分享，不再让我们的网站内容形成孤立，可以很快的增加网站的浏览数量，提升搜索排名，这对一个刚刚建立起来的新站来说，外链的数量基本上可以成为这个网站快速提升流量的关键所在，不过对于后期优化，外链的发布一定要以质量为主，数量为辅。

## nofollow

- **作用：**超链接a标签中的一个属性（还有meta标签），意思是不向这个页面导出权重，也就是说，加上了nofollow标签的链接，目的是告诉搜索引擎不要跟踪加这个链接，不要传递链接权重，不要在搜索算法中计算这个链接。通过设置nofollow标签，我们可以控制网站权重的流动（也就是说常说的“集中权重”），避免链接指向垃圾页面，这样就可以让网站主要的页面能够更快的获得关键词排名了。有的公司在导航上有在线留言，这个就可以加上，因为这个是没有用的，属于垃圾页面，页面会向这个页面导出权重，权重就分散了，另外还有类似广告链接等无意义的页面
- **nofollow 和 external nofollow:** 从字面上意思来看，external是“外部的”，nofollow是“不要追踪”，综合来看external nofollow 则表示“外部的不要追踪”，这两根本就是同一个意思，external nofollow 只是nofollow比较规范的书写而已。

```
1 html
2 复制代码
3 <a href="login.html" rel="nofollow"登录</a><a href="login.html" rel="external
  nofollow"登录</a>
```

## 7.数据结构化标记

结构化数据标记并不能直接帮助提高搜索排名，但它能带来很多好处：

### （1）丰富搜索结果

比如搜索 掘金 出现的页面，红色框部分就是结构化标记的成果，从框内可以更好的了解该网站的内容，而且占据了大版面也有利于吸引用户注意。

### （2）拥有知识面板

在右侧会有个知识面板，该面版可提供更高的品牌知名度和权威性，如下图的右侧红框：





Facebook



[All](#) [Videos](#) [Books](#) [Images](#) [News](#) [More](#)

Tools

About 25,140,000,000 results (0.32 seconds)



Facebook  
<https://www.facebook.com>

Facebook - log in or sign up

Create an account or log into Facebook. Connect with friends, family and other people you know. Share photos and videos, send messages and get updates.

Log In

Log into Facebook to start sharing and connecting with your ...

Find Friends browser

Log into Facebook to start sharing and connecting with your ...

Facebook settings

Log into Facebook to start sharing and connecting with your ...

[More results from facebook.com »](#)



Google Play  
<https://play.google.com/store/apps/details?id=com.facebook.katana>

Facebook - Apps on Google Play

Explore, connect and share your interests with real people.

★★★★★ Rating: 3.4 · 137,260,830 votes · Free · Android · Social Networking



Wikipedia  
<https://en.wikipedia.org/wiki/Facebook>

Facebook

Facebook is an online social media and social networking service owned by American technology giant Meta Platforms. Created in 2004 by Mark Zuckerberg with ...

Meta

Technology company



facebook

Meta Platforms, Inc., doing business as Meta, and formerly named Facebook, Inc., and TheFacebook, Inc., is an American multinational technology conglomerate based in Menlo Park, California. The company owns and operates Facebook, Instagram, Threads, and WhatsApp, among other products and services. [Wikipedia](#)

**Founders:** Mark Zuckerberg, Eduardo Saverin, Andrew McCollum, Chris Hughes, Dustin Moskovitz

**Stock price:** META (NASDAQ)

\$329.19 0.00 (0.00%)

Nov 13, 4:00 PM EST - Disclaimer

**Headquarters:** Menlo Park, CA

**Founded:** February 2004, Cambridge, MA

**CTO:** Andrew Bosworth

**CEO:** Mark Zuckerberg (Jul 2004–)

**Revenue:** 32 billion USD (2023 Q2)

**Subsidiaries:** WhatsApp, Reality Labs, Meta Platforms Ireland, MORE

[Disclaimer](#)

### (3) 支持语义搜索

不再是关键词的匹配搜索，而是通过关键词去寻找问题的答案。当你搜索如下模糊内容时，Google 会通过这种方式设法返回合适的结果：



马斯克的公司有哪些



[All](#)

[News](#)

[Images](#)

[Videos](#)

[Books](#)

[More](#)

Tools

About 16,300,000 results (0.49 seconds)

他是SpaceX创始人、董事长、执行长、首席工程师，特斯拉投资人、执行长、产品设计师、前董事长，无厘公司创始人，Neuralink、OpenAI联合创始人，同时也是X公司的首席技术官、董事长。2022年马斯克以2190亿美元财富成为世界首富。



Wikipedia

<https://zh.wikipedia.org/wiki/伊隆·马斯克>

埃隆·马斯克- 维基百科，自由的百科全书

[About featured snippets](#) • [Feedback](#)



维基百科

<https://zh.wikipedia.org/zh-hans>

埃隆·马斯克- 维基百科，自由的百科全书 - 维基百科

SpaceX创始人 董事长 首席执行官 首席工程师 特斯拉首席执行官 产品设计师 前董事长

### (4) 体现 E-A-T



**E-A-T** 意思是专业 (Expertise) , 权威 (Authoritativeness) 和可信 (Trustworthiness) 的缩写, E-A-T是谷歌算法的一部分, 并被写入[谷歌搜索质量评估指南](#)中。

## 结构化数据形式

结构化数据有三种形式: JSON-LD, Microdata和RDFa。个人之前一直用的JSON-LD, 如下:

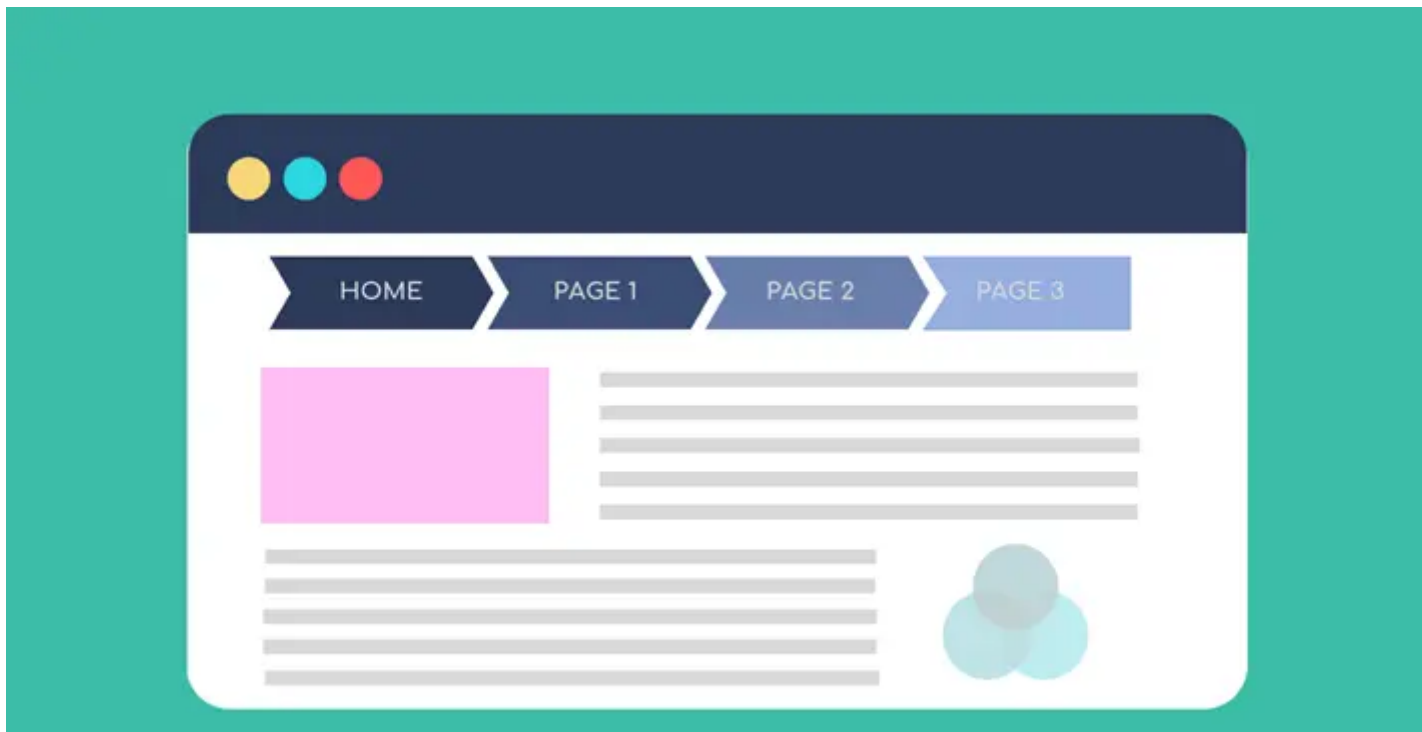
```
1 js
2 复制代码
3 <script type="application/ld+json"    {      "@context": "https://schema.org",
      "@type": "NewsArticle",      "headline": "Title of a News Article",
      "image": [      "https://example.com/photos/1x1/photo.jpg",
      "https://example.com/photos/4x3/photo.jpg",
      "https://example.com/photos/16x9/photo.jpg"      ],      "datePublished":
      "2015-02-05T08:00:00+08:00",      "dateModified": "2015-02-05T09:20:00+08:00",
      "author": [{      "@type": "Person",      "name": "Jane Doe",
      "url": "https://example.com/profile/janedoe123"      }, {
      "@type": "Person",      "name": "John Doe",      "url":
      "https://example.com/profile/johndoe123"      }]    }</script>
```

## Tips

这段代码可以放在 `<head>` 或 `<body>` HTML 部分的任意位置, 更多的JSON-LD款式可以参考[Google官方](#)

## 8.面包屑导航

1. 告诉爬虫当前所处的位置, 提供抓取路径的引导, 让其更快速的了解网站的整体架构, 在抓取网页时也能清楚知道网页的层级及分类索引, 有助于提升搜索结果的排名。此外, 设定面包屑之后, 网站在搜索结果页上也会呈现网址路径, 因此可以适时在面包屑的名称中加入关键词, 增加用户点击率
2. 优化用户体验, 让用户清楚的知道自己所在的位置, 更容易筛选自己需要的信息



## 9.服务端渲染（SSR）

爬虫只会爬取源码，不会执行网站的Js脚本,使用了Vue或者React之类的框架后，页面大多数DOM元素都是在客户端根据js动态生成，可供爬虫抓取分析的内容大大减少。另外，浏览器爬虫不会等待我们的数据完成之后再去抓取我们的页面数据。服务端渲染返回给客户端的是已经获取了异步数据并执行Js脚本的最终HTML，网络爬中就可以抓取到完整页面的信息，所以就要用到SSR渲染了。

如果是项目开始的时候，就知道要做SEO，建议使用比较成熟的SSR框架

1. 基于Vue的 [Nuxtjs](#)
2. 基于Reactde [Nextjs](#)

如果是项目已经成形或者只想优化部分页面，建议使用一些插件来实现，比如 [prerender-spa-plugin](#)、[vue-server-renderer](#) (适用于Vue)

**Tips：** 爬虫不会抓取iFrame里的内容，所以尽量避免使用

## 10.网站地址

1. 和产品关键词或者公司名字高度相关的域名，能够有效提升网站的SEO排名。
2. 各个页面的路径地址尽量不要过长，以及不能出现中文，否则会影响收录。
3. 由于网站改版或者其他因素，可能会出现部分链接地址生效，这时候可以进行301重定向到新的页面，避免出现过多的死链（404页面），如果死链太多，搜索引擎可能认为网站不可靠，从而影响网站的排名和流量。

4. 下面两个链接都是百度页面，像这种有多个URL为相同页面的情况，就是网址规范化问题。无论是对搜索引擎还是网站来说，多个URL为相同页面，浪费了抓取资源，而且如果多个网址内容相同或者高度相同，可能会认为网站重复页面过多一类的因素导致影响收录甚至降权。在一个页面有多个URL时，还会对页面找出权重分散，降低了页面排名能力，网站无意义收录等问题 当页面出现URL规范问题，搜索引擎也会自动判断并选择一个认为是规范的页面进行索引。

- [baidu.com](http://baidu.com)
- [www.baidu.com](http://www.baidu.com)

解决方案：可通过在每个非规范版本的 HTML 网页的 `<head>` 部分中，添加一个 `rel="canonical"` 链接来进行指定规范网址。

```
1  html
2  复制代码
3      <!-- href填写要作为标准的网址 -->      <link rel="canonical"
        href="http://baidu.com"
```

## 11.网站性能

网站打开速度越快，识别效果越好，否则爬虫会认为该网站对用户不友好，降低爬取效率，这时候就要考虑压缩文件体积之类的性能优化了。

## 5. 使用https

Google和其他搜索引擎已经明确表示，他们更喜欢使用HTTPS，因为它提供更高的安全性和更好的用户体验。当您的网站使用HTTPS时，搜索引擎会将其视为更可信和更安全的网站，从而为其排名增加积极因素。

## 13.提交站点收录

将创建好的网站地图提交给搜索引擎，以便搜索引擎能够更快更及时地抓取和索引网站。

1. [百度](#)
2. [google](#)
3. [Bing](#)

## 网站访问数据

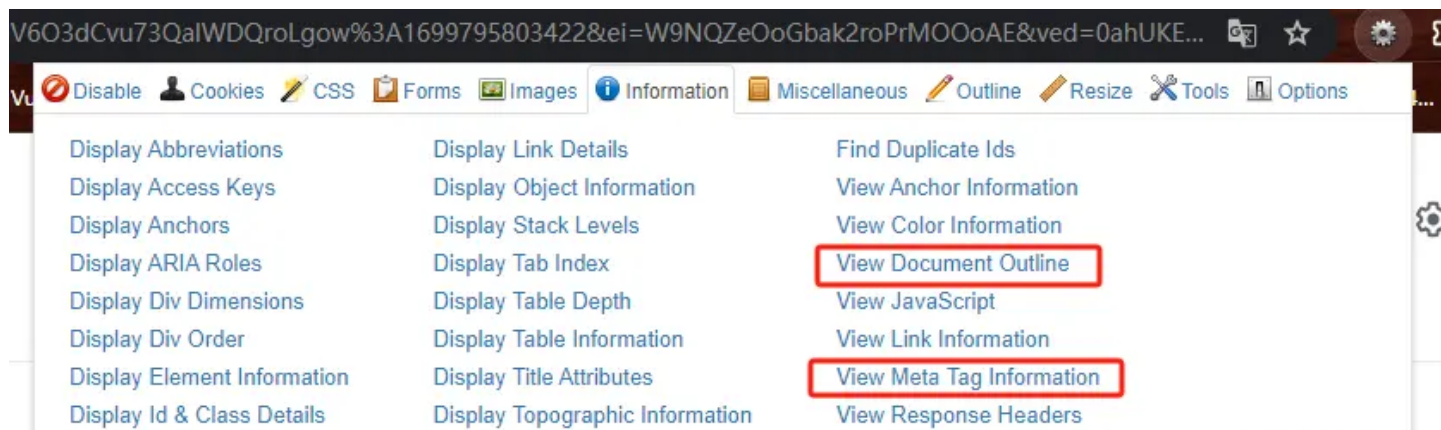
后期维护需要一定的网站访问数据做分析，比如流量来源，页面点击，地域分布等，不同的搜索引擎都有自己的一套分析工具，我们只要按照官网的教程，把埋点的代码嵌入到项目即可。

- [百度统计](#)
- [Google Analytics](#)

## 附赠

1. 检测Title Tag & Meta Description优化
2. 站点地图生成
3. [Web Developer](#), 一个Chrome插件，集成了各种各样的 Web 开发工具，比如可以使用下图红框里的选项（图1），查看网页的Title, Meta, H标签等（图2，3）。

图（1）



图（2）

图（3）

[https://www.google.com.hk/search?q=Web+Develope&oq=Web+Develope&gs\\_lcrp=EgZjaHJvbWUyBggAEElBMgYIBRBFGEyBggGEEUYPNIBBzM3N2owajSoAgCwAgA&sourceid=c](https://www.google.com.hk/search?q=Web+Develope&oq=Web+Develope&gs_lcrp=EgZjaHJvbWUyBggAEElBMgYIBRBFGEyBggGEEUYPNIBBzM3N2owajSoAgCwAgA&sourceid=c)

▣ 26 headings

# <h1> (Missing heading)

## <h2> Accessibility Links

## <h2> Search Modes

# <h1> Search settings

# <h1> Search Results

## <h2> Featured snippet from the web

### <h3> Web Developers and Digital Designers

### <h3> How to Become a Web Developer From Scratch and Find a Job - Kinsta

### <h3> Web Designing: Courses, Admission 2024, Entrance Exam, Syllabus ...

### <h3> Do You Need a Degree to Be a Web Developer? - BrainStation

### <h3> Programming vs Web Development: What's the Difference? - Shiksha

### <h3> What Is a Web Developer? | Skills and Career Paths

### <h3> What Does a Web Developer Do (and How Do I Become ...

## 总结

个人觉得，前端的工作都是集中在项目前期，后面的话主要是一些维护工作，比如公司之前开发了一个支持富文本内容的文章发布平台，然后运营部门就可以自己发一些引流的文章了，和开发关系不大。另外，公司专门请了一个增长黑客（其实就是SEO优化工程师的角色），负责每周给公司写文章的同学提供关键词，分析网站访客数据，提交网站收录，寻找外链资源等。