# Cloudflare Pages 部署项目坑点之 404.html

504 字

## 本文浓缩版

请务必确保你在 Cloudflare Pages 部署的项目中根目录下有 `404.html` 这个文件，这样可以保证用户访问你网站中的无效路径的时候，正常返回 404 响应，否则可能会返回首页的内容，从而产生一些意外的现象。

## 完整版

如果你在 Cloudflare Pages 部署的项目中没有 `404.html`，Cloudflare 会默认这个项目是一个单页应用（SPA），这时所有无效路径都会用根路径的内容进行响应。参考文档：

> *“If your project does not include a top-level `404.html` file, Pages assumes that you are deploying a single-page application. This includes frameworks like React, Vue, and Angular. Pages' default single-page application behavior matches all incoming paths to the root (`/`), allowing you to capture URLs like `/about` or `/help` and respond to them from within your SPA.*
>
> *——Serving Pages | Cloudflare Pages docs”*

也就是说，如果没有 `404.html`，哪怕你用 `https://yourwebsite.com/asihdsebfjkj`这种无效路径访问你的项目，页面仍然可以正常显示。

这会带来什么问题呢？

1. 可能会影响谷歌爬虫对你网站的爬取，如果你的网站中有一些相对路径存在问题，导致谷歌访问到了一个不存在的路径，原本你的网站应当返回 404，但是却正常返回了结果，那么谷歌爬虫会继续对这个页面上的内容进行爬取，从而可能导致一系列错误的抓取路径。
2. 由于谷歌爬虫爬取错误的路径也能得到正常的网页，有可能会一环套一环，产生大量的无效抓取。既影响了谷歌分配给你的网站的抓取资源，也导致你的服务器产生了很多无效流量。

本文还有啰唆版本详细介绍本问题导致谷歌爬虫疯狂访问我的网站，欢迎围观：我的网站被 Googlebot 攻击了，凶手竟是 Cloudflare - 即刻