规范网址：处理重复页面

原作者：Google Search Central
原始网址：https://developers.google.com/search/docs/crawling-indexing/consolidate-duplicate-urls?hl=zh-cn
原文语言：中文；采集日期：2026-10-19
许可：CC BY 4.0；代码示例按原站 Apache 2.0 许可。
保留官方原文与图片，仅移除网站导航、调整排版并将图片本地化。最新内容以原始页面为准。

如何使用 rel="canonical" 及其他方法指定规范网址

若要向 Google 搜索指定重复网页或非常相似网页的规范网址，您可以使用多种方法指明您更愿意使用哪个网址。这些方法按照其对规范化的影响程度排列如下：

- **重定向**：强信号，表明重定向的目标应成为规范网址。
- **rel="canonical" link 注释**：强信号，表明所指定的网址应成为规范网址。
- **站点地图包含**：弱信号，有助于站点地图中包含的网址成为规范网址。

请注意，这些方法可以叠加，因此组合使用会更有效。 这意味着，如果您使用两种或更多种方法，将会增加您的首选规范网址出现在搜索结果中的几率。

虽然我们建议您使用这些方法，但并非硬性要求；即使您不指定首选规范网址，您的网站或许也能表现不错。这是因为，如果您没有指定规范网址，Google 会客观地确定哪个版本的网址最适合在 Google 搜索中向用户显示。

如果您使用了 WordPress、Wix 或 Blogger 等 CMS，可能无法直接修改 HTML。实际上，您的 CMS 可能具有搜索引擎设置页面或其他某种机制，能够将规范网址告知搜索引擎。不妨在 CMS 上搜索有关如何修改网页 <head> 的说明（例如，搜索“wordpress set the canonical element”）。

指定规范网址的原因

虽然指定首选规范网址通常并不重要，但您还是会出于各种原因希望将一组重复或类似网页中的规范网页明确告知 Google：

- **指定您希望用户在搜索结果中看到的网址**。 您可能希望用户通过 `https://www.example.com/dresses/green/green-dress.html`（而非 `https://example.com/dresses/cocktail?gclid=ABCD`）访问您的绿色连衣裙商品页。
- **整合类似网页或重复网页的信号**。指定规范网址可帮助搜索引擎将它们掌握的关于各个网址的信号（例如指向它们的链接）整合到一个首选网址上。这意味着，从其他网站到 `https://example.com/dresses/cocktail?gclid=ABCD` 的信号会整合到指向

`https://www.example.com/dresses/green/green-dress.html`（如果该网址成为规范网址）的链接。

- **简化一段内容的跟踪指标**。如果特定内容可以通过多个网址访问，获取此内容的综合指标的难度会更大。
- **避免花费时间抓取重复网页**。您可能希望 Googlebot 在您的网站上发现尽量多的内容，因此最好让 Googlebot 将时间用于抓取您网站上的新网页（或更新后的网页），而不是抓取相同内容的重复版本。

最佳做法

无论使用哪种规范化方法，都请遵循以下最佳实践：

- **请勿**使用 robots.txt 文件进行规范化。 Google 仍然可能将 robots.txt 中禁止抓取的网址编入索引，但不会收录其内容。
- **请勿**使用网址移除工具进行规范化，它会在搜索结果中隐藏网址的所有版本。
- **请勿**使用不同的规范化方法为同一网页指定不同的规范网址（例如，请勿既在站点地图中为某个网页指定一个规范网址，又使用 `rel="canonical"` 为同一网页另行指定一个规范网址）。
- **请勿**将网址片段指定为规范网址，因为 Google 通常不支持网址片段。
- **请务必**在规范网页本身上添加 `rel="canonical"` 链接（也称为自引用规范）。
- **我们不建议**使用 `noindex` 阻止选择单个网站中的规范网页，因为这样会完全阻止该网页显示在 Google 搜索结果中。`rel="canonical"` `link` 注释是首选解决方案。
- 如果您使用的是 `hreflang` 元素，请务必指定一个采用同一语言的规范网页；如果没有这样的规范网页，请指定一个采用最佳替代语言的规范网页。
- 在网站中提供链接时，请链接到规范网址（而非重复网址）。 始终链接到您认定的规范网址有助于 Google 了解您偏好的网址。
- 如果您使用 JavaScript 进行客户端渲染，请务必确保规范网址的相关信息尽可能清晰。 最好的方法是在 HTML 源代码中指定规范网址，并确保 JavaScript 不会更改规范 link 元素。 如果您无法在 HTML 源代码中设置规范网址，请将其省略，仅使用 JavaScript 进行设置。 这样可确保规范网址的相关信息尽可能清晰。

不同规范化方法之间的比较

下表比较了不同的规范化方法，重点说明了它们在不同场景中在维护和效果方面的优势和劣势。

方法和说明	
`rel="canonical"` link 元素	在所有重复网页的代码中分别添加一个 `<link>` 元素，使其指向规范网页。
	优点： / **缺点：**

	• 可以映射无限多个重复网页。	• 在大型网站或网址经常改变的网站上维护映射可能会比较复杂。 • 仅适用于 HTML 网页，不适用于 PDF 之类的文件。在这类情况下，您可以使用 `rel="canonical"` HTTP 标头。
rel="canonical" HTTP 标头	在网页响应中发送 `rel="canonical"` 标头。 **优点：** • 不会导致网页大小增加。 • 可以映射无限多个重复网页。	**缺点：** • 在大型网站或网址经常改变的网站上维护映射可能会比较复杂。
站点地图	在站点地图中指定您的规范网页。 **优点**： • 易于实施和维护，尤其是在大型网站上。	**缺点**： • Google 仍必须为您在站点地图中声明的所有规范网页确定关联的重复网页。 • 此方法向 Google 发送的信号不如 `rel="canonical"` 映射方法发送的信号强。
重定向	使用永久重定向告知 Google，重定向网址是比其重定向到的网址更差的版本。请仅在弃用重复网页时使用此方法。	
AMP 变体	如果您的某个网页变体是 AMP 网页，请按照 AMP 指南指明规范网页和 AMP 变体。	

使用 `rel="canonical"` `link` 注释

Google 支持明确的 `rel` canonical `link` 注释（如 RFC 6596 中所述）。系统会忽略建议网页替代版本的 `rel="canonical"` 注释；具体来说就是，带有 `hreflang`、`lang`、`media` 和 `type` 属性的 `rel="canonical"` 注释不用于规范化。请改为使用适当的 `link` 注释来指定网页的备用版本；例如用 `link rel="alternate"` `hreflang` 进行语言和国家/地区注释。

您可以通过以下两种方式提供 `rel="canonical"` `link` 注释：

- HTML 中的 `rel="canonical"` `link` 元素
- `rel="canonical"` `link` HTTP 标头

我们建议您从中选择一个使用；在受支持的情况下，同时使用这两种方法更容易出错（例如，您可能会在 HTTP 标头中提供一个网址，在 `rel="canonical"` `link` 元素中提供另一个网址）。

rel="canonical" link 元素

rel="canonical" link 元素（也称为“规范元素”）是指在 HTML 的 head 部分中使用的元素，用于指明另一个网页可体现该网页上的内容。

假设您想将 https://example.com/dresses/green-dresses 设为规范网址（即使有很多个网址指向该内容），那么您可通过执行以下步骤，将此网址指定为规范网址：

1. 将具有 rel="canonical" 属性的 <link> 元素添加到重复网页的 <head> 部分中，并使其指向规范网页。例如：

   ```
   <html>
   <head>
   <title>Explore the world of dresses</title>
   <link rel="canonical" href="https://example.com/dresses/green-dresses" />
   <!-- other elements -->
   </head>
   <!-- rest of the HTML -->
   ```

   我们建议您也向规范网页本身添加此自引用 rel="canonical" link 元素。
2. 如果规范网页有采用不同网址的移动版变体，请为其添加 rel="alternate" link 元素，并使该链接指向此网页的移动版：

   ```
   <html>
   <head>
   <title>Explore the world of dresses</title>
   <link rel="alternate" media="only screen and (max-width: 640px)"
   href="https://m.example.com/dresses/green-dresses">
   <link rel="canonical" href="https://example.com/dresses/green-dresses" />
   <!-- other elements -->
   </head>
   <!-- rest of the HTML -->
   ```

3. 为此网页添加适当的 hreflang 或其他元素。

对于 rel="canonical" link 元素，请使用绝对路径（而非相对路径）。尽管 Google 支持相对路径，但从长远来看，相对路径可能会造成问题（例如，如果您无意中允许抓取您的测试网站），因此我们不建议您这样做。

正面示例： https://www.example.com/dresses/green/green-dress.html

反面示例： /dresses/green/green-dress.html

rel="canonical" link element 仅当出现在 HTML 的 <head> 部分时才被接受，因此请确保至少 <head> 部分是有效的 HTML。

如果您使用 JavaScript 添加 rel="canonical" link 元素，请务必正确注入规范 link 元素。

`rel="canonical"` HTTP 标头

如果您能更改服务器配置，则可使用 `link` HTTP 响应标头（带有 RFC5988 中定义的 `rel="canonical"` 目标属性）（而不是 HTML 元素）为 Google 搜索支持的文档（包括 PDF 文件等非 HTML 文档）指明规范网址。

Google 仅支持在网页搜索结果中使用此方法。

如果您以多种文件格式（例如 PDF 或 Microsoft Word）发布内容，并且均采用自己的网址，那么您可以返回 `rel="canonical"` HTTP 标头，告知 Googlebot 哪个是非 HTML 文件的规范网址。例如，若要指明 PDF 版本的 `.docx` 版本应为规范网址，您可以为内容的 `.docx` 版本添加此 HTTP 标头：

```
HTTP/1.1 200 OK
Content-Length: 19
...
Link: <https://www.example.com/downloads/white-paper.pdf>; rel="canonical"
...
```

与 `rel="canonical"` `link` 元素一样，请在 `rel="canonical"` HTTP 标头中使用绝对网址。

使用站点地图

请分别为您的每个网页选择一个规范网址，然后通过站点地图提交这些规范网址。您在站点地图中列出的所有网页都会被视为向系统建议的规范网页；Google 会根据内容相似度决定哪些网页是重复网页（如果有）。

在站点地图中提供首选规范网址，即可轻松为大型网站指定规范网址。站点地图还能有效地告诉 Google，您认为网站上的哪些页面最为重要。

使用重定向

如果您想移除现有的重复网页，请使用此方法。所有永久重定向方法对 Google 搜索的效果相同，但搜索引擎发现不同重定向方法所用的时间可能有所不同。

要实现最快效果，请使用 HTTP（也称为“服务器端”）重定向。

假定用户可通过以下几种方式访问您的网页：

- `https://example.com/home`
- `https://home.example.com`
- `https://www.example.com`

您可从这些网址中挑选一个作为规范网址，并使用重定向将来自其他网址的流量引导至首选网址。

其他信号

除了明确提供的方法之外，Google 还使用一组规范化信号，这些信号通常基于网站设置：优先选择 HTTPS（而非 HTTP）以及优先选择 `hreflang` 集群中的网址。

优先选择 HTTPS（而非 HTTP）网址作为规范网址

Google 会优先选择 HTTPS 网页（而非等效的 HTTP 网页）作为规范网页，除非存在如下问题或冲突信号：

- HTTPS 网页的 SSL 证书无效。
- HTTPS 网页包含不安全的关联功能（图片除外）。
- HTTPS 网页会将用户重定向至 HTTP 网页或通过 HTTP 网页重定向用户。
- HTTPS 网页包含指向 HTTP 网页的 `rel="canonical" link`。

虽然我们的系统在默认情况下会优先选择 HTTPS 网页（而非 HTTP 网页），但您可通过执行以下任一操作来确保此行为始终都会发生：

- 添加从 HTTP 网页指向 HTTPS 网页的重定向。
- 添加从 HTTP 网页指向 HTTPS 网页的 `rel="canonical" link`。
- 实施 HSTS。

为防止 Google 误将 HTTP 网页选为规范网页，请**避免**以下几种做法：

- 避免使用错误的 TLS/SSL 证书和 HTTPS 到 HTTP 重定向，因为这会导致 Google 非常倾向于选择 HTTP 网页。即使实施 HSTS 也无法消除这种强烈的倾向。
- 请勿在站点地图或 `hreflang` 注释中包含网页的 HTTP 版本（而不是 HTTPS 版本）。
- 避免为错误的主机版本实施 SSL/TLS 证书。例如，在 `example.com` 上提供 `subdomain.example.com` 的证书。该证书必须与您的完整网站网址匹配，或者必须是可用于同一网域上多个子网域的通配证书。

优先选择 `hreflang` 集群中的网址

为了方便网站进行本地化，出于规范化目的，Google 会优先选择 `hreflang` 集群中的网址。例如，如果 `https://example.com/de-de/cats` 和 `https://example.com/de-ch/cats` 通过 `hreflang` 注释指向对方，而不是指向 `https://example.com/de-at/cats`，`de-de` 和 `de-ch` 的网页将成为首选规范网页（而不是未出现在 `hreflang` 集群中的 `/de-at/` 网页）。

详细了解如何排查和解决规范化问题。