



# 使用规范网址

本文介绍如何使用规范网址，以便为可通过多个网址结构或整合访问的内容改善链接和排名信号。

在内容管理系统和在线购物系统中，我们经常通过多个网址提供相同的内容。而借助内容转载，我们也可以非常轻松地将内容分发到完全不同的网址和网域。例如：

由于用户会话和/或搜索偏好设置的不同，同一商品页会有多个动态网址。	<a href="https://www.example.com/products?category=dresses&amp;color=green">https://www.example.com/products?category=dresses&amp;color=green</a> <a href="https://example.com/dresses/cocktail?gclid=ABCD">https://example.com/dresses/cocktail?gclid=ABCD</a> <a href="https://www.example.com/dresses/green/greendress.html">https://www.example.com/dresses/green/greendress.html</a>
当您同时将同一篇博文同时放在多个内容版块下时，您的博客系统会自动保存多个网址。	<a href="https://blog.example.com/dresses/green-dresses-are-awesome/">https://blog.example.com/dresses/green-dresses-are-awesome/</a> <a href="https://blog.example.com/green-things/green-dresses-are-awesome/">https://blog.example.com/green-things/green-dresses-are-awesome/</a>
您的服务器已配置为针对 <b>www</b> 子网域或 <b>http</b> 协议投放相同的内容。	<a href="http://example.com/green-dresses">http://example.com/green-dresses</a> <a href="https://example.com/green-dresses">https://example.com/green-dresses</a> <a href="http://www.example.com/green-dresses">http://www.example.com/green-dresses</a>
您在此博客上提供的要转载到其他网站的内容与这些网域中的原有内容完全重复或部分重复。	<a href="https://news.example.com/green-dresses-for-every-day-155672.html">https://news.example.com/green-dresses-for-every-day-155672.html</a> （转载博文） <a href="https://blog.example.com/dresses/green-dresses-are-awesome/3245/">https://blog.example.com/dresses/green-dresses-are-awesome/3245/</a> （原始博文）

尽管这些系统更便于开发和分发内容，但当用户使用搜索引擎访问您的网页时，这些系统也会带来一些麻烦。例如：

- 合并重复内容或类似内容的链接信号。这些系统可帮助搜索引擎将针对各个网址（例如，指向它们的链接）的信息合并到单个首选网址。也就是说，从其他网站指向 <http://example.com/dresses/cocktail?gclid=ABCD> 的链接会与指向 <https://www.example.com/dresses/green/greendress.html> 的链接合并。
- 跟踪单个商品/主题的指标。当某个内容片断有多个网址时，它的综合指标获取难度会更大。

- 确定您希望用户看到的网址。您希望用户通过 <https://www.example.com/dresses/green/greendress.html>（而非 <https://example.com/dresses/cocktail?gclid=ABCD>）访问您的绿色礼服商品页。
- 指定转载内容的网页排名。如果您是为了将内容发布到其他网域而对其进行转载，那您就需要将网页排名合并到首选网址。

为了解决上述问题，我们建议您为可通过多个网址访问的内容（或等同内容）定义一个规范网址。您可通过以下几种方式为 **Google** 搜索定义此项设置：

- [设置您的首选网域](#)
- [使用 `rel="canonical"` 链接元素指明首选网址](#)
- [使用站点地图为相同的内容设置首选网址](#)
- [对不规范的网址使用 301 重定向](#)
- [指明如何处理动态参数](#)
- [在您的 HTTP 标头中指定一个规范链接](#)
- [优先选择 HTTPS 而不是 HTTP 作为规范链接](#)



上述方法只是我们的建议，并非硬性要求。如果您未指明规范网址，我们会自行确定一个最佳版本。

请勿使用 `robots.txt` 文件进行规范化。

请勿使用网址移除工具进行规范化：它会从搜索结果中移除相应网址的所有版本。

请勿为同一网页指定不同的规范网址（例如，在站点地图中指定了一个网址，同时还使用 `rel="canonical"` 为同一网页指定了另一个网址）。

## 设置您的首选网域

告知 **Google** 您希望为自己的网域使用哪个版本的网站网址：

- <https://www.example.com>
- <https://example.com>

如果您将首选网域设置为 <https://example.com>，**Google** 便会将指向 <https://www.example.com> 的链接与指向 <https://example.com> 的链接视为完全相同。

有关详情，请参阅[设置您的首选网域](#)。

## 使用 `rel="canonical"` 链接元素指明首选网址

假设您希望将 <https://blog.example.com/dresses/green-dresses-are-awesome/> 设置为首选网址（即使有多个网址指向该内容也是如此），那么您可通过以下方式将这项设置告知搜索引擎：

- 使用 `rel="canonical"` 链接元素标记规范网页和任何其他版本。  
将属性为 `rel="canonical"` 的 `<link>` 元素添加到这些网页的 `<head>` 部分：

```
<link rel="canonical" href="https://blog.example.com/dresses/green-dresses-are-awesome" />
```

这即指明了用于访问绿色礼服博文的首选网址，因此搜索结果将更有可能向用户显示该网址结构。（请注意：我们会尽力而为，但无法保证在任何情况下均会遵循这项设置。）

**避免错误：**使用绝对路径，而非包含 `rel="canonical"` 链接元素的相对路径。

建议使用的结构：<https://www.example.com/dresses/green/greendresss.html>

不建议使用的结构：</dresses/green/greendress.html>。

## 使用站点地图为相同的内容指明首选网址

为您的每个商品页选择一个规范（首选）网址，然后通过[站点地图](#)提交这些规范网址，让我们知道您的偏好设置。

我们无法保证一定会使用您通过站点地图提交的网址，但提交网址是一种可让 **Google** 了解您认为自己网站上哪些网页最重要的好方法。

## 针对不规范的网址使用301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>

您最好从这些网址中挑选一个作为首选（规范）目标网址，并使用 **301** 重定向将来自其他网址的流量发送到您的首选网址。服务器端 **301** 重定向是确保将用户和搜索引擎定向到正确网页的最佳方式。**301** 状态代码表示某网页已永久迁移至新位置。

## 指明如何处理动态参数

使用[参数处理方式](#)将您希望忽略的参数告知 **Google**。忽略某些参数既可以减少 **Google** 索引中的重复内容，也可以使您的网站更易于抓取。例如，如果您指定应忽略参数 `sessionid`，那么 **Google** 会将 <https://www.example.com/dresses/green.php?sessionid=273749> 与 <https://www.example.com/dresses/green.php> 视为相同。

## 在您的 HTTP 标头中指定规范链接

如果您能配置自己的服务器，则可使用 `rel="canonical"` [HTTP 标头](#)为 HTML 文档和其他文件（例如 PDF）指明规范网址。假设您的网站通过下面这些不同的网址提供相同的 **PDF**（例如出于跟踪目的）：

<https://www.example.com/downloads/white-paper.pdf>

<https://www.example.com/downloads/partner-1/white-paper.pdf>

<https://www.example.com/downloads/partner-2/white-paper.pdf>

<https://www.example.com/downloads/partner-3/white-paper.pdf>

`https://www.example.com/downloads/partner-0/white-paper.pdf`

在这种情况下，您可以使用 `rel="canonical"` HTTP 标头向 Google 指定该 PDF 文件的规范网址，如下所示：

Link: <`http://www.example.com/downloads/white-paper.pdf`>; `rel="canonical"`

目前，Google 仅在网页搜索中支持这些链接标头元素。

## 优先选择 HTTPS 而不是 HTTP 作为规范网址

Google 会优先选择 HTTPS 网页而不是等效的 HTTP 网页作为规范网址，除非存在如下冲突信号：

- HTTPS 网页的 SSL 证书无效。
- HTTPS 网页包含不安全的关联功能。
- HTTPS 网页由漫游器抓取（HTTP 网页则不然）。
- HTTPS 网页将用户重定向至 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"` 链接。
- HTTPS 网页包含 `noindex robots` 元标记。

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

- 使用可将用户从 HTTP 网页转到 HTTPS 网页的 301 重定向、302 重定向、Javascript 重定向或元重定向。
- 添加从 HTTP 网页指向 HTTPS 网页的 `rel="canonical"` 链接。
- 实施 [HSTS](#) [🔗](#)。

为防止 Google 误将 HTTP 网页视为规范网页，您应避免以下几种做法：

- 使用有误的 SSL 证书和 HTTPS-HTTP 重定向。这会非常容易导致我们优先选择 HTTP，即使实施 HSTS 也无法替换这种强烈的偏好设置。
- 在站点地图或 [hreflang 条目](#) [🔗](#) 中包含 HTTP 网页（而不是 HTTPS 版本）。
- 为错误的主机版本实施 SSL/TLS 证书：例如，在 `example.com` 上为 `www.example.com` 提供证书。该证书必须与完整的网站网址相符，或者必须是可用于某个网域上的多个子网域的通配证书。



如果您使用 `robots.txt` 文件屏蔽某个资源，请务必同时屏蔽该资源的 HTTP 版本和 HTTPS 版本。

是

否

## 其他内容注意事项

Google 可以识别的元标记

保持简单的网址结构

使用规范网址

对特定链接使用rel="nofollow"

显示分页内容

将网站标记为面向儿童的内容

