Functional Specification

目录
按请求类型划分的逻辑
构建 %%protocol%% 变量
构建 %%hostname%% 变量
构建 %%path%% 变量
包含上级目录
祖先类型偏好设定
处理多个祖先页面
构建 %%date%% 变量
其他注意事项
当未使用 “美观固定链接 ”时
尾部斜杠
分页
路径净化
复合类型
无索引页面
错误场景
用户指定的规范 URL

任何有效的、可索引的页面（即返回 200 HTTP 状态码且 不包含 noindex指令的请求）都应在文档的 <head> 部分包含一个规范 URL 标签。

有效的规范 URL 标签采用以下格式： <link rel="canonical" href="{{URL}}" />

以下部分将说明如何构建标签中 {{URL}} 组件的值。

按请求类型划分的逻辑
下表描述了我们会考虑的所有请求类型以及每种情况下规范 URL
值的结构。请注意，这并未涵盖所有可能的情况或边缘情况，因为我们只期望本规范通常在内容管理系统或网
站的约束范围内应用。

另请注意，全文使用的 “页面 ”一词指代响应请求 URL 而提供的任何有效、可索引的结果。这包括一些通常被视为 “非页面 ”的结果，例如 RSS 订阅源。

请求类型	备注	规范 URL 结构	示例
首页	假设首页位于网站根目录。否则，请参考单页。	`%%protocol%%//%%hostname%%/`	`https://www.example.com/`
单页	例如 “关于我们 ”页面或博客文章。	`%%protocol%%//%%hostname%%/%path%`	`https://www.example.com/about-us`
分类索引页	任何基于分类法或类似方式列出选定文章或页面的模板。例如，分类页面、标签页面，或列出特定作者所有文章的页面。	`%%protocol%%//%%hostname%%/%path%`	`https://www.example.com/widgets`, `https://www.example.com/author/john-smith`, `https://www.example.com/color/blue`, `https://www.example.com/category/food`, `https://www.example.com/tag/size/small`
日期索引页	任何按日期筛选列出选定文章或页面的模板。例如，在特定年份发布的所有文章，或在特定日期 /某天添加的所有页面。	`%%protocol%%//%%hostname%%/%date%`	`https://www.example.com/2006`, `https://www.example.com/2012-08-22`, `https://www.example.com/2019/01/30`, `https://www.example.com/date/01-20-2018`, `https://www.example.com/time/1562149536`
“数据 ”页面	由系统生成的 RSS/Atom/XML 视图。例如，文章或分类的评论订阅源，或 XML 站点地图。	`%%protocol%%//%%hostname%%/%path%`	`https://www.example.com/category/post/feed`, `https://www.example.com/sitemap.xml`, `https://www.example.com/blog/feed/atom`

构建 %%protocol%% 变量

如果应用程序知晓网站的首选协议（例如，开发者或用户已为站点地址定义了包含 `http` 或 `https` 的选项），则应获取协议值，并在变量后附加 `:` 字符（以生成如 `https:` 的结果）。

如果应用程序无法确定正在使用的协议，则应在固定链接中省略此参数（从而生成协议无关的 URL）。

构建 %%hostname%% 变量

如果应用程序知晓其运行网站的首选地址（主机名）（例如，开发者或用户已定义网站地址选项，如 www.example.com），则应返回该值。 若没有首选域名的证据，则应返回请求主机名。

构建 %%path%% 变量
决定 %%path%% 变量包含哪些组件的逻辑可能因内容类型、网站、平台和用户偏好 /配置而异。

此外，资源通常可能通过多种不同的路径查询和结构被请求，且几乎没有证据表明哪种方式最 “正确 ” 。即使使用看似可靠的平台方法，例如 WordPress 的 get_permalink()
函数（以及分类索引等的类似方法），返回的 URL 是否应作为规范 URL 值使用，仍可能存在一些模糊性。

鉴于这一挑战，以下章节将探讨基于多种场景返回最优 %%path%% 组件所需的具体逻辑。

包含上级目录

某些页面的 URL 可能包含父级目录。如果这些目录的包含对于返回目标页面是必需的，那么 %%path%% 变量也应包含这些目录。例如：

新 WordPress 安装中的分类索引可能具有路径 category/blue-widgets。在这种情况下，
%%path%% 也应为 category/blue-widgets。
内容类型（如 “artists”）的索引可能产生 URL 模式，例如 artists/picasso。在这种情况下，
%%path%% 也应为 artists/picasso。
关于乐队 Pink Floyd 1970 年代摇滚专辑《 The Dark Side of the Moon
》的单个页面可能嵌套在多层级分类结构中，因此 URL 为 category/music/pink-floyd/dark-side-
of-the-moon 。在这种情况下， %%path%% 也应为 category/music/pink-floyd/dark-side-of-the-
moon 。

注意 ：Yoast SEO 插件包含移除 WordPress 路径中 category/ “基础 ”组件的功能。如果启用此功能，
category/ 组件也应从 %%path%% 中移除。

祖先类型偏好设定

在以下各种情况中，均假设用户已明确指定应使用哪些祖先元素来构建 URL。例如，在 WordPress
中，用户可以定义 “固定链接结构 ”，并决定所有文章链接前都应包含其 标签 （而非 分类 ），或统一添加
“widgets”字样。

若已设定此类偏好，则规范的 %%path%% 应始终遵循此设定。

若某个页面可通过多种路径访问（例如 category/widgets/example-post 与 tag/cats/example-post
均指向同一页面）且未设定祖先类型偏好，则 %%path%%
应在条件允许时按以下优先级顺序选择规范祖先（若无精确匹配则采用适配 /可类比概念）：

分类 ；例如 category/blue-widgets/page
日期 ；例如 2010/12/31/page
作者 ；例如 author/john-smith/page
标签 ；例如 tag/color/blue/page
其他分类法 ；例如 cats/siamese/page
资源的 “原始 ”查询 ；例如?p=123

处理多个祖先页面

在上述每种场景中，请求的页面都可能存在 多个
有效祖先页面（例如，一篇博文属于多个分类，或某个分类本身又属于多个上级分类）。

如果用户已声明偏好某个特定祖先页面（例如 “主要分类 ”），则应在 %%path%% 中使用该祖先页面。

当没有明确偏好时，默认行为是按字母顺序使用首个有效选项。例如：

一篇同时属于 “cats”和 “dogs”分类的博文（当 分类 是首选祖先结构时），其 %%path%% 应为 category/cats/example-post。
使用自定义分类法 “location”标记为 “Spain”和“Europe”的博文（当 location
是首选祖先结构，且假设 “location”的路径根目录为 location/ 时），其 %%path%% 应为 location/europe/example-post。
与 “milk”相关的页面分类法索引（其父级分类法同时包含 “foods”和“liquids”
，且该分类法的路径根目录为 things/ 时），其 %%path%% 应为 things/foods/milk。

构建 %%date%% 变量
日期结构用于日期索引（例如 “2016年发布的所有文章 ”），以及当用户希望所有文章 /
页面都带有日期根目录时。

在这两种情况下，我们假设用户已定义首选日期结构 /格式，该格式可以是任何形式（例如 Y/m/d、 U、 d-M-y 、 date/d-m-Y 等）。如果是这样，则 %%path%% 应以前述格式作为前缀（例如 date/22-05-2018/post-name ）。

若未发现首选格式（且页面可能通过多种格式访问，例如原始时间戳和 “美观 ”结构）， %%path%%
应在有效时按以下优先级顺序选择规范日期格式（若无精确匹配则采用 /使用可比概念）：

Y/m/d，例如 2018/05/22
Y-m-d，例如 2018-05-22
Ymd或 ymd，例如 180522
U，例如 1526947200

其他注意事项

当未使用 “美观固定链接 ”时

某些系统配置（或偏好设置）会导致有效的 URL 格式和请求仅限于 “原始 ”对象查询。例如：

?post_id=6，而非 category/example-post
?pagename=about，而非 about
?category=cats&page=3，而非 category/cats/page/3

在此情况下，规范的 %%path%% 组件应匹配上述原始查询 ID（但应省略任何不改变查询结果的参数）。

请注意，如果 “美观固定链接 ”可用，这些 “原始 ”查询应执行 301 重定向到其对应的规范链接。

尾部斜杠

本文档中的所有示例均假设请求不包含或不强制要求尾部斜杠。实际上，这因系统和用户偏好而异。在大多数
情况下，有效页面（除 “数据 ”页面外） 确实 包含尾部斜杠（并通过 301 重定向强制要求）。

因此，所有 %%path%% 变量应根据以下场景有条件地包含或排除尾部斜杠：

所有请求均强制要求尾部斜杠 ，因此 %%path%% 应附加尾部斜杠。
页面 可以 附加尾部斜杠，但未强制要求 （通过 301 重定向）。在这种情况下， %%path%% 应
附加尾部斜杠。
添加尾部斜杠时页面无法解析 。在这种情况下， %%path%% 不应 附加尾部斜杠。
以尾部斜杠结尾的请求代表不同的资源 （例如， example/page 与 example/folder/
）。在这种情况下， %%path%% 仅当尾部斜杠是资源有效请求的一部分时才输出（例如，在
example/folder/ 上）。

此外，位于域根目录的主页应始终包含尾部斜杠。

分页

某些请求可能接受额外参数以返回分页状态。例如， category/page/3（或?category=cats&page=3）。当返回有效的分页响应时，分页组件应附加到 %%path%%
变量。在添加分页组件不会产生分页结果的情况下（例如，请求超出范围），则不应附加。

路径净化

路径值在输出前应始终经过以下处理：

强制转换为小写
编码非 ASCII UTF-8 字符
移除连续重复的斜杠
删除本文档未涉及的查询、排序、筛选或任意参数（例如 category/page/2/?a=b 应净化为 category/page/2）
移除任何冗余后缀（例如当 category/post/randomstring 返回 category/post
所代表的页面时，应移除 randomstring 部分）

复合类型

某些请求类型可能需要组合多个 %%path%% 组件才能返回有效结果。例如，按日期筛选的分类（
category/cats/2012 或 category/cats/2012/06/02），或基于日期的作者索引的分页状态（
writers/george/2006/page/2）。这些组件应添加到 %%path%% 中。

如果这些组件可以按多种顺序应用（并返回相同且有效的响应），则 %%path%%
应按以下顺序构建（当组件存在时）： %%taxonomy%%/%%date%%/%%pagination%%。

无索引页面

具有 noindex指令的元机器人标签（或 x-robots-tag HTTP 标头）的页面应完全省略规范 URL
标签。这包括模板逻辑指定 noindex属性的情况（如搜索结果页），以及用户将特定页面设置为 noindex 的情况。

错误场景

导致错误的请求（所有产生 4xx 或 5xx 范围 HTTP 响应的场景）应完全省略规范 URL 标签。

用户指定的规范 URL
当用户为页面手动指定了规范 URL
时，此设置应覆盖所有其他逻辑。我们应原样输出该值，绕过上述定义的清理和处理规则。

此规则同样适用于用户指定的跨域规范 URL。

任何冲突、错误或中断的风险都应在发布工作流程中通过内联警告或验证错误来解决。

版本号 #2
由 壳壳虫 创建于 2026-03-26 17:08:40 UTC
由 壳壳虫 更新于 2026-03-26 17:08:40 UTC