

Network Working Group  
Internet-Draft  
Intended status: Informational  
Expires: April 16, 2012

M. Ohye  
J. Kupke  
October 14, 2011

The Canonical Link Relation  
draft-ohye-canonical-link-relation-04

## Abstract

[RFC5988] specified a way to define relationships between links on the web. This document describes a new type of such relationship, "canonical," which designates the preferred URI from a set of identical or vastly similar ones.

## Editorial Note (To be removed by RFC Editor)

Distribution of this document is unlimited. Comments should be sent to the IETF Apps-Discuss mailing list (see <https://www.ietf.org/mailman/listinfo/apps-discuss>).

## Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 16, 2012.

## Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents

Internet-Draft

The Canonical Link Relation

October 2011

carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## [1.](#) Introduction

The canonical link relation specifies the preferred URI from a set of URIs that return identical or vastly similar content, making it possible for references to the context URI to be updated to reference the target URI.

The most common application of the canonical link relation includes specifying the preferred version of a URI from duplicate content pages created with the addition of parameters (e.g. session IDs, tracking IDs, category, or sort information).

## [2.](#) Notational Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [[RFC2119](#)].

## [3.](#) The Canonical Link Relation

The canonical (target) URI MUST identify content that duplicates, is extremely similar, or is a superset of the content at the context (referring) URI. Authors who declare the canonical link relation ought to anticipate that applications such as search engines can:

- o Index content only from the canonical target (i.e. content from the context URIs will be likely disregarded as duplicative)
- o Consolidate URI properties, such as link popularity, to the canonical
- o Display the canonical target as the representative URI

A resource SHOULD NOT specify more than one canonical link relation.

The target/canonical URI MAY:

- o Specify a relative URI (see [\[RFC3986\] Section 4.2](#))
- o Be self-referential (context URI identical to target URI)

- o Exist on a different hostname or domain
- o Have different scheme names, such as "http" to "https," or "gopher" to "ftp"
- o Be a superset of the content of the context URI
  - \* For example, "page1" of a multi-page article may specify the canonical target as the "view-all" URI because "view-all" is a superset of page1's content. However, "page2" SHOULD NOT designate "page1" as the canonical because the content of page1 is not inclusive of page2.
- o Be the source URI of a temporary redirect. For HTTP, this refers to status codes 302, 303, or 307 (Sections [10.3.3](#), [10.3.4](#), and [10.3.8](#), respectively, of [\[RFC2616\]](#)).

The target/canonical URI SHOULD NOT designate:

- o The source URI of a permanent redirect (for HTTP, this refers to 300 and 301 response codes, defined in Sections [10.3.1](#) and [10.3.2](#) of [\[RFC2616\]](#))
- o A URI that also specifies a canonical link relation to a URI other than itself
- o A URI that returns an error code, such as 4xx response in HTTP ([Section 10.4 of \[RFC2616\]](#))
- o The first page of a multi-page article or multi-page listing of items (since the first page is not a duplicate or a superset of the context URI). For example, page2 and page3 of an article SHOULD NOT specify page1 as the canonical.

#### [4.](#) Examples

The following example illustrates:

- o Three URIs that serve nearly the exact same content
- o One URI which is the canonical or "preferred version"
- o Two URIs with additional query parameters, making them the non-preferred version of the content (duplicates). The canonical link relation is therefore specified on these duplicates.

If the preferred version of a URI and its content exists at:  
`http://www.example.com/page.php?item=purse`

Then duplicate content URIs such as:

`http://www.example.com/page.php?item=purse&category=bags`

`http://www.example.com/page.php?item=purse&category=bags&sid=1234`

may designate the canonical link relation in HTML as specified in  
[\[REC-html401-19991224\]](#):

```
<link rel="canonical"
      href="http://www.example.com/page.php?item=purse">
```

or as a relative URI:

```
<link rel="canonical" href="page.php?item=purse">
```

or alternatively, in the HTTP header field as specified in [Section 5 of \[RFC5988\]](#):

```
Link: <http://www.example.com/page.php?item=purse>; rel="canonical"
```

This signals to automated programs, such as search engines, that these are duplicates of the canonical URI:

`http://www.example.com/page.php?item=purse.`

Automated programs may then select the canonical value as the display URI (such as in search results), and additional URI properties such as indexing and ranking signals, can be transferred as well.

## [5.](#) Recommendations

Before adding the canonical link relation, verification of the following is recommended:

1. The content of the context URI is identical with, similar to, or a subset of the content of the canonical.
2. For HTTP, Permanent HTTP redirects ([Section 10.3.2 of \[RFC2616\]](#)), the traditional strong indicator that a URI's content has been permanently moved, could not be implemented in place of the canonical link relation.
3. In the case where the canonical target is a superset of content from the context URI (e.g. page1 or page2 to view-all), that the user experience is strongly taken into consideration, both in regard to possible increased load time and potential complexity in navigation.

## [6.](#) IANA Considerations

IANA is asked to register the Canonical Link Relation below as per [\[RFC5988\]](#).

Ohye & Kupke

Expires April 16, 2012

[Page 4]

---

Internet-Draft

The Canonical Link Relation

October 2011

Relation Name:

CANONICAL

Description:

Designates the preferred version of a resource (the URI and its contents).

Reference:

This specification.

Notes:

None.

Application Data:

None.

## [7.](#) Security Considerations

When a site is compromised, the canonical link relation can be implemented with malicious intent to designate the attacker's URI as the preferred version of the content. While this technique is largely unnoticeable to humans, automated programs may cluster the compromised resource as duplicative of the attacker's designated canonical, transferring properties such as link popularity away from the resource to the attacker's URI.

## 8. Internationalisation Considerations

In designating a canonical URI, please see [[RFC3986](#)] for information on URI encoding.

## 9. Normative References

[REC-html401-19991224] Le Hors, A., Raggett, D., and I. Jacobs, "HTML 4.01 Specification", W3C Recommendation REC-html401-19991224, December 1999, <<http://www.w3.org/TR/1999/REC-html401-19991224>>.

Latest version available at  
<<http://www.w3.org/TR/html401>>.

[RFC2119] Bradner, S., "Key words for use in RFCs to

Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.

[RFC2616] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1", [RFC 2616](#), June 1999.

[RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, [RFC 3986](#), January 2005.

[RFC5988] Nottingham, M., "Web Linking", [RFC 5988](#), October 2010.

## Appendix A. Implementations

Automated programs that implement functionality with regard for the canonical link relation include:

- o Google, canonical link relation HTML and HTTP header support, within the same domain and across domains:
  - \* <<http://googlewebmastercentral.blogspot.com/2009/02/specify-your-canonical.html>>
  - \* <<http://googlewebmastercentral.blogspot.com/2011/06/supporting-relcanonical-http-headers.html>>
  - \* <<http://googlewebmastercentral.blogspot.com/2009/12/handling-legitimate-cross-domain.html>>
- o Yahoo, canonical link relation HTML support within the same domain:
  - \* <<http://www.ysearchblog.com/2009/02/12/fighting-duplication-adding-more-arrows-to-your-quiver/>>
- o Bing, canonical link relation HTML support within the same domain:
  - \* <[http://www.bing.com/community/site\\_blogs/b/webmaster/archive/2009/02/12/partnering-to-help-solve-duplicate-content-issues.aspx](http://www.bing.com/community/site_blogs/b/webmaster/archive/2009/02/12/partnering-to-help-solve-duplicate-content-issues.aspx)>

### Authors' Addresses

Maile Ohye

EMail: [maileohye@gmail.com](mailto:maileohye@gmail.com)

URI: <http://maileohye.com/>

Joachim Kupke

EMail: joachim@kupke.za.net