

Internet Engineering Task Force	S. Johnston	
Internet-Draft	Google	
Intended status: Experimental	July 5, 2010	
Expires: January 6, 2011		

[TOC](#)

Web Categories

draft-johnston-http-category-header-01

Abstract

This document specifies the Category header-field for HyperText Transfer Protocol (HTTP), which enables the sending of taxonomy information in HTTP headers.

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 January 6, 2011.

Copyright Notice

Copyright (c) 2010 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 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.

Table of Contents

- [1.](#) Introduction
 - [1.1.](#) Requirements Language
- [2.](#) Categories
- [3.](#) The Category Header Field
 - [3.1.](#) Examples
- [4.](#) IANA Considerations
 - [4.1.](#) Category Header Registration
- [5.](#) Security Considerations
- [6.](#) Internationalisation Considerations
- [7.](#) References
 - [7.1.](#) Normative References
 - [7.2.](#) Informative References
- [Appendix A.](#) Notes on use with HTML
- [Appendix B.](#) Notes on use with Atom
- [Appendix C.](#) Acknowledgements
- [Appendix D.](#) Change Log (to be removed by RFC Editor before publication)
 - [D.1.](#) draft-johnston-http-category-header-00
 - [D.2.](#) draft-johnston-http-category-header-01
- [Appendix E.](#) Outstanding Issues
- [§](#) Author's Address

1. Introduction

[TOC](#)

A means of indicating categories for resources on the web has been defined by Atom [\[RFC4287\] \(Nottingham, M., Ed. and R. Sayre, Ed., "The Atom Syndication Format," December 2005.\)](#). This document defines a framework for exposing category information in the same format via HTTP headers.

The atom:category element conveys information about a category associated with an entry or feed. A given atom:feed or atom:entry element MAY have zero or more categories which MUST have a "term" attribute (a string that identifies the category to which the entry or feed belongs) and MAY also have a scheme attribute (an IRI that identifies a categorization scheme) and/or a label attribute (a human-readable label for display in end-user applications).

Similarly a web resource may be associated with zero or more categories as indicated in the Category header-field(s). These categories may be divided into separate vocabularies or "schemes" and/or accompanied with human-friendly labels.

[Feedback is welcome on the ietf-http-wg@w3.org mailing list, although this is NOT a work item of the HTTPBIS WG.]

1.1. Requirements Language

[TOC](#)

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 BCP 14, [\[RFC2119\] \(Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," March 1997.\)](#), as scoped to those conformance targets.

This document uses the Augmented Backus-Naur Form (ABNF) notation of [\[RFC2616\] \(Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1," June 1999.\)](#), and explicitly includes the following rules from it: quoted-string, token. Additionally, the following rules are included from [\[RFC3986\] \(Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier \(URI\): Generic Syntax," January 2005.\)](#): URI.

2. Categories

[TOC](#)

In this specification, a category is a grouping of resources by 'term', from a vocabulary ('scheme') identified by an IRI [\[RFC3987\] \(Duerst, M. and M. Suignard, "Internationalized Resource Identifiers \(IRIs\)," January 2005.\)](#). It is comprised of:

- *A "term" which is a string that identifies the category to which the resource belongs.

- *A "scheme" which is an IRI that identifies a categorization scheme (optional).

- *An "label" which is a human-readable label for display in end-user applications (optional).

A category can be viewed as a statement of the form "resource is from the {term} category of {scheme}, to be displayed as {label}", for example "'Löwchen' is from the 'dog' category of 'animals', to be displayed as 'Canine'".

3. The Category Header Field

[TOC](#)

The Category entity-header provides a means for serialising one or more categories in HTTP headers. It is semantically equivalent to the atom:category element in Atom [\[RFC4287\] \(Nottingham, M., Ed. and R. Sayre, Ed., "The Atom Syndication Format," December 2005.\)](#).

```

Category          = "Category" ":" #category-value
category-value    = term *( ";" category-param )
category-param    = ( ( "scheme" "=" <"> scheme <"> )
                    | ( "label" "=" quoted-string )
                    | ( "label*" "=" enc2231-string )
                    | ( category-extension ) )
category-extension = token [ "=" ( token | quoted-string ) ]
enc2231-string    = <extended-value, see [RFC2231], Section 7>
term              = token
scheme            = URI

```

Each category-value conveys exactly one category but there may be multiple category-values for each header-field and/or multiple header-fields per [\[RFC2616\] \(Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1," June 1999.\)](#).

Note that schemes are REQUIRED to be absolute URLs in Category headers, and MUST be quoted if they contain a semicolon (";") or comma (",") as these characters are used to separate category-params and category-values respectively.

The "label" parameter is used to label the category such that it can be used as a human-readable identifier (e.g. a menu entry). Alternately, the "label*" parameter MAY be used encode this label in a different character set, and/or contain language information as per [\[RFC2231\] \(Freed, N. and K. Moore, "MIME Parameter Value and Encoded Word Extensions: Character Sets, Languages, and Continuations," November 1997.\)](#). When using the enc2231-string syntax, producers MUST NOT use a charset value other than 'ISO-8859-1' or 'UTF-8'.

3.1. Examples

[TOC](#)

NOTE: Non-ASCII characters used in prose for examples are encoded using the format "Backslash-U with Delimiters", defined in Section 5.1 of [\[RFC5137\] \(Klensin, J., "ASCII Escaping of Unicode Characters," February 2008.\)](#).

For example:

```
Category: dog
```

indicates that the resource is in the "dog" category.

```
Category: dog; label="Canine"; scheme="http://purl.org/net/animals"
```

indicates that the resource is in the "dog" category, from the "http://purl.org/net/animals" scheme, and should be displayed as "Canine".

The example below shows an instance of the Category header encoding multiple categories, and also the use of [\[RFC2231\] \(Freed, N. and K. Moore, "MIME Parameter Value and Encoded Word Extensions: Character Sets, Languages, and Continuations," November 1997.\)](#) encoding to represent both non-ASCII characters and language information.

```
Category: dog; label="Canine"; scheme="http://purl.org/net/animals",  
        lowchen; label*=UTF-8'de'L%c3%b6wchen";  
        scheme="http://purl.org/net/animals/dogs"
```

Here, the second category has a label encoded in UTF-8, uses the German language ("de"), and contains the Unicode code point '\u'00F6' ("LATIN SMALL LETTER O WITH DIAERESIS").

4. IANA Considerations

[TOC](#)

4.1. Category Header Registration

[TOC](#)

This specification adds an entry for "Category" in HTTP to the Message Header Registry [\[RFC3864\] \(Klyne, G., Nottingham, M., and J. Mogul, "Registration Procedures for Message Header Fields," September 2004.\)](#) referring to this document:

```
Header Field Name: Category  
Protocol: http  
Status: standard  
Author/change controller:  
    IETF (iesg@ietf.org)  
    Internet Engineering Task Force  
Specification document(s):  
    [ this document ]
```

5. Security Considerations

[TOC](#)

The content of the Category header-field is not secure, private or integrity-guaranteed, and due caution should be exercised when using it.

6. Internationalisation Considerations

[TOC](#)

Category header-fields may be localised depending on the Accept-Language header-field, as defined in section 14.4 of [\[RFC2616\]](#) (Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1," June 1999.).

Scheme IRIs in atom:category elements may need to be converted to URIs in order to express them in serialisations that do not support IRIs, as defined in section 3.1 of [\[RFC3987\]](#) (Duerst, M. and M. Suignard, "Internationalized Resource Identifiers (IRIs)," January 2005.). This includes the Category header-field.

7. References

[TOC](#)

7.1. Normative References

[TOC](#)

[RFC2119]	Bradner, S. , " Key words for use in RFCs to Indicate Requirement Levels ," BCP 14, RFC 2119, March 1997 (TXT , HTML , XML).
[RFC2231]	Freed, N. and K. Moore , " MIME Parameter Value and Encoded Word Extensions: Character Sets, Languages, and Continuations ," RFC 2231, November 1997 (TXT , HTML , XML).
[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 (TXT , PS , PDF , HTML , XML).
[RFC3864]	Klyne, G., Nottingham, M., and J. Mogul, " Registration Procedures for Message Header Fields ," BCP 90, RFC 3864, September 2004 (TXT).
[RFC3986]	Berners-Lee, T. , Fielding, R. , and L. Masinter , " Uniform Resource Identifier (URI): Generic Syntax ," STD 66, RFC 3986, January 2005 (TXT , HTML , XML).
[RFC3987]	Duerst, M. and M. Suignard, " Internationalized Resource Identifiers (IRIs) ," RFC 3987, January 2005 (TXT).
[RFC4287]	Nottingham, M., Ed. and R. Sayre, Ed. , " The Atom Syndication Format ," RFC 4287, December 2005 (TXT , HTML , XML).
[RFC5137]	Klensin, J., " ASCII Escaping of Unicode Characters ," BCP 137, RFC 5137, February 2008 (TXT).

7.2. Informative References

[TOC](#)

[I-D.nottingham-http-link-header]	Nottingham, M., " Web Linking ," draft-nottingham-http-link-header-10 (work in progress), May 2010 (TXT).
[RFC2068]	Fielding, R. , Gettys, J. , Mogul, J. , Nielsen, H. , and T. Berners-Lee , " Hypertext Transfer Protocol -- HTTP/1.1 ," RFC 2068, January 1997 (TXT).
[W3C.REC-html401-19991224]	Hors, A., Jacobs, I., and D. Raggett, " HTML 4.01 Specification ," World Wide Web Consortium Recommendation REC-html401-19991224, December 1999 (HTML).
[W3C.WD-html5-20100624]	Hickson, I., " HTML5 ," World Wide Web Consortium WD WD-html5-20100624, June 2010 (HTML).
[rel-tag-microformat]	Çelik, T., Marks, K., and D. Powazek, " rel="tag" Microformat ."

Appendix A. Notes on use with HTML

[TOC](#)

In the absence of a dedicated category element in HTML 4 [[W3C.REC-html401-19991224](#)] (Hors, A., Jacobs, I., and D. Raggett, "[HTML 4.01 Specification](#)," December 1999.) and HTML 5 [[W3C.WD-html5-20100624](#)] (Hickson, I., "[HTML5](#)," June 2010.), category information (including user supplied folksonomy classifications) MAY be exposed using HTML A and/or LINK elements by concatenating the scheme and term:

```
category-link = scheme term
scheme         = URI
term           = token
```

These category-links MAY form a resolveable "tag space" in which case they SHOULD use the "tag" relation-type per [[rel-tag-microformat](#)] ([Çelik, T., Marks, K., and D. Powazek, "rel="tag" Microformat," .](#)). Alternatively META elements MAY be used:

*where the "name" attribute is "keywords" and the "content" attribute is a comma-separated list of term(s)

*where the "http-equiv" attribute is "Category" and the "content" attribute is a comma-separated list of category-value(s)

Appendix B. Notes on use with Atom

[TOC](#)

Where the cardinality is known to be one (for example, when retrieving an individual resource) it MAY be preferable to render the resource natively over HTTP without Atom structures. In this case the contents of the atom:content element SHOULD be returned as the HTTP entity-body and metadata including the type attribute and atom:category element(s) via HTTP header-field(s).

This approach SHOULD NOT be used where the cardinality is guaranteed to be one (for example, search results which MAY return one result).

Appendix C. Acknowledgements

[TOC](#)

The author would like to thank Mark Nottingham for his work on Web Linking [[I-D.nottingham-http-link-header](#)] ([Nottingham, M., "Web Linking," May 2010.](#)) (on which this document was based), the authors of [[RFC2068](#)] ([Fielding, R., Gettys, J., Mogul, J., Nielsen, H., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1," January 1997.](#)) for specification of the Link: header-field on which this is based and all those who commented upon, encouraged and gave feedback to this draft.

Appendix D. Change Log (to be removed by RFC Editor before publication)

[TOC](#)

D.1. draft-johnston-http-category-header-00

[TOC](#)

Initial draft based on draft-nottingham-http-link-header-05

D.2. draft-johnston-http-category-header-01

[TOC](#)

Updated references, affiliation and tickled for IETF-78.

[TOC](#)

Appendix E. Outstanding Issues

[[to be removed by the RFC editor should document proceed to publication as an RFC.]]

The following issues are outstanding and should be addressed:

1. Is extensibility of Category headers necessary as is the case for Link: headers? If so, what are the use cases?
2. Is supporting multi-lingual representations of the same category(s) necessary? If so, what are the risks of doing so?
3. Is a mechanism for maintaining Category header-fields required? If so, should it use the headers themselves or some other mechanism?
4. Does this proposal conflict with others in the same space? If so, is it an improvement on what exists?

Author's Address

[TOC](#)

	Sam Johnston
	Google
	Brandschenkestrasse, 110
	Zürich 8002
	CH
Email:	sj@google.com