

Content Language Headers

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2002). All Rights Reserved.

Abstract

This document defines a "Content-language:" header, for use in cases where one desires to indicate the language of something that has RFC 822-like headers, like MIME body parts or Web documents, and an "Accept-Language:" header for use in cases where one wishes to indicate one's preferences with regard to language.

1. Introduction

There are a number of languages presently or previously used by human beings in this world.

A great number of these people would prefer to have information presented in a language which they understand.

In some contexts, it is possible to have information available in more than one language, or it might be possible to provide tools (such as dictionaries) to assist in the understanding of a language.

In other cases, it may be desirable to use a computer program to convert information from one format (such as plaintext) into another (such as computer-synthesized speech, or Braille, or high-quality print renderings).

A prerequisite for any such function is a means of labelling the information content with an identifier for the language that is used in this information content, such as is defined by [TAGS]. This document specifies a protocol element for use with protocols that use RFC 822-like headers for carrying language tags as defined in [TAGS].

The keywords "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119].

2. The Content-language header

The "Content-Language" header is intended for use in the case where one desires to indicate the language(s) of something that has RFC 822-like headers, such as MIME body parts or Web documents.

The RFC 822 EBNF of the Content-Language header is:

```
Content-Language = "Content-Language" ":" 1#Language-tag
```

In the more strict RFC 2234 ABNF:

```
Content-Language = "Content-Language" ":" [CFWS] Language-List
Language-List = Language-Tag [CFWS]
                  *("," [CFWS] Language-Tag [CFWS])
```

The Content-Language header may list several languages in a comma-separated list.

The CFWS construct is intended to function like the whitespace convention in RFC 822, which means also that one can place parenthesized comments anywhere in the language sequence, or use continuation lines. A formal definition is given in RFC 2822 [RFC2822].

In keeping with the tradition of RFC 2822, a more liberal "obsolete" grammar is also given:

```
obs-content-language = "Content-Language" *WSP ":"
                      [CFWS] Language-List
```

Like RFC 2822, this specification says that conforming implementations MUST accept the obs-content-language syntax, but MUST NOT generate it; all generated headers MUST conform to the Content-Language syntax.

2.1 Examples of Content-language values

Voice recording from Liverpool downtown

```
Content-type: audio/basic
Content-Language: en-scouse
```

Document in Mingo, an American Indian language which does not have an ISO 639 code:

```
Content-type: text/plain
Content-Language: i-mingo
```

A English-French dictionary

```
Content-type: application/dictionary
Content-Language: en, fr (This is a dictionary)
```

An official European Commission document (in a few of its official languages):

```
Content-type: multipart/alternative
Content-Language: da, de, el, en, fr, it
```

An excerpt from Star Trek

```
Content-type: video/mpeg
Content-Language: i-klingon
```

3. The Accept-Language header

The "Accept-Language" header is intended for use in cases where a user or a process desires to identify the preferred language(s) when RFC 822-like headers, such as MIME body parts or Web documents, are used.

The RFC 822 EBNF of the Accept-Language header is:

```
Accept-Language = "Accept-Language" ":"
                  1#( language-range [ ";" "q" "=" qvalue ] )
```

A slightly more restrictive RFC 2234 ABNF definition is:

```
Accept-Language = "Accept-Language:" [CFWS] language-q
                  *( "," [CFWS] language-q )
language-q = language-range [ ";" [CFWS] "q=" qvalue ] [CFWS]
qvalue      = ( "0" [ "." 0*3DIGIT ] )
              / ( "1" [ "." 0*3("0") ] )
```

A more liberal RFC 2234 ABNF definition is:

```
Obs-accept-language = "Accept-Language" *WSP ":" [CFWS]
  obs-language-q *( "," [CFWS] obs-language-q ) [CFWS]
obs-language-q = language-range
  [ [CFWS] ";" [CFWS] "q" [CFWS] "=" qvalue ]
```

Like RFC 2822, this specification says that conforming implementations MUST accept the obs-accept-language syntax, but MUST NOT generate it; all generated messages MUST conform to the Accept-Language syntax.

The syntax and semantics of language-range is defined in [TAGS]. The Accept-Language header may list several language-ranges in a comma-separated list, and each may include a quality value Q. If no Q values are given, the language-ranges are given in priority order, with the leftmost language-range being the most preferred language; this is an extension to the HTTP/1.1 rules, but matches current practice.

If Q values are given, refer to HTTP/1.1 [RFC 2616] for the details on how to evaluate it.

4. Security Considerations

The only security issue that has been raised with language tags since the publication of RFC 1766, which stated that "Security issues are believed to be irrelevant to this memo", is a concern with language ranges used in content negotiation - that they may be used to infer the nationality of the sender, and thus identify potential targets for surveillance.

This is a special case of the general problem that anything you send is visible to the receiving party; it is useful to be aware that such concerns can exist in some cases.

The exact magnitude of the threat, and any possible countermeasures, is left to each application protocol.

5. Character set considerations

This document adds no new considerations beyond what is mentioned in [TAGS].

6. Acknowledgements

This document has benefited from many rounds of review and comments in various fora of the IETF and the Internet working groups.

Any list of contributors is bound to be incomplete; please regard the following as only a selection from the group of people who have contributed to make this document what it is today.

In alphabetical order:

Tim Berners-Lee, Nathaniel Borenstein, Sean M. Burke, John Clews, Jim Conklin, John Cowan, Dave Crocker, Martin Duerst, Michael Everson, Ned Freed, Tim Goodwin, Dirk-Willem van Gulik, Marion Gunn, Paul Hoffman, Olle Jarnefors, John Klensin, Bruce Lilly, Keith Moore, Chris Newman, Masataka Ohta, Keld Jorn Simonsen, Rhys Weatherley, Misha Wolf, Francois Yergeau and many, many others.

Special thanks must go to Michael Everson, who has served as language tag reviewer for almost the entire period, since the publication of RFC 1766, and has provided a great deal of input to this revision. Bruce Lilly did a special job of reading and commenting on my ABNF definitions.

7. References

- [TAGS] Alvestrand, H., "Tags for the Identification of Languages", BCP 47, RFC 3066

- [ISO 639] ISO 639:1988 (E/F) - Code for the representation of names of languages - The International Organization for Standardization, 1st edition, 1988-04-01 Prepared by ISO/TC 37 - Terminology (principles and coordination). Note that a new version (ISO 639-1:2000) is in preparation at the time of this writing.

- [ISO 639-2] ISO 639-2:1998 - Codes for the representation of names of languages -- Part 2: Alpha-3 code - edition 1, 1998-11-01, 66 pages, prepared by ISO/TC 37/SC 2

- [ISO 3166] ISO 3166:1988 (E/F) - Codes for the representation of names of countries - The International Organization for Standardization, 3rd edition, 1988-08-15.

- [ISO 15924] ISO/DIS 15924 - Codes for the representation of names of scripts (under development by ISO TC46/SC2)

- [RFC 2045] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies", RFC 2045, November 1996.
- [RFC 2046] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types", RFC 2046, November 1996.
- [RFC 2047] Moore, K., "MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text", RFC 2047, November 1996.
- [RFC 2048] Freed, N., Klensin, J. and J. Postel, "Multipurpose Internet Mail Extensions (MIME) Part Four: Registration Procedures", RFC 2048, November 1996.
- [RFC 2049] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Five: Conformance Criteria and Examples", RFC 2049, November 1996.
- [RFC 2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC 2234] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, November 1997.
- [RFC 2616] 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.
- [RFC 2822] Resnick, P., "Internet Message Format", RFC 2822, April 2001.

Appendix A: Changes from RFC 1766

The definition of the language tags has been split, and is now RFC 3066. The differences parameter to multipart/alternative is no longer part of this standard, because no implementations of the function were ever found. Consult RFC 1766 if you need the information.

The ABNF for content-language has been updated to use the RFC 2234 ABNF.

Author's Address

Harald Tveit Alvestrand
Cisco Systems
Weidemanns vei 27
7043 Trondheim
NORWAY

EMail: Harald@Alvestrand.no
Phone: +47 73 50 33 52

Full Copyright Statement

Copyright (C) The Internet Society (2002). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.

