Internet Engineering Task Force (IETF)
Request for Comments: 6657
Updates: 2046
Category: Standards Track
ISSN: 2070-1721

A. Melnikov
Isode Limited
J. Reschke
greenbytes

July 2012

Update to MIME regarding "charset" Parameter Handling

Abstract

This document changes RFC 2046 rules regarding default "charset" parameter values for "text/*" media types to better align with common usage by existing clients and servers.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741[1].

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6657[2].

Copyright Notice

Copyright (c) 2012 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[3]) 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.

[1] https://www.rfc-editor.org/rfc/rfc5741.html#section-2
[2] http://www.rfc-editor.org/info/rfc6657
[3] http://trustee.ietf.org/license-info

Table of Contents

1 Introduction and Overview ... 3

2 Conventions Used in This Document ... 4

3 New Rules for Default "charset" Parameter Values for "text/*" Media Types ... 5

4 Default "charset" Parameter Value for "text/plain" Media Type ... 6

5 Security Considerations ... 7

6 IANA Considerations ... 8

7 References ... 9

7.1 Normative References ... 9

7.2 Informative References ... 9

Appendix A Acknowledgements ... 10

Authors' Addresses ... 11

1. Introduction and Overview

RFC 2046 specified that the default "charset" parameter (i.e., the value used when the parameter is not specified) is "US-ASCII" (Section 4.1.2 of [RFC2046]). RFC 2616 changed the default for use by HTTP (Hypertext Transfer Protocol) to be "ISO-8859-1" (Section 3.7.1 of [RFC2616]). This encoding is not very common for new "text/*" media types and a special rule in the HTTP specification adds confusion about which specification ([RFC2046] or [RFC2616]) is authoritative in regards to the default charset for "text/*" media types.

Many complex text subtypes such as "text/html" [RFC2854] and "text/xml" [RFC3023] have internal (to their format) means of describing the charset. Many existing User Agents ignore the default of "US-ASCII" rule for at least "text/html" and "text/xml".

This document changes RFC 2046 rules regarding default "charset" parameter values for "text/*" media types to better align with common usage by existing clients and servers. It does not change the defaults for any currently registered media type.

2. Conventions Used in This Document

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. New Rules for Default "charset" Parameter Values for "text/*" Media Types

Section 4.1.2 of [RFC2046] says:

> *The default character set, which must be assumed in the absence of a charset parameter, is US-ASCII.*

As explained in the Introduction section, this rule is considered outdated, so this document replaces it with the following set of rules:

Each subtype of the "text" media type that uses the "charset" parameter can define its own default value for the "charset" parameter, including the absence of any default.

In order to improve interoperability with deployed agents, "text/*" media type registrations SHOULD either

a. specify that the "charset" parameter is not used for the defined subtype, because the charset information is transported inside the payload (such as in "text/xml"), or

b. require explicit unconditional inclusion of the "charset" parameter, eliminating the need for a default value.

In accordance with option (a) above, registrations for "text/*" media types that can transport charset information inside the corresponding payloads (such as "text/html" and "text/xml") SHOULD NOT specify the use of a "charset" parameter, nor any default value, in order to avoid conflicting interpretations should the "charset" parameter value and the value specified in the payload disagree.

Thus, new subtypes of the "text" media type SHOULD NOT define a default "charset" value. If there is a strong reason to do so despite this advice, they SHOULD use the "UTF-8" [RFC3629] charset as the default.

Regardless of what approach is chosen, all new "text/*" registrations MUST clearly specify how the charset is determined; relying on the default defined in Section 4.1.2 of [RFC2046] is no longer permitted. However, existing "text/*" registrations that fail to specify how the charset is determined still default to US-ASCII.

Specifications covering the "charset" parameter, and what default value, if any, is used, are subtype-specific, NOT protocol-specific. Protocols that use MIME, therefore, MUST NOT override default charset values for "text/*" media types to be different for their specific protocol. The protocol definitions MUST leave that to the subtype definitions.

4. Default "charset" Parameter Value for "text/plain" Media Type

The default "charset" parameter value for "text/plain" is unchanged from [RFC2046] and remains as "US-ASCII".

5. Security Considerations

Guessing of the "charset" parameter can lead to security issues such as content buffer overflows, denial of services, or bypass of filtering mechanisms. However, this document does not promote guessing, but encourages use of charset information that is specified by the sender.

Conflicting information in-band vs. out-of-band can also lead to similar security problems, and this document recommends the use of charset information that is more likely to be correct (for example, in-band over out-of-band).

6. IANA Considerations

IANA has updated the "text" subregistry of the Media Types registry (<http://www.iana.org/assignments/media-types/text/>) to add the following preamble: "See [RFC6657] for information about 'charset' parameter handling for text media types."

Also, IANA has added this RFC to the list of references at the beginning of the Application for Media Type (<http://www.iana.org/form/media-types>).

7. References

7.1. Normative References

[RFC2046] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types", RFC 2046, November 1996.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, RFC 3629, November 2003.

7.2. Informative References

[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.

[RFC2854] Connolly, D. and L. Masinter, "The 'text/html' Media Type", RFC 2854, June 2000.

[RFC3023] Murata, M., St. Laurent, S., and D. Kohn, "XML Media Types", RFC 3023, January 2001.

Appendix A. Acknowledgements

Many thanks to Ned Freed and John Klensin for comments and ideas that motivated creation of this document, and to Carsten Bormann, Murray S. Kucherawy, Barry Leiba, and Henri Sivonen for feedback and text suggestions.

Authors' Addresses

Alexey Melnikov
Isode Limited
5 Castle Business Village
36 Station Road
Hampton, Middlesex TW12 2BX
UK
Email: Alexey.Melnikov@isode.com

Julian F. Reschke
greenbytes GmbH
Hafenweg 16
Muenster, NW 48155
Germany
Email: julian.reschke@greenbytes.de
URI: http://greenbytes.de/tech/webdav/