Multiple Language Content Type

Abstract

This document defines the 'multipart/multilingual' content type, which is an addition to the Multipurpose Internet Mail Extensions (MIME) standard. This content type makes it possible to send one message that contains multiple language versions of the same information. The translations would be identified by a language tag and selected by the email client based on a user’s language settings.

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 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc8255.

Copyright Notice

Copyright (c) 2017 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 (https://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.
1. Introduction

Since the invention of email and the rapid spread of the Internet, more and more people have been able to communicate in more and more countries and in more and more languages. But during this time of technological evolution, email has remained a single-language communication tool, whether it is English to English, Spanish to Spanish, or Japanese to Japanese.

Also during this time, many corporations have established their offices in multicultural cities and have formed departments and teams that span continents, cultures, and languages. Thus, the need to communicate efficiently with little margin for miscommunication has grown significantly.

This document defines the ‘multipart/multilingual’ content type, which is an addition to the Multipurpose Internet Mail Extensions (MIME) standard specified in [RFC2045], [RFC2046], [RFC2047],
This content type makes it possible to send a single message to a group of people in such a way that all of the recipients can read the email in their preferred language. The methods of translation of the message content are beyond the scope of this document, but the structure of the email itself is defined herein.

This document depends on the identification of language in message parts for non-real-time communication. [HUMAN-LANG] is concerned with a similar problem for real-time communication.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. The Content-Type Header Field

The ‘multipart/multilingual’ Media Subtype allows the sending of a message in a number of different languages with the different language versions embedded in the same message. This Media Subtype helps the receiving email client make sense of the message structure.

The multipart subtype ‘multipart/multilingual’ has similar semantics to ‘multipart/alternative’ (as discussed in RFC 2046 [RFC2046]) in that each of the message parts is an alternative version of the same information. The primary difference between ‘multipart/multilingual’ and ‘multipart/alternative’ is that when using ‘multipart/multilingual’, the message part to select for rendering is chosen based on the values of the Content-Language field and optionally the Content-Translation-Type field instead of the ordering of the parts and the Content-Types.

The syntax for this multipart subtype conforms to the common syntax for subtypes of multipart given in Section 5.1.1. of RFC 2046 [RFC2046]. An example ‘multipart/multilingual’ Content-Type header field would look like this:

Content-Type: multipart/multilingual; boundary=01189998819991197253
3. The Message Parts

A ‘multipart/multilingual’ message will have a number of message parts: exactly one multilingual preface, one or more language message parts, and zero or one language-independent message part. The details of these are described below.

3.1. The Multilingual Preface

In order for the message to be received and displayed in non-conforming email clients, the message SHOULD contain an explanatory message part that MUST NOT be marked with a Content-Language field and MUST be the first of the message parts. For maximum support in the most basic of non-conforming email clients, it SHOULD have a Content-Type of ‘text/plain’. Because non-conforming email clients are expected to treat a message with an unknown multipart type as ‘multipart/mixed’ (in accordance with Sections 5.1.3 and 5.1.7 of RFC 2046 [RFC2046]), they may show all of the message parts sequentially or as attachments. Including and showing this explanatory part will help the message recipient understand the message structure.

This initial message part SHOULD briefly explain to the recipient that the message contains multiple languages, and the parts may be rendered sequentially or as attachments. This SHOULD be presented in the same languages that are provided in the subsequent language message parts.

As this explanatory section is likely to contain languages using scripts that require non-US-ASCII characters, it is RECOMMENDED that a UTF-8 charset be used for this message part. See RFC 3629 [RFC3629] for details of UTF-8.

Whilst this section of the message is useful for backward compatibility, it will normally only be shown when rendered by a non-conforming email client. This is because conforming email clients SHOULD only show the single language message part identified by the user’s preferred language and the language message part’s Content-Language.

For the correct display of the multilingual preface in a non-conforming email client, the sender MAY use the Content-Disposition field with a value of ‘inline’ in conformance with RFC 2183 [RFC2183] (which defines the Content-Disposition field). If provided, this SHOULD be placed at the ‘multipart/multilingual’ level and in the multilingual preface. This makes it clear to a non-conforming email client that the multilingual preface should be displayed immediately
to the recipient, followed by any subsequent parts marked as ‘inline’.

For examples of a multilingual preface, see Section 8.

3.2. The Language Message Parts

The language message parts are typically translations of the same message content. These message parts SHOULD be ordered so that the first part after the multilingual preface is in the language believed to be the most likely to be recognized by the recipient; this will constitute the default part when language negotiation fails and there is no language-independent part. All of the language message parts MUST have a Content-Language field and a Content-Type field; they MAY have a Content-Translation-Type field.

The Content-Type for each individual language message part SHOULD be ‘message/rfc822’ to provide good support with non-conforming email clients. However, an implementation MAY use ‘message/global’ as support for ‘message/global’ becomes more commonplace. (See RFC 6532 [RFC6532] for details of ‘message/global’.) Each language message part should have a Subject field in the appropriate language for that language part. If there is a From field present, its value MUST include the same email address as the top-level From header field, although the display name MAY be a localized version. If there is a mismatch of sender email address, the top-level From header field value SHOULD be used to show to the recipient.

3.3. The Language-Independent Message Part

If there is language-independent content for the recipient to see if they have a preferred language other than one of those specified in the language message parts, and the default language message part is unlikely to be understood, another part MAY be provided. This part could typically include one or more language-independent graphics. When this part is present, it MUST be the last part and MUST have a Content-Language field with a value of "zxx" (as described in BCP 47 [RFC5646]). The part SHOULD have a Content-Type of ‘message/rfc822’ or ‘message/global’ (to match the language message parts).
4. Message Part Selection

The logic for selecting the message part to render and present to the recipient is summarized in the next few paragraphs.

If the email client does not understand ‘multipart/multilingual’, then it will treat the message as if it was ‘multipart/mixed’ and render message parts accordingly (in accordance with Sections 5.1.3 and 5.1.7 of RFC 2046 [RFC2046]).

If the email client does understand ‘multipart/multilingual’, then it SHOULD ignore the multilingual preface and select the best match for the user’s preferred language from the language message parts available. Also, the user may prefer to see the original message content in their second language over a machine translation in their first language. The Content-Translation-Type field value can be used for further selection based on this preference. The selection of the language part may be implemented in a variety of ways, although the matching schemes detailed in RFC 4647 [RFC4647] are RECOMMENDED as a starting point for an implementation. The goal is to render the most appropriate translation for the user.

If there is no match for the user’s preferred language or there is no preferred language information available, the email client SHOULD select the language-independent part (if one exists) or the first language part directly after the multilingual preface if a language-independent part does not exist.

If there is no translation type preference information available, the values of the Content-Translation-Type field may be ignored.

Additionally, interactive implementations MAY offer the user a choice from among the available languages or the option to see them all.

5. The Content-Language Field

The Content-Language field in the individual language message parts is used to identify the language in which the message part is written. Based on the value of this field, a conforming email client can determine which message part to display (given the user’s language settings).

The Content-Language MUST comply with RFC 3282 [RFC3282] (which defines the Content-Language field) and BCP 47 [RFC5646] (which defines the structure and semantics for the language tag values).
Examples of this field could look like the following:

Content-Language: en-GB

Content-Language: de

Content-Language: es-MX, fr

Content-Language: sr-Cyrl

6. The Content-Translation-Type Field

The Content-Translation-Type field can be used in the individual language message parts to identify the type of translation. Based on the value of this field and the user’s preferences, a conforming email client can determine which message part to display.

This field can have one of three possible values: ‘original’, ‘human’, or ‘automated’ (although other values may be added in the future). A value of ‘original’ is given in the language message part that is in the original language. A value of ‘human’ is used when a language message part is translated by a human translator or a human has checked and corrected an automated translation. A value of ‘automated’ is used when a language message part has been translated by an electronic agent without proofreading or subsequent correction. New values of the Content-Translation-Type header field ("translTypeExt" in the ABNF) are added according to the procedure specified in Section 9.3.

Examples of this field include:

Content-Translation-Type: original

Content-Translation-Type: human

The syntax of the Content-Translation-Type field in ABNF [RFC5234] is:

Content-Translation-Type = [FWS] translationtype

FWS = <Defined in RFC 5322>
translationtype = "original" / "human" / "automated" /
translTypeExt
translTypeExt = 1*atext
atext = <Defined in RFC 5322>

This references RFC 5322 [RFC5322] for the predefined rules ‘folding white space (FWS)’ and ‘atext’.
7. The Subject Field in the Language Message Parts

On receipt of the message, conforming email clients will need to render the subject in the correct language for the recipient. To enable this, the Subject field SHOULD be provided in each language message part. The value for this field should be a translation of the email subject.

US-ASCII and ‘encoded-word’ examples of this field include:

Subject: A really simple email subject

Subject: =?UTF-8?Q?Un_asunto_de_correo_electr=C3=b3nico_realtmente_sencillo?=  

See RFC 2047 [RFC2047] for the specification of ‘encoded-word’.

The subject to be presented to the recipient SHOULD be selected from the message part identified during the message part selection stage. If no Subject field is found, the top-level Subject header field value should be used.

8. Examples

8.1. An Example of a Simple Multiple-Language Email Message

Below is a minimal example of a multiple-language email message. It has the multilingual preface and two language message parts.

From: Nik@example.com  
To: Nathaniel@example.com  
Subject: Example of a message in Spanish and English  
Date: Thu, 7 Apr 2017 21:28:00 +0100  
MIME-Version: 1.0  
Content-Type: multipart/multilingual;  
boundary="01189998819991197253"

--01189998819991197253  
Content-Type: text/plain; charset="UTF-8"  
Content-Disposition: inline  
Content-Transfer-Encoding: quoted-printable

This is a message in multiple languages. It says the same thing in each language. If you can read it in one language, you can ignore the other translations. The other translations may be presented as attachments or grouped together.
Este es un mensaje en varios idiomas. Dice lo mismo en cada idioma. Si puede leerlo en un idioma, puede ignorar las otras traducciones. Las otras traducciones pueden presentarse como archivos adjuntos o agrupados.

--01189998819991197253
Content-Type: message/rfc822
Content-Language: en-GB
Content-Translation-Type: original
Content-Disposition: inline

Subject: Example of a message in Spanish and English
Content-Type: text/plain; charset="US-ASCII"
Content-Transfer-Encoding: 7bit
MIME-Version: 1.0

Hello, this message content is provided in your language.

--01189998819991197253
Content-Type: message/rfc822
Content-Language: es
Content-Translation-Type: human
Content-Disposition: inline

Content-Type: text/plain; charset="US-ASCII"
Content-Transfer-Encoding: 7bit
MIME-Version: 1.0

Hola, el contenido de este mensaje esta disponible en su idioma.

--01189998819991197253--

8.2.  An Example of a Multiple-Language Email Message with a Language-Independent Part

Below is an example of a multiple-language email message that has the multilingual preface followed by two language message parts and then a language-independent png image.

From: Nik@example.com
To: Nathaniel@example.com
Subject: Example of a message in Spanish and English
Date: Thu, 7 Apr 2017 21:08:00 +0100
MIME-Version: 1.0
Content-Type: multipart/multilingual;
boundary="01189998819991197253"
This is a message in multiple languages. It says the same thing in each language. If you can read it in one language, you can ignore the other translations. The other translations may be presented as attachments or grouped together.

Este es un mensaje en varios idiomas. Dice lo mismo en cada idioma. Si puede leerlo en un idioma, puede ignorar las otras traducciones. Las otras traducciones pueden presentarse como archivos adjuntos o agrupados.

Hello, this message content is provided in your language.

Hola, el contenido de este mensaje está disponible en su idioma.
8.3. An Example of a Complex Multiple-Language Email Message with a Language-Independent Part

Below is an example of a more complex multiple-language email message. It has the multilingual preface and two language message parts and then a language-independent png image. The language message parts have 'multipart/alternative' contents and would therefore require further processing to determine the content to display.

From: Nik@example.com
To: Nathaniel@example.com
Subject: Example of a message in Spanish and English
Date: Thu, 7 Apr 2017 20:55:00 +0100
MIME-Version: 1.0
Content-Type: multipart/multilingual;
            boundary="01189998819991197253"

--01189998819991197253
Content-Type: text/plain; charset="UTF-8"
Content-Disposition: inline
Content-Transfer-Encoding: quoted-printable

This is a message in multiple languages. It says the same thing in each language. If you can read it in one language, you can ignore the other translations. The other translations may be presented as attachments or grouped together.

Este es un mensaje en varios idiomas. Dice lo mismo en cada idioma. Si puede leerlo en un idioma, puede ignorar las otras traducciones. Las otras traducciones pueden presentarse como archivos adjuntos o agrupados.

--01189998819991197253--
Subject: Example of a message in Spanish and English

Content-Type: multipart/alternative;
boundary="72530118999911999881"; charset="US-ASCII"
MIME-Version: 1.0

--72530118999911999881
Content-Type: text/plain; charset="US-ASCII"
Content-Transfer-Encoding: 7bit

Hello, this message content is provided in your language.

--72530118999911999881
Content-Type: text/html; charset="US-ASCII"
Content-Transfer-Encoding: 7bit

<html><body>Hello, this message content is <b>provided</b>/i> in <i>your</i> language.</body></html>
9. IANA Considerations

9.1. The ‘multipart/multilingual’ Media Type

The ‘multipart/multilingual’ Media Type has been registered with IANA. This is the registration template based on the template specified in [RFC6838]:

Media Type name: multipart

Media subtype name: multilingual

Required parameters: boundary (defined in RFC 2046)

Optional parameters: N/A
Encoding considerations:
There are no encoding considerations for this multipart other
than that of the embedded body parts. The embedded body parts
(typically one 'text/plain' plus one or more 'message/*') may
contain 7-bit, 8-bit, or binary encodings.

Security considerations:
See the Security Considerations section in RFC 8255

Interoperability considerations:
Existing systems that do not treat unknown multipart subtypes
as 'multipart/mixed' may not correctly render a
'multipart/multilingual' type. These systems would also be non-
compliant with MIME.

Published specification: RFC 8255

Applications that use this media type:
Mail Transfer Agents, Mail User Agents, spam detection,
virus detection modules, and message authentication modules.

Fragment identifier considerations: N/A

Additional information:
Deprecated alias names for this type: N/A
Magic number(s): N/A
File extension(s): N/A
Macintosh file type code(s): N/A

Person & email address to contact for further information:
Nik Tomkinson
rfc.nik.tomkinson@gmail.com

Nathaniel Borenstein
nsb@mimicast.com

Intended usage: COMMON

Restrictions on usage: N/A

Author/Change controller: IETF
9.2. The Content-Translation-Type Field

The Content-Translation-Type field has been added to the IANA "Permanent Message Header Field Names" registry. That entry references this document. This registration template is below:

Header field name: Content-Translation-Type
Applicable protocol: MIME
Status: standard
Author/Change controller: IETF
Specification document(s): RFC 8255
Related information: none

9.3. The Content-Translation-Type Header Field Values

IANA has created a new registry titled "Content-Translation-Type Header Field Values". New values must be registered using the "Specification Required" [RFC8126] IANA registration procedure. Registrations must include a translation type value, a short description, and a reference to the specification.

This document also registers three initial values specified below.

Value: original
Description: Content in the original language
Reference: RFC 8255

Value: human
Description: Content that has been translated by a human translator or a human has checked and corrected an automated translation
Reference: RFC 8255

Value: automated
Description: Content that has been translated by an electronic agent without proofreading or subsequent correction
Reference: RFC 8255
10. Security Considerations

Whilst it is intended that each language message part is a direct
translation of the original message, this may not always be the case;
these parts could contain undesirable content. Therefore, there is a
possible risk that undesirable text or images could be shown to the
recipient if the message is passed through a spam filter that does
not check all of the message parts. The risk should be minimal due
to the fact that an unknown multipart subtype should be treated as
'multipart/mixed'; thus, each message part should be subsequently
scanned.

If the email contains undesirable content in a language that the
recipient cannot understand and this unknown content is assumed to be
a direct translation of the content that the recipient can
understand, the recipient may unintentionally forward undesirable
text or images to a recipient that can understand it. To mitigate this
risk, an interactive implementation may allow the recipient to see
all of the translations for comparison.

Because the language message parts have a Content-Type of 'message/
rfc822' or 'message/global', they might contain From fields that
could have different values from that of the top-level From field,
and they may not reflect the actual sender. The inconsistent From
field values might get shown to the recipient in a non-conforming
email client and may mislead the recipient into thinking that the
email came from someone other than the real sender.

11. References

11.1. Normative References

[RFC2045] Freed, N. and N. Borenstein, "Multipurpose Internet Mail
Extensions (MIME) Part One: Format of Internet Message
Bodies", RFC 2045, DOI 10.17487/RFC2045, November 1996,

[RFC2046] Freed, N. and N. Borenstein, "Multipurpose Internet Mail
Extensions (MIME) Part Two: Media Types", RFC 2046,
DOI 10.17487/RFC2046, November 1996,

[RFC2047] Moore, K., "MIME (Multipurpose Internet Mail Extensions)
Part Three: Message Header Extensions for Non-ASCII Text",
RFC 2047, DOI 10.17487/RFC2047, November 1996,


11.2.  Informative References

[HUMAN-LANG]

Acknowledgements

The authors are grateful for the helpful input received from many people but would especially like to acknowledge the help of Harald Alvestrand, Stephane Bortzmeyer, Eric Burger, Ben Campbell, Mark Davis, Doug Ewell, Ned Freed, Randall Gellens, Gunnar Hellstrom, Mirja Kuehlewind, Barry Leiba, Sean Leonard, John Levine, Alexey Melnikov, Addison Phillips, Julian Reschke, Pete Resnick, Adam Roach, Brian Rosen, Fiona Tomkinson, Simon Tyler, and Daniel Vargha.

The authors would also like to thank Fernando Alvaro and Luis de Pablo for their work on the Spanish translations.

Authors’ Addresses

Nik Tomkinson
Mimecast, Ltd.
CityPoint, One Ropemaker Street
London EC2Y 9AW
United Kingdom

Email: rfc.nik.tomkinson@gmail.com

Nathaniel Borenstein
Mimecast, Ltd.
480 Pleasant Street
Watertown, MA 02472
United States of America

Email: nsb@mimecast.com