IETF
Internet-Draft
Intended status: Standards Track
Expires: April 17, 2016

N. Tomkinson
N. Borenstein
Mimecast Ltd

October 15, 2015

Multiple Language Content Type
draft-tomkinson-slim-multilangcontent-02

Abstract

This document defines an addition to the Multipurpose Internet Mail Extensions (MIME) standard to make it possible to send one message that contains multiple language versions of the same information. The translations would be identified by a language code and selected by the email client based on a user's language settings or locale.

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 17, 2016.

Copyright Notice

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 multi-cultural cities and formed departments and teams that span continents, cultures and languages, so the need to communicate efficiently with little margin for miscommunication has grown exponentially.

The objective of this document is to define an addition to the Multipurpose Internet Mail Extensions (MIME) standard, to make 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.

Whilst this document depends on identification of language in message parts for non-real-time communication, there is a companion document that is concerned with a similar problem for real-time communication: [I-D.gellens-slim-negotiating-human-language]

1.1. Requirements Language

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 RFC 2119 [RFC2119].

2. The Content-Type Header Field

The "multipart/multilingual" MIME subtype allows the sending of a message in a number of different languages with the translations embedded in the same message. This MIME 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

Translation-Type parameter of the Content-Language 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 unmatched 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 which MUST NOT be marked with a Content-Language field and MUST be the first of the message parts. Because non-conforming email clients are expected to treat the message 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 explain briefly 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.

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, because conforming email clients SHOULD only show the single language message part identified by the user’s preferred language (or locale) 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 an example of a multilingual preface, see the examples in Section 8.

3.2. The Language Message Parts

The language message parts are translations of the same message content. These message parts MAY be ordered so that the first part after the multilingual preface is in the language believed to be the most likely to be recognised by the recipient. All of the language message parts MUST have a Content-Language field and a Content-Type field, they SHOULD have a Subject field and MAY have a Translation-Type parameter applied to the Content-Language field.

The Content-Type for each individual language part MAY be any MIME type (including multipart subtypes such as multipart/alternative). However, it is RECOMMENDED that the Content-Type of the language parts is kept as simple as possible for interoperability with existing email clients. The language parts are not required to have matching Content-Types or multipart structures. For example, there might be an English part of type "text/html" followed by a Spanish part of type "application/pdf" followed by a Chinese part of type "image/jpeg". Whatever the content-type, the contents SHOULD be composed for optimal viewing in the specified language.

For a non-multipart type, it is RECOMMENDED that the sender applies a Name parameter to the Content-Type field. This will help the recipient identify the translations when the translations are rendered as attachments by a non-conforming email client.

An example of this parameter is as follows:

Content-Type: text/plain; name="english.txt"

3.3. The Unmatched Message Part

If there is content intended for the recipient to see if they have a preferred language other than one of those specified in the language parts, another part MAY be provided. This would also be useful when a language independent graphic is available. When this unmatched part is present, it MUST be the last part, MUST NOT have a Content-Language field and SHOULD NOT have a Subject field.

4. Message Part Selection

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

Firstly, if the email client does not understand multipart/multilingual then it SHOULD treat the message as if it was multipart/mixed and render message parts accordingly.

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 Translation-Type parameter of the Content-Language field value can be used for further selection based on this preference. The selection of language part may be implemented in a variety of ways and is a matter for the email client and its user preferences. The goal is to render the most appropriate translation for the user. Similarly, the subject to display (for example in a message listing) should be chosen from the selected language message part if it is available.

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 unmatched part (if one exists) or the first language part (directly after the multilingual preface) if an unmatched part does not exist. The top-level Subject header field value should be used whenever a suitable translation cannot be identified.

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

Additionally, interactive implementations MAY offer the user a choice from among the available languages.

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 or locale).

The Content-Language MUST comply with RFC 3282 [RFC3282] (which defines the Content-Language field) and BCP 47/RFC 5646 [RFC5646] (which defines the structure and semantics for the language code

values). While RFC 5646 provides a mechanism accommodating increasingly fine-grained distinctions, in the interest of maximum interoperability, each Content-Language value SHOULD be restricted to the largest granularity of language tags; in other words, it is RECOMMENDED to specify only a Primary-subtag and NOT to include subtags (e.g., for region or dialect) unless the languages might be mutually incomprehensible without them. Examples of this field for English, German and an instruction manual in Spanish and French, could look like the following:

Content-Language: en

Content-Language: de

Content-Language: es, fr

6. The Translation-Type Parameter

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

This parameter 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.

Examples of this parameter include:

Content-Language: en; translation-type=original

Content-Language: fr; translation-type=human

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: =?iso-8859-1?Q?un_asunto_de_correo_electr=F3nico_sencillo?=

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

8. Examples

8.1. An Example of a Simple Multiple language email message

```
From: Nik
To: Nathaniel
Subject:  example of a message in Spanish and English
Content-Type: multipart/multilingual; boundary=01189998819991197253
Content-Disposition: inline

--01189998819991197253
Content-Disposition: inline

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-Language: en; translation-type=original
Content-Type: text/plain; name="english.txt"
Content-Disposition: inline
Subject: example of a message in Spanish and English

Hello, this message content is provided in your language.

--01189998819991197253
Content-Language: es; translation-type=human
Content-Type: text/plain; name="espanol.txt"
Content-Disposition: inline
Subject: =?iso-8859-1?Q?ejemplo_pr=E1ctico_de_mensaje_
en_espa=F1ol_e_ingl=E9s?=

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

--01189998819991197253
Content-Type: image/gif
Content-Disposition: inline

..GIF image showing iconic or language-independent content here..

--01189998819991197253--
```

8.2. An Example of a Complex Multiple language email message

Below is an example of a more complex multiple language email message formatted using the method detailed in this document. Note that the language message parts have multipart contents and would therefore require further processing to determine the content to display.

```
From: Nik
To: Nathaniel
Subject:  example of a message in Spanish and English
Content-Type: multipart/multilingual; boundary=01189998819991197253
Content-Disposition: inline

--01189998819991197253
Content-Disposition: inline

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-Language: en; translation-type=original
Content-Type: multipart/alternative; boundary=multipartaltboundary
Subject: example of a message in Spanish and English

--multipartaltboundary
Content-Type: text/plain; name="english.txt"

Hello, this message content is provided in your language.

--multipartaltboundary
Content-Type: text/html; name="english.html"

<html><body><p>Hello, this message content is provided in your
language.<p></body></html>

--multipartaltboundary--

--01189998819991197253
Content-Language: es; translation-type=human
Content-Type: multipart/mixed; boundary=multipartmixboundary
Subject: =?iso-8859-1?Q?ejemplo_pr=E1ctico_de_mensaje_
```

```
   en_espa=F1ol_e_ingl=E9s?=

   --multipartmixboundary
   Content-Type:application/pdf; name="espanol.pdf"

   ..PDF file in Spanish here..

   --multipartmixboundary
   Content-Type:image/jpeg; name="espanol.jpg"

   ..JPEG image showing Spanish content here..

   --multipartmixboundary--

   --01189998819991197253
   Content-Type: image/gif
   Content-Disposition: inline

   ..GIF image showing iconic or language-independent content here..

   --01189998819991197253--
```

9. Changes from Previous Versions

9.1. Changes from draft-tomkinson-multilangcontent-01 to draft-tomkinson-slim-multilangcontent-00

- File name and version number changed to reflect the proposed WG name SLIM (Selection of Language for Internet Media).
- Replaced the Subject-Translation field in the language message parts with Subject and provided US-ASCII and non-US-ASCII examples.
- Introduced the language-independent unmatched message part.
- Many wording improvements and clarifications throughout the document.

9.2. Changes from draft-tomkinson-slim-multilangcontent-00 to draft-tomkinson-slim-multilangcontent-01

- Added Translation-Type in each language message part to identify the source of the translation (original/human/automated).

9.3. Changes from draft-tomkinson-slim-multilangcontent-01 to draft-tomkinson-slim-multilangcontent-02

- Changed Translation-Type to be a parameter for the Content-Language field rather than a new separate field.
- Added a paragraph about using Content-Disposition field to help non-conforming mail clients correctly render the multilingual preface.
- Recommended using a Name parameter on the language part Content-Type to help the recipient identify the translations in non-conforming mail clients.
- Many wording improvements and clarifications throughout the document.

10. 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, Mark Davis, Doug Ewell, Randall Gellens, Gunnar Hellstrom, Alexey Melnikov, 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.

11. IANA Considerations

The multipart/multilingual MIME type will be registered with IANA.

12. Security Considerations

This document has no additional security considerations beyond those that apply to the standards and procedures on which it is built.

13. References

13.1. Normative References

[RFC2046] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types", RFC 2046, DOI 10.17487/RFC2046, November 1996, <http://www.rfc-editor.org/info/rfc2046>.

[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, <http://www.rfc-editor.org/info/rfc2047>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC2183] Troost, R., Dorner, S., and K. Moore, Ed., "Communicating Presentation Information in Internet Messages: The Content-Disposition Header Field", RFC 2183, DOI 10.17487/RFC2183, August 1997, <http://www.rfc-editor.org/info/rfc2183>.

[RFC3282] Alvestrand, H., "Content Language Headers", RFC 3282, DOI 10.17487/RFC3282, May 2002, <http://www.rfc-editor.org/info/rfc3282>.

[RFC5646] Phillips, A., Ed. and M. Davis, Ed., "Tags for Identifying Languages", BCP 47, RFC 5646, DOI 10.17487/RFC5646, September 2009, <http://www.rfc-editor.org/info/rfc5646>.

13.2. Informational References

[I-D.gellens-slim-negotiating-human-language] Gellens, R., "Negotiating Human Language in Real-Time Communications", draft-gellens-slim-negotiating-human-language-02 (work in progress), July 2015.

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
North America

Email: nsb@mimecast.com