IETF
Internet-Draft
Intended status: Standards Track
Expires: February 18, 2017

N. Tomkinson
N. Borenstein
Mimecast Ltd

August 17, 2016

Multiple Language Content Type
draft-ietf-slim-multilangcontent-05

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 tag and selected by the email client based on a user's language settings.

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 February 18, 2017.

Copyright Notice

Copyright (c) 2016 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.

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 significantly.

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.ietf-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 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 which 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 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.

As this explanatory section is likely to contain languages using scripts that require non-US-ASCII characters, it is RECOMMENDED that UTF-8 encoding is used for this message part.

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 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 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 recognised by the recipient as 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 and MAY have a Translation-Type parameter applied to the Content-Language 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 although the display name MAY be a localised version.

3.3. The Language Independent Message Part

If there is language independent content intended 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 could typically be a language independent graphic. When this part is present, it MUST be the last part, MUST have a Content-Language field with a value of "zxx" (as described in BCP 47/ RFC 5646 [RFC5646]) and SHOULD NOT have a Subject field and SHOULD NOT have a From field. 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 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, 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 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).

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).

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: =?UTF-8?Q?Un_asunto_de_correo_electr=C3=b3nico_
        realmente_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 (for example if the language independent part is selected) 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 Jul 2016 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; 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; translation-type=human
   Content-Disposition: inline

   Subject: =?UTF-8?Q?Ejemplo_pr=C3=A1ctico_de_mensaje_?=
    =?UTF-8?Q?en_espa=C3=B1ol_e_ingl=C3=A9s?=
   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 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 Jul 2016 21:08:00 +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; 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; translation-type=human
Content-Disposition: inline

Subject: =?UTF-8?Q?Ejemplo_pr=C3=A1ctico_de_mensaje_?=
 =?UTF-8?Q?en_espa=C3=B1ol_e_ingl=C3=A9s?=
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
Content-Type: message/rfc822; name="Icon"
Content-Language: zxx
Content-Disposition: inline

Content-Type: multipart/mixed;
        boundary="99911972530118999881"; charset="US-ASCII"
Content-Transfer-Encoding: 7bit
MIME-Version: 1.0

--99911972530118999881
Content-Type: image/png; name="icon.png"
Content-Disposition: inline
Content-Transfer-Encoding: base64

iVBORw0KGgoAAAANSUhEUgAAADAAAAAwCAYAAABXAvmHAAAKQ2lDQ1BJQ0MgUHJvZmlsZQAA
SA2dlndUU1... shortened for brevity ...7yxfd1SNsEy+OXr76qr
997zF2hvZYeDEP5ftGV6Xzx2o9MAAAAASUVORK5CYII=

--99911972530118999881--
--01189998819991197253--
```

8.3. An Example of a complex Multiple language email message with 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 Jul 2016 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
Content-Type: message/rfc822
Content-Language: en; translation-type=original
Content-Disposition: inline

Subject: Example of a message in Spanish and English
Content-Type: multipart/alternative;
        boundary="72530118999911999881"; charset="US-ASCII"
Content-Transfer-Encoding: 7bit
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> in <i>your</i>
language.</body></html>
```

```
--72530118999911999881--
--01189998819991197253
Content-Type: message/rfc822
Content-Language: es; translation-type=human
Content-Disposition: inline

Subject: =?UTF-8?Q?Ejemplo_pr=C3=A1ctico_de_mensaje_?=
 =?UTF-8?Q?en_espa=C3=B1ol_e_ingl=C3=A9s?=
Content-Type: multipart/alternative;
        boundary="530118999989991197281"; charset="US-ASCII"
Content-Transfer-Encoding: 7bit
MIME-Version: 1.0

--530118999989991197281
Content-Type: text/plain; charset="US-ASCII"
Content-Transfer-Encoding: 7bit

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

--530118999989991197281
Content-Type: text/html; charset="US-ASCII"
Content-Transfer-Encoding: 7bit

<html><body>Hola, el contenido de este <b>mensaje</b> <i>esta</i>
disponible en su idioma.</body></html>

--530118999989991197281--
--01189998819991197253
Content-Type: message/rfc822; name="Icon"
Content-Language: zxx
Content-Disposition: inline

Content-Type: multipart/mixed;
        boundary="999119725301189998881"; charset="US-ASCII"
Content-Transfer-Encoding: 7bit
MIME-Version: 1.0

--99911972530118999881
Content-Type: image/png; name="icon.png"
Content-Disposition: inline
Content-Transfer-Encoding: base64

iVBORw0KGgoAAAANSUhEUgAAADAAAAAwCAYAAABXAvmHAAAKQ2lDQ1BJQ0MgUHJvZmlsZQAA
SA2dlndUU1... shortened for brevity ...7yxfd1SNsEy+OXr76qr
997zF2hvZYeDEP5ftGV6Xzx2o9MAAAAASUVORK5CYII=

--99911972530118999881--
--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 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.

9.4. Changes from draft-tomkinson-slim-multilangcontent-02 to draft-ietf-slim-multilangcontent-00

- Name change to reflect the draft being accepted into SLIM as a working group document.
- Updated examples to use UTF-8 encoding where required.

- Removed references to 'locale' for identifying language preference.
- Recommended language matching schemes from RFC 4647 [RFC4647].
- Renamed the unmatched part to language independent part to reinforce its intended purpose.
- Added requirement for using Content-Language: zxx in the language independent part.
- Many wording improvements and clarifications throughout the document.

9.5. Changes from draft-ietf-slim-multilangcontent-00 to draft-ietf-slim-multilangcontent-01

- Changed the inner content type to require message/rfc822 or message/global.
- Updated the examples to reflect the new inner content types.
- Added to the security considerations to highlight the risk from insufficient spam filters.

9.6. Changes from draft-ietf-slim-multilangcontent-01 to draft-ietf-slim-multilangcontent-02

- Restricted the use of a From field in the language message parts and the language independent part.
- Updated the security considerations to highlight the risk of an unmatched sender addresses that could be set in the language message parts.

9.7. Changes from draft-ietf-slim-multilangcontent-02 to draft-ietf-slim-multilangcontent-03

- Relaxed the restriction on the use of the From field in the language message parts to allow a localised version of the sender's display name.

9.8. Changes from draft-ietf-slim-multilangcontent-03 to draft-ietf-slim-multilangcontent-04

- Updated the wording of the security considerations section to reflect the relaxation of the use of the From field in the language message parts.

9.9. Changes from draft-ietf-slim-multilangcontent-04 to draft-ietf-slim-multilangcontent-05

- Referenced the RFC for message/global in Language Message Parts section.
- Removed RFC 2119 keyword in the Message Part Selection section.
- Included full email addresses in all examples.
- Updated reference name of real-time companion document in the Introduction.
- Removed paragraph warning of over use of language sub-tags.
- Changed 'exponential' to 'significantly' in Introduction.

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, Sean Leonard, John Levine, Alexey Melnikov, Addison Phillips, Pete Resnick, 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.

11. IANA Considerations

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

12. 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 and 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 and so each message part should be subsequently scanned.

Because the language message parts have a Content-Type of message/rfc822 or message/global, they might contain From fields which could have different values to that of the top-level From field and 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.

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>.

[RFC4647] Phillips, A. and M. Davis, "Matching of Language Tags", BCP 47, RFC 4647, DOI 10.17487/RFC4647, September 2006, <http://www.rfc-editor.org/info/rfc4647>.

[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>.

[RFC6532] Yang, A., Steele, S., and N. Freed, "Internationalized Email Headers", RFC 6532, DOI 10.17487/RFC6532, February 2012, <http://www.rfc-editor.org/info/rfc6532>.

13.2. Informational References

[I-D.ietf-slim-negotiating-human-language]
Gellens, R., "Negotiating Human Language in Real-Time Communications", draft-ietf-slim-negotiating-human-language-04 (work in progress), July 2016.

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