Guidelines for Internationalized Email Clients
draft-dainow-eai-email-clients-00

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

This Internet-Draft will expire on July 18, 2008.

Copyright Notice

Copyright (C) The IETF Trust (2008).

Abstract

This document provides some guidelines for electronic mail clients that support Email Address Internationalization (EAI). A number of interoperability cases between different versions of email components are reviewed. Some User Interface recommendations are made that will minimize discrepancies between the display of composed and received email in different language environments.
Table of Contents

1. Conventions used in this document........................................3
2. Introduction.................................................................................3
   2.1. Terminology............................................................................3
3. Version Interoperability............................................................4
   3.1. Sending..................................................................................4
       3.1.1. Downgrade......................................................................4
   3.2. Receiving................................................................................7
       3.2.1. Display of Downgraded Headers......................................7
   3.3. Version Checks........................................................................8
4. Character Encoding.......................................................................8
5. Normalization..............................................................................8
6. Alternate Addresses.....................................................................9
   6.1. Sender....................................................................................9
   6.2. Recipients..............................................................................9
7. Security Considerations............................................................10
8. IANA Considerations.....................................................................10
9. Acknowledgments........................................................................10
10. References..................................................................................10
    10.1. Normative References..........................................................10
    10.2. Informative References........................................................11
Author’s Addresses..........................................................................12
Intellectual Property Statement.....................................................12
Disclaimer of Validity......................................................................13
1. 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].

These key words will be used when the statement derives from a normative reference. Use of these key words in lower case is to be interpreted as a recommendation within this informative guideline where the implementer is not bound by the statement.

2. Introduction

RFC 4952, Overview and Framework for Internationalized Email [RFC4952], describes changes to electronic mail (email) to fully support internationalized characters. The fundamental change is to remove the ASCII only restriction on email addresses and allow them to contain UTF-8 characters. Additional documents provide detailed specifications for the extensions required to email headers [I-D.ietf-eai-utf8headers] and to the protocols SMTP [I-D.ietf-eai-smtpext], POP [I-D.ietf-eai-pop] and IMAP [I-D.ietf-eai-imap-utf8].

This document provides guidelines for email clients that support these specifications for Email Address Internationalization (EAI).

2.1. Terminology

A number of different acronyms are typically used to describe the major functional components of email.

- Mail User Agent (MUA)
- Message Submission Agent (MSA)
- Message Transfer Agent (MTA)
- Message Delivery Agent (MDA)
- Message Store (MS)

The architecture of modern email systems can range from simple, with all components running on one server, to very complex, with components being distributed across multiple, geographically dispersed machines. Nevertheless, the above terminology is generally sufficient to represent different architectures from a functional point of view. For a comprehensive description of email architecture see [I-D.crocker-email-arch].
sender -> MUA -> MSA -> MTA

... 

MTA -> MDA -> MS -> FPI -> MUA -> recipient

(where ... represents possible additional MTA relays)

In this context, an "Email Client" is an MUA that has an interface to an MSA to send email and an interface to the MS to retrieve email. The interface to retrieve mail (FPI) is either direct access to the file system or to a POP or IMAP server. The MUA also provides a User Interface (UI) that allows an end user to read (display) and write (compose) their email.

A common email architecture includes the MSA function within the MTA. An improved architecture that better addresses security concerns is a separate MSA component as shown here \[RFC4409\], \[RFC5068\].

"UTF8SMTP" is used to indicate email address internationalization as specified by \[RFC4952\] and related documents.

"message/global" is an email message that contains UTF-8 characters beyond 7 bit ASCII (note that UTF-8 includes the ASCII character set) in message headers and/or body parts \[I-D.ietf-eai-utf8headers\].

"message/rfc822" is an email message that contains only 7 bit ASCII and does not use any UTF8SMTP extensions. Note that the original message (as composed by the user) may contain non-ASCII characters that have been encoded into ASCII using IDNA \[RFC3490\], MIME body encoding \[RFC2045\] or MIME header encoding \[RFC2047\].

3. Version Interoperability

3.1. Sending

For an MUA that supports UTF8SMTP, there is a matrix of possibilities based on whether the email envelope and the message contain non-ASCII characters and whether the MSA supports the UTF8SMTP extensions or not. The following table shows all the possible combinations. Y/N indicates that the MSA does/does not support UTF8SMTP.
The Envelope and the Message type are considered separately because the Envelope may contain, for example, email addresses that are all ASCII but the Subject or other header fields in the Message may contain non-ASCII (Cases 3, 7). Such email MUST NOT be sent to an MSA which does not support UTF8SMTP.

Messages containing non-ASCII characters should be sent using the UTF8SMTP extensions in preference to older encoding methods, such as IDNA [RFC3490] and MIME header encoding [RFC2047]. If the message body contains non-ASCII characters, it should be sent using 8BITMIME [RFC1652] instead of MIME body encoding such as quoted-printable or base64 [RFC2045].

This could be considered a configuration error. If the MSA does not support UTF8SMTP, the user should upgrade the MSA or else use a traditional email client. This is a reasonable approach in the case of an organization, where an IT group would be expected to synchronize MUA and MSA versions.

However, home users may end up in this situation when they have a computer that came with UTF8SMTP email client software and their Internet Service Provider (ISP) does not support UTF8SMTP. Many such users would not know how to locate an appropriate "traditional" email client and install it.

The MUA MUST NOT submit a message with UTF8SMTP headers if the MSA does not support the UTF8SMTP extensions. There are four options for this case described in section 2.2 of [I-D.ietf-eai-smtpext], summarized briefly as:

*TBD are cases 2 and 6 possible*
1. Rewrite the headers as ASCII (if the client contains the MSA).

2. Reject the message.

3. Find an alternate route to the destination (if the client interfaces to the MTA).

4. "Downgrade" the message.

3.1.1. Downgrade

The MUA MAY support the "downgrade" option. This builds a message with all ASCII headers so that it is compatible with email components that don’t support the UTF8SMTP extensions [I-D.ietf-eai-downgrade].

Downgrade is intended to provide a degree of downward compatibility between email systems that do not upgrade to UTF8SMTP at the same time. It is specified as an option for all email components MUA, MSA, MTA and MDA.

Downgrade requires an Alternate Address that is all ASCII delivery for each address that contains non-ASCII characters. This includes both the From and the Recipient addresses. A "From" Alternate Address is required so that if a message is downgraded, there will be a return path for delivery error notifications.

Guidelines for handling Alternate Addresses are discussed further in the section 6.

The following shows how a "From" header in a message sent from a non-ASCII email address with an ASCII Alternate Address would be downgraded.

Original header:

From: EAI Display <eai-addr@domain.idn <alt-addr@domain.ascii>>

Downgrade would change this to:

From: =?UTF-8?Q?EAI Display?= <alt-addr@domain.ascii>  
Downgraded-From: =?UTF-8?Q?EAI Display?=  
=?UTF-8?Q?<eai-addr@domain.idn?> <alt-addr@domain.ascii>>

Complete examples of Downgrade are shown in the Appendix of [I-D.ietf-eai-downgrade].
3.2. Receiving

The matrix of possibilities is based on the email message type and whether IMAP/POP and the MUA support the UTF8SMTP extensions or not (Y/N) [I-D.ietf-eai-imap-utf8], [I-D.ietf-eai-pop].

<table>
<thead>
<tr>
<th>Case</th>
<th>Message</th>
<th>IMAP</th>
<th>MUA</th>
<th>Transfered</th>
<th>Displayed</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>global</td>
<td>Y</td>
<td>Y</td>
<td>global</td>
<td>global</td>
<td>All UTF8SMTP</td>
</tr>
<tr>
<td>2</td>
<td>global</td>
<td>N</td>
<td>N</td>
<td>-</td>
<td>-</td>
<td>Configuration error</td>
</tr>
<tr>
<td>3</td>
<td>global</td>
<td>Y</td>
<td>N</td>
<td>downgraded</td>
<td>downgraded</td>
<td>POP/IMAP downgrade</td>
</tr>
<tr>
<td>4</td>
<td>downgraded</td>
<td>Y/N</td>
<td>Y</td>
<td>downgraded</td>
<td>global</td>
<td>Display original</td>
</tr>
<tr>
<td>5</td>
<td>downgraded</td>
<td>Y/N</td>
<td>N</td>
<td>downgraded</td>
<td>downgraded</td>
<td>Traditional</td>
</tr>
<tr>
<td>6</td>
<td>rfc822</td>
<td>Y/N</td>
<td>Y/N</td>
<td>rfc822</td>
<td>rfc822</td>
<td>Traditional</td>
</tr>
</tbody>
</table>

Cases 2-3

If the MUA does not support UTF8SMTP, direct maildrop access with message/global is prohibited. IMAP or POP must support Downgrade for this configuration.

Cases 4-6

An ASCII message may be received from either a UTF8SMTP or an ASCII interface.

It is possible that the original message was a UTF8SMTP message that got downgraded to ASCII in transit. A message can be identified as downgraded because it will have one or more headers that are prefixed with "Downgraded-".

The MUA may apply a conversion to restore the information contained in the "Downgraded-" headers of a downgraded message. This is separated out as Case 4.

3.2.1. Display of Downgraded Headers

*TBD include document here or keep as a separate reference* [I-D.fujiwara-eai-downgraded-display].

Support for conversion of "Downgraded-" headers is a separate consideration from support for Downgrade. If User A replies to a downgraded message from User B without converting, the reply will be sent to B’s Alternate Address. If "Downgraded-" headers are converted, then the reply will be sent to B’s UTF8SMTP address.
Assuming the UTF8SMTP address is the primary mail address of the sender, then it is preferable to convert and use the email address from the Downgraded headers.

*TBD* The MUA should remove all "Downgraded-" header fields when replying to a downgraded message, even if it does not decode downgraded messages.

3.3. Version Checks

When the configurations for the MSA/MTA or IMAP/POP interfaces are set up or changed, these connections should be checked to see if UTF8SMTP is supported. If not, a message should be given to users so that they have advance warning that internationalized email cannot be sent and/or received.

4. Character Encoding

Email text messages (message bodies) may be composed and displayed in many different character encoding schemes. Numerous character encodings have been developed over time in order to best represent different language scripts. In recent years there has been a trend to prefer Unicode as a "universal" character set and UTF-8 as the preferred encoding method.

A good general principle to follow is to minimize character conversions. This will reduce the chance that the received message is displayed differently from how it was composed. So displaying received mail and forwarding/replying to mail should use the character encoding of the received mail.

A UTF8SMTP compliant MUA should use UTF-8 as the default encoding for mail message bodies. If email clients utilize this default, then character conversions will be minimized and there will be less chance that someone will receive mail in an unrecognized encoding.

Notwithstanding the above, there may be cases where the defaults do not work well. The MUA should consider support for some local encodings for the mail body and encoded-word encodings per each destination because older MUAs may not be able to parse UTF-8. There should be options for the user to reset the default character encoding. There should also be options to change the encoding when reading or writing individual email messages.

5. Normalization

*This section TBD*
Normalization of UTF-8 text streams is specified in [I-D.klensin-net-utf8].

Email addresses need additional considerations. The domain part of the email address should be normalized by IDNA (dot-mapping and NAMEPREP).

Special consideration is needed for the "@" symbol used to separate the local name from the domain name. In the Japanese environment, FULLWIDTH COMMERCIAL AT (U+FF20) needs to be treated/replaced as "@".

The MUA should normalize all non-ASCII email addresses. This would apply to email addresses transmitted and addresses stored by the MUA (such as in the address book).

6. Alternate Addresses

Even if the MUA does not implement Downgrade, it should provide for Alternate Addresses. Otherwise, Downgrade cannot be done anywhere on the path to the recipient if a non-UTF8SMTP email component is encountered. Without Alternate Addresses for both the Sender and the Recipient, Downgrade cannot be done.

6.1. Sender

The sender’s email address is usually configured once and saved, often under an "Account Settings" option. Thereafter it is automatically used as the From address in mail that is sent. A configuration field for "Alternate Address" should be added. Only ASCII addresses are allowed in this field. If the sender’s address is ASCII, then entering an address in the Alternate Address field should not be allowed.

The configured value of this field is used as an ALT-ADDRESS parameter on the SMTP command "MAIL FROM:" and in From: message headers.

6.2. Recipients

Many email clients provide a way for the user to save a list of email addresses along with related information, typically called an "Address Book".

A field for an Alternate Address should be added to each address book entry. Only ASCII addresses are allowed in this field. A value for this field does not have to be provided by the user as it is not a requirement that a UTF8SMTP email address have an alternate ASCII
address. If the main email address is ASCII, then entering an address in the Alternate Address field should not be allowed.

When an email address that has an Alternate Address is selected from the Address Book and entered in an email message, such as in the To: field, it would be natural to display it in UTF8SMTP "double angle bracket" format, for example:

    Display Name <eai-addr@domain.idn <alt-addr@domain.ascii>>

The user should also be able to enter and edit the Alternate Address in an email message before it is sent.

The recipient Alternate Address, if provided in an email composition, is used as an ALT-ADDRESS parameter on the SMTP command "RCPT TO:" and in message headers where the recipient address is used.

7. Security Considerations

As an Informational document, this does not introduce any additional security considerations beyond those covered by the normative references for Email Address Internationalization (EAI).

8. IANA Considerations

IANA changes are covered by the normative references for Email Address Internationalization (EAI).

9. Acknowledgments

10. References

10.1. Normative References


[I-D.ietf-eai-utf8headers] Yeh, J., "Internationalized Email Headers", draft-ietf-eai-utf8headers-07 (work in progress), September 2007.


10.2. Informative References


Author’s Addresses

Ernie Dainow
Afilias Canada
4141 Yonge Street, Suite 204
Toronto, Ontario
Canada M2P 2A8

Email: edainow@ca.afilias.info

Kazunori Fujiwara
JPRS
Chiyoda First Bldg. East 13F, 3-8-1 Nishi-Kanda
Chiyoda-ku, Tokyo 101-0065
Japan

Phone: +81 3 5215 8451
Email: fujiwara@jprs.co.jp

Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.
Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM 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.

Copyright Statement

Copyright (C) The IETF Trust (2008).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

Acknowledgment

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