

Network Working Group
Internet Draft
Intended status: Standards Track
Expires: December 2007

A. Melnikov (Ed.)
Isode Limited
June 14, 2007

SMTP Language Extension
draft-melnikov-smtp-lang-07.txt

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/lid-abstracts.html>

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

Copyright Notice

Copyright (C) The IETF Trust (2007).

Abstract

The Simple Mail Transfer Protocol ([RFC 2821](#)) allows server responses to include human-readable text that in many cases needs to be presented to the user. This document specifies a way for a client to negotiate which language the server should use when sending human-readable text. It also extends the UTF-8 Delivery Status Notifications format to include language field for the human-readable text.

0. Meta Information on this draft

This information is intended to facilitate discussion.

<<NOTE to RFC Editor: please remove this section>>

The protocol discussed in this document is experimental and subject to change. Persons planning on either implementing or using this protocol are STRONGLY URGED to get in touch with the author before embarking on such a project.

ToDo List:

1). Martin Duerst wrote:

On the other hand, for the LANG parameter for the MAIL command should allow a language priority list. The reason for this is that (if my understanding is correct), this parameter is passed on along the relay chain of SMTP servers, and is supposed to go back to the original sender, and using a list increases the chance that there is something at the relevant server that can be understood by the originator.

2). Martin Duerst wrote:

For language fallback, I suggest you have a look at <http://www.ietf.org/internet-drafts/draft-ietf-ltru-matching-15.txt> (also IESG-approved). This gives you the (basic) "language-range" ABNF construct that includes the "*" wildcard.

Also, it seems that the matching going on on the server when the client issues a LANG command (e.g. in Example 5) is very close to and can (and should) be described in terms of [Section 3.4](#), Lookup, of the above draft. The only difference I see is that [Section 3.4](#) requires a solution (maybe default) in all cases, whereas in your case, the default if no matching language is found is not i-default, but "no change" or in other words, "previously selected language".

3). Greg Vaudreuil wrote:

The interaction between the LANG verb and the LANG Mail tag is not specified. The goal of the MAIL FROM tag is to get responses useful to the message sender. Those responses may come in the form of the DSN itself (covered) or in the SMTP reply converted to a DSN by the client SMTP. If the SMTP client requests french SMTP dialogue but the message sender requests German for an error message, the SMTP reply code text should be in German, that is, the MAIL-FROM LANG tag should override the LANG verb for the SMTP responses used to generate a DSN on the client SMTP side.

4). Chris Newman:

SMTP responses should be allowed to contain multiple languages: the first is always English, followed by a special delimiter, followed by text in another language (maybe allow for more than

2 languages). This can be done as a multiline response, e.g.

```
250-Command has succeeded
250 <delimiter, language mark><text in another language>
```

<<Followup discussion with Pete: don't specify exact syntax, but add some text saying that an implementation MAY return error text in other languages>>

5). Stephane Bortzmeyer wrote:

Otherwise, in the security section, I suggest to include text like:

Languages and language variations such as scripts are often closely associated with specific social, national, religious or ethnic affinities. Thus, language tags used in content negotiation, like other information exchanged on the Internet, might be a source of concern because they might be used to infer information about the sender and thus identify potential targets for surveillance.

If, for instance, the same program is both a Web browser and a Mail User Agent, the fact that the user configured his Web browser to request pages in a specific language should not automatically imply that his mail client broadcasts this preference to every Usenet newsgroup or mailing list.

[Rationale: in countries like Moldavia or the former Yugoslavia, asking for the cyrillic or the latin script is not innocent and is often tied to political views.]

Changes since -00

- 1). Corrected grammar error in LANG command description [section](#)
- [2](#)). Included Mark Crispin's suggestion of allowing the server to substitute a primary language if the sublanguage asked for is not available.
- 3). Added [section 5](#) that describes extended LANG reply
- 4). Corrected example, more examples
- 5). Added extension mechanism
- 6). Specified interaction with [RFC-2034](#) ("SMTP Service Extension for Returning Enhanced Error Codes")
- 7). LANG command must always have language-tag as a parameter. Only EHLO response could be used to examine list of supported languages.

Changes since -01

- 1). Corrected ABNF for CR
- 2). Updated Copyright [section](#)
- 3). Other minor bugfixes

Changes since -02

- 1). Extended DSN format to include language tag
- 2). Fixed few typos.

Changes since -03

- 1). Changed DSN format to include language tag and translation of text part of diagnostic-code-field. Don't use diagnostic-code-field for a non English text.
- 2). Added LANG parameter to MAIL FROM.

Changes since -04

- 1). Updated boilerplate.
- 2). Updated references. Split references into Normative and Informative.
- 3). Updated ABNF (ABNF for UTF-8 responses, allow for multiple language tags)
- 4). Clarified that LANGUAGE extension is applicable to both MTAs and MSAs.
- 5). Added '*' language (to match EAI POP3 draft).
- 6). Fixed EHLO capabilities in examples to match HELP output.
- 7). Added new section on line length limit.
- 8). Copied Security Considerations from EAI POP3 document.
- 9). Removed any mentioning of extensions to the LANG command.
- 10). Made extended data prefix optional (currently it is only used in response to the LANG command)
- 11). Many minor editorial changes.

Changes since -05

- 1). Updated boilerplate.
- 2). Replaced reference to [RFC 3066](#) with [draft-ietf-ltru-registry](#), removed some text as a result of this change.
- 3). Clarified that if the receiving MTA doesn't support the LANGUAGE extension, then the LANG parameter must be silently ignored.
- 4). Updated DSN format ABNF as per comment from Harald Alvestrand.

Open issues

- 0). Open issues are enclosed in <<>>.

1. Conventions used in this document

In examples, "C:" and "S:" indicate lines sent by the client and server respectively. If such lines are wrapped without a new "C:" or "S:" label, then the wrapping is for editorial clarity and is not part of the command.

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](#) [[KEYWORDS](#)].

2. Framework for the Language SMTP service extension

The Language SMTP service extension uses the SMTP service extension mechanism described in [[ESMTP](#)]. The following SMTP service extension is therefore defined:

- (1) The name of the SMTP service extension is "Language". This extension is applicable to regular SMTP [[ESMTP](#)], Message Submission Protocol [[SUBMIT](#)] and LMTP [[LMTP](#)].
- (2) The EHLO keyword value associated with this service extension is "LANGUAGE".
- (3) The LANGUAGE EHLO keyword will have zero or more space separated arguments, each containing a name of supported language tags. If no arguments are specified, this means that server is unable to enumerate the list of languages it supports.
- (4) A new SMTP verb "LANG" is defined by this document.
- (5) One optional parameter is added to the MAIL command:

An optional parameter for the MAIL command, using the esmtp-keyword "LANG", (used to propagate a language that should be

used in human readable part and/or localized-diagnostic-text-field field of "message/delivery-status" part (see [section 6.](#)) of a delivery status notification for the message), is defined in [section 7.](#)

3. LANG Command

LANG 1*(SP language-tag)

Arguments:

one or more language tag as defined in [[LANG-TAGS](#)].

Restrictions:

The LANG command is permitted throughout a mail connection.

Reply Codes:

Success:

250 LANG command completed successfully

Error:

504 None of the specified languages is supported

421 <domain> Service not available, closing transmission channel

Discussion:

The LANG command requests that human-readable text emitted by the server be localized to one of the language specified in the argument. If multiple language tags are specified, they are specified in the decreasing order of preference. This is called "the language priority list" in [[LANG-MATCH](#)].

If the SMTP client is an MTA, the LANG command can be used to return potential error messages in the language requested by the original sender.

If a sublanguage was asked for and not available but the primary language is available, the server SHOULD switch to the primary language and MUST use an extended LANG reply containing the identifier of the primary language it switched to as described in [section 5.](#)

<<Should the fallback to primary languages be removed?>>

<<Need to describe that the server first picks a language from the list, if it can't find any, it should try to use sublanguages for the specified languages>>

Any server that supports this extension MUST support the language "i-default". It SHOULD <<MUST?>> use the language "i-default" as described in [[CHARSET-POLICY](#)] as its default language until another supported language is negotiated by the client. If a server is able to enumerate supported languages it MUST include "i-default" in the EHLO response. Otherwise it

MUST NOT return any language in the LANGUAGE EHLO response.

The special "*" language range argument indicates a request to use a language designated as preferred by the server administrator. The preferred language MAY vary based on the currently authenticated user.

If the server can't find any language (or sublanguage) that it supports, the server returns the 504 reply code. Servers supporting the Enhanced Error Codes extension [[RFC-2034](#)] SHOULD use the 5.3.3 "System not capable of selected features" [[ENHANCED-SC](#)] error code in this case.

If the command succeeds, the server will return human-readable responses in the specified language starting with the successful 250 response to the LANG command. These responses will be in UTF-8 [[UTF-8](#)]. In particular, LANG command MAY affect the result of a HELP command. The successful 250 response to the LANG command MUST use the extended reply as described in [section 5](#). This reply will communicate the selected language to the client.

If the command fails, the server will continue to return human-readable responses in the language it was previously using.

Example 1:

```
< The server defaults to using responses in "i-default" language
  until the user explicitly changes the language. >
```

```
S: 220 smtp.example.com ESMTP server ready
C: EHLO main.example.com
S: 250-smtp.example.com
S: 250-AUTH CRAM-MD5 DIGEST-MD5
S: 250-EXPN
S: 250-VERFY
S: 250-DSN
S: 250 LANGUAGE EN FR RU i-default
C: HELP
S: 214-This is Bukamail version X.X.X
S: 214-Topics:
S: 214-   HELO   EHLO   MAIL   RCPT   DATA
S: 214-   RSET   NOOP   QUIT   HELP   VRFY
S: 214-   EXPN   VERB   AUTH   DSN
S: 214-For more info use "HELP <topic>".
S: 214 End of HELP info
```

```
< Once the client changes the language, all responses will be in
  that language starting with 250 response to the LANG command.
  Note that in the following examples accented characters
  are not shown, as they are not allowed in RFCs. However the
```

accented characters, encoded in UTF-8, would be present in the actual protocol exchange. >

```
C: LANG FR
S: 250 La commande LANG a ete executee avec succes

C: HELP
S: 214-Ici le programme Bukamail version X.X.X
S: 214-Topics:
S: 214-   HELO   EHLO   MAIL   RCPT   DATA
S: 214-   RSET   NOOP   QUIT   HELP   VRFY
S: 214-   EXPN   VERB   AUTH   DSN
S: 214-Pour obtenir de l'information supplementaire, utilisez
      "HELP <topic>".
S: 214 Fin de l'information
```

< If a server does not support the requested language, responses will continue to be returned in the current language the server is using. >

```
C: LANG DE
S: 504 Cette langue n'est pas acceptee
```

Example 2:

< The client tries to select a language that is not supported >

```
C: LANG be-Latn
S: 504 The specified language is not supported.
```

Example 3:

< The client tries to use a LANG extension not supported by the server >

```
C: LANG i-default (blah blah)
S: 504 LANG extension blah is not recognized.
```

4. SMTP Line length limit

This extension doesn't affect the maximum text line length as defined by [ESMTP] and extended by other ESMTP extensions. However ESMTP maximum line length is defined in octets and *not* Unicode characters. If a server implementation truncates (or splits to create multiline response) UTF-8 response text to conform to such ESMTP limits, it **MUST** truncate the response at the Unicode character boundary. (E.g. if a response encoded in UTF-8 would be 1002 octets including the terminating CRLF and the last Unicode character before CRLF is represented as 3 UTF-8 octets, the response must be truncated after 997 octets, with the terminating CRLF added after that).

5. "LANG" extended reply

Extended reply is the reply that contains additional information in the text part. Extended reply allows to pass additional information from server to client. Client may choose to ignore additional information in an extended reply. Thus client that doesn't recognize an extended reply would treat it as a regular SMTP reply.

Example 4:

```
< The client tries to select the language, but it is
  unavailable. However primary language is available>
```

```
C: LANG FR-ca
S: 250 [LANG FR]La commande LANG a ete executee avec succes
```

Client that supports LANGUAGE extension MUST recognize Enhanced Error Codes defined in [\[RFC-2034\]](#). When server supports both LANGUAGE and ENHANCEDSTATUSCODES extensions, Extended reply data MUST follow any Enhanced Error Code in reply.

Example 5:

```
< The server supports both LANGUAGE and ENHANCEDSTATUSCODES>
```

```
S: 220 smtp.example.com ESMTP server ready
C: EHLO main.example.com
S: 250-smtp.example.com
S: 250-LANGUAGE EN FR RU i-default
S: 250 ENHANCEDSTATUSCODES
C: LANG FR-ca
S: 250 2.0.0 [LANG FR]La Language commande a ete execute avec
  success
```

6. The LANG parameter of the ESMTP MAIL command

Then LANG esmtp-keyword on the extended MAIL command specifies what language should be used in human readable part and/or localized-diagnostic-text-field field of "message/delivery-status" part (see [section 7.](#)) of a delivery status notification for the message.

If the LANG esmtp-keyword is used, it MUST have an associated esmtp-value. The ABNF for the LANG parameter is:

```
lang-parameter = "LANG=" language-tag
```

If the message is relayed to another SMTP server that supports LANGUAGE ESMTP extension, the MTA acting as the client MUST check if the receiving MTA lists the language specified in lang-param

("requested language") in the list of supported language tags in LANGUAGE EHLO response. If the receiving MTA either lists the requested language or doesn't list any language tag (i.e. the receiving MTA is unable to list languages it supports) the sender MUST issue LANG command for the requested language. After that, regardless of the result of LANG command, the client MTA MUST specify LANG parameter in MAIL command.

The receiving MTA SHOULD use the language specified in LANG parameter if it has to generate a DSN for the message. Human readable part in generated DSN SHOULD contain the description of the event in both i-default and the requested language. If the receiving MTA doesn't support the requested language, it MUST act as if the client didn't specify any LANG parameter in the MAIL command.

If the receiving MTA doesn't support the LANGUAGE extension, the sending MTA MUST behave as if the LANG parameter was not specified, i.e. the LANG parameter MUST be silently dropped.

Example 6:

```
< Relaying of the message >

S: 220 smtp.example.com ESMTP server ready
C: EHLO main.example.com
S: 250-smtp.example.com
S: 250-DSN
S: 250-8BITMIME
S: 250 LANGUAGE
C: LANG RU
S: 504 Unsupported language
C: MAIL FROM:<Katerina@example.ru> LANG=ru
S: 250 <Katerina@example.ru> sender ok
C: DATA
S: 354 okay, send message
C: (message goes here)
C: .
S: 250 message accepted
C: QUIT
S: 221 goodbye
```

7. Delivery status notifications and extension

This section updates [[EAI-DSN](#)].

The format of International delivery status notifications (message/utf-8-delivery-status content type) is specified in [[EAI-DSN](#)]. This memo extends the per-recipient-fields of [[EAI-DSN](#)] [[DSN](#)] to include one new field, Localized-Diagnostic, that is equivalent to text part of Diagnostic-Code but includes a language

tag and contains text in the specified language. In the Augmented BNF [ABNF], per-recipient-fields is therefore extended as follows:

```
per-recipient-fields =  
    [ original-recipient-field CRLF ]  
    final-recipient-field CRLF  
    action-field CRLF  
    status-field CRLF  
    [ remote-mta-field CRLF ]  
    [ *(localized-diagnostic-text-field CRLF)  
      diagnostic-code-field CRLF ]  
    [ last-attempt-date-field CRLF ]  
    [ will-retry-until-field CRLF ]  
    *( extension-field CRLF )
```

<<Should the localized-diagnostic-text-field be moved to the very end, in order to improve interoperability?>>

```
localized-diagnostic-text-field = "Localized-Diagnostic" ":"  
                                language ";" *utf8-text
```

where language is a language tag as described in [LANG-TAGS] and ABNF for utf8-text is specified in [Section 8](#) of this document.

An SMTP server that supports both DSN and LANGUAGE extensions SHOULD include localized-diagnostic-text-field. Note that multiple Localized-Diagnostic DSN fields are allowed, but they MUST use different language tags. diagnostic-code-field MUST NOT contain text in any language other than English.

8. Formal Syntax

The following syntax specification uses the augmented Backus-Naur Form (BNF) as described in [ABNF]. Non-terminals referenced but not defined below are as defined by [ESMTP] or [ABNF].

Except as noted otherwise, all alphabetic characters are case-insensitive. The use of upper or lower case characters to define token strings is for editorial clarity only. Implementations MUST accept these strings in a case-insensitive fashion.

```
LANG-Command = "LANG" 1*(SP language-tag) CRLF
```

```
LANGUAGE-List = "LANGUAGE" *(SP <language-tag>) CRLF  
    ; Note 1: When "i-default" is used, all responses MUST be  
    ; entirely in ASCII.  
    ;  
    ; Note 2: See [LANG-TAGS] for the list of allowed  
    ; language tags.
```

```
language-tag = <language-tag> as defined in [LANG-TAGS]
```

```

Reply-line |= Lang-Reply-line
    ; Reply-line is defined in [ESMTP].
    ; See section 5 for description of Lang-Reply-line

Lang-Reply-line = Reply-code [ SP ext-text ] CRLF
    ; Reply line for a LANG command

ext-text = [ext-data] utf8-text

utf8-text = 1*utf8-no-cr-lf

utf8-no-cr-lf = %d1-9 /           ; Any Unicode character except for NUL,
                %d11 /          ; CR and LF
                %d12 /
                %d14-127 /
                UTFMB

UTFMB   = UTF2 / UTF3 / UTF4

UTF0    = %x80-BF

UTF2    = %xC2-DF UTF0

UTF3    = %xE0 %xA0-BF UTF0 / %xE1-EC 2(UTF0) /
          %xED %x80-9F UTF0 / %xEE-EF 2(UTF0)

UTF4    = %xF0 %x90-BF 2(UTF0) / %xF1-F3 3(UTF0) /
          %xF4 %x80-8F 2(UTF0)

ext-data = "[" ext-name SP ext-value "]"
    ; Note 1: In the case of multiline response the same ext-data
    ; SHOULD appear on every line.
    ;
    ; Note 2: In case when server also supports "SMTP Service
    ; Extension for Returning Enhanced Error Codes" [RFC-2034],
    ; ext-data MUST follow any Enhanced Error Code.

ext-name = "LANG"

ext-value = Primary-tag
    ; Primary tag as defined by [LANG-TAGS]

```

9. Security Considerations

It is possible for a man-in-the-middle attacker to insert a LANG command in the command stream thus making protocol-level diagnostic responses unintelligible to the user. A mechanism to integrity protect the session, such as TLS [[TLS](#)] can be used to defeat such attacks.

10. References

10.1. Normative References

[ESMTP] Klensin, J., "Simple Mail Transfer Protocol", [RFC 2821](#), April 2001.

[SUBMIT] Gellens, R. and J. Klensin, "Message Submission for Mail", [RFC 4409](#), April 2006.

[LMTP] Myers, J., "Local Mail Transfer Protocol", Carnegie-Mellon University, [RFC 2033](#), October 1996.

[LANG-TAGS] Phillips, A. and M. Davis, "Tags for Identifying Languages", [RFC 4646](#), September 2006.

[LANG-MATCH] Phillips, A. and M. Davis, "Matching of Language Tags", [RFC 4647](#), September 2006.

[CHARSET-POLICY] Alvestrand, H., "IETF Policy on Character Sets and Languages", [RFC 2277](#), January 1998.

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

[KEYWORDS] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [RFC 2119](#), March 1997.

[ABNF] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", [RFC 4234](#), October 2005.

[RFC-2034] Freed, N., "SMTP Service Extension for Returning Enhanced Error Codes", [RFC 2034](#), Innosoft, October 1996.

[ENHANCED-SC] Vaudreuil, G., "Enhanced Mail System Status Codes", [RFC 3463](#), January 2003.

[EAI-DSN] Newman, C., "International Delivery and Disposition Notifications", work in progress, [draft-ietf-eai-dsn-XX.txt](#)

[DSN] Moore, K., Vaudreuil, G., "An Extensible Message Format for Delivery Status Notifications", University of Tennessee, Lucent Technologies, [RFC 3464](#), January 2003.

10.2. Informative References

[TLS] Dierks, T. and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.1", [RFC 4346](#), April 2006.

11. Acknowledgments

This document is based on the early version of [draft-gahrns-imap-language-xx.txt](#). Thus the work of Andrew McCown is appreciated.

Many thanks to the following people who gave feedback on the document: Chris Newman, Brad Knowles, Paul Hoffman, Harald Alvestrand, Martin Duerst, Greg Vaudreuil and Stephane Bortzmeyer.

12. Author's Address

Mike Gahrns
Microsoft
One Microsoft Way
Redmond, WA, 98072

Phone: (425) 936-9833
Email: mikega@microsoft.com

Alexey Melnikov (Editor)
Isode Limited
5 Castle Business Village
36 Station Road
Hampton, Middlesex
TW12 2BX,
United Kingdom

E-Mail: Alexey.Melnikov@isode.com

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.

Copyright Statement

Copyright (C) The IETF Trust (2007).

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.

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.

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).