

Internet Engineering Task Force (IETF)
Request for Comments: 6856
Obsoletes: 5721
Category: Standards Track
ISSN: 2070-1721

R. Gellens
QUALCOMM Incorporated
C. Newman
Oracle
J. Yao
CNNIC
K. Fujiwara
JPRS
March 2013

Post Office Protocol Version 3 (POP3) Support for UTF-8

Abstract

This specification extends the Post Office Protocol version 3 (POP3) to support international strings encoded in UTF-8 in usernames, passwords, mail addresses, message headers, and protocol-level text strings.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at <http://www.rfc-editor.org/info/rfc6856>.

Copyright Notice

Copyright (c) 2013 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.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1.	Introduction	3
1.1.	Conventions Used in This Document	3
2.	"UTF8" Capability	4
2.1.	The "UTF8" Command	5
2.2.	USER Argument to "UTF8" Capability	6
3.	"LANG" Capability	7
3.1.	Definition	7
3.2.	Discussion	7
3.3.	Examples	8
4.	Non-ASCII Character Maildrops	9
5.	"UTF8" Response Code	10
6.	IANA Considerations	10
7.	Security Considerations	10
8.	References	11
8.1.	Normative References	11
8.2.	Informative References	12
	Appendix A. Design Rationale	13
	Appendix B. Acknowledgments	13

1. Introduction

This document forms part of the Email Address Internationalization protocols described in the Email Address Internationalization Framework document [RFC6530]. As part of the overall Email Address Internationalization work, email messages can be transmitted and delivered containing a Unicode string encoded in UTF-8 in the header and/or body, and maildrops that are accessed using POP3 [RFC1939] might natively store Unicode characters.

This specification extends POP3 using the POP3 extension mechanism [RFC2449] to permit un-encoded UTF-8 [RFC3629] in headers and bodies (e.g., transferred using 8-bit content-transfer-encoding) as described in "Internationalized Email Headers" [RFC6532]. It also adds a mechanism to support login names and passwords containing a UTF-8 string (see Section 1.1 below), a mechanism to support UTF-8 strings in protocol-level response strings, and the ability to negotiate a language for such response strings.

This specification also adds a new response code to indicate that a message was not delivered because it required UTF-8 mode (as discussed in Section 2) and the server was unable or unwilling to create and deliver a surrogate form of the message as discussed in Section 7 of "IMAP Support for UTF-8" [RFC6855].

This specification replaces an earlier, experimental, approach to the same problem [RFC5721].

1.1. 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 "Key words for use in RFCs to Indicate Requirement Levels" [RFC2119].

The terms "UTF-8 string" or "UTF-8 character" are used to refer to Unicode characters, which may or may not be members of the ASCII repertoire, encoded in UTF-8 [RFC3629], a standard Unicode encoding form. All other specialized terms used in this specification are defined in the Email Address Internationalization framework document.

In examples, "C:" and "S:" indicate lines sent by the client and server, respectively. If a single "C:" or "S:" label applies to multiple lines, then the line breaks between those lines are for editorial clarity only and are not part of the actual protocol exchange.

Note that examples always use ASCII characters due to limitations of the RFC format; otherwise, some examples for the "LANG" command would have appeared incorrectly.

2. "UTF8" Capability

This specification adds a new POP3 Extension [RFC2449] capability response tag and command to specify support for header field information outside the ASCII repertoire. The capability tag and new command and functionality are described below.

CAPA tag:

UTF8

Arguments with CAPA tag:

USER

Added Commands:

UTF8

Standard commands affected:

USER, PASS, APOP, LIST, TOP, RETR

Announced states / possible differences:

both / no

Commands valid in states:

AUTHORIZATION

Specification reference:

this document

Discussion:

This capability adds the "UTF8" command to POP3. The "UTF8" command switches the session from the ASCII-only mode of POP3 [RFC1939] to UTF-8 mode. The UTF-8 mode means that all messages transmitted between servers and clients are UTF-8 strings, and both servers and clients can send and accept UTF-8 strings.

2.1. The "UTF8" Command

The "UTF8" command enables UTF-8 mode. The "UTF8" command has no parameters.

UTF-8 mode has no effect on messages in an ASCII-only maildrop. Messages in native Unicode maildrops can be encoded in UTF-8 using internationalized headers [RFC6532], in 8bit content-transfer-encoding (see Section 2.8 of MIME [RFC2045]), in ASCII, or in any combination of these options. In UTF-8 mode, if the character encoding format of maildrops is UTF-8 or ASCII, the messages are sent to the client as is; if the character encoding format of maildrops is a format other than UTF-8 or ASCII, the messages' encoding format SHOULD be converted to be UTF-8 before they are sent to the client. When UTF-8 mode has not been enabled, character strings outside the ASCII repertoire MUST NOT be sent to the client as is. If a client requests a UTF-8 message when UTF-8 mode is not enabled, the server MUST either send the client a surrogate message that complies with unextended POP and Internet Mail Format without UTF-8 mode support, or fail the request with an -ERR response. See Section 7 of "IMAP Support for UTF-8" [RFC6855] for information about creating a surrogate message and for a discussion of potential issues. Section 5 of this document discusses "UTF8" response codes. The server MAY respond to the "UTF8" command with an -ERR response.

Note that even in UTF-8 mode, MIME binary content-transfer-encoding as defined in Section 6.2 of MIME [RFC2045] is still not permitted. MIME 8bit content-transfer-encoding (8BITMIME) [RFC6152] is obviously allowed.

The octet count (size) of a message reported in a response to the "LIST" command SHOULD match the actual number of octets sent in a "RETR" response (not counting byte-stuffing). Sizes reported elsewhere, such as in "STAT" responses and non-standardized, free-form text in positive status indicators (following "+OK") need not be accurate, but it is preferable if they are.

Normal operation for maildrops that natively support non-ASCII characters will be for both servers and clients to support the extension discussed in this specification. Upgrading both clients and servers is the only fully satisfactory way to support the capabilities offered by the "UTF8" extension and SMTPUTF8 mail more generally. Servers must, however, anticipate the possibility of a client attempting to access a message that requires this extension without having issued the "UTF8" command. There are no completely satisfactory responses for this case other than upgrading the client to support this specification. One solution, unsatisfactory because

the user may be confused by being able to access the message through some means and not others, is that a server MAY choose to reject the command to retrieve the message as discussed in Section 5. Other alternatives, including the possibility of creating and delivering a surrogate form of the message, are discussed in Section 7 of "IMAP Support for UTF-8" [RFC6855].

Clients MUST NOT issue the "STLS" command [RFC2595] after issuing UTF8; servers MAY (but are not required to) enforce this by rejecting with an -ERR response an "STLS" command issued subsequent to a successful "UTF8" command. (Because this is a protocol error as opposed to a failure based on conditions, an extended response code [RFC2449] is not specified.)

2.2. USER Argument to "UTF8" Capability

If the USER argument is included with this capability, it indicates that the server accepts UTF-8 usernames and passwords.

Servers that include the USER argument in the "UTF8" capability response SHOULD apply SASLprep [RFC4013] or one of its Standards Track successors to the arguments of the "USER" and "PASS" commands.

A client or server that supports APOP and permits UTF-8 in usernames or passwords MUST apply SASLprep or one of its Standards Track successors to the username and password used to compute the APOP digest.

When applying SASLprep, servers MUST reject UTF-8 usernames or passwords that contain a UTF-8 character listed in Section 2.3 of SASLprep. When applying SASLprep to the USER argument, the PASS argument, or the APOP username argument, a compliant server or client MUST treat them as a query string [RFC3454]. When applying SASLprep to the APOP password argument, a compliant server or client MUST treat them as a stored string [RFC3454].

If the server includes the USER argument in the UTF8 capability response, the client MAY use UTF-8 characters with a "USER", "PASS", or "APOP" command; the client MAY do so before issuing the "UTF8" command. Clients MUST NOT use UTF-8 characters when authenticating if the server did not include the USER argument in the UTF8 capability response.

The server MUST reject UTF-8 usernames or passwords that fail to comply with the formal syntax in UTF-8 [RFC3629].

Use of UTF-8 strings in the "AUTH" command is governed by the POP3 SASL [RFC5034] mechanism.

3. "LANG" Capability

This document adds a new POP3 extension [RFC2449] capability response tag to indicate support for a new command: "LANG".

3.1. Definition

The capability tag and new command are described below.

CAPA tag:

LANG

Arguments with CAPA tag:

none

Added Commands:

LANG

Standard commands affected:

All

Announced states / possible differences:

both / no

Commands valid in states:

AUTHORIZATION, TRANSACTION

Specification reference:

this document

3.2. Discussion

POP3 allows most +OK and -ERR server responses to include human-readable text that, in some cases, might be presented to the user. But that text is limited to ASCII by the POP3 specification [RFC1939]. The "LANG" capability and command permit a POP3 client to negotiate which language the server uses when sending human-readable text.

The "LANG" command requests that human-readable text included in all subsequent +OK and -ERR responses be localized to a language matching the language range argument (the "basic language range" as described by the "Matching of Language Tags" [RFC4647]). If the command succeeds, the server returns a +OK response followed by a single space, the exact language tag selected, and another space. Human-readable text in the appropriate language then appears in the rest of the line. This, and subsequent protocol-level human-readable text, is encoded in the UTF-8 charset.

If the command fails, the server returns an -ERR response and subsequent human-readable response text continues to use the language that was previously used.

If the client issues a "LANG" command with the special "*" language range argument, it indicates a request to use a language designated as preferred by the server administrator. The preferred language MAY vary based on the currently active user.

If no argument is given and the POP3 server issues a positive response, that response will usually consist of multiple lines. After the initial +OK, for each language tag the server supports, the POP3 server responds with a line for that language. This line is called a "language listing".

In order to simplify parsing, all POP3 servers are required to use a certain format for language listings. A language listing consists of the language tag [RFC5646] of the message, optionally followed by a single space and a human-readable description of the language in the language itself, using the UTF-8 charset. There is no specific order to the listing of languages; the order may depend on configuration or implementation.

3.3. Examples

Examples for "LANG" capability usage are shown below.

Note that some examples do not include the correct character accents due to limitations of the RFC format.

```
C: USER karen
S: +OK Hello, karen
C: PASS password
S: +OK karen's maildrop contains 2 messages (320 octets)
```

Client requests deprecated MUL language [ISO639-2]. Server replies with -ERR response.

```
C: LANG MUL
S: -ERR invalid language MUL
```


A LANG command with no parameters is a request for a language listing.

```
C: LANG
S: +OK Language listing follows:
S: en English
S: en-boont English Boontling dialect
S: de Deutsch
S: it Italiano
S: es Espanol
S: sv Svenska
S: .
```

A request for a language listing might fail.

```
C: LANG
S: -ERR Server is unable to list languages
```

Once the client selects the language, all responses will be in that language, starting with the response to the "LANG" command.

```
C: LANG es
S: +OK es Idioma cambiado
```

If a server returns an -ERR response to a "LANG" command that specifies a primary language, the current language for responses remains in effect.

```
C: LANG uga
S: -ERR es Idioma <<UGA>> no es conocido
```

```
C: LANG sv
S: +OK sv Kommandot "LANG" lyckades
```

```
C: LANG *
S: +OK es Idioma cambiado
```

4. Non-ASCII Character Maildrops

When a POP3 server uses a native non-ASCII character maildrop, it is the responsibility of the server to comply with the POP3 base specification [RFC1939] and Internet Message Format [RFC5322] when not in UTF-8 mode. When the server is not in UTF-8 mode and the message requires that mode, requests to download the message MAY be rejected (as specified in the next section) or the various alternatives outlined in Section 2.1 above, including creation and delivery of surrogates for the original message, MAY be considered.

5. "UTF8" Response Code

Per "POP3 Extension Mechanism" [RFC2449], this document adds a new response code: UTF8, described below.

Complete response code:
UTF8

Valid for responses:
-ERR

Valid for commands:
LIST, TOP, RETR

Response code meaning and expected client behavior:

The "UTF8" response code indicates that a failure is due to a request for message content that contains a UTF-8 string when the client is not in UTF-8 mode.

The client MAY reissue the command after entering UTF-8 mode.

6. IANA Considerations

Sections 2 and 3 of this specification update two capabilities ("UTF8" and "LANG") in the POP3 capability registry [RFC2449].

Section 5 of this specification adds one new response code ("UTF8") to the POP3 response codes registry [RFC2449].

7. Security Considerations

The security considerations of UTF-8 [RFC3629], SASLprep [RFC4013], and the Unicode Format for Network Interchange [RFC5198] apply to this specification, particularly with respect to use of UTF-8 strings in usernames and passwords.

The "LANG *" command might reveal the existence and preferred language of a user to an active attacker probing the system if the active language changes in response to the "USER", "PASS", or "APOP" commands prior to validating the user's credentials. Servers are strongly advised to implement a configuration to prevent this exposure.

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 protect the

integrity of the session can be used to defeat such attacks. For example, a client can issue the "STLS" command [RFC2595] before issuing the "LANG" command.

As with other internationalization upgrades, modifications to server authentication code (in this case, to support non-ASCII strings) need to be done with care to avoid introducing vulnerabilities (for example, in string parsing or matching). This is particularly important if the native databases or mailstore of the operating system use some character set or encoding other than Unicode in UTF-8.

8. References

8.1. Normative References

- [RFC1939] Myers, J. and M. Rose, "Post Office Protocol - Version 3", STD 53, RFC 1939, May 1996.
- [RFC2045] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies", RFC 2045, November 1996.
- [RFC2047] Moore, K., "MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text", RFC 2047, November 1996.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2449] Gellens, R., Newman, C., and L. Lundblade, "POP3 Extension Mechanism", RFC 2449, November 1998.
- [RFC3454] Hoffman, P. and M. Blanchet, "Preparation of Internationalized Strings ("stringprep")", RFC 3454, December 2002.
- [RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, RFC 3629, November 2003.
- [RFC4013] Zeilenga, K., "SASLprep: Stringprep Profile for User Names and Passwords", RFC 4013, February 2005.
- [RFC4647] Phillips, A. and M. Davis, "Matching of Language Tags", BCP 47, RFC 4647, September 2006.
- [RFC5198] Klensin, J. and M. Padlipsky, "Unicode Format for Network Interchange", RFC 5198, March 2008.

- [RFC5322] Resnick, P., Ed., "Internet Message Format", RFC 5322, October 2008.
- [RFC5646] Phillips, A. and M. Davis, "Tags for Identifying Languages", BCP 47, RFC 5646, September 2009.
- [RFC6152] Klensin, J., Freed, N., Rose, M., and D. Crocker, "SMTP Service Extension for 8-bit MIME Transport", STD 71, RFC 6152, March 2011.
- [RFC6530] Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", RFC 6530, February 2012.
- [RFC6532] Yang, A., Steele, S., and N. Freed, "Internationalized Email Headers", RFC 6532, February 2012.
- [RFC6855] Resnick, P., Newman, C., and S. Shen, "IMAP Support for UTF-8", RFC 6855, March 2013.

8.2. Informative References

- [ISO639-2] International Organization for Standardization, "ISO 639-2:1998. Codes for the representation of names of languages -- Part 2: Alpha-3 code", October 1998.
- [RFC2231] Freed, N. and K. Moore, "MIME Parameter Value and Encoded Word Extensions: Character Sets, Languages, and Continuations", RFC 2231, November 1997.
- [RFC2595] Newman, C., "Using TLS with IMAP, POP3 and ACAP", RFC 2595, June 1999.
- [RFC5034] Siemborski, R. and A. Menon-Sen, "The Post Office Protocol (POP3) Simple Authentication and Security Layer (SASL) Authentication Mechanism", RFC 5034, July 2007.
- [RFC5721] Gellens, R. and C. Newman, "POP3 Support for UTF-8", RFC 5721, February 2010.

Appendix A. Design Rationale

This non-normative section discusses the reasons behind some of the design choices in this specification.

Due to interoperability problems with the MIME Message Header Extensions [RFC2047] and limited deployment of the extended MIME parameter encodings [RFC2231], it is hoped these 7-bit encoding mechanisms can be deprecated in the future when UTF-8 header support becomes prevalent.

The USER capability (Section 2.2) and hence the upgraded "USER" command and additional support for non-ASCII credentials, are optional because the implementation burden of SASLprep [RFC4013] is not well understood, and mandating such support in all cases could negatively impact deployment.

Appendix B. Acknowledgments

Thanks to John Klensin, Joseph Yee, Tony Hansen, Alexey Melnikov, and other Email Address Internationalization working group participants who provided helpful suggestions and interesting debate that improved this specification.

Authors' Addresses

Randall Gellens
QUALCOMM Incorporated
5775 Morehouse Drive
San Diego, CA 92651
USA

E-Mail: rg+ietf@qualcomm.com

Chris Newman
Oracle
800 Royal Oaks
Monrovia, CA 91016-6347
USA

E-Mail: chris.newman@oracle.com

Jiankang YAO
CNNIC
No.4 South 4th Street, Zhongguancun
Beijing
China

Phone: +86 10 58813007
E-Mail: yaojk@cnnic.cn

Kazunori Fujiwara
Japan Registry Services Co., Ltd.
Chiyoda First Bldg. East 13F, 3-8-1 Nishi-Kanda
Tokyo
Japan

Phone: +81 3 5215 8451
E-Mail: fujiwara@jprs.co.jp