

Internet Engineering Task Force  
Internet-Draft  
Obsoletes: [RFC5738](#) (if approved)  
Intended status: Standards Track  
Expires: June 28, 2012

P. Resnick, Ed.  
Qualcomm Incorporated  
C. Newman, Ed.  
Oracle  
S. Shen, Ed.  
CNNIC  
December 26, 2011

**IMAP Support for UTF-8**  
**draft-ietf-eai-5738bis-03**

Abstract

This specification extends the Internet Message Access Protocol version 4rev1 (IMAP4rev1) to support UTF-8 encoded international characters in user names, mail addresses and message headers. This specification replaces [RFC 5738](#).

Status of This Memo

This Internet-Draft is submitted to IETF 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 June 28, 2012.

Copyright Notice

Copyright (c) 2011 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.

## Table of Contents

<a href="#">1.</a>	<a href="#">Introduction . . . . .</a>	<a href="#">3</a>
<a href="#">2.</a>	<a href="#">Conventions Used in this Document . . . . .</a>	<a href="#">3</a>
<a href="#">3.</a>	<a href="#">UTF8=ACCEPT IMAP Capability . . . . .</a>	<a href="#">3</a>
<a href="#">3.1.</a>	<a href="#">IMAP UTF-8 Quoted Strings . . . . .</a>	<a href="#">4</a>
<a href="#">3.2.</a>	<a href="#">UTF8 Parameter to SELECT and EXAMINE . . . . .</a>	<a href="#">5</a>
<a href="#">3.3.</a>	<a href="#">UTF-8 LIST and LSUB Responses . . . . .</a>	<a href="#">6</a>
<a href="#">3.4.</a>	<a href="#">UTF-8 Interaction with IMAP4 LIST Command Extensions . . . . .</a>	<a href="#">6</a>
<a href="#">3.4.1.</a>	<a href="#">UTF8 LIST Selection Option . . . . .</a>	<a href="#">6</a>
<a href="#">3.4.2.</a>	<a href="#">UTF8 LIST Return Option . . . . .</a>	<a href="#">7</a>
<a href="#">4.</a>	<a href="#">IMAP UTF8 Append Data Extension . . . . .</a>	<a href="#">8</a>
<a href="#">5.</a>	<a href="#">LOGIN Command and UTF-8 . . . . .</a>	<a href="#">9</a>
<a href="#">6.</a>	<a href="#">UTF8=ONLY Capability . . . . .</a>	<a href="#">9</a>
<a href="#">7.</a>	<a href="#">Issues with UTF-8 Header Mailstore . . . . .</a>	<a href="#">9</a>
<a href="#">8.</a>	<a href="#">IANA Considerations . . . . .</a>	<a href="#">10</a>
<a href="#">9.</a>	<a href="#">Security Considerations . . . . .</a>	<a href="#">11</a>
<a href="#">10.</a>	<a href="#">References . . . . .</a>	<a href="#">11</a>
<a href="#">10.1.</a>	<a href="#">Normative References . . . . .</a>	<a href="#">11</a>
<a href="#">10.2.</a>	<a href="#">Informative References . . . . .</a>	<a href="#">12</a>
<a href="#">Appendix A.</a>	<a href="#">Appendix A. Design Rationale . . . . .</a>	<a href="#">13</a>
<a href="#">Appendix B.</a>	<a href="#">Appendix B. Acknowledgments . . . . .</a>	<a href="#">13</a>



## 1. Introduction

This specification extends IMAP4rev1 [RFC3501] to permit UTF-8 [RFC3629] in headers as described in "Internationalized Email Headers" [I-D.ietf-eai-rfc5335bis]. It also adds a mechanism to support mailbox names, login names, and passwords using the UTF-8 charset. This specification creates two new IMAP capabilities to allow servers to advertise these new extensions, along with one new IMAP LIST selection option and a new IMAP LIST return option.

This specification permits implementation of an IMAP server that hides mailboxes with internationalized email messages from IMAP clients that do not support this extension. Implementation of "Post-delivery Message Downgrading for Internationalized Email Messages" [popimap-downgrade] is necessary for an MDA to make mailboxes with internationalized email messages visible to IMAP clients that do not support this extension.

This specification replaces an earlier, experimental, approach to the same problem [RFC5738].

## 2. Conventions Used in this Document

The key words "MUST", "MUST NOT", "SHOULD", "SHOULD NOT", and "MAY" in this document are to be interpreted as defined in "Key words for use in RFCs to Indicate Requirement Levels" [RFC2119].

The formal syntax uses the Augmented Backus-Naur Form (ABNF) [RFC5234] notation including the core rules defined in [Appendix B of \[RFC5234\]](#). In addition, rules from IMAP4rev1 [RFC3501], UTF-8 [RFC3629], "Collected Extensions to IMAP4 ABNF" [RFC4466], and IMAP4 LIST Command Extensions [RFC5258] are also referenced.

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.

## 3. UTF8=ACCEPT IMAP Capability

The "UTF8=ACCEPT" capability indicates that the server supports UTF-8 quoted strings, the "UTF8" parameter to SELECT and EXAMINE, and UTF-8 responses from the LIST and LSUB commands.

A client MUST use the "ENABLE UTF8=ACCEPT" command (defined in [RFC5161]) to indicate to the server that the client accepts UTF-8 quoted-strings. The "ENABLE UTF8=ACCEPT" command MUST only be used



in the authenticated state. (Note that the "UTF8=ONLY" capability described in [Section 6](#) imply the "UTF8=ACCEPT" capability. See additional information in these sections.)

### 3.1. IMAP UTF-8 Quoted Strings

The IMAP4rev1 [[RFC3501](#)] base specification forbids the use of 8-bit characters in atoms or quoted strings. Thus, a UTF-8 string can only be sent as a literal. This can be inconvenient from a coding standpoint, and unless the server offers IMAP4 non-synchronizing literals [[RFC2088](#)], this requires an extra round trip for each UTF-8 string sent by the client. When the IMAP server advertises the "UTF8=ACCEPT" capability, it informs the client that it supports native UTF-8 quoted-strings with the following syntax:

```
string          =/ uQuoted

uQuoted         = "\"" DQUOTE *uQUOTED-CHAR DQUOTE
                  ; referred as 'utf8-quote' in this document

DQUOTE         = <Defined in appendix B.1 of RFC 5234>

uQUOTED-CHAR    = QUOTED-CHAR / UTF8-2 / UTF8-3 / UTF8-4

UTF8-2         = <Defined in Section 4 of RFC3629>

UTF8-3         = <Defined in Section 4 of RFC3629>

UTF8-4         = <Defined in Section 4 of RFC3629>
```

When this quoting mechanism is used by the client (specifically an octet sequence beginning with "\*" and ending with "), then the server MUST reject octet sequences with the high bit set that fail to comply with the formal syntax in [[RFC3629](#)] with a BAD response.

The IMAP server MUST NOT send utf8-quoted syntax to the client unless the client has indicated support for that syntax by using the "ENABLE UTF8=ACCEPT" command.

If the server advertises the "UTF8=ACCEPT" capability, the client MAY use utf8-quoted syntax with any IMAP argument that permits a string (including astring and nstring). However, if characters outside the US-ASCII repertoire are used in an inappropriate place, the results would be the same as if other syntactically valid but semantically invalid characters were used. Specific cases where UTF-8 characters are permitted or not permitted are described in the following paragraphs.



All IMAP servers that advertise the "UTF8=ACCEPT" capability SHOULD accept UTF-8 in mailbox names, and those that also support the "Mailbox International Naming Convention" described in [RFC 3501](#), [Section 5.1.3](#) MUST accept utf8-quoted mailbox names and convert them to the appropriate internal format. Mailbox names MUST comply with the Net-Unicode Definition ([Section 2 of \[RFC5198\]](#)) with the specific exception that they MUST NOT contain control characters (0000-001F, 0080-009F), delete (007F), line separator (2028), or paragraph separator (2029).

An IMAP client MUST NOT issue a SEARCH command that uses a mixture of utf8-quoted syntax and a SEARCH CHARSET other than UTF-8. If an IMAP server receives such a SEARCH command, it SHOULD reject the command with a BAD response (due to the conflicting charset labels).

### **[3.2.](#) UTF8 Parameter to SELECT and EXAMINE**

The "UTF8=ACCEPT" capability also indicates that the server supports the "UTF8" parameter to SELECT and EXAMINE. When a mailbox is selected with the "UTF8" parameter, it alters the behavior of all IMAP commands related to message sizes, message headers, and MIME body headers so they refer to the message with UTF-8 headers.

Servers MAY include mailboxes that can only be selected or examined if the "UTF8" parameter is provided. However, such mailboxes MUST NOT be included in the output of LIST, LSUB, or equivalent command unless the "UTF8" parameter is present. If a client attempts to SELECT or EXAMINE such mailboxes without the "UTF8" parameter, the server MUST reject the command with a [UTF-8-ONLY] response code. As a result, such mailboxes will not be accessible by IMAP clients written prior to this specification and are discouraged unless the server advertises "UTF8=ONLY".

utf8-select-param = "UTF8" ;; Conforms to select-param from [RFC 4466](#)

C: a SELECT newmailbox (UTF8)

S: ...

S: a OK SELECT completed

C: b FETCH 1 (SIZE ENVELOPE BODY)



S: ... UTF-8 header native results

S: b OK FETCH completed

C: c EXAMINE legacymailbox (UTF8)

S: c NO [NOT-UTF-8] Mailbox does not support UTF-8 access

C: d SELECT funky-new-mailbox

S: d NO [UTF-8-ONLY] Mailbox requires UTF-8 client

### **[3.3.](#) UTF-8 LIST and LSUB Responses**

After an IMAP client successfully issues an "ENABLE UTF8=ACCEPT" command, the server MUST NOT return in LIST results any mailbox names to the client following the IMAP4 Mailbox International Naming Convention. Instead, the server MUST return any mailbox names with characters using either utf8-quoted or literal syntax. (The IMAP4 Mailbox International Naming Convention has proved problematic in the past, so the desire is to make this syntax obsolete as quickly as possible.)

### **[3.4.](#) UTF-8 Interaction with IMAP4 LIST Command Extensions**

When an IMAP server advertises the "UTF8=ACCEPT" capability the server MUST implement a "UTF8" LIST selection option and LIST return option. These options are compatible with the "LIST-EXTENDED" [[RFC5258](#)] capability.

When an IMAP server advertises both the "UTF8=ACCEPT" capability and the "LIST-EXTENDED" [[RFC5258](#)] capability, the server MUST support the LIST extensions described in this section.

#### **[3.4.1.](#) UTF8 LIST Selection Option**

The "UTF8" LIST selection option tells the server to include mailboxes that only support UTF-8 headers in the output of the list command. Use of this selection option will also result in UTF-8 mailbox names in the result as described in [Section 3.3](#) and implies the "UTF8" List return option described in [Section 3.4.2](#).



### **3.4.2. UTF8 LIST Return Option**

If the client supplies the "UTF8" LIST return option, then the server MUST include either the "\NoUTF8" or the "\UTF8Only" mailbox attribute as appropriate. The "\NoUTF8" mailbox attribute indicates that an attempt to SELECT or EXAMINE that mailbox with the "UTF8" parameter will fail with a [NOT-UTF-8] response code. The "\UTF8Only" mailbox attribute indicates that an attempt to SELECT or EXAMINE that mailbox without the "UTF8" parameter will fail with a [UTF-8-ONLY] response code. Note that computing this information may be expensive on some server implementations, so this return option should not be used unless necessary.

The ABNF [[RFC5234](#)] for these LIST extensions follows:

```
List-select-independent-opt =/ "UTF8"
    ; List-select-independent-opt is defined in RFC 5258 Section 6

list-select-base-opt        =/ "UTF8ONLY"
    ; list-select-base-opt is defined in RFC 5258 Section 6

return-option               =/ "UTF8"
    ; return-option is defined in RFC 5258 Section 6

mbx-list-oflag              =/ "\NoUTF8" / "\UTF8Only"
    ; mbx-list-oflag is defined in RFC 3501 Section 9

resp-text-code              =/ "NOT-UTF-8" / "UTF-8-ONLY"
    ; resp-text-code is defined in RFC 3501 Section 9
```

In the event that the server advertises UTF8=ACCEPT and does not also advertise LIST-EXTENDED [[RFC 5258](#)], then the first four of the above five ABNF rules are replaced with the following ABNF:



```
list =      "LIST" [SP list-select-opts] SP mailbox SP list-mailbox
           [SP list-return-opts]
           ; replaces "list" rule in RFC 3501 Section 9

list-select-opts = "(" [*(list-select-opt SP) list-select-opt] ")"

list-select-opt  = "SUBSCRIBED" / "UTF8"

list-return-opts = "RETURN" SP
                  "(" [return-option *(SP return-option)] ")"

return-option = "SUBSCRIBED" / "UTF8"

mbx-list-sflag =/ "\NonExistent"

mbx-list-oflag =/ "\NoUTF8" / "\UTF8Only" / "\Subscribed"
                ; mbx-list-oflag is defined in RFC 3501 Section 9
```

In this case, the "SUBSCRIBED" LIST select option and "SUBSCRIBED" LIST return option work as described in LIST Command Extensions [RFC 5258]. Any IMAP client that uses LIST selection options or LIST return options (including "UTF8" and "SUBSCRIBED") MUST be able to interoperate with a server that advertises both LIST-EXTENDED and UTF8=ACCEPT. Such servers MAY include LIST response extensions as described in LIST Command Extensions [[RFC 5258](#)].

#### **4. IMAP UTF8 Append Data Extension**

If the "UTF8=ACCEPT" capability is advertised, then the server accepts UTF-8 headers in the APPEND command message argument. A client that sends a message with UTF-8 headers to the server MUST send them using the "UTF8" APPEND data extension. If the server also advertises the CATENATE capability (as specified in [[RFC4469](#)]), the client can use the same data extension to include such a message in a CATENATE message part. The ABNF for the APPEND data extension and CATENATE extension follows:

```
utf8-literal    = "UTF8" SP "(" literal8 ")"

append-data     =/ utf8-literal

cat-part        =/ utf8-literal
```

A server that advertises "UTF8=ACCEPT" MAY fail for \NotUTF8 mailboxes with a NOT-UTF-8 response code. If this command does not fail, it MAY follow the requirements of the IMAP base specification and [[RFC5322](#)] for message fetching. Mechanisms for 7-bit downgrading to help comply with the standards are discussed in



[\[popimap-downgrade\]](#).

IMAP servers that advertise support for "UTF8=ACCEPT" or "UTF8=ONLY" MUST reject an APPEND command that includes any 8-bit in the message headers with a "NO" response, when IMAP clients do not issue "ENABLE UTF8=ACCEPT" or "ENABLE UTF8=ONLY".

Note that the "UTF8=ONLY" capability described in [Section 6](#) implies the "UTF8=ACCEPT" capability. See additional information in that section.

## **5. LOGIN Command and UTF-8**

This specification doesn't extend the IMAP LOGIN command [\[RFC3501\]](#) to support UTF-8 usernames and passwords. Whenever a client needs to use UTF-8 username/passwords, it MUST use the IMAP AUTHENTICATE command which is already capable of passing UTF-8 user names and credentials.

Although this makes it syntactically legal to have a UTF-8 user name or password, there is no guarantee the user provisioning system used by the IMAP server will allow such identities. This is an implementation decision and MAY depend on what identity system the IMAP server is configured to use.

## **6. UTF8=ONLY Capability**

The "UTF8=ONLY" capability permits an IMAP server to advertise that it does not support the international mailbox name convention (modified UTF-7), and does not permit selection or examination of any mailbox unless the "UTF8" parameter is provided. As this is an incompatible change to IMAP, a clear warning is necessary. IMAP clients that find implementation of the "UTF8=ONLY" capability problematic are encouraged to at least detect the "UTF8=ONLY" capability and provide an informative error message to the end-user.

The "UTF8=ONLY" capability implies the "UTF8=ACCEPT" capability. UTF8=ACCEPT and UTF8=ONLY SHOULD be mutually exclusive. An IMAP server can advertise one of them, but never both.

## **7. Issues with UTF-8 Header Mailstore**

When an IMAP server uses a mailbox format that supports UTF-8 headers and it permits selection or examination of that mailbox without the "UTF8" parameter, it is the responsibility of the server to comply with the IMAP4rev1 base specification [\[RFC3501\]](#) and [\[RFC5322\]](#) with respect to all header information transmitted over the wire. Mechanisms for 7-bit downgrading to help comply with the standards



are discussed in [[popimap-downgrade](#)].

## 8. IANA Considerations

This document adds two new capabilities ("UTF8=ACCEPT" and "UTF8=ONLY") to the IMAP4rev1 Capabilities registry [[RFC3501](#)]. Three other IMAP capabilities that were described in the experimental predecessor to this document (UTF8=ALL, UTF8=APPEND, UTF8=USER) are to be marked OBSOLETE in the registry.

This document adds two new IMAP4 list selection options and one new IMAP4 list return option.

### 1. LIST-EXTENDED option name: UTF8

LIST-EXTENDED option type: SELECTION

Implied return options(s): UTF8

LIST-EXTENDED option description: Causes the LIST response to include mailboxes that mandate the UTF8 SELECT/EXAMINE parameter.

Published specification: RFC 5738bis, [Section 3.4.1](#)

Security considerations: RFC 5738bis, [Section 9](#)

Intended usage: COMMON

Person and email address to contact for further information: see the Authors' Addresses at the end of this specification

Owner/Change controller: iesg@ietf.org

### 2. LIST-EXTENDED option name: UTF8ONLY

LIST-EXTENDED option type: SELECTION

Implied return options(s): UTF8

LIST-EXTENDED option description: Part of previous experiment, no longer used.

Published specification: RFC 5738bis, [Section 3.4.1](#)

Security considerations: RFC 5738bis, [section 9](#)



Intended usage: OBSOLETE

Person and email address to contact for further information: see the Authors' Addresses at the end of this specification

Owner/Change controller: iesg@ietf.org

### 3. LIST-EXTENDED option name: UTF8

LIST-EXTENDED option type: RETURN

Implied return options(s): none

LIST-EXTENDED option description: Causes the LIST response to include \NoUTF8 and \UTF8only mailbox attributes.

Published specification: RFC 5738bis, [Section 3.4.1](#)

Security considerations: RFC 5738bis, [section 9](#)

Intended usage: COMMON

Person and email address to contact for further information: see the Authors' Addresses at the end of this specification

Owner/Change controller: iesg@ietf.org

## 9. Security Considerations

The security considerations of UTF-8 [[RFC3629](#)] and SASLprep [[RFC4013](#)] apply to this specification, particularly with respect to use of UTF-8 in user names and passwords. Otherwise, this is not believed to alter the security considerations of IMAP4rev1.

This document does not address downgrading scenarios, the security issues are discussed in [[popimap-downgrade](#)]

## 10. References

### 10.1. Normative References

- |           |  |
|-----------|--|
| [RFC2119] | Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", <a href="#">BCP 14</a> , <a href="#">RFC 2119</a> , March 1997. |
| [RFC3501] | Crispin, M., "INTERNET MESSAGE ACCESS PROTOCOL - VERSION 4rev1", <a href="#">RFC 3501</a> , March 2003.                                  |



- [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.
- [RFC4466] Melnikov, A. and C. Daboo, "Collected Extensions to IMAP4 ABNF", [RFC 4466](#), April 2006.
- [RFC4469] Resnick, P., "Internet Message Access Protocol (IMAP) CATENATE Extension", [RFC 4469](#), April 2006.
- [RFC5161] Gulbrandsen, A. and A. Melnikov, "The IMAP ENABLE Extension", [RFC 5161](#), March 2008.
- [RFC5198] Klensin, J. and M. Padlipsky, "Unicode Format for Network Interchange", [RFC 5198](#), March 2008.
- [RFC5234] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, [RFC 5234](#), January 2008.
- [RFC5258] Leiba, B. and A. Melnikov, "Internet Message Access Protocol version 4 - LIST Command Extensions", [RFC 5258](#), June 2008.
- [RFC5322] Resnick, P., Ed., "Internet Message Format", [RFC 5322](#), October 2008.
- [I-D.ietf-eai-rfc5335bis] Yang, A., Steele, S., and N. Freed, "Internationalized Email Headers", [draft-ietf-eai-rfc5335bis-13](#) (work in progress), October 2011.

## **[10.2.](#) Informative References**

- [RFC2088] Myers, J., "IMAP4 non-synchronizing literals", [RFC 2088](#), January 1997.
- [RFC5738] Resnick, P. and C. Newman, "IMAP Support for UTF-8", [RFC 5738](#), March 2010.
- [popimap-downgrade] Fujiwara, K., "Post-delivery Message



Downgrading for Internationalized Email Messages",  
[draft-ietf-eai-popimap-downgrade-03](#) (work in progress), October 2010.

## **Appendix A. Appendix A. Design Rationale**

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

The basic approach of advertising the ability to access a mailbox in UTF-8 mode is intended to permit graceful upgrade, including servers that support multiple mailbox formats. In particular, it would be undesirable to force conversion of an entire server mailstore to UTF-8 headers, so being able to phase-in support for new mailboxes and gradually migrate old mailboxes is permitted by this design.

The "UTF8=ONLY" mechanism simplifies diagnosis of interoperability problems when legacy support goes away. In the situation where backwards compatibility is broken anyway, just-send-UTF-8 IMAP has the advantage that it might work with some legacy clients. However, the difficulty of diagnosing interoperability problems caused by a just-send-UTF-8 IMAP mechanism is the reason the "UTF8=ONLY" capability mechanism was chosen.

## **Appendix B. Appendix B. Acknowledgments**

The authors wish to thank the participants of the EAI working group for their contributions to this document with particular thanks to Harald Alvestrand, David Black, Randall Gellens, Arnt Gulbrandsen, Kari Hurtt, John Klensin, Xiaodong Lee, Charles Lindsey, Alexey Melnikov, Subramanian Moonesamy, Shawn Steele, Daniel Taharlev, and Joseph Yee for their specific contributions to the discussion.

### **Authors' Addresses**

Pete Resnick (editor)  
Qualcomm Incorporated  
5775 Morehouse Drive  
San Diego, CA 92121-1714  
US

Phone: +1 858 651 4478  
EMail: [presnick@qualcomm.com](mailto:presnick@qualcomm.com)



Chris Newman (editor)

Oracle

800 Royal Oaks

Monrovia, CA 91016

USA

Phone:

E-Mail: [chris.newman@oracle.com](mailto:chris.newman@oracle.com)

Sean Shen (editor)

CNNIC

No.4 South 4th Zhongguancun Street

Beijing, 100190

China

Phone: +86 10-58813038

E-Mail: [shenshuo@cnnic.cn](mailto:shenshuo@cnnic.cn)

