

mailmaint  
Internet-Draft  
Intended status: Standards Track  
Expires: 8 January 2026

A. Gulbrandsen  
ICANN  
J. Yao  
CNNIC  
7 July 2025

SMTPUTF8 address syntax  
draft-ietf-mailmaint-smtputf8-syntax-01

## Abstract

This document specifies rules for email addresses that are flexible enough to express the addresses typically used with SMTPUTF8, while avoiding confusing or risky elements.

This is one of a pair of documents: This is simple to implement, contains only globally viable rules and is intended to be usable for software such as an MTA. Its companion defines has more complex rules, takes regional usage into account and aims to allow only addresses that are readable and cut-and-pastable in some community.

## Discussion Venues

This note is to be removed before publishing as an RFC.

Discussion of this document takes place on the Mail Maintenance Working Group mailing list ([mailmaint@ietf.org](mailto:mailmaint@ietf.org)), which is archived at <https://mailarchive.ietf.org/arch/browse/mailmaint/>.

Source for this draft and an issue tracker can be found at <https://github.com/arnt/mailmaint-smtputf8>.

## 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 <https://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 8 January 2026.

## Copyright Notice

Copyright (c) 2025 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 (<https://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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

## Table of Contents

1. Introduction . . . . .	2
2. Requirements Language . . . . .	3
3. Terminology . . . . .	3
4. Rules . . . . .	4
5. Examples . . . . .	4
6. IANA Considerations . . . . .	5
7. Security Considerations . . . . .	5
8. References . . . . .	5
8.1. Normative References . . . . .	5
8.2. Informative References . . . . .	6
Appendix A. Acknowledgments . . . . .	7
Appendix B. Instructions to the RFC editor . . . . .	7
Appendix C. Open issues . . . . .	7
Authors' Addresses . . . . .	7

## 1. Introduction

[RFC6530]-[RFC6533] and [RFC6854]-[RFC6858] extend various aspects of the email system to support non-ASCII both in localparts and domain parts. In addition, some email software supports unicode in domain parts by using encoded domain parts in the SMTP transaction ("RCPT TO:info@xn--dmi-0na.fo (mailto:info@xn--dmi-0na.fo)") and presenting the unicode version (d淡mi.fo in this case) in the user interface.

The email address syntax extension is in [RFC6532], and allows almost all UTF8 strings as localparts. While this certainly allows everything users want to use, it is also flexible enough to allow many things that users and implementers find surprising and sometimes worrying.

The flexibility has caused considerable reluctance to support the full syntax in contexts such as web form address validation.

This document attempts to describe rules that:

1. includes the addresses that users generally want to use for themselves and organizations want to provision for their employees.
2. excludes things that have been described as security risks.
3. Looks safe at first glance to implementers (including ones with little unicode expertise) and are fairly easy to use in unit tests.
4. Contain no regional rules.

These goals are somewhat aspirational.

## 2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## 3. Terminology

Script, in this document, refers to the unicode script property (see [UAX24]). Each code point is assigned to one script ("a" is Latin), except that some are assigned to "Common" or a few other special values. Fraktur and /etc/rc.local aren't scripts in this document, but Latin is.

Latin refers those code points that have the script property "Latin" in Unicode. Orléans in France and Münster in Germany both have Latin names in this document. It also refers to combinations of those code points and combining characters, and to strings that contain no code points from other scripts.

Han, Cyrillic etc. refer to those code points that have the respective script property in Unicode, as well as to strings that contain no code points from other scripts.

ASCII refers to the first 128 code points within unicode, which includes the letters A-Z but not or . It also refers to strings that contain only ASCII code points.

Non-ASCII refers to unicode code points except the first 128, and also to strings that contain at least one such code point.

By way of example, the address `info@dmí.fo` is latin and non-ASCII, its localpart is latin and ASCII, and its domain part is latin and non-ASCII. 中国 is a Han string in this document, but 阿Q正 is neither a Latin string nor a Han string, because it contains a Latin Q and three Han code points.

#### 4. Rules

Based on the above goals, the following rules are formulated:

1. An address **MUST NOT** contain an a-label (e.g. `xn--dmí-0na`).
2. An address **MUST** contain only code points in the PRECIS IdentifierClass.
3. An address **MUST** consist entirely of a sequence of composite characters, ZWJ and ZWNJ. ("c" followed by "combining hook below" is an example of a composite character, "d" is another example; see [RFC6365] for the definition.)
4. An address **MOT NOT** contain more than one script, disregarding ASCII. (Disregarding ASCII, the word Orlans contains only an , which is one script, namely Latin.)

#### 5. Examples

`example@example.com` is permitted, because 1) it does not contain any a-label, 2) it consists entirely of permissible code points, 4) it consists of 19 composite characters, and 4) it contains no non-ASCII code points at all.

The address `dmí@dmí.fo` is permitted, because 1) it does not contain any a-label, 2) does not apply, 3) it consists entirely of permissible code points, 4) it consists of 12 composite characters, 5) does not apply and 6) it consists entirely of 'Latin' and 'Common' code points (and ./@).

The address `U+200E '@' U+200F '.'` U+200E is not permitted, because 4) U+200E and U+200F are not parts of composite characters.

阿Q正@阿Q正.example is permitted because it contains ASCII and Han, `dmí@dmí.fo` is legal because it contains ASCII and Latin, but 阿Q正@dmí.fo is illegal because it contains Han 阿 and the Latin non-ASCII letter .

TODO: add more examples and rationales again.

## 6. IANA Considerations

This document does not require any actions from the IANA.

## 7. Security Considerations

When a program renders a unicode string on-screen or audibly and includes a substring supplied by a potentially malevolent source, the included substring can affect the rendering of a surprisingly large part of the overall string.

This document describes rules that make it difficult for an attacker to use email addresses for such an attack. Implementers should be aware of other possible vectors for the same kind of attack, such as subject fields and email address display-names.

If an address is signed using DKIM and (against the rules of this document) mixes left-to-right and right-to-left writing, parts of both the localpart and the domain part can be rendered on the same side of the '@'. This can create the appearance that a different domain signed the message.

The rules in this document permit a number of code points that can make it difficult to cut and paste.

## 8. References

### 8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/rfc/rfc2119>>.
- [RFC5322] Resnick, P., Ed., "Internet Message Format", RFC 5322, DOI 10.17487/RFC5322, October 2008, <<https://www.rfc-editor.org/rfc/rfc5322>>.
- [RFC6365] Hoffman, P. and J. Klensin, "Terminology Used in Internationalization in the IETF", BCP 166, RFC 6365, DOI 10.17487/RFC6365, September 2011, <<https://www.rfc-editor.org/rfc/rfc6365>>.
- [RFC6530] Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", RFC 6530, DOI 10.17487/RFC6530, February 2012, <<https://www.rfc-editor.org/rfc/rfc6530>>.

- [RFC6532] Yang, A., Steele, S., and N. Freed, "Internationalized Email Headers", RFC 6532, DOI 10.17487/RFC6532, February 2012, <<https://www.rfc-editor.org/rfc/rfc6532>>.
- [RFC6533] Hansen, T., Ed., Newman, C., and A. Melnikov, "Internationalized Delivery Status and Disposition Notifications", RFC 6533, DOI 10.17487/RFC6533, February 2012, <<https://www.rfc-editor.org/rfc/rfc6533>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/rfc/rfc8174>>.

## 8.2. Informative References

- [RFC3490] Faltstrom, P., Hoffman, P., and A. Costello, "Internationalizing Domain Names in Applications (IDNA)", RFC 3490, DOI 10.17487/RFC3490, March 2003, <<https://www.rfc-editor.org/rfc/rfc3490>>.
- [RFC5891] Klensin, J., "Internationalized Domain Names in Applications (IDNA): Protocol", RFC 5891, DOI 10.17487/RFC5891, August 2010, <<https://www.rfc-editor.org/rfc/rfc5891>>.
- [RFC6854] Leiba, B., "Update to Internet Message Format to Allow Group Syntax in the "From:" and "Sender:" Header Fields", RFC 6854, DOI 10.17487/RFC6854, March 2013, <<https://www.rfc-editor.org/rfc/rfc6854>>.
- [RFC6858] Gulbrandsen, A., "Simplified POP and IMAP Downgrading for Internationalized Email", RFC 6858, DOI 10.17487/RFC6858, March 2013, <<https://www.rfc-editor.org/rfc/rfc6858>>.
- [UAX24] Whistler, K., "Unicode Script Property", n.d., <<https://unicode.org/reports/tr24>>.
- [UMLAUT] "Metal Umlaut", n.d., <[https://en.wikipedia.org/wiki/Metal\\_umlaut](https://en.wikipedia.org/wiki/Metal_umlaut)>.
- [TYPE\_EMAIL] "WHATWG input type=email", n.d., <[https://html.spec.whatwg.org/multipage/input.html#email-state-\(type=email\)](https://html.spec.whatwg.org/multipage/input.html#email-state-(type=email))>.

## Appendix A. Acknowledgments

The authors wish to thank John C. Klensin, (your name here, please)

(Wow, the ack section is already outdated)

Dmi.fo and 例子.中国 are reserved by nic.fo and CNNIC for use in examples and documentation.

阿Q正@ is a famous Chinese novella, 阿Q is the main character.

## Appendix B. Instructions to the RFC editor

Please remove all mentions of the Protocol Police before publication (including this sentence).

Please remove the Open Issues section.

## Appendix C. Open issues

1. PRECIS IdentifierClass?
2. More examples.
3. Wording to identify destiny; I think this should probably become a proposed standard and modify a couple of RFCs, but I'm uncertain about some details and left that open now.
4. More words on the relationship between this and the companion. There are several parallel differences, maybe this warrants a section of its own.
5. Should this even mention the requirements placed on domains by IDNA, ICANN, web browsers and others?

## Authors' Addresses

Arnt Gulbrandsen  
ICANN  
6 Rond Point Schumann, Bd. 1  
1040 Brussels  
Belgium  
Email: arnt@gulbrandsen.priv.no

Jiankang Yao  
CNNIC  
No.4 South 4th Zhongguancun Street

Beijing  
100190  
China  
Email: yaojk@cnnic.cn