Network Working Group Internet-Draft Intended status: Informational Expires: August 18, 2014 H. Flanagan RFC Editor February 14, 2014

The Use of Non-ASCII Characters in RFCs draft-flanagan-nonascii-00

Abstract

This document lays out the requirements regarding the use of non-ASCII characters in RFCs. It includes examples for the different sections of an RFC.

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 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 August 18, 2014.

Copyright Notice

Copyright (c) 2014 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

l.	Inti	roductio	n.	•		•		•			•			•	•	•		•							•	2
2.	Basi	ic requi	reme	nts																						2
3.	Rule	es for t	he u	se (of	no	n-2	ASC	CII	Ε (cha	ara	act	er	ŝ											3
3	.1.	General	usa	ge ·	thr	ou	gh	out	t a	a c	oot	cun	ner	ıτ	•	•	•	•	•	•	•	•		•	•	3
3	.2.	Author	Name	s																						3
3	.3.	Body of	the	do	cun	en.	t		•		•			•	•	•	•	•	•	•	•	•		•	•	4
3	.4.	Tables		•					•		•			•	•	•	•	•	•	•	•	•		•	•	5
3	.5.	Code co	mpon	ent	s.																					5
3	.6.	Bibliog	raph	ic ·	tex	t	•	•	•		•		•	•		•		•	•		•		•	•		5
3	.7.	Keyword	s.	•			•	•	•		•		•	•		•		•	•		•		•	•		6
1.	Norr	nalizati	on F	orm	s.		•	•	•		•		•	•		•		•	•		•		•	•		6
5.	IANA	A Consid	erat	ion	s.		•	•	•		•		•	•		•		•	•		•		•	•		6
5.	Inte	ernation	aliz	ati	on	Co	ns:	ide	era	ati	Lor	ns	•	•		•		•	•		•		•	•		6
7.	Secu	ırity Co	nsid	era	tic	ns	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	6
3.	Ackr	nowledge	ment	s			•	•	•		•		•	•		•		•	•		•		•	•		6
9.	Norr	native R	efer	enc	es		•	•	•		•		•	•		•		•	•		•		•	•		7
10.	Info	ormative	Ref	ere	nce	S	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	7
11.	Refe	erences		•			•	•	•		•		•	•		•		•	•		•		•	•		7
Aut	hor's	s Addres	s.																							7

1. Introduction

For much of the history of the RFC Series, the character encoding used has been US-ASCII [ASCII]. This was a sensible choice at the time: the language of the Series is English, a language which only uses US-ASCII-encoded characters (ignoring for a moment words borrowed from more richly decorated alphabets). US-ASCII is the "lowest common denominator" for character encoding, making crossplatform viewing trivial.

There are limits to US-ASCII, however, which hinder its continued use as the exclusive character encoding for the Series. The increasing need for easily readable, internationalized content suggests it is time to allow non-ASCII characters in RFCs where necessary. Given the continuing goal of maximum readability across platforms, the use of non-ASCII should not be gratuitous in a document. This RFC describes the rules under which non-ASCII characters may be used in an RFC. These rules will be applied as the necessary changes are made to submission checking and editorial tools.

2. Basic requirements

Three fundamental requirements inform the guidance and examples provided in this document. They are:

- o Searches of an index need to be able to find multiple ways of writing an author's name;
- o People whose system does not have the fonts needed to display a particular RFC need to be able to read the non-canonical HTML, text, or PDF RFC correctly.
- 3. Rules for the use of non-ASCII characters
- 3.1. General usage throughout a document

The mention, as opposed to the use, of non-ASCII characters requires the use of Unicode identifiers. Including non-ASCII characters in the text is encouraged to make the mention clearer to readers with devices that can render the non-ASCII text. Including Unicode character names is allowed.

The distinction is between an occasion in which a word appears in a sentence in order to convey the meaning of that word (use), and an occasion in which a word appears in order to make some point about that word itself (mention).

Spelling for words commonly used in the English language will follow the guidance in the Mirriam-Webster dictionary. Other uses, including archaic or untraditional spellings, are not allowed.

Example:

Use:

SIP/2.0 200 = 2**3 * 5**2 но сто девяносто девять - простое Via: SIP/2.0/UDP 192.0.2.198;branch=z9hG4bK1324923

Call-ID: unreason.1234ksdfak3j2erwedfsASdf

CSeq: 35 INVITE

From: sip:user@example.com;tag=11141343

To: sip:user@example.edu;tag=2229

Content-Length: 154

Content-Type: application/sdp

Mention:

For example, the characters "STFETER" (U+13DA U+13A2 U+13B5 U+13AC U+13A2 U+13AC U+13D2) from the Cherokee block look similar to the ASCII characters "STPETER" as they might look when presented using a "creative" font family.

3.2. Author Names

Valid Unicode is required, and for non-ASCII names, an ASCII-only identifier is required.

Example for the header:

Network Working Group

Request for Comments: 2611

BCP: 33

Category: Best Current Practice

ISIS/CEO, JRC Ispra

R. Iannella

DSTC Pty Ltd

P. Fältström (P. Faltstrom)

Tele2/Swipnet

June 1999

Example for the Acknowledgements:

OLD:

The following people contributed significant text to early versions of this draft: Patrik Faltstrom, William Chan, and Fred Baker.

PROPOSED/NEW:

The following people contributed significant text to early versions of this draft: Patrik Fältström (Patrik Faltstrom), 陈智昌 (William Chan), and Fred Baker.

3.3. Body of the document

Non-ASCII characters may be used only where necessary to clearly express the intended meaning of the text. When such characters are mentioned in the text, the following rules apply:

- o Non-ASCII characters will require identifying the codepoint (e.g. U+0394)
- o Use of the actual UTF-8 character (e.g., Δ) is encouraged so that a reader can more easily see what the character is, if their device can render the text.
- o If preferred by the author(s), the use of the official character names like "Greek Capital Letter Delta" is allowed.

Examples:

OLD (draft-ietf-precis-framework):

However, the problem is made more serious by introducing the full range of Unicode code points into protocol strings. For example, the characters U+13DA U+13A2 U+13B5 U+13AC U+13A2 U+13AC U+13D2 from the Cherokee block look similar to the ASCII characters "STPETER" as they might look when presented using a "creative" font family.

NEW/ALLOWED:

However, the problem is made more serious by introducing the full range of Unicode code points into protocol strings. For example, the characters U+13DA U+13A2 U+13B5 U+13AC U+13A2 U+13AC U+13D2 (STFETER) from the Cherokee block look similar to the ASCII characters "STPETER" as they might look when presented using a "creative" font family.

ALSO ACCEPTABLE:

However, the problem is made more serious by introducing the full range of Unicode code points into protocol strings. For example, **the characters** "STPETER" (U+13DA U+13A2 U+13B5 U+13AC U+13A2 U+13AC U+13D2) from the Cherokee block look similar to the ASCII characters "STPETER" as they might look when presented using a "creative" font family.

3.4. Tables

Tables follow the same rules for identifiers and characters as the body. If it is sensible (i.e., more understandable for a reader) for a given document to have two tables, one including the identifiers and characters, one with just the characters, that will be allowed on a case by case basis.

Example: TBD

3.5. Code components

Use the U+ notation except within a code component where you must follow the rules of the programming language in which you are writing the code

Example:

TBD

3.6. Bibliographic text

The reference entry must be in English; whatever subfields are present must be available in ASCII. As long as good sense is used, they may also include non-ASCII characters at author discretion. This applies to both normative and informative references.

Example:

[GOST3410] "Information technology. Cryptographic data security. Signature and verification processes of [electronic] digital signature.", GOST R 34.10-2001, Gosudarstvennyi Standard of Russian Federation, Government Committee of Russia for Standards, 2001. (In Russian)

Allowable addition to the above citation:

"Информационная технология. Криптографическая защита информации. Процессы формирования и проверки электронной цифровой подписи ", GOST R 34.10-2001, Государственный стандарт Российской Федерации, 2001.

3.7. Keywords

Keywords must be US-ASCII only.

4. Normalization Forms

If the normalization matters to the content, the authors must submit in a normalization-resistant form. In other words, authors should not expect normalization forms to be preserved.

5. IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an $\ensuremath{\mathsf{RFC}}$.

6. Internationalization Considerations

TBD

- 7. Security Considerations
- 8. Acknowledgements

With many thanks to the members of the IAB i18n program.

Internet-Draft Non-ASCII RFCs February 2014

9. Normative References

10. Informative References

[ASCII] American National Standard for Information Systems - Coded Character Sets - 7-Bit American National Standard Code for Information Interchange (7-Bit ASCII), ANSI X3.4- 1986, American National Standards Institute, Inc., March 26, 1986.

11. References

Author's Address

Heather Flanagan RFC Editor

Email: rse@rfc-editor.org