

Network Working Group
Internet-Draft
Intended status: Informational
Expires: January 1, 2015

Y. YONEYA
JPRS
T. Nemoto
Keio University
June 30, 2014

**Mapping characters for PRECIS classes
draft-ietf-precis-mappings-08**

Abstract

The framework for preparation and comparison of internationalized strings ("PRECIS") defines several classes of strings for preparation and comparison. Case mapping is defined because many protocols perform case-sensitive or case-insensitive string comparison and so preparation of the string is mandatory. The Internationalized Domain Names in Applications (IDNA) and the PRECIS problem statement describes mappings for internationalized strings that are not limited to case, but include width mapping and mapping of delimiters and other specials that can be taken into consideration. This document provides guidelines for authors of protocol profiles of the PRECIS framework and describes several mappings that can be applied between receiving user input and passing permitted code points to internationalized protocols. The mappings described here are expected to be applied as an additional mapping and locale-/context-dependent case mapping.

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 January 1, 2015.

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

1.	Introduction	3
2.	Protocol dependent mappings	3
2.1.	Delimiter mapping	3
2.2.	Special mapping	4
2.3.	Local case mapping	4
3.	Order of operations	5
4.	Security Considerations	5
5.	IANA Considerations	5
6.	Acknowledgment	5
7.	References	6
7.1.	Normative References	6
7.2.	Informative References	6
Appendix A.	Mapping type list each protocol	7
A.1.	Mapping type list for each protocol	7
Appendix B.	The reason why local case mapping is alternative to case mapping in PRECIS framework	8
Appendix C.	Limitation to local case mapping	8
Appendix D.	Change Log	9
D.1.	Changes since -00	9
D.2.	Changes since -01	9
D.3.	Changes since -02	9
D.4.	Changes since -03	10
D.5.	Changes since -04	10
D.6.	Changes since -05	10
D.7.	Changes since -06	10
D.8.	Changes since -07	11
	Authors' Addresses	11

1. Introduction

In many cases, user input of internationalized strings is generated through the use of an input method editor ("IME") or through copy-and-paste from free text. Users generally do not care about the case and/or width of input characters because they consider those characters to be functionally equivalent or visually identical. Furthermore, users rarely switch the IME state to input special characters such as protocol elements. For Internationalized Domain Names ("IDNs"), the IDNA Mapping specification [[RFC5895](#)] describes methods for handling these issues. For PRECIS strings, case mapping and width mapping are defined in the PRECIS framework specification [[I-D.ietf-precis-framework](#)]. Further, the handling of mappings other than case and width, such as delimiter, special, and local case, are also important in order to increase the probability that strings match as users expect. This document provides guidelines for authors of protocol profiles of the PRECIS framework and describes several mappings that can be applied between receiving user input and passing permitted code points to internationalized protocols. The delimiter mapping and special mapping rules described here are applied as "additional mappings" beyond those defined in the PRECIS framework, whereas the "local case mapping" rule provides an alternative to the case mapping rule specified in the PRECIS framework since it handles some locale-dependent and context-dependent mappings.

2. Protocol dependent mappings

The PRECIS framework defines several protocol-independent mappings. The additional mappings and local case mapping defined in this document are protocol-dependent, i.e., they depend on the rules for a particular application protocol.

2.1. Delimiter mapping

Some application protocols define delimiters for their own use, resulting in the fact that the delimiters are different for each protocol. The delimiter mapping table should therefore be based on a well-defined mapping table for each protocol.

Delimiter mapping is used to map characters that are similar to protocol delimiters into the canonical delimiter characters. For example, there are width-compatible characters that correspond to the '@' in email addresses and the ':' and '/' in URIs. The '+', '-', '<' and '>' characters are other common delimiters that might require such mapping. For the FULL STOP character (U+002E), a delimiter in the visual presentation of domain names, some IMEs produce a character such as IDEOGRAPHIC FULL STOP (U+3002) when a user types FULL STOP on the keyboard. In all these cases, the visually similar

characters that can come from user input need to be mapped to the correct protocol delimiter characters before the string is passed to the protocol.

2.2. Special mapping

Aside from delimiter characters, certain protocols have characters which need to be mapped in ways that are different from the rules specified in the PRECIS framework (e.g., mapping non-ASCII space characters to ASCII space). In this document, these mappings are called "special mappings". They are different for each protocol. Therefore, the special mapping table should be based on a well-defined mapping table for each protocol. Examples of special mapping are the following;

- o White spaces are mapped to SPACE (U+0020)
- o Some characters such as control characters are mapped to nothing (Deletion)

As examples, EAP [[RFC3748](#)], SASLprep [[RFC4013](#)], IMAP4 ACL [[RFC4314](#)] and LDAPprep [[RFC4518](#)] define the rule that some codepoints for the non-ASCII space are mapped to SPACE (U+0020).

2.3. Local case mapping

The purpose of local case mapping is to increase the probability of a matching result from the comparison between uppercase and lowercase characters, targeting characters which mapping depends on locale or locale and context.

As an example of locale and context-dependent mapping, LATIN CAPITAL LETTER I ("I", U+0049) is normally mapped to LATIN SMALL LETTER I ("i", U+0069); however, if the case of Turkish (or one of several other languages), unless an I is before a dot_{above}, the character should be mapped to LATIN SMALL LETTER DOTLESS I (U+0131).

Case mapping using Unicode Default Case Folding in PRECIS framework does not consider such locale or context because it is a common framework for internationalization. Local case mapping defined in this document corresponds to demands from applications which supports users' locale and/or context. The target characters of local case mapping are characters defined in the SpecialCasing.txt [[Specialcasing](#)] file in [section 3.13](#) of the Unicode Standard [[Unicode](#)].

The case folding method for a target character is to map into lower case as defined in SpecialCasing.txt. The case folding method for

all other, non-target characters is as specified in [Section 4.1.3](#) of the PRECIS framework (i.e., It is RECOMMENDED to use Unicode Default Case Folding for all non-target characters). If an application supports users' locale and/or context, local case mapping can increase the probability of getting matching-results from the comparison between strings.

If Unicode Default Case Folding is selected as "Case Mapping" in PRECIS profiles registry, PRECIS profile designers may consider whether local case mapping can be applied. And if it can be applied, it is better to add "local case mapping is applicable alternatively" after "Unicode Default Case Folding" for note to application developers. The reason why local case mapping is alternative to Unicode Default Case Folding is written in the [Appendix B](#).

[3.](#) Order of operations

Delimiter mapping and special mapping described in this document are expected to be applied as additional mappings in the PRECIS framework. The mappings described in this document could be applied in any order. This section specifies a particular order to minimize the effect of codepoint changes introduced by the mappings. This mapping order is very general and has been designed to be acceptable to the widest user community.

1. Delimiter mapping
2. Special mapping

[4.](#) Security Considerations

As well as Mapping Characters for IDNA2008 [[RFC5895](#)], this document suggests creating mappings that might cause confusion for some users while alleviating confusion in other users. Such confusion is not covered in any depth in this document.

[5.](#) IANA Considerations

This document has no actions for the IANA.

[6.](#) Acknowledgment

Martin Duerst suggested a need for the case folding about the mapping (map final sigma to sigma, German sz to ss,.).

Alexey Melnikov, Andrew Sullivan, Barry Leiba, Heather Flanagan, Joe Hildebrand, John Klensin, Marc Blanchet, Pete Resnick and Peter

Saint-Andre, et al. gave important suggestion for this document during at WG meeting and WG LC.

7. References

7.1. Normative References

[I-D.ietf-precis-framework]

Saint-Andre, P. and M. Blanchet, "PRECIS Framework: Preparation and Comparison of Internationalized Strings in Application Protocols", [draft-ietf-precis-framework-14](#) (work in progress), February 2014.

[Unicode] The Unicode Consortium, "The Unicode Standard, Version 6.3.0", <<http://www.unicode.org/versions/Unicode6.3.0/>>, 2012.

[Casefolding]

The Unicode Consortium, "CaseFolding-6.3.0.txt", Unicode Character Database, July 2011, <<http://www.unicode.org/Public/6.3.0/ucd/CaseFolding.txt>>, .

[Specialcasing]

The Unicode Consortium, "SpecialCasing-6.3.0.txt", Unicode Character Database, July 2011, <<http://www.unicode.org/Public/6.3.0/ucd/SpecialCasing.txt>>, .

7.2. Informative References

[RFC3454] Hoffman, P. and M. Blanchet, "Preparation of Internationalized Strings ("stringprep")", [RFC 3454](#), December 2002.

[RFC3490] Faltstrom, P., Hoffman, P., and A. Costello, "Internationalizing Domain Names in Applications (IDNA)", [RFC 3490](#), March 2003.

[RFC3491] Hoffman, P. and M. Blanchet, "Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN)", [RFC 3491](#), March 2003.

[RFC3722] Bakke, M., "String Profile for Internet Small Computer Systems Interface (iSCSI) Names", [RFC 3722](#), April 2004.

- [RFC3748] Aboba, B., Blunk, L., Vollbrecht, J., Carlson, J., and H. Levkowetz, "Extensible Authentication Protocol (EAP)", [RFC 3748](#), June 2004.
- [RFC4013] Zeilenga, K., "SASLprep: Stringprep Profile for User Names and Passwords", [RFC 4013](#), February 2005.
- [RFC4314] Melnikov, A., "IMAP4 Access Control List (ACL) Extension", [RFC 4314](#), December 2005.
- [RFC4518] Zeilenga, K., "Lightweight Directory Access Protocol (LDAP): Internationalized String Preparation", [RFC 4518](#), June 2006.
- [RFC5895] Resnick, P. and P. Hoffman, "Mapping Characters for Internationalized Domain Names in Applications (IDNA) 2008", [RFC 5895](#), September 2010.
- [RFC6122] Saint-Andre, P., "Extensible Messaging and Presence Protocol (XMPP): Address Format", [RFC 6122](#), March 2011.
- [RFC6885] Blanchet, M. and A. Sullivan, "Stringprep Revision and Problem Statement for the Preparation and Comparison of Internationalized Strings (PRECIS)", [RFC 6885](#), March 2013.
- [ISO.3166-1]
International Organization for Standardization, "Codes for the representation of names of countries and their subdivisions - Part 1: Country codes", ISO Standard 3166-1:1997, 1997.

[Appendix A.](#) Mapping type list each protocol

[A.1.](#) Mapping type list for each protocol

This table is the mapping type list for each protocol. Values marked "o" indicate that the protocol use the type of mapping. Values marked "-" indicate that the protocol doesn't use the type of mapping.

Protocol and mapping RFC	Width (NFKC)	Delimiter	Case	Special
IDNA (RFC 3490)	-	0	-	-
IDNA (RFC 3491)	0	-	0	-
iSCSI (RFC 3722)	0	-	0	-
EAP (RFC 3748)	0	-	-	0
SASL (RFC 4013)	0	-	-	0
IMAP (RFC 4314)	0	-	-	0
LDAP (RFC 4518)	0	-	0	0
XMPP (RFC 6120)	-	-	0	-

Appendix B. The reason why local case mapping is alternative to case mapping in PRECIS framework

One outstanding issue regarding full case folding for characters is, the character "LATIN SMALL LETTER SHARP S" (U+00DF) (hereinafter referred to as "eszett") becomes two "LATIN SMALL LETTER S"s (U+0073 U+0073) by performing the case mapping using Unicode Default Case Folding in the PRECIS framework. If local case mapping in this document is not an alternative to case mapping in PRECIS framework, PRECIS profile designers can select both mappings, therefore, German's eszett can not keep the locale if the case mapping in the PRECIS framework was performed after the local case mapping.

Appendix C. Limitation to local case mapping

As described in section [Section 2.3](#), target characters of local case mapping are characters defined in SpecialCasing.txt. The Unicode Standard (at least, up to version 6.3.0) does not define mappings between "GREEK SMALL LETTER SIGMA" (U+03C3) (hereinafter referred to as "small sigma") and "GREEK SMALL LETTER FINAL SIGMA" (U+03C2) (hereinafter referred to as "final sigma") depend on context. Thus, final sigma is always mapped to small sigma by local case mapping. (Cf. Followings are comments in SpecialCasing.txt.)

```
# Note: the following cases are not included, since they would
case-fold in lowercasing
# 03C3; 03C2; 03A3; 03A3; Final_Sigma; # GREEK SMALL LETTER SIGMA
# 03C2; 03C3; 03A3; 03A3; Not_Final_Sigma; # GREEK SMALL LETTER
```

Local case mapping follows Unicode definition, so mapping of small sigma and final sigma is up to the definition.

[Appendix D](#). Change Log

[D.1](#). Changes since -00

- o Modify the [Section 4.3](#) "Local case mapping" to specify the method to calculate codepoints that local case mapping targets.
- o Add the [Section 6](#) "Open issues".
- o Modify the [Section 7](#) "IANA Considerations".
- o Modify the [Section 8](#) "Security Considerations".
- o Remove the "The initial PRECIS local case mapping registrations".
- o Add the [Appendix C](#) "Code points list for local case mapping".
- o Add the [Appendix D](#) "Change Log".

[D.2](#). Changes since -01

- o Unified PRECIS notation in all capital letters as well as other documents.
- o Removed the [Section 1](#) "Types of mapping" and the [Section 2](#) "Protocol independent mapping" because width mapping is now in framework document.
- o Added relationship between the framework document and this document in the [Section 3](#) "Order of operations".
- o Updated the [Section 4](#) "Open issues" to address new issue raised on mailing list.
- o Move the [Section 6](#) "IANA Considerations" after the [Section 5](#) "Security Considerations".
- o Remove the [Appendix B](#) "Codepoints which need special mapping" and mentioned related documents in the [Section 2.2](#) .

[D.3](#). Changes since -02

- o Removed the "Open issues".

D.4. Changes since -03

- o Modify the [Section 1](#) "Introduction" in more clear text.
- o Modify the [Section 2.3](#) "Local case mapping" to clarify the purpose of the local case mapping and an example, and add restriction to use with PRECIS framework.
- o Change the format in the [Appendix B](#) "Code points list for local case mapping".
- o Split the [Section 7](#) "References" into "Normative References" and "Informative References"
- o Update the Unicode version 6.2 to 6.3 in this document.

D.5. Changes since -04

- o Correct a sentence in the [Section 2.3](#) "Local case mapping".

D.6. Changes since -05

- o Correct some sentences in this document.
- o Modify the local case mapping's rule and target characters in [Section 2.3](#) "Local case mapping". This is to avoid user's confusion towards Greek's final sigma and German's eszett.
- o Add the [Section 4](#) "Open issues".
- o Modify the [Section 8](#) "Security Considerations".
- o Modify the table format in the [Appendix A](#). "Mapping type list each protocol".
- o Removed the [Appendix B](#) "Code points list for local case mapping".
- o Add the [Appendix B](#) "Local case mapping vs Case mapping".

D.7. Changes since -06

- o Removed the [Section 4](#) "Open issues".
- o Change the title of the [Appendix B](#) "Local case mapping vs Case mapping" to "The reason why local case mapping is alternative to case mapping in PRECIS framework".
- o Add the [Appendix C](#) "Limitation to local case mapping".

D.8. Changes since -07

- o Modify the [Section 1](#) "Introduction".
- o Modify the local case mapping's rule and target characters in [Section 2.3](#) "Local case mapping".
- o Modify the [Section 3](#) "Order of operations".

Authors' Addresses

Yoshiro YONEYA

JPRS

Chiyoda First Bldg. East 13F

3-8-1 Nishi-Kanda

Chiyoda-ku, Tokyo 101-0065

Japan

Phone: +81 3 5215 8451

Email: yoshiro.yoneya@jprs.co.jp

Takahiro Nemoto

Keio University

Graduate School of Media Design

4-1-1 Hiyoshi, Kohoku-ku

Yokohama, Kanagawa 223-8526

Japan

Phone: +81 45 564 2517

Email: t.nemo10@kmd.keio.ac.jp

