

Network Working Group
Internet-Draft
Intended status: Informational
Expires: August 13, 2012

Y. YONEYA
JPRS
T. NEMOTO
Keio University
February 10, 2012

**Mapping characters for PRECIS classes
draft-yoneya-precis-mappings-01**

Abstract

Preparation and comparison of internationalized strings ("PRECIS") Framework [[I-D.ietf-precis-framework](#)] is defining several classes of strings for preparation and comparison. In the document, case mapping is defined because many of protocols handle case sensitive or case insensitive string comparison and therefore preparation of string is mandatory. As described in IDNA mapping [[RFC5895](#)] and PRECIS problem statement [[I-D.ietf-precis-problem-statement](#)], mappings in internationalized strings are not limited to case, but also width, delimiters and/or other specials are taken into consideration. This document considers mappings other than case mapping in PRECIS context.

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 13, 2012.

Copyright Notice

Copyright (c) 2012 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.

1. Introduction

In many cases, user input of internationalized strings is generated by input method editor ("IME") or copy-and-paste from free text. Usually users do not care case and/or width of input characters because they are identical for users' eyes. Further, users rarely switch IME state to input special characters such as protocol elements. For Internationalized Domain Names ("IDNs"), IDNA Mapping [[RFC5895](#)] describes methods to treat these issues. For PRECIS strings, case mapping is defined as a process in PRECIS Framework [[I-D.ietf-precis-framework](#)], but width mapping, delimiter mapping and/or special mapping are not defined. Handling of mappings other than case is also important to increase chance of strings match as users expect. This document considers such mappings in PRECIS context.

[2.](#) Type of mappings

[2.1.](#) Width mapping

Fullwidth and halfwidth characters (those defined with Decomposition Types <wide> and <narrow>) are mapped to their decomposition mappings as shown in the Unicode character database [[Unicode](#)]. This mapping should be performed before case mapping because fullwidth/halfwidth characters includes both upper case and lower case letters.

Width mapping will increase backward compatibility with Stringprep [[RFC3454](#)] and PRECIS Framework [[I-D.ietf-precis-framework](#)]. Because in a Stringprep profile which specifies Unicode normalization form KC (NFKC) for normalization method, fullwidth/halfwidth characters are mapped into its compatible form. If a PRECIS Framework profile specified NFKC (which is not recommended), width mapping might not be useful.

[2.2.](#) Delimiter mapping

Definitions of delimiters in certain protocols are differ from each other. Therefore, delimiter mapping table should be based on well defined mapping table for each protocols. This mapping should be performed after width mapping because some punctuations have fullwidth form.

One of the most useful case of delimiter mapping is when FULL STOP character (U+002E) is a delimiter as well as domain name. Some of IME generates FULL STOP compatible characters such as IDEOGRAPHIC FULL STOP (U+3002) when users type FULL STOP on the keyboard.

[2.3.](#) Special mapping

Certain protocols defined special mapping. And they are differ from each other. Therefore, special mapping table should be based on well defined mapping table for each protocols. For example, LDAPprep[RFC4518] defines the rule that some codepoints(Appendix B.4) are mapped to SPACE (U+0020). This mapping should be performed after width mapping because some punctuations have fullwidth form. But, there is no preferred order of delimiter mapping and special mapping. See [Section 3](#) for more detail.

3. Discussion

There are several points for discussion on this topic.

- o Is delimiter mapping a part of special mapping? If it is not a part of special mapping, whether order of handling about special mapping should be performed before or after delimiter mapping? Are delimiters and other specials mutually orthogonal? If they are, their order of handling is not important. But if not, how is their order of handling made?
- o Is additional case mapping considered? Does the case folding for special characters (final sigma(U+03C2), German sz(U+00DF), Turkish I with dot above(U+0130), or dotless i(U+0131) ...) need special handling?
- o Whether mappings other than case are targets of PRECIS or not? If they are target, are they a part of PRECIS Framework [[I-D.ietf-precis-framework](#)] or separate ones like IDNA Mapping [[RFC5895](#)] specification?
- o Are there another mappings not described in this document? For example, migration from Stringprep [[RFC3454](#)] to PRECIS Framework [[I-D.ietf-precis-framework](#)] needs some special treatment?

4. IANA Considerations

TBD.

5. Security Considerations

TBD.

6. Acknowledgment

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

7. 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. Levkowitz, "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.
- [I-D.ietf-precis-framework]
Blanchet, M. and P. Saint-Andre, "PRECIS Framework: Handling Internationalized Strings in Protocols", [draft-ietf-precis-framework-01](#) (work in progress), October 2011.
- [I-D.ietf-precis-problem-statement]
Blanchet, M. and A. Sullivan, "Stringprep Revision Problem Statement", [draft-ietf-precis-problem-statement-04](#) (work in progress), January 2012.

[Unicode] The Unicode Consortium, "The Unicode Standard, Version 6.1.0", <http://www.unicode.org/versions/Unicode6.1.0/>, 2012.

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

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

This table is the mapping type list for each protocols. 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.

\ Type of mapping	Width	Delimiter	Case	Special
RFC \	(NFKC)			
3490	-	o	-	-
3491	o	-	o	-
3722	o	-	o	-
3748	o	-	-	o
4013	o	-	-	o
4314	o	-	-	o
4518	o	-	o	o
6120	-	-	o	-

[Appendix B](#). Codepoints which need special mapping

[B.1](#). [RFC3748](#)

Non-ASCII space characters [StringPrep, C.1.2] that can be mapped to SPACE (U+0020).

[B.2](#). [RFC4013](#)

Non-ASCII space characters [StringPrep, C.1.2] that can be mapped to SPACE (U+0020).

[B.3](#). [RFC4314](#)

Non-ASCII space characters [StringPrep, C.1.2] that can be mapped to SPACE (U+0020).

[B.4](#). [RFC4518](#)

Codepoints mapped to SPACE (U+0020) are following;

- U+0009 (CHARACTER TABULATION)
- U+000A (LINE FEED (LF))
- U+000B (LINE TABULATION)
- U+000C (FORM FEED (FF))
- U+000D (CARRIAGE RETURN (CR))
- U+0085 (NEXT LINE (NEL))
- U+0020 (SPACE)
- U+00A0 (NO-BREAK SPACE)
- U+1680 (OGHAM SPACE MARK)
- U+2000 (EN QUAD)
- U+2001 (EM QUAD)
- U+2002 (EN SPACE)
- U+2003 (EM SPACE)
- U+2004 (THREE-PER-EM SPACE)
- U+2005 (FOUR-PER-EM SPACE)
- U+2006 (SIX-PER-EM SPACE)
- U+2007 (FIGURE SPACE)
- U+2008 (PUNCTUATION SPACE)
- U+2009 (THIN SPACE)
- U+200A (HAIR SPACE)
- U+2028 (Line Separator)
- U+2029 (Paragraph Separator)
- U+202F (NARROW NO-BREAK SPACE)
- U+205F (MEDIUM MATHEMATICAL SPACE)
- U+3000 (IDEOGRAPHIC SPACE)

All other control code (e.g., Cc) points or code points with a

control function (e.g., Cf) are mapped to nothing. Codepoints mapped to nothing that aren't specified by Stringprep are following;

U+0000-0008
U+000E-001F
U+007F-0084
U+0086-009F
U+06DD
U+070F
U+180E
U+200E-200F
U+202A-202E
U+2061-2063
U+206A-206F
U+FFF9-FFFB
U+1D173-1D17A
U+E0001
U+E0020-E007F

[Appendix C](#). Change Log

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

- o Add the [Section 2.3](#) "Special mapping" in [Section 2](#) Type of mappings.
- o Add the topic about the special mapping and additional case mapping in [Section 3](#) Discussion.
- o Add Appendices;
[Appendix A](#) "Mapping type list each protocols"
[Appendix B](#) "Code point list is need special mapping"
[Appendix C](#) "Change Log"
- o Add the [Section 6](#) Acknowledgment.

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
4-1-1 Hiyoshi, Kohoku-ku
Yokohama, Kanagawa 223-8526
Japan

Phone: +81 45 564 2517
Email: t.nemo10@kmd.keio.ac.jp

