Network Working Group Y. YONEYA

Internet-Draft JPRS

Intended status: Informational T. Nemoto

Expires: April 24, 2014 Keio University

October 21, 2013

Mapping characters for PRECIS classes

draft-ietf-precis-mappings-05

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 in

the PRECIS framework.

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 April 24, 2014.

Copyright Notice

Copyright (c) 2013 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 . . . . . . . . . . . . . . . . . . . . . . . . 2

2. Protocol dependent mappings . . . . . . . . . . . . . . . . . 3

2.1. Delimiter mapping . . . . . . . . . . . . . . . . . . . . 3

2.2. Special mapping . . . . . . . . . . . . . . . . . . . . . 3

2.3. Local case mapping . . . . . . . . . . . . . . . . . . . 4

3. Order of operations . . . . . . . . . . . . . . . . . . . . . 4

4. Security Considerations . . . . . . . . . . . . . . . . . . . 5

5. IANA Considerations . . . . . . . . . . . . . . . . . . . . . 5

6. Acknowledgment . . . . . . . . . . . . . . . . . . . . . . . 5

7. References . . . . . . . . . . . . . . . . . . . . . . . . . 5

7.1. Normative References . . . . . . . . . . . . . . . . . . 5

7.2. Informative References . . . . . . . . . . . . . . . . . 6

Appendix A. Mapping type list each protocol . . . . . . . . . . 7

A.1. Mapping type list for each protocol . . . . . . . . . . . 7

Appendix B. Code points list for local case mapping . . . . . . 7

B.1. Unicode 6.3 . . . . . . . . . . . . . . . . . . . . . . . 7

Appendix C. Change Log . . . . . . . . . . . . . . . . . . . . . 8

C.1. Changes since -00 . . . . . . . . . . . . . . . . . . . . 8

C.2. Changes since -01 . . . . . . . . . . . . . . . . . . . . 8

C.3. Changes since -02 . . . . . . . . . . . . . . . . . . . . 9

C.4. Changes since -03 . . . . . . . . . . . . . . . . . . . . 9

C.5. Changes since -04 . . . . . . . . . . . . . . . . . . . . 9

Authors' Addresses . . . . . . . . . . . . . . . . . . . . . . . 9

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]. Delimiter mapping, special mapping,

and language dependent mapping, however, are not defined. The handling of mappings

other than case and width is also important in order to increase the chance that

strings match as users expect.

This document provides guidelines for

authors of protocol profiles of the PRECIS framework and describes

mappings that can be applied between receiving user input and passing

permitted code points to internationalized protocols. The mappings

described in this document are expected to be applied as additional

mappings in the PRECIS framework.

2. Protocol dependent mappings

The PRECIS framework defines several protocol-independent mappings.

The additional mappings 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 supposed to map compatable delimiter characters to canonical characters. For example, '@' in

electronic mail addresses or ':' and '/' in URIs have width compatible characters.

The '+', '-', '<' and '>' may also be classified as delimiter characters that can be mapped to a set of canonical characters. Another example is

the FULL STOP character (“.”, U+002E) which is a delimiter in the visual

presentation of domain names. Some IMEs generate semantic or width

compatible character of FULL STOP such as IDEOGRAPHIC FULL STOP

(“。”, U+3002) when a user types FULL STOP on the keyboard. Such FULL STOP

compatible characters need to be mapped to the FULL STOP before

passing the string 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 whose mapping depends on locale or

on 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 language is Turkish (or one of several

other languages), unless an I is before a dot\_above, then the

character should be mapped to LATIN SMALL LETTER DOTLESS I (“ı”, U+0131).

To solve such problems for PRECIS framework, this document defines

characters that need local case mapping based on the

SpecialCasing.txt [Specialcasing] file in section 3.13 of The Unicode

Standard [Unicode].

The following are the methods to calculate codepoints that local case

mapping targets. Here Casefolding() means full casefolding described

in the CaseFolding.txt file [Casefolding], and Specialcasing() means

specialcasing described in the SpecialCasing.txt file

[Specialcasing].

If Casefolding(Specialcasing(cp)) != Casefolding(cp)

Then cp is a target

Else cp is not a target;

Local case mapping can be selected only when case mapping is selected

using the PRECIS Framework profile. Application developers should

calculate the codepoints of local case mapping targets by using the

latest Casefolding.txt and SpecialCasing.txt. Appendix B "Code

points list for local case mapping" lists codepoints in Unicode 6.3

calculated by this method.

3. Order of operations

The mappings 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

3. Local case 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, 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-11

(work in progress), October 2013.

[Unicode] The Unicode Consortium, "The Unicode Standard, Version

6.3.0", <http://www.unicode.org/versions/Unicode6.3.0/>,

2012.

[Casefolding]

YONEYA & Nemoto Expires April 24, 2014 [Page 5]

Internet-Draft precis mapping October 2013

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.

YONEYA & Nemoto Expires April 24, 2014 [Page 6]

Internet-Draft precis mapping October 2013

[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.

+----------------------+-------------+-----------+------+---------+

| \ 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. Code points list for local case mapping

Followings are a list of characters that need Local case mapping.

Format:

<Codepoint>; <Lowercase>; <Language> <Condition>; <Comments>

<Language> means the alpha-2 codes in [ISO.3166-1].

B.1. Unicode 6.3

0049; 0069 0307; lt More\_Above; # LATIN CAPITAL LETTER I

004A; 006A 0307; lt More\_Above; # LATIN CAPITAL LETTER J

00CC; 0069 0307 0300; lt; # LATIN CAPITAL LETTER I WITH GRAVE

00CD; 0069 0307 0301; lt; # LATIN CAPITAL LETTER I WITH ACUTE

0128; 0069 0307 0303; lt; # LATIN CAPITAL LETTER I WITH TILDE

YONEYA & Nemoto Expires April 24, 2014 [Page 7]

Internet-Draft precis mapping October 2013

012E; 012F 0307; lt More\_Above; # LATIN CAPITAL LETTER I WITH

OGONEK

0307; 0307; lt After\_Soft\_Dotted; # COMBINING DOT ABOVE

0049; 0131; tr Not\_Before\_Dot; # LATIN CAPITAL LETTER I

0130; 0069; tr; # LATIN CAPITAL LETTER I WITH DOT ABOVE

0307; ; tr After\_I; # COMBINING DOT ABOVE

0049; 0131; az Not\_Before\_Dot; # LATIN CAPITAL LETTER I

0130; 0069; az; # LATIN CAPITAL LETTER I WITH DOT ABOVE

0307; ; az After\_I; # COMBINING DOT ABOVE

Appendix C. Change Log

C.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".

C.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".

YONEYA & Nemoto Expires April 24, 2014 [Page 8]

Internet-Draft precis mapping October 2013

o Remove the Appendix B "Codepoints which need special mapping" and

mentioned related documents in the Section 2.2 .

C.3. Changes since -02

o Removed the "Open issues".

C.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.

C.5. Changes since -04

o Correct a sentence in the Section 2.3 "Local case mapping".

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

YONEYA & Nemoto Expires April 24, 2014 [Page 9]

Internet-Draft precis mapping October 2013

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