

PRECIS  
Internet-Draft  
Intended status: Standards Track  
Expires: May 30, 2015

P. Saint-Andre  
&yet  
November 26, 2014

Preparation, Enforcement, and Comparison of Internationalized Strings  
Representing Nicknames  
draft-ietf-precis-nickname-13

## Abstract

This document describes methods for handling Unicode strings representing nicknames.

## 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 May 30, 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

<a href="#">1.</a>	Introduction . . . . .	<a href="#">2</a>
<a href="#">1.1.</a>	Overview . . . . .	<a href="#">2</a>
<a href="#">1.2.</a>	Terminology . . . . .	<a href="#">3</a>
<a href="#">2.</a>	Nickname Profile . . . . .	<a href="#">3</a>
<a href="#">2.1.</a>	Preparation . . . . .	<a href="#">3</a>
<a href="#">2.2.</a>	Enforcement . . . . .	<a href="#">3</a>
<a href="#">2.3.</a>	Comparison . . . . .	<a href="#">4</a>
<a href="#">3.</a>	Examples . . . . .	<a href="#">4</a>
<a href="#">4.</a>	Use in Application Protocols . . . . .	<a href="#">6</a>
<a href="#">5.</a>	IANA Considerations . . . . .	<a href="#">6</a>
<a href="#">6.</a>	Security Considerations . . . . .	<a href="#">7</a>
<a href="#">6.1.</a>	Reuse of PRECIS . . . . .	<a href="#">7</a>
<a href="#">6.2.</a>	Reuse of Unicode . . . . .	<a href="#">7</a>
<a href="#">6.3.</a>	Visually Similar Characters . . . . .	<a href="#">7</a>
<a href="#">7.</a>	References . . . . .	<a href="#">8</a>
<a href="#">7.1.</a>	Normative References . . . . .	<a href="#">8</a>
<a href="#">7.2.</a>	Informative References . . . . .	<a href="#">8</a>
<a href="#">Appendix A.</a>	Acknowledgements . . . . .	<a href="#">9</a>
	Author's Address . . . . .	<a href="#">9</a>

## [1.](#) Introduction

### [1.1.](#) Overview

Technologies for textual chatrooms customarily enable participants to specify a nickname for use in the room; e.g., this is true of Internet Relay Chat [[RFC2811](#)] as well as multi-party chat technologies based on the Extensible Messaging and Presence Protocol (XMPP) [[RFC6120](#)] [[XEP-0045](#)], the Message Session Relay Protocol (MSRP) [[RFC4975](#)] [[I-D.ietf-simple-chat](#)], and Centralized Conferencing (XCON) [[RFC5239](#)] [[I-D.boulton-xcon-session-chat](#)]. Recent chatroom technologies also allow internationalized nicknames because they support characters from outside the ASCII range [[RFC20](#)], typically by means of the Unicode character set [[UNICODE](#)]. Although such nicknames tend to be used primarily for display purposes, they are sometimes used for programmatic purposes as well (e.g., kicking users or avoiding nickname conflicts). Note too that nicknames can be used not only in chatrooms but also more generally as a user's preferred display name (see for instance [[XEP-0172](#)]).

Nicknames (also called "petnames") are also used in contexts other than messaging, such as petnames for devices, bank accounts, and the like. The rules specified in this document can also be applied to such nicknames.

To increase the likelihood that nicknames will work in ways that make sense for typical users throughout the world, this document defines rules for preparing, enforcing, and comparing internationalized nicknames.

## 1.2. Terminology

Many important terms used in this document are defined in [\[I-D.ietf-precis-framework\]](#), [\[RFC6365\]](#), and [\[UNICODE\]](#).

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 [RFC 2119](#) [\[RFC2119\]](#).

## 2. Nickname Profile

Detailed rules for the preparation, enforcement, and comparison of nicknames are provided in the following sections, which define the Nickname profile of the PRECIS FreeformClass (on the distinction between these actions, refer to [\[I-D.ietf-precis-framework\]](#)).

### 2.1. Preparation

An entity that prepares a string according to this profile MUST ensure that the string consists only of Unicode code points that conform to the "FreeformClass" base string class defined in [\[I-D.ietf-precis-framework\]](#). In addition, the string MUST be encoded as UTF-8 [\[RFC3629\]](#).

### 2.2. Enforcement

An entity that performs enforcement according to this profile MUST prepare a string as described in the previous section and MUST also apply the rules specified below for the Nickname profile (these rules MUST be applied in the order shown).

1. Width Mapping Rule: There is no width-mapping rule (such a rule is not necessary because width mapping is performed as part of normalization using NFKC as specified below).
2. Additional Mapping Rule: The additional mapping rule consists of the following sub-rules.
  1. Any instances of non-ASCII space MUST be mapped to ASCII space (U+0020); a non-ASCII space is any Unicode code point having a general category of "Zs", naturally with the exception of U+0020.

2. Leading and trailing whitespace (i.e., one or more instances of the ASCII space character at the beginning or end of a nickname) MUST be removed (e.g., "stpeter " is mapped to "stpeter").
3. Interior sequences of more than one ASCII space character MUST be mapped to a single ASCII space character (e.g., "St Peter" is mapped to "St Peter").
3. Case Mapping Rule: Uppercase and titlecase characters MUST be mapped to their lowercase equivalents using Unicode Default Case Folding. In applications that prohibit conflicting nicknames, this rule helps to reduce the possibility of confusion by ensuring that nicknames differing only by case (e.g., "stpeter" vs. "StPeter") would not be presented to a human user at the same time.
4. Normalization Rule: The string MUST be normalized using Unicode Normalization Form KC (NFKC). Because NFKC is more "aggressive" in finding matches than other normalization forms (in the terminology of Unicode, it performs both canonical and compatibility decomposition before recomposing code points), this rule helps to reduce the possibility of confusion by increasing the number of characters that would match (e.g., U+2163 ROMAN NUMERAL FOUR would match the combination of U+0049 LATIN CAPITAL LETTER I and U+0056 LATIN CAPITAL LETTER V).
5. Directionality Rule: Applications MUST apply the "Bidi Rule" defined in [[RFC5893](#)] (i.e., each of the six conditions of the Bidi Rule must be satisfied).

### [2.3.](#) Comparison

An entity that performs comparison of two strings according to this profile MUST prepare each string and enforce the normalization, case-mapping, and width-mapping rules specified in the previous two sections. The two strings are to be considered equivalent if they are an exact octet-for-octet match (sometimes called "bit-string identity").

### [3.](#) Examples

The following examples illustrate a small number of nicknames that are consistent with the format defined above, along with the output string resulting from application of the PRECIS rules, which would be used for comparison purposes (note that the characters < and > are used to delineate the actual nickname and are not part of the nickname strings).

Table 1: A sample of legal nicknames

#	Nickname	Output for Comparison
1	<Foo>	<foo>
2	<foo>	<foo>
3	<Foo Bar>	<foo bar>
4	<foo bar>	<foo bar>
5	<&#x3A3;>	GREEK SMALL LETTER SIGMA (U+03C3)
6	<&#x3C3;>	GREEK SMALL LETTER SIGMA (U+03C3)
7	<&#x3C2;>	GREEK SMALL LETTER FINAL SIGMA (U+03C2)
8	<&#x265A;>	BLACK CHESS KING (U+265A)

Regarding examples 5, 6, and 7: case-mapping of GREEK CAPITAL LETTER SIGMA (U+03A3) to lowercase (i.e., to GREEK SMALL LETTER SIGMA, U+03C3) during comparison would result in matching the nicknames in examples 5 and 6; however, because the PRECIS mapping rules do not account for the special status of GREEK SMALL LETTER FINAL SIGMA (U+03C2), the nicknames in examples 5 and 7 or examples 6 and 7 would not be matched. Regarding example 8: symbol characters such as BLACK CHESS KING (U+265A) are allowed by the PRECIS FreeformClass and thus can be used in nicknames.

The following examples illustrate strings that are not valid nicknames because they violate the format defined above.

Table 2: A sample of strings that violate the nickname rules

#	Non-Nickname string	Notes
9	< foo >	Leading spaces
10	<foo      bar>	Multiple spaces
10	<>	Zero-length string

#### 4. Use in Application Protocols

This specification defines only the PRECIS-based rules for handling of nickname strings. It is the responsibility of an application protocol (e.g., MSRP, XCON, or XMPP) to specify the protocol slots in which nickname strings can appear, as well as the entities that are expected to enforce the rules governing nickname strings in that protocol.

Above and beyond the PRECIS-based rules specified here, application protocols can also define application-specific rules governing nickname strings (rules regarding the minimum or maximum length of nicknames, further restrictions on allowable characters or character ranges, safeguards to mitigate the effects of visually similar characters, etc.).

Naturally, application protocols can also specify rules governing the actual use of nicknames in applications (reserved nicknames, authorization requirements for using nicknames, whether certain nicknames can be prohibited, handling of duplicates, the relationship between nicknames and underlying identifiers such as SIP URIs or Jabber IDs, etc.).

Entities that enforce the rules specified in this document are encouraged to be liberal in what they accept by following this procedure:

1. Where possible, map characters (e.g, through width mapping, additional mapping, special mapping, case mapping, or normalization) and accept the mapped string.
2. If mapping is not possible (e.g., because a character is disallowed in the FreeformClass), reject the string.

#### 5. IANA Considerations

The IANA shall add the following entry to the PRECIS Profiles Registry:

Name: Nickname.

Base Class: FreeformClass.

Applicability: Nicknames in messaging and text conferencing technologies such as MSRP, XCON, and XMPP.

Replaces: None.

Width Mapping Rule: None (handled via NFKC).

Additional Mapping Rule: Map non-ASCII space characters to ASCII space, strip leading and trailing space characters, map interior sequences of multiple space characters to a single ASCII space.

Case Mapping Rule: For comparison purposes, map uppercase and titlecase characters to lowercase using Unicode Default Case Folding.

Normalization Rule: NFKC.

Directionality Rule: The "Bidi Rule" defined in [RFC 5893](#) applies.

Enforcement: To be specified by applications.

Specification: RFC XXXX. [Note to RFC Editor: please change "XXXX" to the RFC number issued for this specification.]

## [6.](#) Security Considerations

### [6.1.](#) Reuse of PRECIS

The security considerations described in [[I-D.ietf-precis-framework](#)] apply to the "FreeformClass" string class used in this document for nicknames.

### [6.2.](#) Reuse of Unicode

The security considerations described in [[UTS39](#)] apply to the use of Unicode characters in nicknames.

### [6.3.](#) Visually Similar Characters

[[I-D.ietf-precis-framework](#)] describes some of the security considerations related to visually similar characters, also called "confusable characters" or "confusables".

Although the mapping rules defined under [Section 2](#) of this document are designed in part to reduce the possibility of confusion about nicknames, this document does not provide more detailed recommendations regarding the handling of visually similar characters, such as those provided in [[UTS39](#)].

## 7. References

### 7.1. Normative References

- [I-D.ietf-precis-framework]  
Saint-Andre, P. and M. Blanchet, "Precis Framework: Handling Internationalized Strings in Protocols", [draft-ietf-precis-framework-20](#) (work in progress), November 2014.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, [RFC 3629](#), November 2003.
- [RFC5893] Alvestrand, H. and C. Karp, "Right-to-Left Scripts for Internationalized Domain Names for Applications (IDNA)", [RFC 5893](#), August 2010.
- [UNICODE] The Unicode Consortium, "The Unicode Standard, Version 6.3", 2013,  
<<http://www.unicode.org/versions/Unicode6.3.0/>>.
- [UTS39] The Unicode Consortium, "Unicode Technical Standard #39: Unicode Security Mechanisms", November 2013,  
<<http://unicode.org/reports/tr39/>>.

### 7.2. Informative References

- [I-D.boulton-xcon-session-chat]  
Barnes, M., Boulton, C., and S. Loreto, "Chatrooms within a Centralized Conferencing (XCON) System", [draft-boulton-xcon-session-chat-08](#) (work in progress), July 2011.
- [I-D.ietf-simple-chat]  
Niemi, A., Garcia, M., and G. Sandbakken, "Multi-party Chat Using the Message Session Relay Protocol (MSRP)", [draft-ietf-simple-chat-18](#) (work in progress), January 2013.
- [RFC20] Cerf, V., "ASCII format for network interchange", [RFC 20](#), October 1969.
- [RFC2811] Kalt, C., "Internet Relay Chat: Channel Management", [RFC 2811](#), April 2000.

- [RFC4975] Campbell, B., Mahy, R., and C. Jennings, "The Message Session Relay Protocol (MSRP)", [RFC 4975](#), September 2007.
- [RFC5239] Barnes, M., Boulton, C., and O. Levin, "A Framework for Centralized Conferencing", [RFC 5239](#), June 2008.
- [RFC6120] Saint-Andre, P., "Extensible Messaging and Presence Protocol (XMPP): Core", [RFC 6120](#), March 2011.
- [RFC6365] Hoffman, P. and J. Klensin, "Terminology Used in Internationalization in the IETF", [BCP 166](#), [RFC 6365](#), September 2011.
- [XEP-0045] Saint-Andre, P., "Multi-User Chat", XSF XEP 0045, February 2012.
- [XEP-0172] Saint-Andre, P. and V. Mercier, "User Nickname", XSF XEP 0172, March 2012.

#### [Appendix A](#). Acknowledgements

Thanks to Kim Alvefur, Mary Barnes, Dave Cridland, Miguel Garcia, Salvatore Loreto, and Enrico Marocco for their reviews and comments.

Peter Saint-Andre wishes to acknowledge Cisco Systems, Inc., for employing him during his work on earlier versions of this document.

#### Author's Address

Peter Saint-Andre  
&yet

Email: [peter@andyet.com](mailto:peter@andyet.com)  
URI: <https://andyet.com/>