

Internet Draft
[draft-ietf-idn-nameprep-10.txt](#)
May 17, 2002
Expires in six months

Paul Hoffman
IMC & VPNC
Marc Blanchet
ViaGenie

Nameprep: A Stringprep Profile for Internationalized Domain Names

Status of this memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC2026](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

To view the list Internet-Draft Shadow Directories, see <http://www.ietf.org/shadow.html>.

Abstract

This document describes how to prepare internationalized domain name labels in order to increase the likelihood that name input and name comparison work in ways that make sense for typical users throughout the world. This profile of the stringprep protocol is used as part of a suite of on-the-wire protocols for internationalizing the DNS.

[1](#). Introduction

This document specifies processing rules that will allow users to enter internationalized domain names in applications and have the highest chance of getting the content of the strings correct. It is a profile of stringprep [[STRINGPREP](#)]. These processing rules are only intended for internationalized domain names, not for arbitrary text.

This profile defines the following, as required by [[STRINGPREP](#)]

- The intended applicability of the profile: internationalized domain names processed by IDNA
- The character repertoire that is the input and output to stringprep: Unicode 3.1, specified in [Section 2](#)
- The mappings used: specified in [Section 3](#)

- The Unicode normalization used: specified in [Section 4](#)
- The characters that are prohibited as output: specified in [section 5](#)

[1.1](#) Interaction of protocol parts

Nameprep is used by the IDNA [[IDNA](#)] protocol for preparing domain names; it is not designed for any other purpose. It is explicitly not designed for processing arbitrary free text and SHOULD NOT be used for that purpose. Nameprep is a profile of Stringprep [[STRINGPREP](#)]. Implementations of Nameprep MUST fully implement Stringprep.

Nameprep is used to process domain name labels, not domain names. IDNA calls nameprep for each label in a domain name, not for the whole domain name.

[1.2](#) Terminology

The key words "MUST", "MUST NOT", "SHOULD", "SHOULD NOT", and "MAY" in this document are to be interpreted as described in [RFC 2119](#) [[RFC2119](#)].

[2](#). Character Repertoire

This profile uses Unicode 3.1, as defined in [[STRINGPREP](#)] [Appendix A.1](#).

[3](#). Mapping

This profile specifies mapping using the following tables from [[STRINGPREP](#)]:

Table B.1

Table B.2

[4](#). Normalization

This profile specifies using Unicode normalization form KC, as described in [[STRINGPREP](#)].

[5](#). Prohibited Output

This profile specifies prohibiting using the following tables from [[STRINGPREP](#)]:

Table C.1

Table C.2

Table C.3

Table C.4

Table C.5

Table C.6
Table C.7
Table C.8
Table C.9

IMPORTANT NOTE: This profile MUST be used with the IDNA protocol. The IDNA protocol has additional prohibitions that are checked outside of this profile.

In addition, this profile adds the prohibitions. Thus, the full set of prohibited characters are those from the tables above plus those listed individually below.

[5.1](#) Inappropriate characters from common input mechanisms

U+3002 is used as if it were U+002E in many domain name input mechanisms used by applications, particularly in Asia. Thus, U+3002 is prohibited in domain names by this specification.

3002; IDEOGRAPHIC FULL STOP

[6](#). Unassigned Code Points in Internationalized Domain Names

If the processing in [\[IDNA\]](#) specifies that a list of unassigned code points be used, the system uses table A.1 from [\[STRINGPREP\]](#) as its list of unassigned code points.

[7](#). References

[7.1](#) Normative references

[IDNA] Patrik Faltstrom, Paul Hoffman, and Adam M. Costello, "Internationalizing Domain Names in Applications (IDNA)", [draft-ietf-idn-idna](#), work-in-progress.

[STRINGPREP] Paul Hoffman and Marc Blanchet, "Preparation of Internationalized Strings ("stringprep")", [draft-hoffman-stringprep](#), work in progress.

[7.2](#) Informative references

[RFC2119] Scott Bradner, "Key words for use in RFCs to Indicate Requirement Levels", March 1997, [RFC 2119](#).

[STD13] Paul Mockapetris, "Domain names - concepts and facilities" (RFC 1034) and "Domain names - implementation and specification" ([RFC 1035](#), STD 13, November 1987.

[8. Security Considerations](#)

The Unicode and ISO/IEC 10646 repertoires have many characters that look similar. In many cases, users of security protocols might do visual matching, such as when comparing the names of trusted third parties. Because it is impossible to map similar-looking characters without a great deal of context such as knowing the fonts used, stringprep does nothing to map similar-looking characters together nor to prohibit some characters because they look like others.

Security on the Internet partly relies on the DNS. Thus, any change to the characteristics of the DNS can change the security of much of the Internet.

Domain names are used by users to connect to Internet servers. The security of the Internet would be compromised if a user entering a single internationalized name could be connected to different servers based on different interpretations of the internationalized domain name.

Current applications might assume that the characters allowed in domain names will always be the same as they are in [\[STD13\]](#). This document vastly increases the number of characters available in domain names. Every program that uses "special" characters in conjunction with domain names may be vulnerable to attack based on the new characters allowed by this specification.

[9. IANA Considerations](#)

This is a profile of stringprep. When it becomes an RFC, it should be registered in the stringprep profile registry.

[10. Acknowledgements](#)

Many people from the IETF IDN Working Group and the Unicode Technical Committee contributed ideas that went into the first draft of this document.

The IDN namprep design team made many useful changes to the first draft. That team and its advisors include:

Asmus Freytag
Cathy Wissink
Francois Yergeau
James Seng
Marc Blanchet
Mark Davis
Martin Duerst
Patrik Faltstrom
Paul Hoffman

Additional significant improvements were proposed by:

Jonathan Rosenne
Kent Karlsson
Scott Hollenbeck
Dave Crocker
Erik Nordmark

11. Author Contact Information

Paul Hoffman
Internet Mail Consortium and VPN Consortium
127 Segre Place
Santa Cruz, CA 95060 USA
paul.hoffman@imc.org and paul.hoffman@vpnc.org

Marc Blanchet
Viagenie inc.
2875 boul. Laurier, bur. 300
Ste-Foy, Quebec, Canada, G1V 2M2
Marc.Blanchet@viagenie.qc.ca