                    Nameprep: A Stringprep Profile for
                    Internationalized Domain Names (IDN)

Status of this Memo

   This document specifies an Internet standards track protocol for the
   Internet community, and requests discussion and suggestions for
   improvements. Please refer to the current edition of the "Internet
   Official Protocol Standards" (STD 1) for the standardization state
   and status of this protocol. Distribution of this memo is unlimited.

Abstract

   This document describes how to prepare internationalized domain name
   (IDN) 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 Domain Name System (DNS).

1. Introduction

   This document specifies processing rules that will allow users to
   enter internationalized domain names (IDNs) into 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 3.2, 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.

   - Bidirectional character handling: specified in section 6.

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 BCP 14, RFC
   2119 [RFC2119].

2. Character Repertoire

   This profile uses Unicode 3.2, as defined in [STRINGPREP] Appendix A.

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.1.2
   Table C.2.2
   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.

6. Bidirectional characters

   This profile specifies checking bidirectional strings as described in
   [STRINGPREP] section 6.

7. 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.

8. References

8.1 Normative References

   [RFC2119]   Bradner, S., "Key words for use in RFCs to Indicate
                 Requirement Levels", BCP 14, RFC 2119, March 1997.

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

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

8.2 Informative references

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

9. 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.

10. IANA Considerations

   This is a profile of stringprep. It has been registered by the IANA
   in the stringprep profile registry
   (www.iana.org/assignments/stringprep-profiles).

      Name of this profile:
         Nameprep

      RFC in which the profile is defined:
         This document.

      Indicator whether or not this is the newest version of the
      profile:
         This is the first version of Nameprep.

11. Acknowledgements

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

   The IDN Nameprep design team made many useful changes to the
   document. 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
      Matitiahu Allouche

12. Authors' Addresses

    Paul Hoffman
    Internet Mail Consortium and VPN Consortium
    127 Segre Place
    Santa Cruz, CA 95060 USA

    EMail: paul.hoffman@imc.org and paul.hoffman@vpnc.org

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

    EMail: Marc.Blanchet@viagenie.qc.ca

13. Full Copyright Statement

Acknowledgement