

IDNABIS
Internet-Draft
Intended status: Standards Track
Expires: March 7, 2010

P. Resnick
Qualcomm Incorporated
P. Hoffman
VPN Consortium
September 3, 2009

Mapping Characters in IDNA
draft-ietf-idnabis-mappings-04

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#). This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on March 7, 2010.

Copyright Notice

Copyright (c) 2009 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 in effect on the date of publication of this document (<http://trustee.ietf.org/license-info>). Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Abstract

In the original version of the Internationalized Domain Names in Applications (IDNA) protocol, any Unicode code points taken from user input were mapped into a set of Unicode code points that "make sense", which were then encoded and passed to the domain name system (DNS). The current version of IDNA presumes that the input to the protocol comes from a set of "permitted" code points, which it then encodes and passes to the DNS, but does not specify what to do with the result of user input. This document describes the actions that can be taken by an implementation between user input and passing permitted code points to the new IDNA protocol.

1. Introduction

This document describes the operations that can be applied to user input in order to get it into a form acceptable by the Internationalized Domain Names in Applications (IDNA) protocol [[I-D.ietf-idnabis-protocol](#)]. A general implementation procedure for mapping is given in [section 2](#).

It should be noted that this document does not specify the behavior of a protocol that appears "on the wire". It describes an operation that is to be applied to user input in order to prepare that user input for use in an "on the network" protocol. As unusual as this may be for an IETF protocol document, it is a necessary operation to maintain interoperability.

2. The General Procedure

This section defines a general algorithm that applications ought to implement in order to produce Unicode code points that will be valid under the IDNA protocol. An application might implement the full mapping as described below, or can choose a different mapping. In fact, an application might want to implement a full mapping that is substantially compatible with the original IDNA protocol instead of the algorithm given here.

The general algorithm that an application (or the input method provided by an operating system) ought to use is relatively

straightforward:

1. Upper case characters are mapped to their lower case equivalents by using the algorithm for mapping case in Unicode characters.
2. Full-width and half-width characters (those defined with Decomposition Types <wide> and <narrow>) are mapped to their decomposition mappings as shown in the Unicode character database.
3. All characters are mapped using Unicode Normalization Form C (NFC).
4. [[I-D.ietf-idnabis-protocol](#)] is specified such that the protocol acts on the individual labels of the domain name. If an implementation of this mapping is also performing the step of separation of the parts of a domain name into labels by using the FULL STOP character (U+002E), the following character can be mapped to the FULL STOP before label separation occurs:

* IDEOGRAPHIC FULL STOP (U+3002)

There are other characters that are used as "full stops" that one could consider mapping as label separators, but their use as such has not been investigated thoroughly.

Definitions for the rules in this algorithm can be found in [[Unicode51](#)]. Specifically:

- o Unicode Normalization Form C can be found in Annex #15 of [[Unicode51](#)].
- o In order to map upper case characters to their lower case equivalents (defined in section 3.13 of [[Unicode51](#)]), first map characters to the "Lowercase_Mapping" property (the "<lower>" entry in the second column) in <<http://www.unicode.org/Public/UNIDATA/SpecialCasing.txt>>, if any. Then, map characters to the "Simple_Lowercase_Mapping" property (the fourteenth column) in <<http://www.unicode.org/Public/UNIDATA/UnicodeData.txt>>, if any.
- o In order to map full-width and half-width characters to their decomposition mappings, map any character whose "Decomposition_Type" (contained in the first part of of the sixth column) in <<http://www.unicode.org/Public/UNIDATA/UnicodeData.txt>> is either "<wide>" or "<narrow>" to the "Decomposition_Mapping" of that character (contained in the second part of the sixth column) in <<http://www.unicode.org/Public/UNIDATA/UnicodeData.txt>>.

- o The <<http://www.unicode.org/Public/UNIDATA/UCD.html>> web page has useful descriptions of the contents of these files.

If the mappings in this document are applied to versions of Unicode later than Unicode 5.1, the later versions of the Unicode Standard should be consulted.

These form a minimal set of mappings that an application should strongly consider doing. Of course, there are many others that might be done.

3. IANA Considerations

This memo includes no request to IANA.

4. Security Considerations

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 (nor in the other IDNA-related documents).

5. Acknowledgements

This document is the product of the IDNabis Working Group and contains contributions from many people in the Working Group.

6. Normative References

[I-D.ietf-idnabis-protocol]

Klensin, J., "Internationalized Domain Names in Applications (IDNA): Protocol",
[draft-ietf-idnabis-protocol-15](#) (work in progress),
September 2009.

[Unicode51]

The Unicode Consortium, "The Unicode Standard, Version 5.1.0", 2008.

defined by: The Unicode Standard, Version 5.0, Boston, MA, Addison-Wesley, 2007, ISBN 0-321-48091-0, as amended by Unicode 5.1.0
(<http://www.unicode.org/versions/Unicode5.1.0/>).

Authors' Addresses

Peter W. Resnick
Qualcomm Incorporated
5775 Morehouse Drive
San Diego, CA 92121-1714
US

Phone: +1 858 651 4478
Email: presnick@qualcomm.com
URI: <http://www.qualcomm.com/~presnick/>

Paul Hoffman
VPN Consortium
127 Segre Place
Santa Cruz, CA 95060
US

Phone: 1-831-426-9827
Email: paul.hoffman@vpnc.org

