

Network Working Group
Internet-Draft
Intended status: BCP
Expires: September 10, 2009

Y. YONEYA
JPRS
X. LEE
CNNIC
YJ. Suh
NIDA
E. Chen
TWNIC
March 9, 2009

CJK local mapping in IDNA2008
draft-jet-idnabis-cjk-localmapping-01

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 September 10, 2009.

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](http://trustee.ietf.org/license-info) 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

Development of IDNA2008 is now in final stage. It will cause incompatibilities for Chinese, Japanese and Korean (CJK) scripts and languages. To avoid incompatibilities with IDNA2008 and current IDNA (IDNA2003), definition of specific local mapping (pre process of IDNA to be performed to IDN candidate string) for CJK is recommended.

Table of Contents

1.	Introduction	4
1.1.	Positioning of this document	4
1.2.	Why CJK?	4
2.	Terminology	4
3.	List of incompatibilities of CJK between IDNA2008 and IDNA2003	5
3.1.	Label separators	5
3.2.	Compatibility characters	6
3.3.	Exceptions	6
4.	Solutions for incompatibilities	7
4.1.	Label saparators	7
4.2.	Compatibility characters	7
4.3.	Exceptions	7
5.	Guideline to keep compatibility in registration protocol . . .	7
6.	Guideline to keep compatibility in domain name lookup protocol	7
7.	IANA Considerations	7
8.	Security Considerations	8
9.	Acknowledgements	8
10.	References	8
10.1.	Normative References	8
10.2.	Informative References	9
Appendix A.	Change Log	9
A.1.	Version -01	9
	Authors' Addresses	10

1. Introduction

1.1. Positioning of this document

IDNA protocol is going to be revised by IDNA2008 ([[I-D.ietf-idnabis-rationale](#)] [[I-D.ietf-idnabis-defs](#)] [[I-D.ietf-idnabis-protocol](#)] [[I-D.ietf-idnabis-tables](#)] [[I-D.ietf-idnabis-bidi](#)]) which is in the final stage includes incompatibilities with IDNA2003 ([[RFC3490](#)] [[RFC3491](#)] [[RFC3492](#)]) in some cases. Due to those incompatibilities, name resolution of existing registered IDNs are possible to fail. To avoid such incompatibilities, IDNA2008 recommends to perform local mapping before registration and domain name lookup of IDNA processing, but it does not mention any specific method. This document defines local mapping for IDNA2008 regarding CJK to avoid incompatibilities between IDNA2008 and IDNA2003.

1.2. Why CJK?

CJK shares some scripts such as Han and punctuations. Therefore, it is useful to have common local mapping definition in areas and / or languages that share scripts. Furthermore, ccTLDs in CJK area are initiative of IDN because they have been actively working for IDN development and deployment since IDNA2003 start up, had published JET Guideline together, and accumulated and shared their experiences and knowledge of IDN registration and operation. Definition developed by such ccTLDs depending on their experiences is useful for community.

2. Terminology

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

CJK

The term "CJK" stands for "Chinese, Japanese and Korean".

CJK IDN

The term "CJK IDN" stands for "Chinese IDN" or "Japanese IDN" or "Korean IDN".

CJK scripts

Script Name	Code(Range)	C	J	K
CJK Symbols and Punctuation	U+3000-U+3007	Y	Y	
Hiragana	U+3040-U+309F		Y	
Katakana	U+30A0-U+30FF		Y	
CJK Unified Ideographs Extension A	U+3400-U+4DFF	Y		
CJK Unified Ideographs	U+4E00-U+9FFF	Y	Y	
Hangul Syllables	U+AC00-U+D7A3			Y
CJK Compatibility Ideographs	U+F900-U+FAFF	Y		
Halfwidth and Fullwidth Forms	U+FF00-U+FFEF	Y	Y	
CJK Unified Ideographs Extension B	U+20000-U+2A6D6	Y		
CJK Compatibility Ideographs Supplement	U+2F800-U+2FA1F	Y		

Chinese IDN

The term "Chinese IDN" stands for "IDN consists from CJK scripts marked with 'Y' in 'C' column and LDH". Permitted characters in Chinese IDN are listed in [[IANA-IDN-Language-zh-CN](#)] and [[IANA-IDN-Language-zh-Tw](#)].

Japanese IDN

The term "Japanese IDN" stands for "IDN consists from CJK scripts marked with 'Y' in 'J' column and LDH". Permitted characters in Japanese IDN are listed in [[IANA-IDN-Language-ja-JP](#)].

Korean IDN

The term "Korean IDN" stands for "IDN consists from CJK scripts marked with 'Y' in 'K' column and LDH". Permitted characters in Korean IDN are listed in [[IANA-IDN-Language-ko-KR](#)].

Other terms defined in [[I-D.ietf-idnabis-defs](#)] are used.

3. List of incompatibilities of CJK between IDNA2008 and IDNA2003

3.1. Label separators

Following characters are defined as label separator in IDNA2003, but not defined in IDNA2008.

Script Name	Code	C J K
IDEOGRAPHIC FULL STOP	U+3002	Y Y
HALFWIDTH IDEOGRAPHIC FULL STOP	U+FF61	Y Y
FULLWIDTH FULL STOP	U+FF0E	Y Y

CJK IDN which includes those characters are valid in IDNA2003 but invalid in IDNA2008.

3.2. Compatibility characters

Compatibility characters that are mapped to canonical (valid) characters by Unicode Normalization Form KC (NFKC) [[Unicode](#)] [[UAX15](#)] in IDNA2003 are invalid in IDNA2008.

Script Name	Code(Range)	C J K
FULLWIDTH DIGITS	U+FF10-U+FF19	Y Y
FULLWIDTH LATIN CAPITAL LETTERS	U+FF21-U+FF3A	Y Y
FULLWIDTH LATIN SMALL LETTERS	U+FF41-U+FF5A	Y Y
HALFWIDTH KATAKANA LETTERS	U+FF65-U+FF9F	Y Y

CJK IDN which includes those characters are valid in IDNA2003 but invalid in IDNA2008.

3.3. Exceptions

Some of mark characters of quasi-Han or quasi-Kana are exceptions in IDNA2008 that have position and / or adjacent character property limitations are valid in IDNA2003.

Script Name	Code	C J K
IDEOGRAPHIC ITERATION MARK	U+3006	Y Y
KATAKANA MIDDLEDOT	U+30FB	Y Y

CJK IDN which includes those characters are valid in IDNA2003 but happen to be invalid in IDNA2008.

4. Solutions for incompatibilities

4.1. Label saparators

If characters listed in section [Section 3.1](#) is included in CJK IDN candidate string, those characters were mapped to FULL STOP (U+002E) in local mapping processing.

4.2. Compatibility characters

If characters listed in section [Section 3.2](#) is included in CJK IDN candidate string, those characters were mapped to canonical characters by NFKC in local mapping processing.

4.3. Exceptions

If characters listed in section [Section 3.3](#) is included in CJK IDN candidate string, those characters were treated as PROTOCOL VALID instead of CONTEXT0. [NOTE: this is not local mapping, but requires property change of those characters]

5. Guideline to keep compatibility in registration protocol

Registries who treat CJK IDN must implement solutions described in section [Section 4](#).

6. Guideline to keep compatibility in domain name lookup protocol

Application softwares which treat CJK IDN must implement solutions described in section [Section 4](#). Local mappings or corresponding pre-processing must be performed on user interface (input/output) processing, such as:

- User typing or pasting in input area
- String extraction from free context in contents area
- Displaing on dialogue and / or address bar

Note that those strings are used for domain name lookup.

7. IANA Considerations

None.

8. Security Considerations

None.

9. Acknowledgements

Many suggestions and advices were given from JET members, especially Yao Jiankang, ...

10. References

10.1. Normative References

[I-D.ietf-idnabis-bidi]

Alvestrand, H. and C. Karp, "An updated IDNA criterion for right-to-left scripts", [draft-ietf-idnabis-bidi-03](#) (work in progress), November 2008.

[I-D.ietf-idnabis-defs]

Klensin, J., "Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework", [draft-ietf-idnabis-defs-07](#) (work in progress), March 2009.

[I-D.ietf-idnabis-protocol]

Klensin, J., "Internationalized Domain Names in Applications (IDNA): Protocol", [draft-ietf-idnabis-protocol-10](#) (work in progress), March 2009.

[I-D.ietf-idnabis-rationale]

Klensin, J., "Internationalized Domain Names for Applications (IDNA): Background, Explanation, and Rationale", [draft-ietf-idnabis-rationale-08](#) (work in progress), March 2009.

[I-D.ietf-idnabis-tables]

Faltstrom, P., "The Unicode code points and IDNA", [draft-ietf-idnabis-tables-05](#) (work in progress), December 2008.

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

[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.
- [RFC3492] Costello, A., "Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)", [RFC 3492](#), March 2003.
- [UAX15] The Unicode Consortium, "Unicode Normalization Forms", Unicode Standard Annex #15, <http://www.unicode.org/reports/tr15/>, 2008.
- [Unicode] The Unicode Consortium, "The Unicode Standard, Version 5.0", Boston, MA, USA: Addiso-Wesley. ISBN 0-321-48091-0, 2007.

10.2. Informative References

- [IANA-IDN-Language-ja-JP]
JPRS, ".JP Japanese", IANA IDN language tables, http://www.iana.org/domains/idn-tables/tables/jp_ja-jp_1.2.html, August 2005.
- [IANA-IDN-Language-ko-KR]
KRNIC, ".KR Korean", IANA IDN language tables, http://www.iana.org/domains/idn-tables/tables/kr_ko-kr_1.0.html, March 2004.
- [IANA-IDN-Language-zh-CN]
CNNIC, ".CN Chinese", IANA IDN language tables, http://www.iana.org/domains/idn-tables/tables/cn_zh-cn_4.0.html, March 2005.
- [IANA-IDN-Language-zh-TW]
TWNIC, ".TW Chinese", IANA IDN language tables, http://www.iana.org/domains/idn-tables/tables/tw_zh-tw_4.0.html, March 2005.

Appendix A. Change Log

A.1. Version -01

- o Typo correction: one of "CJK Unified Ideographs Extension A" to "CJK Unified Ideographs Extension B"
- o Update version number of normative references.

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: yone@jprs.co.jp

Xiaodong LEE
CNNIC
No.4 South 4th Street, Zhongguancun
Beijing 100080
Peoples Republic of China

Phone: +86 10 58813020
Email: lee@cnnic.cn

YungJin Suh
NIDA
11F, 398, Seocho-no
Deocho-gu, Seoul 137-857
South Korea

Phone: +82 2 2186 4562
Email: yjsuh@nida.or.kr

Erin Chen
TWNIC
4F-2, No.9 Sec. 2, Roosevelt Rd.
Taipei 100
Taiwan

Phone: +886 2 23411313#502
Email: erin@twNIC.net.tw

