

Network Working Group
Request for Comments: 4713
Category: Informational

X. Lee
W. Mao
CNNIC
E. Chen
N. Hsu
TWNIC
J. Klensin
October 2006

Registration and Administration Recommendations for Chinese Domain Names

Status of This Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2006).

IESG Note

This RFC is not a candidate for any level of Internet Standard. The IETF disclaims any knowledge of the fitness of this RFC for any purpose and in particular notes that the decision to publish is not based on IETF review for such things as security, congestion control, or inappropriate interaction with deployed protocols. The RFC Editor has chosen to publish this document at its discretion. Readers of this document should exercise caution in evaluating its value for implementation and deployment. See RFC 3932 for more information.

Abstract

Many Chinese characters in common use have variants, which makes most of the Chinese Domain Names (CDNs) have at least two different forms. The equivalence between Simplified Chinese (SC) and Traditional Chinese (TC) characters is very important for CDN registration. This memo builds on the basic concepts, general guidelines, and framework of RFC 3743 to specify proposed registration and administration procedures for Chinese domain names. The document provides the information needed for understanding and using the tables defined in the IANA table registrations for Simplified and Traditional Chinese.

Table of Contents

1. Introduction	2
2. Terminology	3
2.1. Chinese Characters	3
2.2. Chinese Domain Name Label (CDNL)	3
2.3. Simplified Chinese Variant Table (SCVT)	4
2.4. Traditional Chinese Variant Table (TCVT)	4
2.5. Original Chinese Domain Name Label (OCDNL)	4
3. Procedure for Registration of Chinese Domain Name Labels	4
3.1. Terminology and Context	4
3.2. Procedure in Terms of the RFC 3743 Model	4
3.3. RFC 3743 Optional Registry Processing	5
4. Security Considerations	5
5. Acknowledgements	6
6. References	6
6.1. Normative References	6
6.2. Informative References	7

1. Introduction

With the standardization of Internationalized Domain Names for Application (IDNA, described in [RFC3490], [RFC3491], and [RFC3492]), internationalized domain names (IDNs), i.e., those that contain non-ASCII characters, are included in the DNS, and users can access the Internet with their native languages, most of which are not English. However, many languages have special requirements, which are not addressed in the IDNA RFCs. One way to deal with some of the remaining issues involves grouping characters that could be confused together as "variants". The variant approach is discussed in RFC 4290 [RFC4290] and specifically for documents written in Chinese, Japanese, or Korean (CJK documents), in the so-called "JET Guidelines" RFC 3743 [RFC3743]. Readers of this document are assumed to be familiar with the concepts and terminology of the latter. The guidelines specified in this document provide a set of specific tables and methods required to apply the JET Guidelines to Chinese characters. For example, changes were made in the forms of a large number of Chinese characters during the last century to simplify writing and reading. These "Simplified" characters have been adopted in some Chinese-speaking communities, while others continue to use the "Traditional" forms. On the global Internet, if IDNA were used alone, there would be considerable potential for confusion if the two forms were not considered together. Consequently, effective use of Chinese Domain Names (CDNs) requires variant equivalence, as described in RFC 3743, to handle character differences between Simplified and Traditional Chinese forms.

Chinese variant equivalence itself is very complicated in principle (please read [C2C] for further information). When it comes to the usage of Chinese domain names, the basic requirement is to match the user perception of Chinese characters between Simplified Chinese (SC) and Traditional Chinese (TC) forms. When users register SC or TC domain names, they will wish to obtain the other forms (Traditional or Simplified, respectively) as well, and expect others to be able to access the website or other resources in both forms.

This document specifies a solution for Chinese domain name registration and administration that has been adopted and deployed by CNNIC (the top-level domain registry for "CN") and TWNIC (the top-level domain registry for "TW") to manage Simplified Chinese and Traditional Chinese domain name equivalence. In the terminology of RFC 3743, this solution is based on Internationalized Domain Labels (IDLs).

2. Terminology

This document adopts the terminologies that are defined in RFC 3743. It is not possible to understand this document without first understanding the concepts and terminology of RFC 3743, including terminology introduced in its examples. Additional terminology is defined later in this document.

2.1. Chinese Characters

This document suggests permitting only a subset of Chinese characters in Chinese Domain Names (CDNs) and hence in the DNS. When this document discusses Chinese characters, it only refers to the subset of the characters in the first column of the current IANA registration tables for Chinese as discussed in Section 2.3 and Section 2.4. These are defined, in detail, in [LVT-SC] and [LVT-TC]. Of course, characters excluded from these tables are still valid Chinese characters. However, this document strongly suggests that registries do not permit any registration of Chinese characters that are not listed in the tables. The tables themselves will be updated in the future if necessary.

2.2. Chinese Domain Name Label (CDNL)

If an IDN label includes at least one Chinese character, it is called a Chinese Domain Name (CDN) Label. CDN labels may contain characters from the traditional letter-digit-hyphen (LDH) set as well as Chinese characters.

2.3. Simplified Chinese Variant Table (SCVT)

Based on RFC 3743 [RFC3743], a language table for Simplified Chinese has been defined [LVT-SC]. It can be used for the registration of Simplified Chinese domain names. The key feature of this table is that the preferred variant is the SC character, which is used by Chinese mainland users or defined in Chinese-related standards.

2.4. Traditional Chinese Variant Table (TCVT)

Similarly, a language table has been defined for Traditional Chinese [LVT-TC]. It is also based on the rules of RFC 3743. It can be used for registration of Traditional Chinese domain names. The preferred variant is the TC character, which is used by Taiwan users or defined in related standards.

2.5. Original Chinese Domain Name Label (OCDNL)

The Chinese Domain Name Label that users submit for registration.

3. Procedure for Registration of Chinese Domain Name Labels

3.1. Terminology and Context

This document adopts the same procedure for Chinese Domain Name Label (CDNL) registration as the one defined for more general IDN labels in section 3.2.3 of RFC 3743 [RFC3743]. The terminology and notation used below, and the steps that are mentioned, derive from that document. In particular, "CV" is the character variant associated with an input character ("IN") and a language table. The language tables used here are those for Chinese as spoken and written in the Chinese mainland (ZH-CN) and on Taiwan (ZH-TW). "PV" is the selected Preferred Variant.

3.2. Procedure in Terms of the RFC 3743 Model

The first column of the Simplified Chinese Variant Table (SCVT) is the same as the first column of the corresponding Traditional Chinese Variant Table (TCVT) and so are the third columns of both tables. Consequently, the CV(IN, ZH-CN) will be same as the CV(IN, ZH-TW) after Step 3; the PV(IN, ZH-CN) is in SC form, and the PV(IN, ZH-TW) is in TC form. As a result, there will not be more than three records (i.e., for the original label (OCDNL), the Simplified Chinese (SC) form, and the Traditional Chinese (TC) form) to be added into the zone file after applying this procedure. In other words, the procedure does not generate labels that contain a mixture of Simplified and Traditional Chinese as variants.

The set of languages associated with the input (IN) is both ZH-CN and ZH-TW by default. The procedure for CDNL registration uses the optional registry-defined rules provided in RFC 3743 for optional processing, with the understanding that the rules may vary for different registries supporting CDNs. The motivation for such rules is described below.

The preferred variant(s) is/are TC in TCVT, and SC in SCVT. There may be more than one preferred variant for a given valid character.

3.3. RFC 3743 Optional Registry Processing

In actuality, while IDNA, and hence RFC 3743, process characters one at a time, the actual relationship between the valid code point and the preferred variant is contextual: whether one character can be substituted for another depends on the characters with which it is associated in a label or, more generally, in a phrase. In particular, some of the preferred variants make no sense in combination with other characters; therefore, those combinations should not be added into the Zone file (described as "ZV" or zone variants in RFC 3743). If desired, it should be possible to define and implement rules to reduce the preferred variant labels to only plausible ones. This could be done, for example, with some artificial intelligence tools, or with feedback from the registrant, or with selection based on frequency of occurrence in other texts. To illustrate one possibility, the OCDNL could be required to be TC-only or SC-only, and if there is more than one preferred variant, the OCDNL will be used as the PV, instead of the PV produced by the algorithm.

To reemphasize, the tables in [LVT-SC] and [LVT-TC] follow the table format and terminologies defined in [RFC3743]. If one intends to implement Chinese domain name registrations based on these two tables or ones similar to them, a complete understanding of RFC 3743 is needed for the proper use of those tables.

4. Security Considerations

This document is subject to the same security considerations as RFC 3743, which defines the table formats and operations. As with that base document, part of its intent is to reduce the security problems that might be caused by confusion among characters with similar appearances or meanings. While it will not introduce any additional security issues, additional registration restrictions such as those outlined in Section 3 may further reduce potential problems.

5. Acknowledgements

Thanks to these people for their suggestions and for their efforts to bring this tough work to conclusion and to promote the results: WANG YanFeng, Ai-Chin LU, Shian-Shyong TSENG, QIAN HuaLin, and Li-Ming TSENG.

The authors especially thank Joe ZHANG and XiaoMing WANG for their outstanding contributions on SCVT in [LVT-SC]. Also, thanks to Kenny HUANG, Zheng-Wei LIN, Shi-Xiong TSENG, Lie-Neng WU, Cheng-Wu PAN, Lin-Mei WEI, and Qi-Qing HSU for their efforts and contributions on editing the TCVT in [LVT-TC]. These experts provided basic materials or gave very crucial suggestions and principles to accomplish these two variant tables.

The authors also gratefully acknowledge the contributions of those who commented and made suggestions on this document, including James SENG, and other JET members.

6. References

6.1. Normative References

- [LVT-SC] QIAN, H. and X. LEE, ".CN Chinese Character Table", IANA IDN Languages Tables, March 2005.
- [LVT-TC] LU, A., ".TW Traditional Chinese Character Table", IANA IDN Languages Tables, March 2005.
- [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.
- [RFC3743] Konishi, K., Huang, K., Qian, H., and Y. Ko, "Joint Engineering Team (JET) Guidelines for Internationalized Domain Names (IDN) Registration and Administration for Chinese, Japanese, and Korean", RFC 3743, April 2004.

6.2. Informative References

- [C2C] Halpern, J. and J. Kerman, "Pitfalls and Complexities of Chinese to Chinese Conversion", International Unicode Conference (14th) in Boston, March 1999.
- [RFC4290] Klensin, J., "Suggested Practices for Registration of Internationalized Domain Names (IDN)", RFC 4290, December 2005.

Authors' Addresses

LEE Xiaodong
CNNIC, No.4 South 4th Street, Zhongguancun
Beijing 100080
Phone: +86 10 58813020

EEmail: lee@cnnic.cn
URI: <http://www.cnnic.cn>

MAO Wei
CNNIC, No.4 South 4th Street, Zhongguancun
Beijing 100080

Phone: +86 10 58813055
EEmail: mao@cnnic.cn
URI: <http://www.cnnic.cn>

Erin CHEN
TWNIC, 4F-2, No. 9, Sec. 2, Roosevelt Rd.
Taipei 100
Phone: +886 2 23411313

EEmail: erin@twNIC.net.tw
URI: <http://www.twNIC.net.tw>

Nai-Wen HSU
TWNIC, 4F-2, No. 9, Sec. 2, Roosevelt Rd.
Taipei 100

Phone: +886 2 23411313
EEmail: snw@twNIC.net.tw
URI: <http://www.twNIC.net.tw>

John C KLENSIN
1770 Massachusetts Ave, #322
Cambridge, MA 02140
USA

Phone: +1 617 491 5735
EEmail: john+ietf@jck.com

Full Copyright Statement

Copyright (C) The Internet Society (2006).

This document is subject to the rights, licenses and restrictions contained in BCP 78 and at www.rfc-editor.org/copyright.html, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgement

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).

