

Internet Draft
<draft-deng-idn-icdn-00.txt>
July , 2001
Expires in six months

Authors: Xiang Deng

The Implementation of Chinese character in IDN

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."

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>

Terminology

The key words "MUST", "SHALL", "REQUIRED", "SHOULD", "RECOMMENDED", and "MAY" in this document are to be interpreted as described in [RFC 2119](#) [[RFC2119](#)].

Abstract

This document mainly talks about Chinese characters and two proposed schemes of implementation based on [[IDNREQ](#)] and [[NAMEPREP](#)], though there are some differences among them. The distinction between these two schemes is the position of the implementation function:

- client side processing

or

- server side processing

In China, the most popular character set are [GBK],[BIG5],[GB18030], while in this document, all examples are based on [UCS].

1. Characteristics of Chinese characters and Chinese language

1.1 The context dependent semantics of Chinese characters

In [UCS], each Chinese character is a codepoint, which is composed of two bytes.

Chinese character can be classified as two groups. In one group, each character does its own meaning (notional character) while that of the other group has not (empty characters). Both notional characters and empty characters can be made words by combining with other character(s), even sentences. The notional character is the basic unit of Chinese language which has meaning similar to phonemes.

1.2 A Chinese characters may have several writing forms.

Chinese characters were continuously evolved and widely spread during 5,000-year-long Chinese history. They were also largely introduced into other countries and became a major component of their languages. Therefore, it is inevitable for a Chinese character has many other writing forms. In Unicode encoding standards, the criterion for distributing codepoint is the shape of character. So the different glyph of the same Chinese character have several different codepoint according to the international encoding standard.

Currently, there are two forms of writing Chinese character:

- simplified character(SC): mainland of China
- traditional character(TC): Taiwan, Hongkong, Macao

Except for some special writing forms of certain character, their meaning had also been changed in the long history. Generally different writing forms of a Chinese character can substituted by each other without changing the meaning of the word (phrase).

1.3 The Usage of Appellation in China

In China, Generally speaking, every companies, organizations and people have two names: full name and abbreviation.

The abbreviated name is easy to remember and to communicate. The full name is a formal name which is used in formal document, situation.

To the name owners, the two names are equal necessary and important. So, in domain name registration, they usually register both full name and the corresponding abbreviations in order to permit people to access the same domain name by typing the full name or the abbreviated name. Some of the full names are quite long, that's why the length of domain name is important for Chinese user.

2. Chinese characters in DNS

1.1 Traditional and Simplified Chinese Conversion has 3 forms:

- 1-1 mapping: one traditional character(TC) maps to ONLY one simplified character(SC).**
- 1-n mapping: one TC has several SC writing forms**
- n-1 mapping: one SC has several forms of TC**

1.2 Delimiter folding

The full stop in chinese is "íú". Therefore, the "íú" in CDNS is equal to the dot "." as the delimiter.

1.3 Label sequence

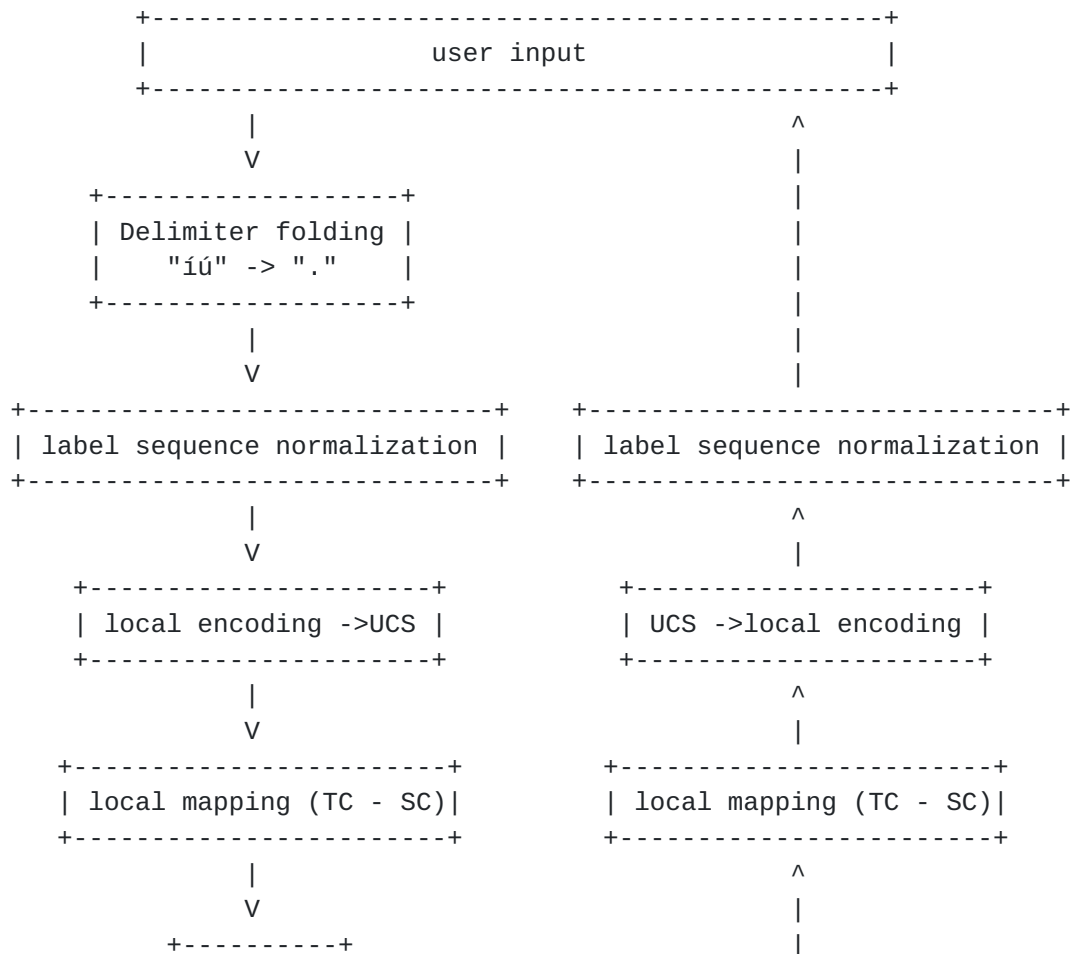
Currently, the label sequence of LDH domain name is from left to right, (e.g.:abc.def.ghi.net), the subdomain is to the left and the superset of the subdomain is to the right.

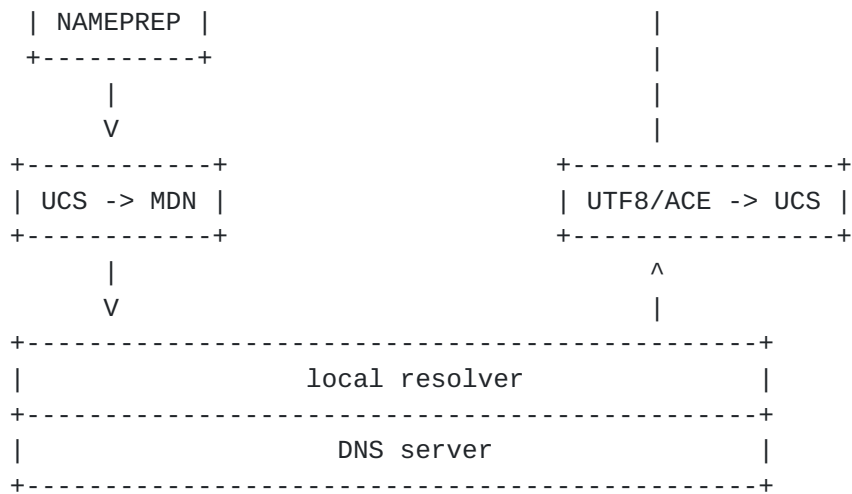
In China, user has reverse convention of language. Considering the culture different between the east and the west, it's necessary for people to access the Internet with the convention of using their native languages. for example:

```
abc.com.cn
perfer to :
cn.com.abc
```

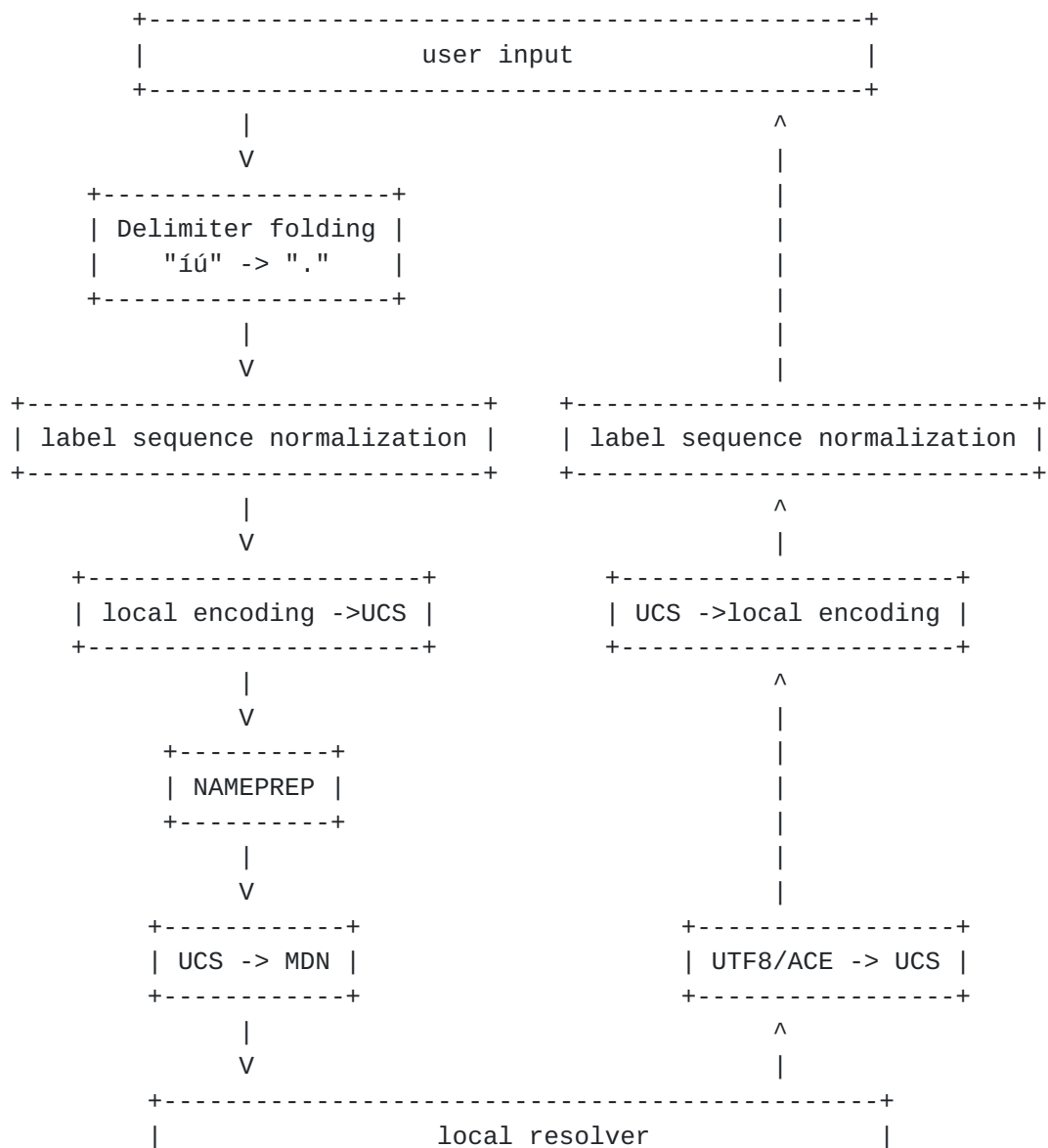
3. Solutions

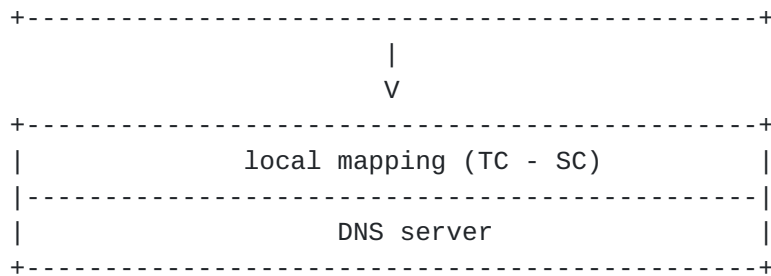
3.1 Client side solution





3.1 Server side solution





6 Authors' Address

xiang deng
China Internet Network Information Center
NO.4 South 4th ST. Beijing, P.R.China, 100080, PO BOX 349
Tel: +86-10-62619750

7 References

- [IDNREQ] Requirements of Internationalized Domain Names, Zita Wenzel,
James Seng, [draft-ietf-idn-requirements](#)
- [NAMEPREP] Paul Hoffman & Marc Blanchet, Preparation of
Internationalized Host Names, [draft-ietf-idn-nameprep](#)
- [RFC2119] Scott Bradner, Key words for use in RFCs to Indicate
Requirement Levels, March 1997, [RFC 2119](#).
- [STD13] Paul Mockapetris, Domain names - implementation and
specification, November 1987, STD 13 ([RFC 1034](#) and 1035).
- [UNAME] Internationalized Domain Names and Unique Identifiers/Names
Li Ming TSENG, Jan Ming HO, Hua Lin QIAN, Kenny HUANG
[draft-ietf-idn-uname](#)
- [TS CONV] Traditional and Simplified Chinese Conversion
Xiao Dong Lee, Nai Wen Hsu, Erin Chen, Guo Nian Sun
[draft-ietf-idn-tsconv](#)
- [ISO10646] ISO/IEC 10646-1:2000. International Standard -- Information
technology -- Universal Multiple-Octet Coded Character Set
(UCS) -- Part 1: Architecture and Basic Multilingual Plane.
- [Unicode3] The Unicode Consortium, "The Unicode Standard -- Version3.0",
ISBN 0-201-61633-5.