

# Report on Chinese Variants in Internationalized Top-Level Domains

*James Seng*

*on the behalf of the Chinese Case  
Study Team*



# Chinese Case Study Team

Name	Role
Xiaodong Lee	Case Study Coordinator
Chris Dillon	Team Member
Hong Xue	Team Member
James Seng	Team Member
Jian Zhang	Team Member
Jonathan Shea	Team Member
Joseph Yee	Team Member
June Seo	Team Member
Shian-Shyong Tseng	Team Member
Wei Wang	Team Member
Yangwoo Ko	Team Member
Yoshiro Yoneya	Team Member
Zhoucai Zhang	Team Member
Edmon Chung	Observer
Yang Yu	Observer
Steve Sheng	Case Study Liaison
Francisco Arias	Subject Matter Expert (Registry Operations)
Kim Davies	Subject Matter Expert (Security)
Nicholas Ostler	Subject Matter Expert (Linguistics)
Andrew Sullivan	Subject Matter Expert (Protocols)

# Scope of Work

- Focus on Chinese Variants at the TLD
  - Lower Levels
  - User Expectations
- Focus on Chinese Variants
  - *Might* have implications for Japanese and Korean
  - We include experts from Japan and Korea
- Focus on Unicode
  - Not included in Unicode is considered out of scope

# Han Script

## Han Script

---

Chinese Hanzi

漢字/汉字

Japanese Kanji

漢字

Korean Hanja

漢字/한자

- Hanzi, Kanji and Hanja are ideographs, each is a graphic symbol that represent an idea or concept
- In Unicode, Han script is unified into “CJK Unified Ideographs” (~26,000) and Extension A, B, C and D (~44,000)

# Chinese Hanzi

漢 汉  
字 字

- Chinese Hanzi originate from pictographs evolves into ideographs



- In 1964 & 1986, China introduces more than 2,000 Simplified Hanzi

Simplified Component

寶 → 宝

Simplified Radical

言 → 讠

Complex Component

龍 → 龙

Simplified Hanzi

聾 → 聋      課 → 课

- Resulted in two writing system using ONE script:
  - Simplified Chinese (SC) and Traditional Chinese (TC)
  - But not not always 1-to-1

發 → 发  
髮 → 发

# Japanese Kanji

書

- Japanese Kanji (漢字) were imported from China
- Japanese introduce “new character form”(新字体) derived from the “old character form”(旧字体) in 1923-49

鐵 → 鉄

學 → 学

國 → 国

寶 → 宝

- New and old character form are recognized as variants except when uses to express name of persons and places
  - New and old form is considered distinct in domain names

# Korean Hanja

漢字  
한자

- Korean Hanja (漢字/한자) were also import from China
- Hanja is no longer widely used in South Korea. Modern Korean uses Hanguel (한글) as a writing system.
  - As of 14<sup>th</sup> April 2011, all government documents can only be written in Hanguel, unless allowed by Presidential decree.
- Registry operator for .KR do not allow Hanja
- *Interesting Note:* Korean has also simplified Hanja known as Yakja (略字/약자).

無 → ㄴ

# Define Chinese Variants

“characters with different visual forms but with the same pronunciations and with the same meanings ...”

- Includes:

Simplified/Traditional

簡體 / 简体

Z-Variant

戶 / 户 黃 / 黄

- Does not includes:
  - Different spellings/form
  - Different translation, transliteration



# Chinese User Expectations

- Chinese regards Simplified and Traditional as equivalent and interchangeable
  - A mixture of Simplified and Traditional DNS queries on .中国/.中國
- Chinese Variant is important to Chinese IDNs

NIC	IDNs	Has Variants
CNNIC	320,000	77%
TWNIC	40,000	83%
HKNIC	24,000	85%

# Chinese & Japanese Variants Handling at the TLD

- TLD provides certain contextual indicator

学会.cn / 學會.cn

Expected as variant and the same

学会.jp / 學會.jp

Not variant and expected to be distinct

- But no contextual indicator for .学会 and .學會
  - Both Chinese and Japanese experts agreed to take a more conservative approach.
- An application for {学会} would results {學會, 學會, 学会, 学会, 学会, 学会} variants to be reserved. {學會} may be also be delegated for the Chinese.

# Issues with Language Variant Tables

- The need for Chinese Language Variant Tables for the Root Zone
- Whether IDN variants at TLD level should be based on language or script
- Considerations for a process to define the root Variant Tables
- The standard format for Variant Tables

# Evaluation, Allocation, Delegation and Operation of Chinese IDN Variant TLDs

- Evaluation Issues (when ICANN evaluate the TLD application), such as string similarity, conflict with geographic names, discovery of variants, etc
- Contention/Objection/Dispute Issues (when ICANN rules that there is a contention for the applied-for gTLD, or when objections or disputes are received for the new TLD)
- Allocation Issues (when ICANN decides which string(s) is/are to be allocated for the new TLD)
- Delegation Issues (when ICANN decides whether the applicant is the right authority to be delegated the new TLD string(s) and whether there is a fee involved)

# Other Impacts

Organizations	Considerations
IANA	Variant Tables, Variant Resolution Mechanism, IDN WHOIS standards
Root Server	Number of entries in the root zone
Registrars	Fees/Charges for Variant, EPP changes
Registrants	Registration Policy, Management of Variants
DNS Service Providers	Application Protocol (SMTP/HTTP/IMAP), Infrastructure support, etc
Software/ Application	Configuration for variants, Tools for variants handling, Provisions etc

# Questions

One World  
One Internet



**Dakar**  
SÉNÉGAL  
N°42 23 - 28 October 2011



Report Available at:

[http://www.icann.org/en/  
topics/new-gtlds/chinese-vip-  
issues-report-03oct11-en.pdf](http://www.icann.org/en/topics/new-gtlds/chinese-vip-issues-report-03oct11-en.pdf)

Public Comment Box:

[http://www.icann.org/en/public-  
comment/idn-vip-  
chinese-03oct11-en.htm](http://www.icann.org/en/public-comment/idn-vip-chinese-03oct11-en.htm)



# Thank You

