

Chinese Conversion The Wiki Way

paper by Zhengzhu
presentation by Theodoranian

Uniqueness of Chinese Char.

- Chinese writing system use Logogram, and each character has its unique meaning.
- ex. 魚 / 鱼 (yu) for Fish

“New” Chinese Char.

After the end of the Chinese Civil War, the “new Chinese” government introduced a whole new set of simplified characters in 1950s.

ex.

魚 → 鱼 (Fish)

中國 → 中国 (China)

Differs by Regions

- Non-Chinese words are commonly translated into Chinese by pronunciation. Sometimes by meaning.
- The translation norm may differ from different Chinese-speaking regions.
- ex. Laser:
 - Mainland China: 激光 (Ji Guang) Stimulated Light
 - Taiwan: 雷射 (Laser) Thunder Shot

Why Need Conversion?

- Reading different sets of characters is understandable, but maybe inconvenient.

Not Only 1-to-1

- Most Simp. Char. to Trad. are 1-to-1.
- However, there are about 200 commonly used Simp. Char. having multiple corresponding Trad. Char. depending on meanings.
- ex. 后(simp.) → 后 (Trad.) Queen ; 後(Trad.) Back

How Does It Work?

- Conversion depends on what the user has specified regardless of the original entries.
- For 1-to-1 conversion, a mapping table is used.
- For multiple correspondence, we add the phrase into the mapping table.
 - 1-to-1:后 = 後
 - Multi.:王后 = 王后 (prevent 后 from being convert to 後)

Customizable Table

- The initial mapping table is extracted from several open source Chinese input systems.
- Moreover, there is a policy for users to request for more conversion and the sysops edit the table.
- By users' requests, the conversion table is getting more and more complete.

Getting to “Perfect” Manual Conversion

- The auto-conversion table is not enough, because language grows and has many variabilities.
- Therefore, we use some markups to manually convert certain phrases and names.
- Markup: `-{zh-cn:aa;zh-tw:bb;zh-hk:cc}-`
- ex. Guinness Book of Records:

`-{zh-cn:吉尼斯世界纪录;zh-tw:金氏世界紀錄;zh-hk:健力士世界紀錄;
zh-sg:健力士世界纪录}-`

Deployment

- 📌 Zh.Wikipedia
- 📌 Zh.Wikibooks
- 📌 Zh.Wiktioary
- 📌 Zh.Wikiquote
- 📌 all projects after mediawiki upgrade 1.4+

To Other Languages

- 📌 Serbian Wikipedia with two writing systems may also implement this method.

Conclusion

- With more and more people participating in the process, we believe that wikipedia will provide better service in the future.
- It's the wiki way.

Thanx & Contact Info

- 📌 zhengzhu: zhengzhu@gmail.com
- 📌 theodoranian: theodoranian@gmail.com
- 📌 Please feel free to contact with him or me.