

Experience on Whois Database



Nai-Wen Hsu

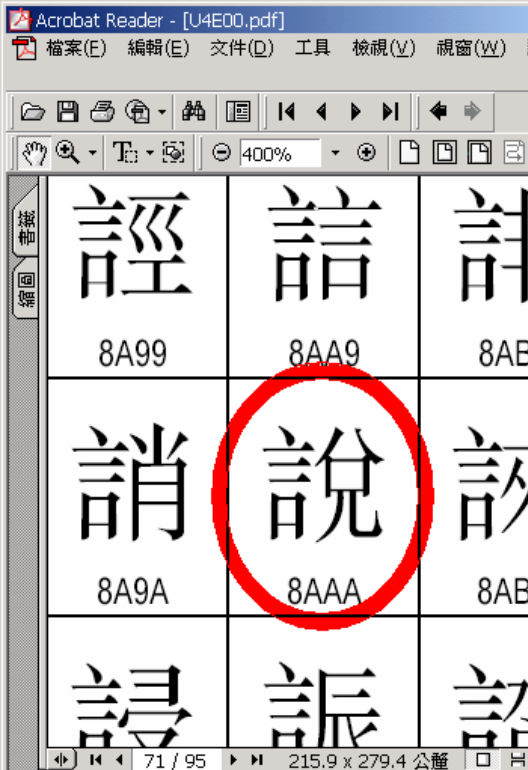
TWNIC

Jul. 21, 2004



Contents

- Variant table
- IDN Registration System/Policy
- IDN Whois Display
- IDN Whois Issues



There are 5 '說' in Unicode, 4 in Traditional Chinese, 1 in Simplified Chinese



Chinese Variant Table

- Valid Code Point (VCP)
- TW Preferred Variant (TWPV)
- CN Preferred Variant (CNPV)
- Character Variant (CV)

VCP	TWPV	CNPV	CV
說	說	说	说說說說說
說	說	说	说說說說說
说	說	说	说說說說說
說	說	说	说說說說說
說	說	说	说說說說說



TWNIC IDN Registration System

- Follow IDN guideline (RFC3447) applying language variant table
- Chinese Language variant table describes the mapping between Simplified Chinese and Traditional Chinese characters
- Taiwan, China, Japan and Korea(TCJK) each has its local encoding, and Taiwan local encoding based applications only display Traditional Chinese characters

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4E00	一 TCJK	丁 TCJK	丂 C	七 TCJK	上 C	丅 C	丆 C	万 TCJK	丈 TCJK	三 TCJK	上 TCJK	下 TCJK	丌 TC	不 TCJK	与 TCJ	丏 TC
4E10	丒 TCJ	丑 TCJK	专 C	专 C	且 TCJK	丕 TCJK	世 TCJK	卅 CJ	丘 TCJK	丙 TCJK	业 C	丛 C	东 C	丝 C	丞 TCJK	丢 TC
4E20	北 C	兩 CJ	丢 C	丕 C	两 C	严 C	並 TCJ	喪 C	丨 CJ	乚 C	个 CJ	丫 TC	丩 C	中 TCJK	乚 TC	丰 C
4E30	丰 TC	卯 TCJ	串 TCJK	弗 TC	临 C	業 C	、 CJ	丷 C	丸 TCJK	丹 TCJK	为 C	主 TCJK	井 TCJ	丽 C	举 C	丿 CJ
4E40	乚 C	乚 C	乚 TCJK	乃 TCJK	乚 C	久 TCJK	乚 C	毛 TC	么 TC	义 C	丩 C	之 TCJK	乌 C	乍 TCJK	乎 TCJK	乏 TCJK
4E50	乐 C	采 C	兵 TC	兵 TC	乔 C	厖 CJ	乖 TCJK	乘 CJ	乘 TCJK	乙 TCJK	乚 C	丩 C	乚 TC	九 TCJK	乞 TCJK	也 TCJK
4E60	习 C	乡 C	此 CJ	纟 C	丩 C	乚 C	书 C		乱 C	乚 TC	乚 C	乚 CK		乚 CK	乚 C	乚 C
4E70	买 C	乱 CJ	姿 C	乳 TCJK	乚 C	乚 C	乚 CK	乚 CK	乚 C	乚 C	乚 C	乚 C		乚 C	乾 TCJK	亂 TC
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4E80	龜 CJ	乾 C	亂 TCJK	𪔐 TC	𪔑 TC	丨 CJ	了 TCJK	尔 C	予 TCJK	争 CJ	事 CJ	事 TCJK	二 TCJK	丕 TC	于 TCJK	
4E90	丂 CK	云 TCJK	互 TCJK	丕 TC	五 TCJK	井 TCJK	三 C	𪔒 C	巨 CJK	互 TCJ	丕 C	些 TCJK	𪔓 CJ	𪔔 C	亞 TCJK	𪔕 TCJ



TWNIC IDN Registration System(cont.)

- EPP can support non-ASCII data by using UTF-8 encoding
- Store non-ASCII string and Domain name in database
- Translate IDN to punycode when build zone file from database

TWNIC WHOIS Elements

	Company/Organization Registrant		Individual Registrant	
	DB Element	Displayed by Whois	DB Element	Whois Display
Domain Name	V	V	V	V
Company ID	V			
Company Name -- Chinese	V	V		
Company Name-- English	V	V		
Registrant ID	V		V	
Registrant Chinese Name	V	V (Option)	V	V (Option)
Registrant English Name	V	V	V	V
Registrant Company Address	V	V		
Registrant English Address	V	V		
Registrant Chinese Address			V	V (Option)
Registrant TEL	V	V	V	V (Option)
Registrant FAX	V	V		
Registrant E-mail	V	V	V	V
DN Application Date	V	V	V	V
DN Expire Date	V	V	V	V
DNS Record	V	V	V	V
Registrar	V	V	V	V



Chinese Variant Table

- Valid Code Point (VCP)
- TW Preferred Variant (TWPV)
- CN Preferred Variant (CNPV)
- Character Variant (CV)

VCP	TWPV	CNPV	CV
台	台	台	台檯臺臺颱
網	網	网	網网
中	中	中	中
心	心	心	心

Percentage of 1-to-n mapping (1 VCP to N CV)

Number of character variant(s)	N=1	N=2	N=3	N=4	N=5	N=6	N=7	N=8
Number of Characters	13888 66.4%	5156 24.7%	1158 5.5%	424 2.0%	165 0.79%	60 0.29%	35 0.17%	16 0.08%

Each Valid Code Point may has more than one character variant



Chinese IDN Registration Policy- Chinese IDN Example

Original form : 台網中心.tw (punycode : xn--fiq43lrrlz83a.tw)

Traditional form : 台網中心.tw (punycode : xn--fiq43lrrlz83a.tw)

Simplified form : 台网中心.tw (punycode: xn--fiq43lrrlfy5a.tw)

Character Variant DN : 檯網中心.tw 檯网中心.tw 籐網中心.tw
籐网中心.tw 臺網中心.tw 臺网中心.tw
颯網中心.tw 颯网中心.tw

原型域名: 台網中心.tw (xn--fiq43lrrlz83a.tw)
繁體對照字域名: 台網中心.tw (xn--fiq43lrrlz83a.tw)
簡體對照字域名: 台網中心.tw (xn--fiq43lrrlfy5a.tw)
保留字域名(由下列相關字組合之域名):

台 檯 臺 颱
網 网
中
心

Registrant:

台灣網路資訊中心

Taiwan Network Infotmation Center

4F-2, No. 9, Sec. 2, Roosevelt Rd., Taipei

台北市羅斯福路二段9號4樓之2

Domain Name: 台網中心.tw

Contact:

許乃文

Nai-Wen Hsu snw@twNIC.net.tw

TEL: (02)23413300

FAX:

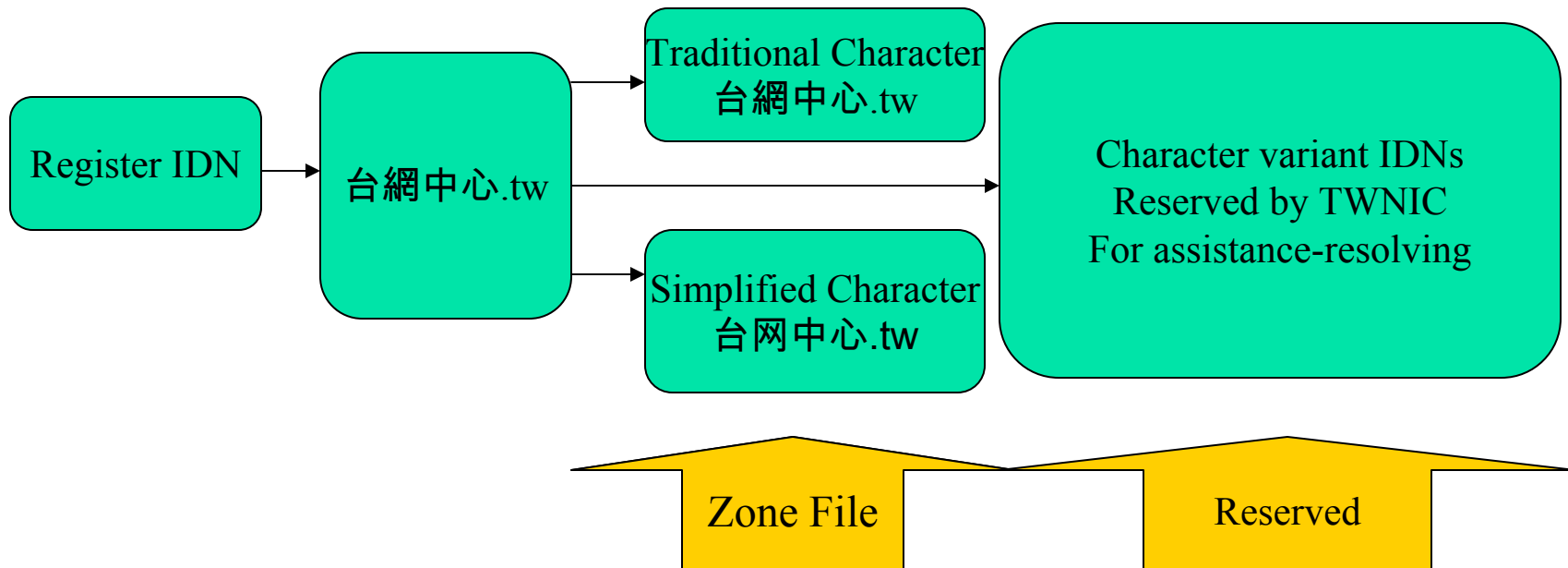
Record expires on (DD-MM-YYYY)

Record created on 14-02-2001 (DD-MM-YYYY)

Registrar: [TWNIC](#)

(reserved word)

IDN Registration



- 台網中心.tw, 台网中心.tw : put into zone file for resolving
- 檯網中心.tw, 檯网中心.tw, 籃網中心.tw, 籃网中心.tw, 臺網中心.tw, 臺网中心.tw, 颱網中心.tw, 颱网中心.tw: not put into zone file, reserved by TWNIC, other can not register these domain names



Whois client

- Most command line whois clients can handle Chinese characters, so users can read the output of whois on the terminal supporting Chinese character
- Traditional Chinese characters can only display on terminal supporting Chinese characters
- It is a problem when display the characters out of local encoding on command line mode



Whois client (cont.)

- Rwhois (RFC1714) is a multilingual aware protocol that has ‘-language’ option to tell server which language is preferred
- But most rwhois clients have not implemented the language option



Whois server

- To query a domain name that can be input in local encoding or punycode
- Whois output contain both Chinese and English characters



Issues of IDN Whois

- Each Whois query needs to check variant table
- Need to Display TWPV, CNPV and CV domain name simultaneously
- Lack of IDN Whois standard
- Should local encoding must be provided for Query/Display



Thank You!
