# $ISO/IEC\ JTC1/SC2/WG2/IRG\ N2333$

**Doc Type:** Ideographic Rapporteur Group Document

Title: Report on the encoding situation of Founder Population Information Font

Source: Kushim Jiang (姜兆勤)

Status: Individual Contribution to IRG #51

Action: For consideration by JTC1/SC2/WG2/IRG and UTC

Date: 2018-8-30

#### 1. Introduction

Entrusted by the *Ministry of Public Security* (公安部治安管理局), to solve the inclusion of uncommon characters in the *Basic Information Database of the Population* (人口基本信息库), Founder has made the special font and input method. (site)

The typeface named *Founder Song - Population Information* (方正宋体-人口信息) includes two fonts for characters in BMP and SIP. Beside the whole Extension B, the font for SIP also includes a small amount of characters in Extension C to E which can be found in *Table of General Standard Chinese Characters* (通用规范汉字表). And the font for BMP includes the whole URO and Extension A.

Font Block Number Font Block Number Extension A 6582 / 6582 **PUA** 4776 / 6400 **URO FZSong** 20902 / 20902 **FZSong** Extension B 42711 / 42711 **RKXX** URO+ 22 / 74 RKXX Extension C 44 / 4149 (BMP) **PUA** 4898 / 6400 (SIP) Extension D 8 / 222 **CJK** Compability 22 / 472 108 / 5762 Extension E

Table 1. Distribution of characters in the fonts

The PUA characters in BMP font include three parts:

- · All the PUA characters in SIP font;
- · All the PUA characters in GBK 1.0;
- · Some components.

After deleting the repeated glyphs, 4789 characters in PUA are considered.

#### 2. Duplicate Characters

#### 2.1. Duplicate Characters with UCS

The production of the font began in 2004, so some characters are duplicate with UCS. After the IDS check, most of the characters has been encoded or will be encoded.

Table 2. Distribution of characters in various blocks in UCS

Block	Number	Rate
Radical Supplement	2	0.04 %
URO	4	0.08 %
URO+	10	0.21 %
Extension A	238	4.97 %
Extension B	2126	44.39 %
Extension C	330	6.89 %
Extension D	24	0.50 %
Extension E	427	8.92 %
Extension F	118	2.46 %
Compatibility Supplement	4	0.08 %
WS2015	254	5.30 %
WS2017	366	7.64 %
WS2017 - GDM	120	2.51 %
WS2017 - GXM	162	3.38 %
Not encoded	886	18.50 %

#### 2.2. Duplicate Characters with SJ/T 11239-2001 and IRG N1531

[SJ/T 11239-2001] *Information technology - Chinese ideograms coded character set for information interchange - the 8<sup>th</sup> supplementary set (信息技术 - 信息交换用汉字编码字符集 - 第八辅助集) was published by the Ministry of Information Industry on Dec. 28<sup>th</sup>, 2001 for meeting the special needs of place name. The document was mentioned in the activity report (IRG N2187, Fig.1.6).* 

[IRG N1531] G source withdrew 3215 characters on Nov. 7<sup>th</sup>, 2008. Among them are many characters labeled CYY, meaning *Chinese Academy of Surveying and Mapping* (测研院). They are used for place name.

After the IDS check, the characters in the fonts and those in two documents overlap greatly.

Table 2. Distribution of characters in SJ/T 11239-2001 and IRG N1531

Document	Number	Total Number	Rate
SJ/T 11239-2001	623	2497	24.95 %
IRG N1531	682	3215	21.21 %

#### 2.3. Duplicate Characters with Xieyang Wang's Contribution

Before senting the document, Xieyang Wang (王谢杨) sorts out 183 characters to be submitted. 17 of them can be found in this font with a rate of 9.29 %.

Now there are 869 characters left.

#### 2.2. Duplicate Characters with fonts of Zhonghua Book Company

In SPUA-A plane of Zhonghua Book Company Song (中华书局宋体字库) there are 36493 characters. (<u>site</u>) 70 characters are found in this font.

Glyph in RKXX	Glyph in ZHSJ Song
垃	增
增	觜
示	示
瓘	瓘
蕉	蕉
進	逃
月阝	月肾
復	復
劫	敖
学	学

弦	就
嬞	嫜
琏	亞
瑹	臻
康	康
斌	址
羽虫	羽虫
鏦	鏦
圩	圦
塚	塚
破	砑
函	函
們	們
媄	僕
虓	虓
	-
翌	翌
径	径
維	催
叡	叡
杕	太
消し	消L
烺	烺

理	瑁
南瓦	南瓦
畝	畝
旺	臣
腔	腔
研	研
倔	砙
碼	碼
進	進
筆	筆
艠	艥
节	节
蒜	蒜
蘇	蔬
詰	詣
街	街
襌	翼
適	適
鎵	鋖
鑵	鑵
泥	阨
額	額
烋	烋
	•

俣	倶
熒	爽
荥	荥
洃	洨
隽	烤
寯	寯
珲	琏
爭	琤
瑶	瑶
枞	拟
翰	翰
描	祜
砍	砍
룉	相里

The glyphs collected in Zhonghua Book Company Song have their evidences from antient books.

## 3. Modifiable Characters

## 3.1. Unification

49 characters are considered to be unified or put to IVD.

Glyph in Font	Unifiable or IVD Glyph
叆	叆
壑	壑
埬	埭
凛	漂

或	彧
跃	跃
洩	洩
磜	磜
李	学
釗	侴
谡	误
谖	谖
堂	坐
法	瀅
嘉	巍
橪	烝
族鸟	跨
立	詝
滘	湾
<b>食習</b>	鳛
兠	兠
垫	垫
如	如
嫔	姓
宣長	寰
拴	拴
樑	樑

款	縠
濛	濴
<b></b>	汤玉
塚	琢
雑	<b>家</b> 生
賞	竟
謞	記
莊莊	<del></del>
<b></b> 复	复
莲	瑳
瀬	瀟
巉	巉
媢	妇
嫊	뿛
嫌	濂
桗	樑
對	羖
或	彧
滢	滢
貹	炷
媱	媱

The judgement is only based on glyphs, so further analysis should be carried out. Also, some original glyphs can be corrected.

## 3.2. Analogical simplification and complication

199 characters have simplified or traditional form characters in UCS.

Glyph in Font	Simplified or Traditional Glyph
漗	濶
靓	<b>清</b> 見
<b></b>	勲
玦	琜
琺	瑾
亚	斑
踌	璹
载	蔌
选	遶
钗	錢
铸	鍊
铓	鑅
镗	鑑
造	鎧
形	馬飛
嗣	
E	开始 飛飛
仉	倪
历	萬
凤	鳳

.     .       . </th <th></th> <th></th>		
湯     議       養     滿       時     體       體     問       問     問       選     滅       趣     類       好     好       好     女       好     女       女 </td <td><b></b></td> <td><b></b></td>	<b></b>	<b></b>
養     滿       满     中       愷     問       問     門       問     灣       水     滅       返     奶       好     好       好     好       好     好       好     好       好     好       好     媽       城     城       城     城       城     城       城     城	诗	許
满     滿       吃     億       問     問       周     灣       減     週       域     媒       好     媒       域     媒       域     媒       域     媒       域     媒       域     媒       域     媒       域     媒       域     媒	逷	言勿
時     障       億     6       周     周       海     溝       瀬     週       坂     域       妖     女       女     域       城     域       城     域       城     域       城     域       城     域	龚	<b>泛</b> 住
遺     饋       問     問       溶     溝       減     週       域     媒       好     域       好     域       好     域       域     域       域     域       域     域       域     域       域     域       域     域       域     域	满	瀛
周     問       商     閣       準     滅       遊     妈       奶     嫌       好     嫌       好     娥       奶     娥       奶     娥       媽     媾       媽     媾       媽     媾       媽     媽       媽     媽       媽     媽       媽     媽       媽     媽       媽     媽       媽     媽       媽     媽       媽     媽       媽     媽       財     婦       財     婦       財     婦       財     婦       財     婦       財     財       財     財       財     財       財     財       財     財       財     財       財     財       財     財       財     財       財     財       財     財       財     財       財     財       財     財       財     財       財     財       財     財       財     財       財 </td <td>庤</td> <td>喡</td>	庤	喡
闾     閣       準     溝       滅     選       坂     媽       妖     女       女     基       城     娥       好     媽       城     城       城     城       城     城	造	食豈
洋     溝       滅     週       ፴     ፴       ፴     ፴       ፴     ፴       ፴     ፴       姣     ፴       姣     ፴       Ø     ፴       媽     媾       滿     滿	周	害
減     減       週     類       妖     媒       女     虚       好     媒       好     媒       婚     媒       媾     媾	谱	喜
通     通       姐     媽       嫉     嫌       好     處       好     娥       好     媽       婚     媽       媽     媽       媽     媽	半	澕
姐     媽       姊     姚       妹     娥       婚     娥       娥     娥       婚     媾       媾     媾	潚	濾
姨     姨       好     da       树     烟       好     J       城     城		題
姨     姨       好     da       树     烟       好     J       城     城	奴	妈
好     嬧       娴     嫻       媛     ൃ       媽     ൃ       媽     ൃ	嫉	樂
娴     嫻       嫂     ൃ       媽     娲	姎	婡
嫂   孋     媾   嫋	烬	嬧
<b>媽</b>	如闵	嫂
	嫪	燦羅
<b> </b>	妨	女滿
/ /	娃	瑩
<b></b>	- 現	馬異
9來 馬來	9來	 馬來
調調		馬馬

纏	纏
段	璜
玲	瑜
<b></b>	琥
华	璍
顼	瑱
珂	麗
珠	琜
宝	寶
葉	瓅
营	蹭
璜	璜
樣	養
鞱	轁
段	曦
<b></b>	暊
觉	竟
敳	敳
焱风	凝風
规	煈
矫	燆
煋	爗
煌	燈

禕
音 電 龍龍
釲
舒
氫
詠
鉀
銢
鉅
鈴
銝
鏡
鑅
鏛
稦
確烏
為
出售
洪高
蹟
顕
韻
艦

	T
<b></b>	穦
<b>洮</b> 东	漣
鉱	鑑
貯	貯
锌	銉
径	聖
笼	質
浡	·
<b></b>	漫
埃	換
- 場	墤
熽	埔 (sim.)
農	太(sim.)
晕	掌
忉	恒
	悓
<b></b>	藝
脯	月滿
桦	櫸
换	換
汶	瀁
焊	块 (variant)
<b></b>	駂

犡
押岡
琸
瓐
琘
蓝
議
<u></u>
篇
築
義
蔘
評
賚
机
軚
<b>毕</b>
釿
晃
紊
爏
諌
言樂

历3	歷
<b></b>	長
#	莊
蘇	蘇
蝉	嘽
渱	洍
澆	瀗
恻	惧
烁	樂
帧	慎
骄	獢
蚁	嬟
姓	女 蒙
遊	婕
嬣	嬣
類	娯 (variant)
晴	馬青
⊒昷	馬昷
網	網
绺	絡
現	瑪
现	<b></b>
五六	頂

理     環       理     環       現     機       機     機       極     横       極     積       毒     輸       ፍ     育       時     環	
環     環       城     機       横     横       積     積       毒     輸       肓     噴	
嫉     櫦       概     機       極     槓       毒     毒       翰     翰       诉     噴	
機     機       極     槓       毒     轟       翰     翰       诉     噴	
極     槓       毒     轟       翰     翰       唉     噴	
毒   轟     翰   翰     喉   噴	
<b>翰</b> 翰	
<b></b>	
HE	
飏	
<b></b>	
<b>遊</b>	
<b>農</b>	
/ / / / / / / / / / / / / / / / / / /	
<b>浴</b>	
· · · · · · · · · · · · · · · · · · ·	
鸣鴫	
<b></b>	
<b></b>	
· · · · · · · · · · · · · · · · · · ·	
<b>挥</b> 鍕	
<b></b>	
<b></b>	

稀	豨
铸	踳
铝	鍢
湿	録
铝	金色
辏	臻
基	鍖
镖	鑂
学与	鸰
頭	頭
筑	簠
镔	鍭
挦	鎅
锄	劍 (variant)
输	瀭
绝	料图
媊	嫺
珠	馬末

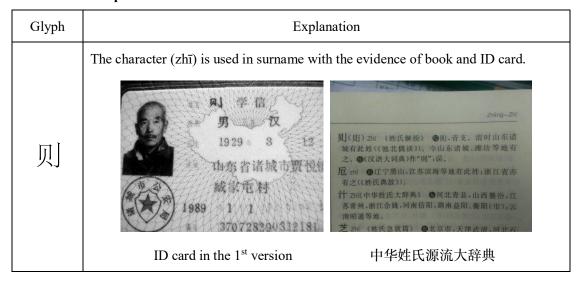
## 3.3. Glyph correction

According to the available information, 1 character needs to be corrected.

Wrong	Correct	Explanation
崔	裎	The character is found in Changyo, Tiantai, Taizhou, Zhejiang. From the picture taken locally, the right one should be the right one. It is submitted in WS2017, UK-10945.

## 4. Encodable Characters

#### 4.1. Characters for person name



#### **4.2. Character Components**

10 components cannot be unified with existing components.

Glyph	Explanation
F	It can be found in 比此毕阰即庇邸玭吡坒妣屁庇批沘纰芘毕姕舭仳毕呰豼跂秕
	昆枇柴毞玭肶奃妝皆毗欪哛眭豈赀毕粃忠箟妛毗傾傾藬廎噸澒嶺澃蹞顄鎮

	It can be found in radical 骨 and so on.
<i>/</i> \	It can be found in
	<b>举</b> 其咫 <u>回</u> 委宾
人	It can be found in 丛众众华晋平匹两坐夾巫纵苁苁纵逊凤纵纵以发來卒怂枞众
	光城赤炭炭队孙帝瓮刺侳俠俩剉匧疭诬郟陝奙浹凾扊座厜唘褭夹帝犖投挫쬗
	野唑 <u>价</u> 、似则 不 网 别
チ	It can be found in 并邢刑妍岍形汧秆併栞弄併枅郱屛帡研进剏型洴盂瓶笲胼惼
	新
	謂 <b>鴘</b> 鮮蘢斷姘弃拼栟
<b>⇒</b>	It can be found in 尹肀 3 君 幸伊 市 异 吚 君 芛 吚 君 莱 伊 哥 東 尹 東 東 東 東 東 東 東 東 東 東 東 東 東 東 東 東 東
	事有那郡姆希蒂伊 <del>斯</del> 東雅峱
车	It can be found in 轧轨轮轩轪轫靰轴钥匦斩转轭轮软郓轩軗钘崅枥钠軞蚔報粒
	轱轲轳轴轵轶轷轸轹轺轻牻钹羚轮钟银轼载轾轿辀辁辂较铫絣谻輆啭堑
	68 车(车) 58
	新华字典 (11 <sup>th</sup> version)
艮	It can be found in
	- - - - - - - - - - - - - - - - - - -
	概暨蛾
畫	It can be found in 畫畫圕書書書書書書書書書書書書書書書書書書書書書書書書書書書書書書書書書書書
	澅忙畫
	[145](主) 93
	新华字典 (11 <sup>th</sup> version)
辛	It can be found in 辦辦辨辨辨辨辨辨辨辨辨辦辦辦辦辦辦辦辦辦辦

### 5. Conclusion

The font for government departments shows that characters with GDM and GXM source are being used in population information database. Judgement should be based on evidence, at the same time, attention should be paid to actual use.