

Universal Multiple-Octet Coded Character Set

UCS

ISO/IEC JTC1/SC2/WG2 IRGN2538R

Date: 2022-02-28

Source:	Macao Special Administrative Region, China
Title:	Submission of One Character for Macao Supplementary Character Set (MSCS)
Meeting:	IRG#58
Status:	Member's submission
Actions required:	To be considered by IRG
Distribution:	IRG Members and Ideographic Experts
Medium:	Electronic
Pages:	3
Appendix	1

1 Submission of One Character for MSCS

Macao Special Administrative Region Government (MSARG) proposes to encode the 𠄎 character which has been used in personal name and published on ID card.

According to the existing unified cases in the character code chart of Unicode 14.0, this character should be unified to U+5F50, as shown in Figure 1. Its glyph is the same as UTC-01005.

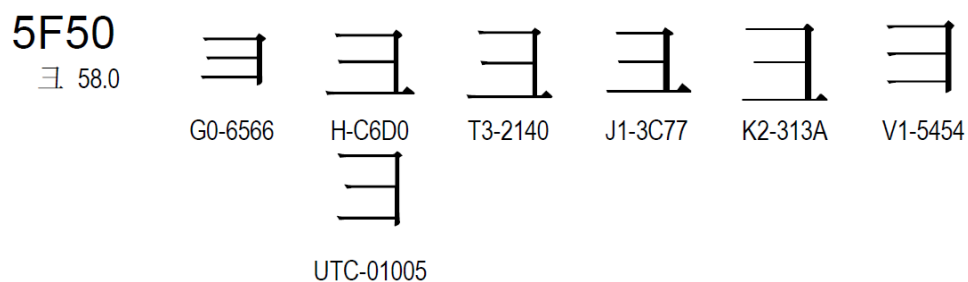
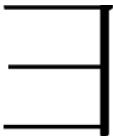


Figure 1 Representative glyphs of U+5F50 in Unicode 14.0 Character Code Chart

However, the Pinyin reading of the 𠄎 character used in Macao is XUE, which indicates it is the simplified form of “Snow” (雪 U+96EA), likely the character in China’s Second Chinese Character Simplification Scheme¹. In this case, MSARG proposes to encode the 𠄎 character as a new character rather than being unified to U+5F50.

The information of the 𠄎 character is as follows:

Source Reference	PUA ²	Macao's Glyph	Remarks
MC-00140	U+EEBE		Usage: personal name Pinyin: XUE

The image and evidence of the character are also provided³.

2 Related IRG Document

The document IRGN2509 also proposed to disunify U+5F50 and encode two new characters with the representative glyphs following the 𠄎 and 𠄏 standards.

3 Contact Information

Any queries about the information of the characters, please contact: Mr. Clement Chau (cchau@safp.gov.mo) and Ms. Leong I Man (leongiman@safp.gov.mo). We greatly appreciate the effort of all IRG members and ideographic experts.

¹ The Second Chinese Character Simplification Scheme (《第二次漢字簡化方案(草案)》) was published by PRC government in 1977 and officially abolished in 1986.

² MSCS assigns a PUA code point in the designated area for interchange in government systems before it is officially encoded in ISO/IEC 10646.

³ This character was approved by the Working Group of Chinese Characters Encoding for the Information Systems of Macao SAR on February 23, 2022 and it is planned to officially publish it in the next version of MSCS.

Appendix – Image and evidence of the character

End of document

Universal Multiple-Octet Coded Character Set
 International Organization for Standardization
 Organisation Internationale de Normalisation
 Международная организация по стандартизации

Doc Type: Ideographic Research Group Document
Title: Feedback on IRGN2538
Source: Eiso Chan (陈永聪, Culture and Art Publishing House)
Status: Individual Contribution to online IRG #58
Action: For consideration by IRG, MSARG & UTC
Date: 2022-02-17

Macao SARG requests for a UNC, which should be disunified several sources from U+5F50 in fact. At least, we know MC-00140 (same as UTC-01005) is needed in mainland China, Macao SAR and Việt Nam.

The evidence provided by Macao SAR includes two pieces of information: the Putonghua reading and the Macao telegraph code. Please see the following picture.

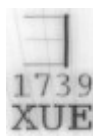


Fig. 1 Submitted evidence

The Putonghua reading without the tone mark is XUE, which is very clear to show it's the variant of U+96EA (雪). This is strong to make it not unify to U+5F50, and I still support to encode a new character.

The current Macao telegraph code system was issued by Vasco de Almeida e Costa (高斯達), the former governor of Macao, as Decreto-Lei n.º 88/85/M in 1985, which the official Portuguese name is *Aprova o silabário codificado de romanização do cantonense* (核准關於密碼及廣州音譯音字彙). This system was derived from the one used in mainland China since the end of Qing Dynasty, but something had been changed and added in this system. The submitted evidence cited from the Macao ID card (Bilhete de Identidade de Residente Macau in Portuguese) shows the telegraph code is 1739. When we check the No. 1739 character in the original Macao telegraph book, we can get the following.

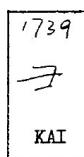


Fig. 2 No. 1739 in Macao telegraph book

The Latin alphabet string at the bottom of the square is used to show the Macao Romanization value of this character. KAI means that No. 1739 is the same as the corresponding one in other Chinese telegraph books in mainland China, Hong Kong SAR and Taiwan Province. It's the real "LONG HAND" as Table 1 in IRGN2509 shows. The unification rule in the telegraph book is different from Unicode/UCS. In telegraph books, the simplified character and the corresponding traditional character should use the same telegraph code. The telegraph code of U+96EA (𠄎) in Macao telegraph book is No. 7185 which is the same as mainland China, Hong Kong SAR and Taiwan Province. Please see the following picture.

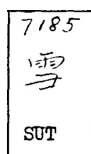


Fig. 3 No. 7185 in Macao telegraph book

The note to MSARG: It's better to help the card-owner update the telegraph code to 7185 or assign a new code for the new character. If the new character been encoded in Unicode / UCS, the related properties and tags for the person name should be updated and added for the daily use.

The note to UTC: It's not better to add the *possible* `kMacaoTelegraph` property value for the new character as 1739, and we should keep 1739 under U+5F50, if any expert requests to add the `kMacaoTelegraph` property to Unihan Database, such as me. If Macao SAR assigns a new telegraph code for the new character, we should use the new code; if not, it's better to keep empty for it.

(End of Document)

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/IRGN2543 and IRGN2538_TCA Feedback

Date: 2022-03-11

Doc. Type:	Member body contribution
Title:	TCA's Feedback to IRGN2543 and IRGN2538
Source:	TCA
Status:	Input to IRG meeting 58
Action:	For consideration by JTC1/SC2/WG2/IRG
Distribution:	IRG Members and Ideographic Experts
No. of pages:	2
Appendixes	None
Medium:	Electronic

TCA has some thoughts on the proposal to separate the following three character (see figure below) into separate codes as mentioned in this document, as described below.

SNOUT	𠂇
HAND	𠂆
LONG HAND	𠂇

Table 1. The characters

5F50 𠂇 58.0

G0-6566 𠂇 H-C6D0 𠂇 T3-2140 𠂇 J1-3C77 𠂇 K2-313A 𠂇 V1-5454

UTC-01005 𠂇

CJK Radicals Supplement

2E95 𠂇 CJK RADICAL SNOUT TWO

→ 5F50 𠂇

IRGN2509 page 4 & ISO/IEC 106464 code chart

For the three characters in the picture above, TCA considered that it is only necessary to split into two codepoint, because the meaning of both characters is the same (hand 𠂆 and long hand 𠂇), both are evolved from the ancient character "又(𠂇)" and the modern Chinese character "手".

U+5F50 was discussed in IRG #57, the decision was made to prefer separate coding, if China and Vietnam agree, then 𠂇(snout) and 𠂇(hand) will be coded separately (Refer to IRGN2500). William T. Nelson proposed that long hand (𠂇) is also used as a variant of snout(𠂇) mentioned in Japanese dictionaries, it can be handled by IVS because the two forms are too similar.

In addition, for IRGN2538, the name urgent character(≡) applied for in Macao should not be a new character, it can be solved by applying for horizontal expansion, either with the new code or with U+5F50.

(End of document)

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2 IRG N2538

Date: 2022-03-13

Source:	China
Meeting:	IRG#58
Title:	Feedback on IRGN2538 by China
Status:	Member's submission
Actions required:	To be considered by IRG
Distribution:	IRG
Author:	TAO Yang
Medium:	Electronic
Pages:	3
Appendixes:	0

Part 1 Traceability of difference causes

It is a good way to explore this problem from the original source.

As it shows in GB 18030-2005 信息技术 中文编码字符集, U+5F50 is the radical character of 彗/彗/彗. This relationship can be seen more clearly in the code chart.

E5	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A		灇 6FC9	灈 6FA7	灉 6FB9	灊 6FB6	灋 6FC2	灌 6FE1	灍 6FEE	灎 6FDE	灏 6FE0	灑 6FEP	灒 701A	灓 7023	灔 701B	灕 7039	灖 7035
B	灗 704F	灘 705E	灙 5B80	灚 5B84	灛 5B95	灜 5B93	灝 5BA5	灟 5BB8	灠 752F	灡 9A9E	灢 6434	灣 5BE4	灤 5BEE	灥 8930	灦 5BF0	灧 8E47
C	灨 8B07	灩 8FB6	灪 8FD3	灭 8FD5	灮 8FE5	灯 8FEE	灰 8FE4	灱 8FE9	灲 8FE6	灳 8FF3	灴 8FE8	灵 9005	灶 9004	灷 900B	灸 9026	灹 9011
D	灺 900D	灻 9016	灼 9021	災 9035	灾 9036	灿 902D	𠄎 902F	𠄏 9044	𠄐 9051	𠄑 9052	𠄒 9050	𠄓 9068	𠄔 9058	𠄕 9062	𠄖 905B	𠄗 66B9
E	𠄘 9074	𠄙 907D	𠄚 9082	𠄛 9088	𠄜 9083	𠄝 908B	彡 5F50	彣 5F57	彤 5F56	彥 5F58	𠄞 5C3B	𠄟 54AB	𠄠 5C50	𠄡 5C59	𠄢 5B71	𠄣 5C63
F	𠄤 5C66	𠄥 7FBC	𠄦 5P2A	𠄧 5F29	𠄨 5F2D	𠄩 8274	𠄪 5F3C	𠄫 9B3B	𠄬 5C6E	𠄭 5981	𠄮 5983	𠄯 598D	𠄰 59A9	𠄱 59AA	𠄲 59A3	

GB 18030-2005 信息技术 中文编码字符集

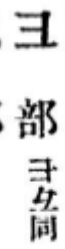
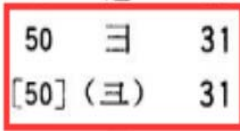
5F50 彡 𠄎	𠄎 𠄎 𠄎 𠄎 𠄎 𠄎	5F58 彥 𠄞 𠄟 𠄠 𠄡 𠄢 𠄣 𠄤 𠄥 𠄦 𠄧 𠄨 𠄩 𠄪 𠄫 𠄬 𠄭 𠄮 𠄯 𠄰 𠄱 𠄲 𠄳 𠄴 𠄵 𠄶 𠄷 𠄸 𠄹 𠄺 𠄻 𠄼 𠄽 𠄾 𠄿	𠄞 𠄟 𠄠 𠄡 𠄢 𠄣 𠄤 𠄥 𠄦 𠄧 𠄨 𠄩 𠄪 𠄫 𠄬 𠄭 𠄮 𠄯 𠄰 𠄱 𠄲 𠄳 𠄴 𠄵 𠄶 𠄷 𠄸 𠄹 𠄺 𠄻 𠄼 𠄽 𠄾 𠄿
----------------	----------------------------	---	--

Code chart of URO

Obviously, only the glyph from G-source is different from other sources. This difference needs to

find an explanation path from the development history of dictionaries.

According to the timeline, the evolution of this radical has also become clear.

Dictionary Name	Radical in Checklists	Head characters																																												
正字通 ZhengZiTong																																														
康熙字典 KangXi Zidian																																														
中华大字典 Zhonghua Dazidian																																														
新华字典 Xinhua Zidian		 <table border="1" style="display: inline-table; vertical-align: top;"> <tr><td>当郎</td><td>1128</td></tr> <tr><td>录</td><td>2175</td></tr> <tr><td>彣</td><td>3025</td></tr> <tr><td>彣</td><td>4469</td></tr> <tr><td>(彣)</td><td>3024</td></tr> <tr><td>(彣)</td><td>6164</td></tr> <tr><td>彣</td><td>6164</td></tr> </table>	当郎	1128	录	2175	彣	3025	彣	4469	(彣)	3024	(彣)	6164	彣	6164																														
当郎	1128																																													
录	2175																																													
彣	3025																																													
彣	4469																																													
(彣)	3024																																													
(彣)	6164																																													
彣	6164																																													
现代汉语词典 Xiandai Hanyu Cidian		 <table border="1" style="display: inline-table; vertical-align: top;"> <tr><td>50</td><td>彡部</td></tr> <tr><td>归</td><td>487</td></tr> <tr><td>彣°</td><td>193</td></tr> <tr><td>当°</td><td>258</td></tr> <tr><td></td><td>261</td></tr> <tr><td>寻</td><td>1483</td></tr> <tr><td>灵</td><td>823</td></tr> <tr><td>彣</td><td>1695</td></tr> <tr><td>彣</td><td>583</td></tr> <tr><td>(寻)</td><td>1483</td></tr> <tr><td>(寻)</td><td>1483</td></tr> <tr><td>(归)°</td><td>487</td></tr> <tr><td>彣</td><td>1606</td></tr> <tr><td>(彣)</td><td>1606</td></tr> </table> <table border="1" style="display: inline-table; vertical-align: top; margin-left: 20px;"> <tr><td>[50]</td><td>彡部</td></tr> <tr><td>彣</td><td>1322</td></tr> <tr><td>彣</td><td>1682</td></tr> <tr><td>(彣)</td><td>580</td></tr> <tr><td>(彣)</td><td>1537</td></tr> <tr><td>彣</td><td>1537</td></tr> <tr><td>彣°</td><td>793</td></tr> <tr><td>彣</td><td>796</td></tr> </table>	50	彡部	归	487	彣°	193	当°	258		261	寻	1483	灵	823	彣	1695	彣	583	(寻)	1483	(寻)	1483	(归)°	487	彣	1606	(彣)	1606	[50]	彡部	彣	1322	彣	1682	(彣)	580	(彣)	1537	彣	1537	彣°	793	彣	796
50	彡部																																													
归	487																																													
彣°	193																																													
当°	258																																													
	261																																													
寻	1483																																													
灵	823																																													
彣	1695																																													
彣	583																																													
(寻)	1483																																													
(寻)	1483																																													
(归)°	487																																													
彣	1606																																													
(彣)	1606																																													
[50]	彡部																																													
彣	1322																																													
彣	1682																																													
(彣)	580																																													
(彣)	1537																																													
彣	1537																																													
彣°	793																																													
彣	796																																													

According to the principle of merging approximate radicals to facilitate shape near glyphs retrieval,

the characters contain components 𠃉/𠃊/ 𠃋/ 𠃌 are integrated into one dictionary catalogue. The traditional dictionaries determined the main radical as 𠃋, then the glyphs of sources except G are in accordance with traditional regulations. But characters have undergone a long period of reform in mainland China, some of the characters contain 𠃋 are integrated to 𠃉, then the quantity of characters with component 𠃉 become a majority, the modern dictionaries of mainland China accept this adjustment and use 𠃉 as the radical glyph. That's why only the G glyph is different to the others.

Part 2 Suggestions on the use of ID in Macao

Returning to the question raised by the representative of Macao.

The glyph 𠃉 has been used as a person name on ID card. Unify the glyph to U+5F50 or add an UNC, bases on whether the civil affairs system of Macao compiles this glyph on code position 5F50. If yes, it would be better to unify to 5F50 and add a M-source. If no, adding to a new code point is feasible.

Part 3 China's opinion on G-source glyph

Previously, China believed that according to the radical rules of traditional dictionaries, the glyph should be modified to 𠃋 to keep the code chart consistent, and 𠃉 should added as a new character on another code point.

However, this idea does not take into account the risks in practical applications.

The technicians of the Ministry noticed also noticed the proposals on dis-unifying U+5F50 𠃉 (IRGN2414, IRGN2509, IRGN2538, IRGN2543) and thought it would be a great impact for affecting 56,000 person names and 400 place names in the ID system. Thus, keeping it unchanged or a compromise is hoped.

(End of document)