

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/IRGN2546RR

Date: 2022-03-14

Doc. Type:	Member body contribution
Title:	Response to normalization and meaning issues on TCA characters in WS2021
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:	5
Appendixes	None
Medium:	Electronic

In the IRG Working Set 2021 review process, regarding the TCA glyphs and the graphic evidence issues, we thank you for your comments and in order not to make the response too scattered, we respond through this document as follows.

1. About the names character's proof

In the WS2021, TCA submits the characters used for personal names. The proof of glyphs is not just a simple table, but the data produced from the Household Service System (see [IRGN2486](#)). Taiwan's household service system has been computerized for a long time, and the name on each person's ID card is printed from the font of this system. [Website link to the table of English codes for each county and city in Taiwan.](#)

In the early days, there was no requirement for dictionary source for the personal name registration if the character used was not yet encoded. At that time, if the computer could not key in the character, it was added by handwriting, and then the character was made up.

After the announcement of the new “Name Act” by the Taiwan government in 2015, as long as **the character does not have a dictionary source, it is usually not allowed to be registered as a personal name character.** Therefore, TCA has no way or difficulty in providing additional information on the personal names that were approved for use before the publication of the “Name Act”.


What TCA is trying to say is that the purpose of coding is not only computerization, but also data exchange. The names submitted by TCA are the names of living people, and it is not only used by a few people, these characters are provided by the Department of Household Registration for higher frequency of use.

UK believes that once a citizen with a unique name character dies, that coded

character is no longer needed. This is a cognitive error; in Taiwan, the law requires that information related to a person's name be retained for seven generations.

The name characters are not for private use, the name system is the basis of all government operations. For example, the school enrollment notices for children are printed and sent to parents and schools through the system; the electoral roll and the notification list of eligible voters are also generated through the system, and notices are sent to the public and printed in the register for confirmation of information when voting. The National Health Insurance Card and Worker's Insurance are all linked to the Household Management System. So, name characters are used in schools, banks, and government documents. If these names cannot be encoded, computerization cannot be completed and data exchange cannot be done.

The name of an ancient person can be coded because it has been recorded in ancient literature or dictionaries; the Jade Emperor, the alias of Yuhuang dadi (玉皇大帝) of Daoist-usage characters(UK-20704), which no one knows, can be coded; the name of a living person, which has practical applications and information exchange needs, is not considered suitable for coding. Then TCA would like to ask, what is the purpose of the current coding? And what is the

high usage rate of the alias Jade Emperor (UK-20704)  UK-20704? Or is the name of a living person used more often?

2. IRG's experts are afraid of the wrong characters

IRG experts no longer have to wonder if the characters used in the names submitted by TCA are a wrong character, because they have already entered the government's public affairs system. There is a case in Hong Kong, the 2017 Hong Kong sister 雷莊兒, the name is wrongly written "兒" to "兜", she also did not ask to change back. In addition, the SAT submitted the characters of WS 2015, WS 2017 and WS2021, there are many characters that are said to be a wrong character in the evidence, but the IRG still accepted the them.

3. About pronunciation and meaning

Taiwan's ID card sample is shown below, the main information on the front is the name of the person, the back is the name of the parents and spouse, no pronunciation information.



ID card front side



ID card back side

Figure 1 Taiwan's ID card (sample)

In addition, TCA is unable to provide information on the pronunciation of some characters, because users did not be required for the pronunciation when they registered, and we do not want to give a pronunciation freely. However, we would like to highlight the fact that sometimes the personal names only have the shape of the character rather than the sound or meaning. For example, in the Kangxi Dictionary, some characters are only explained by the names of people or of places (Figure 2). Nevertheless, we will do our best and add more information if we find it later.



Figure 2 The names of people or of places in Kangxi

4. About glyph normalization

ISO/IEC 10646 is a character-based register, and each character is allowed to

have a different glyph, which is why ISO/IEC 10646 codes are presented in multiple columns side by side.

For TCA, normalization is limited to characters for education use, but otherwise remains the same as its source form. The range of characters used in education includes “國字標準字體表”, “次常用字國字標準字體表”, and characters that have been identified by experts at the Ministry of Education as deserving normalization, such as characters used in chemistry. We have already explained the form of the character "辶" ([IRGN 2347](#)), and now we will take the form of "辶" as an example for the personal names. We cannot change it arbitrarily. If we change the form of the character "辶", it will not be used on the ID card, and it means that we will create a glyphs that no one uses.



Figure 3 Unicode 14.0 version

The parts of 辶 and 辶 are also inconsistent in many member body's' characters, the same as in the case of TCA. Most of the characters discussed in the IRG are no longer used for common purposes, but for special purposes, including personal names, location names, historian studies, and philological purposes. Therefore, the more the original glyphs are introduced, the better!

	部首序號	部首正形	部首變體				
			一	二	三	四	五
◆	140	艸	艸	艸	艸	艸	艸
			艸				

Figure 4 Various forms of 艸

From the viewpoint of encoding, when a character has both a regular form and a variant form, the best way is to give the official code to the regular form, while the variant form is treated as IVS. It is impossible to create a regular

form in order to encode a variant. Take U+2A6DE for example.



When the Macao's application was made for the urgent character ([IRGN2433R](#)), TCA had two similar glyphs (TC-622B and TD-7A2D).

Although the form of TD-7A2D (彳 with 3 strokes) is the same as the Macao glyph, the 彳 form of TC-622B is used as a split stroke, so the code application was finally submitted with TC-622B(彳 with 4 strokes). In the future, maybe TCA will process TD-7A2D (彳 with 3 strokes) as IVS.

The figure shows two screenshots of the CNS 11643 Website. Both screenshots display the character '琛' (琛) in three different fonts: 楷體, 黑體, and 宋體. The top screenshot shows the character '琛' (琛) with the CNS code 12-662B highlighted in a red box. The bottom screenshot shows the character '琛' (琛) with the CNS code 13-7A2D highlighted in a red box. Both screenshots also show the character's phonetic information, including its pronunciation in various systems (e.g., 注音符號, 注音符號二式, 通用, 漢語, 耶魯, 威妥瑪) and its stroke order information (e.g., 筆畫, 倉頡).

Figure 5 CNS 11643 Website

(End of document)