

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

Doc Type: Working Group Document
Title: Comments on encoding Shuowen Small Seal
Source: TCA and China
Status: Proposal
Action: For consideration by JTC1/SC2/WG2 and UTC
Reference N4853_Shuowen Seal Ad Hoc Meeting Resolutions
Date: 2018-06-12

Please refer to WG2/N4853 for Shuowen Seal Ad Hoc Meeting(28-29th Aug, 2017) Resolutions. There are two parts remained to be processed:

1. Q2: How should duplicate glyphs be handled in a standard encoding?
 - There is agreement to define “duplicate” with a typology, using terms such as (A) “exact duplicate”, (B) “possible duplicate”, and (C) “non-duplicate”.
 - There is agreement to handle items (A) and (B) using a VS scheme to be determined. Items in (C) would not be suitable for unification.
 - There is agreement to create tables to document the decisions made with respect to the handling of specific duplicates.
2. Q3: Compatibility with existing reference and research implementations.
 - There is agreement that 藤花樹本 may serve as “baseline” , for the determination of the repertory (the number of rows / code points), and for production of a single-column chart in 藤花樹本 glyph style.
 - There is agreement to create multi-column charts (for reference in ISO/IEC JTC1/SC2/WG2 N4853 developing the Shuowen Seal encoding methodology, repertory, and property data), including columns representing the glyphs of specific important editions of Shuowen. The multi-column chart data will include at minimum 藤花樹本 and 陳昌治本 columns (glyphs and mappings), but may also include data for additional columns as available.
 - There is agreement that issues related to 避諱 (taboo character deformation / glyph substitution) will be addressed in conjunction with

multicolumn chart development (perhaps by means of a VS scheme).

- Details of the encoding methodology, script documentation, character repertory, character properties (print sources, radical assignment, collation, etc), and specific code-chart form will be determined as the result of the above process, and a consensus proposal for the encoding of a new Shuowen Seal block will be submitted to WG2.

After discussion and verification by our experts, according to the basic principle of the recognition of small seal glyphs, in addition to the character sets that have been previously identified as non-duplicate ones, it has been confirmed that there are 15 sets of “exact duplicate” in the duplicate of the Tenghuaxie version, which are only necessary to be encoded one code point, and six sets of “possible duplicate”. In order to retain the integrity of small seal glyph system, they need to be individually encoded. There are eight taboo character deformations; other Shuowen versions will be used as references to restore the glyphs. The detailed description is as follows.

1. Duplicate

WG2N4853's Q2 clearly defines the three categories of the “duplicate” of the Tenghuaxie version: exact duplicates, possible duplicates, and non-duplicates. Among them, non-duplicate is treated as different character; it will be encoded separately. The other two categories are as follows:

- A. **Exact duplicate, a total of 15 sets (see the table below).** These duplicate sets are merely different in line shape and length, and they have no influence on glyph structure. It has been confirmed only one code point will be encoded.

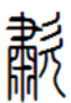
No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning	Note
1	00939	口	右		you4	助也。从口从又。	encode
	02078	又	右		you5	手口相助也。从又从口。	
2	01005	口	否		fou3/pi3	不也。从口从不。	encode

No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning	Note
	08352	不	否		fou3/pi3	不也。从口从 不，不亦聲。	
3	00888	口	吹		chui1	嘘也。从口从 欠。	encode
	06060	欠	吹		chui1/chui1	出气也。从欠从 口。	
4	00980	口	吁		xu1	驚也。从口于 聲。	encode
	03407	亏	吁		xu1	驚語也。从口从 亏，亏亦聲。	
5	02780	放	敖		ao2	出游也。从出从 放。	encode
	04307	出	敖		ao2	游也。从出从 放。	
6	03436	豈	愷		kai3	康也。从心、 豈，豈亦聲。	
	07327	心	愷		kai3	樂也。从心豈 聲。	encode
7	07729	水	湍		tuō1/tuō4	河津也。在西河 西。从水湍聲。	encode
	00881	口	湍		tuō4	口液也。从口垂 聲。唾或从水。	
8	07347	心	愬		zhe2	敬也。从心折 聲。	encode
	00899	口	愬		zhe2	知也。从口折 聲。哲或从心	
9	06118	欠	㞞		you3	愁兒。从欠幼 聲。	encode
	01029	口	㞞		you1	鹿鳴聲也。从口 幼聲。呦或从欠	

No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning	Note
10	10374	車	輟		chuo4	車小缺復合者。从車𠂔聲。	encode
	05346	网	輟		zhuo2	捕鳥覆車也。从网𠂔聲。𠂔或从車	
11	03072	刀	剗		duo2	判也。从刀度聲。	encode
	02259	支	剗		du4	閉也。从支度聲。讀若杜。𠂔或从刀。	
12	03915	木	梃		ni3	木也。實如黎。从木尼聲。	encode
	04191	木	梃		chi4/ni3	簍柄也。从木尸聲。尿或从木尼聲。	
13	08507	手	𢦏		gong3	攤也。从手巩聲。	encode
	02063	𢦏	𢦏		gong3	衰也。从𢦏工聲。𢦏或加手。	
14	09582	虫	螯		mao2	蟹螯也。从虫𢦏聲。	encode
	09707	𧈧	螯		mou2/mao2	蟲食艸根者。从蟲，象其形。吏抵冒取民財則生。𧈧或从𢦏。	
15	10500	𠂔	院		yuan4	堅也。从𠂔完聲。	encode
	05054	𠂔	院		huan2/yuan4	周垣也。从𠂔奐聲。𠂔或从𠂔。又爰眷切。	

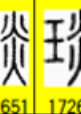
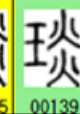
B. Possible duplicate, a total of 6 sets (see the table below). These duplicate character sets have slight differences in line number, line type and component position, including style difference in Guwen or Zhouwen etc. Considering that Shuowen is not only a dictionary but also a comprehensive sorting and build-up of the small seal glyph system; therefore, even if glyphs that have only subtle differences are worth keeping, and they should be encoded individually. The Zhouwen (Chin characters in the Warring States Period) and Gowen (the Six Kingdoms characters in the Warring States Period) collected in the duplicates in Shuowen are valid supplements to the small seal glyph system, they should be preserved as well and be encoded respectively.

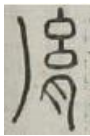
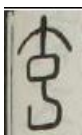
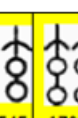
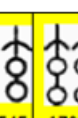
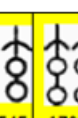
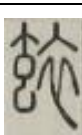
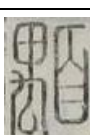
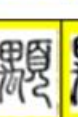
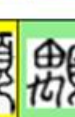
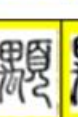
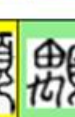
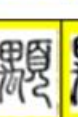
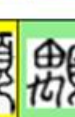
No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning
1	00314	艸	藍		lan2	染青艸也。从艸監聲。
	00644	艸	藍		lan2	瓜蒞也。从艸監聲。
2	00519	艸	難		ran2	艸也。从艸難聲。
	07002	火	難		ran2	燒也。从火狀聲。或从艸難。
3	07909	水	沿		yan2	緣水而下也。从水占聲。《春秋傳》曰：“王沿夏。”
	07647	水	沿		yun3	水。出河東東垣王屋山，東爲涕。从水允聲。古文。
4	03714	臺	臺		guo1	度也，民所度居也。从回，象城口之重，兩亭相對也。或但从口。（音韋。）凡口之屬皆从口。
	09841	土	臺		yong1	城垣也。从土庸聲。古文墉

No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning
5	06088	欠	𪛗		xiao4	吟也。从欠肅聲。《詩》曰：“其𪛗也訶。”
	00932	口	𪛗		yu4	吹聲也。从口肅聲。籀文嘯从欠
6	06019	見	尋		de2	取也。从見从寸。寸，度之，亦手也。
	01350	彳	尋		de2	行有所得也。从彳尋聲。古文省彳。

2. Taboo character deformation

According to expert verification, there are currently 8 taboo character deformations in the Tenghuaxie version, which are presented in two ways, namely changing the character (e.g., number 00023) or omitting the strokes (e.g., number 00140, etc.). The other edition glyph will be used as references to restore the original glyph of the taboo character deformations of the Tenghuaxie version and encoded them as well.

No.	Serial No. in Tenghuaxie version	Original Glyph	TTF Font	Corresp Modern Char	According to the preliminary chart provided by Suzuki Toshiya San
1	00023			禎	孫 陳 N4688 TCA 毛4 華東 北師 逢甲 段         00023 00023 00023 00023 00023 01079 11072 18652 00023
2	00140			琰	孫 陳 N4688 TCA 毛4 華東 北師 逢甲 段         00144 00144 00144 00144 00144 01200 12651 17265 00139

No.	Serial No. in Tenghuaxie version	Original Glyph	TTF Font	Corresp Modern Char	According to the preliminary chart provided by Suzuki Toshiya San																											
3	02931			胤	<table><tr><td>孫</td><td>陳</td><td>N4688</td><td>TCA</td><td>毛4</td><td>華東</td><td>北師</td><td>逢甲</td><td>段</td></tr><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>03027</td><td>03027</td><td>03027</td><td>03027</td><td>03025</td><td>04001</td><td>09106</td><td>20517</td><td>02914</td></tr></table>	孫	陳	N4688	TCA	毛4	華東	北師	逢甲	段										03027	03027	03027	03027	03025	04001	09106	20517	02914
孫	陳	N4688	TCA	毛4	華東	北師	逢甲	段																								
																																
03027	03027	03027	03027	03025	04001	09106	20517	02914																								
4	02773			玄	<table><tr><td>孫</td><td>陳</td><td>N4688</td><td>TCA</td><td>毛4</td><td>華東</td><td>北師</td><td>逢甲</td><td>段</td></tr><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>02868</td><td>02868</td><td>02868</td><td>02868</td><td>02866</td><td>03842</td><td>10518</td><td>17093</td><td>02756</td></tr></table>	孫	陳	N4688	TCA	毛4	華東	北師	逢甲	段										02868	02868	02868	02868	02866	03842	10518	17093	02756
孫	陳	N4688	TCA	毛4	華東	北師	逢甲	段																								
																																
02868	02868	02868	02868	02866	03842	10518	17093	02756																								
5	02775			茲	<table><tr><td>孫</td><td>陳</td><td>N4688</td><td>TCA</td><td>毛4</td><td>華東</td><td>北師</td><td>逢甲</td><td>段</td></tr><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>02870</td><td>02870</td><td>02870</td><td>02870</td><td>02868</td><td>03844</td><td>07545</td><td>17095</td><td>02758</td></tr></table>	孫	陳	N4688	TCA	毛4	華東	北師	逢甲	段										02870	02870	02870	02870	02868	03844	07545	17095	02758
孫	陳	N4688	TCA	毛4	華東	北師	逢甲	段																								
																																
02870	02870	02870	02870	02868	03844	07545	17095	02758																								
6	X096				<table><tr><td>孫</td><td>陳</td><td>N4688</td><td>TCA</td><td>毛4</td><td>華東</td><td>北師</td><td>逢甲</td><td>段</td></tr><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>02871</td><td>02871</td><td>02871</td><td>02871</td><td>02869</td><td>03845</td><td>09562</td><td></td><td></td></tr></table>	孫	陳	N4688	TCA	毛4	華東	北師	逢甲	段										02871	02871	02871	02871	02869	03845	09562		
孫	陳	N4688	TCA	毛4	華東	北師	逢甲	段																								
																																
02871	02871	02871	02871	02869	03845	09562																										
7	x155			曆	<table><tr><td>孫</td><td>陳</td><td>N4688</td><td>TCA</td><td>毛4</td><td>華東</td><td>北師</td><td>逢甲</td><td>段</td></tr><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>04868</td><td>04868</td><td>04868</td><td>04868</td><td>04866</td><td>05841</td><td>17743</td><td></td><td></td></tr></table>	孫	陳	N4688	TCA	毛4	華東	北師	逢甲	段										04868	04868	04868	04868	04866	05841	17743		
孫	陳	N4688	TCA	毛4	華東	北師	逢甲	段																								
																																
04868	04868	04868	04868	04866	05841	17743																										
8	06178			顥	<table><tr><td>孫</td><td>陳</td><td>N4688</td><td>TCA</td><td>毛4</td><td>華東</td><td>北師</td><td>逢甲</td><td>段</td></tr><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>06396</td><td>06396</td><td>06395</td><td>06396</td><td>06394</td><td>07345</td><td>07801</td><td>26515</td><td>06165</td></tr></table>	孫	陳	N4688	TCA	毛4	華東	北師	逢甲	段										06396	06396	06395	06396	06394	07345	07801	26515	06165
孫	陳	N4688	TCA	毛4	華東	北師	逢甲	段																								
																																
06396	06396	06395	06396	06394	07345	07801	26515	06165																								

Suggestions from experts and scholars are welcome. TCA and China will use this document and the feedback to make amendments to the Shuowen small seal list.

Experts of the working group

China: 王寧(Wang Ning), 王立軍(Wang Liun), 王曉明(Wang Xiaoming), 胡佳佳(Hu Jiajia), 陳壯(Chen Zhuang)

TCA: 李璫(Xian Li), 季旭昇(Hsiu-sheng Chi), 許學仁(Xue-jen Hsu), 宋建華(Chien-hua), 魏林梅(Lin-mei Wei), 曾彥玲(Yen-ling, Tseng)

Organization: Chinese Foundation for Digitization Technology, China Electronics Standardization Institute

2017年8月28-29所舉辦的《說文》小篆會議(Shuowen Seal Ad Hoc)，該次會議決議請見WG2/N4853，其中有兩個留待處理的部分：

3. Q2: How should duplicate glyphs be handled in a standard encoding?
 - There is agreement to define “duplicate” with a typology, using terms such as (A) “exact duplicate”, (B) “possible duplicate”, and (C) “non-duplicate”.
 - There is agreement to handle items (A) and (B) using a VS scheme to be determined. Items in (C) would not be suitable for unification.
 - There is agreement to create tables to document the decisions made with respect to the handling of specific duplicates.
4. Q3: Compatibility with existing reference and research implementations.
 - There is agreement that 藤花榭本 may serve as “baseline”, for the determination of the repertory (the number of rows / code points), and for production of a single-column chart in 藤花榭本 glyph style.
 - There is agreement to create multi-column charts (for reference in ISO/IEC JTC1/SC2/WG2 N4853 developing the Shuowen Seal encoding methodology, repertory, and property data), including columns representing the glyphs of specific important editions of Shuowen. The multi-column chart data will include at minimum 藤花榭本 and 陳昌治本 columns (glyphs and mappings), but may also include data for additional columns as available.
 - There is agreement that issues related to 避諱 (taboo character deformation / glyph substitution) will be addressed in conjunction with multicolumn chart development (perhaps by means of a VS scheme).
 - Details of the encoding methodology, script documentation, character repertory, character properties (print sources, radical assignment, collation, etc), and specific code-chart form will be determined as the result of the above process, and a consensus proposal for the encoding of a new Shuowen Seal block will be submitted to WG2.

經過我們專家的討論與查證，根據小篆字形認同的基本原則，除原來已經確認為「非同形」字組外，目前確認了藤花榭本的重文有15組「基本同形」每組字只需編一個碼位；有6組「字形微別」，為保留小篆字形系統的完整性，需分開編碼。另外還有8個避諱字，需參考《說文》其他版本還原字形。詳細說明分述如下。

1. 重文(Duplicate)

WG2N4853的Q2，明確定義了藤花榭本「重文」的三種分類：基本同形(exact duplicate)、字形微別(possible duplicate)、以及非同形(non-duplicate)。其中，

「非同形」視為不同字，會分開編碼。另外兩類說明如下：

- A. 基本同形(exact duplicate)，共有15組(見下表)。這些重文字組只是線條態勢和長短方面的席位差異，對字形結構沒有影響，確定每組只會編一個碼位。

No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning	Note
1	00939	口	右		you4	助也。从口从又。	encode
	02078	又	右		you5	手口相助也。从又从口。	
2	01005	口	否		fou3/pi3	不也。从口从不。	encode
	08352	不	否		fou3/pi3	不也。从口从不，不亦聲。	
3	00888	口	吹		chui1	噓也。从口从欠。	encode
	06060	欠	吹		chui1/chui1	出气也。从欠从口。	
4	00980	口	吁		xu1	驚也。从口于聲。	encode
	03407	亏	吁		xu1	驚語也。从口从亏，亏亦聲。	
5	02780	放	敖		ao2	出游也。从出从放。	encode
	04307	出	敖		ao2	游也。从出从放。	
6	03436	豈	愷		kai3	康也。从心、豈，豈亦聲。	
	07327	心	愷		kai3	樂也。从心豈聲。	encode
7	07729	水	湓		tuo1/tuo4	河津也。在西河西。从水𣶒聲。	encode

No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning	Note
	00881	口	湮		tuo4	口液也。从口垂聲。唾或从水。	
8	07347	心	慙		zhe2	敬也。从心折聲。	encode
	00899	口	慙		zhe2	知也。从口折聲。哲或从心	
9	06118	欠	𢇛		you3	愁兒。从欠幼聲。	encode
	01029	口	𢇛		you1	鹿鳴聲也。从口幼聲。呦或从欠	
10	10374	車	輟		chuo4	車小缺復合者。从車𠂔聲。	encode
	05346	网	輟		zhuo2	捕鳥覆車也。从网𠂔聲。𠂔或从車	
11	03072	刀	剗		duo2	判也。从刀度聲。	encode
	02259	支	剗		du4	閉也。从支度聲。讀若杜。𠂔或从刀。	
12	03915	木	梃		ni3	木也。實如黎。从木尼聲。	encode
	04191	木	梃		chi4/ni3	簍柄也。从木尸聲。尿或从木尼聲。	
13	08507	手	𢇛		gong3	攤也。从手玕聲。	encode
	02063	𠂔	𢇛		gong3	衰也。从𠂔工聲。玕或加手。	

No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning	Note
14	09582	虫	蝥		mao2	蝥蝥也。从虫敕聲。	encode
	09707	𧈧	蝥		mou2/mao2	蟲食艸根者。从蟲，象其形。吏抵冒取民財則生。蝥或从敕。	
15	10500	𠂔	院		yuan4	堅也。从𠂔完聲。	encode
	05054	宀	院		huan2/yuan4	周垣也。从宀奩聲。窳或从𠂔。又爰眷切。	

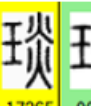
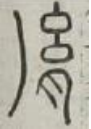
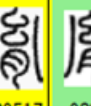
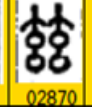
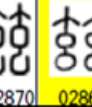
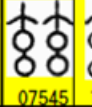
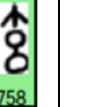
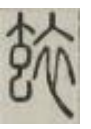
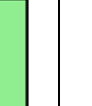
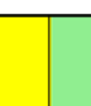
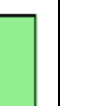
B. 字形微別(possible duplicate)，有6組(見下表)。這些重文字組在線條數、線條類型、部件位置等方面有細微差異，包括古文、籀文等風格上的差異。考慮到《說文》本身並不僅僅是一部字典，而是對小篆字形系統的全面整理和構建，因此即使只有細微差異的字形，也有保留的價值，應分別給予編碼。而其中所收錄的重文中的籀文（戰國時期的秦文字）和古文（戰國時期的六國文字）字形，是對小篆字形系統的有效補充，也應該予以保留，分別給予編碼。

No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning
1	00314	艸	藍		lan2	染青艸也。从艸監聲。
	00644	艸	蒨		lan2	瓜蒨也。从艸監聲。
2	00519	艸	難		ran2	艸也。从艸難聲。
	07002	火	難		ran2	燒也。从火狀聲。或从艸難。

No.	Serial No. in Tenghuaxie version	SW Radical	Corresp Modern Char	TTF Font	pronunciation	meaning
3	07909	水	沿		yan2	緣水而下也。从水占聲。《春秋傳》曰：“王沿夏。”
	07647	水	沿		yun3	水。出河東東垣王屋山，東爲涕。从水允聲。古文。
4	03714	臺	臺		guo1	度也，民所度居也。从回，象城臺之重，兩亭相對也。或但从口。（音韋。）凡臺之屬皆从臺。
	09841	土	臺		yong1	城垣也。从土庸聲。古文墉
5	06088	欠	𩚑		xiao4	吟也。从欠肅聲。《詩》曰：“其𩚑也訶。”
	00932	口	𩚑		yu4	吹聲也。从口肅聲。籀文嘯从欠
6	06019	見	尋		de2	取也。从見从寸。寸，度之，亦手也。
	01350	彳	尋		de2	行有所得也。从彳尋聲。古文省彳。

2. 避諱字

經專家查證，目前確認藤花樹本共有8個避諱字，分別以改字（如編號00023）或省略筆形（如編號00140等）兩種方式呈現。對於藤花樹本的避諱字會參考其他版本的字形，還原其原形，並予以編碼。

No.	Serial No. in Tenghuaxie version	Original Glyph	TTF Font	Corresp Modern Char	According to the preliminary chart provided by Suzuki Toshiya San
1	00023		禎	禎	孫 陳 N4688 TCA 毛4 華東 北師 逢甲 段         00023 00023 00023 00023 00023 01079 11072 18652 00023
2	00140		琰	琰	孫 陳 N4688 TCA 毛4 華東 北師 逢甲 段         00144 00144 00144 00144 00144 01200 12651 17265 00139
3	02931		胤	胤	孫 陳 N4688 TCA 毛4 華東 北師 逢甲 段         03027 03027 03027 03027 03025 04001 09106 20517 02914
4	02773		玄	玄	孫 陳 N4688 TCA 毛4 華東 北師 逢甲 段         02868 02868 02868 02868 02866 03842 10518 17093 02756
5	02775		茲	茲	孫 陳 N4688 TCA 毛4 華東 北師 逢甲 段         02870 02870 02870 02870 02868 03844 07545 17095 02758
6	X096		𪛗		孫 陳 N4688 TCA 毛4 華東 北師 逢甲 段         02871 02871 02871 02871 02869 03845 09562
7	x155		曆	曆	孫 陳 N4688 TCA 毛4 華東 北師 逢甲 段         04868 04868 04868 04868 04866 05841 17743
8	06178		顥	顥	孫 陳 N4688 TCA 毛4 華東 北師 逢甲 段         06396 06396 06395 06396 06394 07345 07801 26515 06165

歡迎各位專家學者給予建議。TCA and China會依據此份文件與回饋意見進行《說文》小篆字表的修正。

Experts of the working group

China: 王寧(Wang Ning), 王立軍(Wang Liun), 王曉明(Wang Xiaoming), 胡佳佳(Hu Jiajia), 陳壯(Chen Zhuang)

TCA: 李璫(Xian Li), 季旭昇(Hsiu-sheng Chi), 許學仁(Xue-jen Hsu), 宋建華(Chien-hua), 魏林梅(Lin-mei Wei), 曾彥玲(Yen-ling, Tseng)

Organization: Chinese Foundation for Digitization Technology, China Electronics Standardization Institute

(end of document)