Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 IRG N1836

Date: 2012-02-23

Title:	Report from the Old Hanzi Expert Group
Source:	Old Hanzi Expert Group
Status:	Input to IRG
Action:	
Distribution:	IRG Members and Ideographic Experts
Reference:	IRG N1810, 1827, 1835, 1842, WG2 N4236, N4245.
No. of pages:	5
Medium:	Electronic

The Old Hanzi Expert Group ad hoc meeting was held at Tokyo from February 20th to 23rd, 2012. The meetings were held at Kikai Shinko Kaikan, Minato-ku, Tokyo, Japan.

This report is organized as follows:

1. WG2#59 meeting report
2. Discussion on the items assigned by IRG N1827
3. Discussion on IRG N1786
4. Discussion on WG2 N4236
5. Other items
6. Open Issues
7. Appreciation

1. WG2#59 meeting report

Chen Zhuang reported the resolution of WG2#59 (2012-02-10/13), M59.19. Old Hanzi expert group accepts it unanimously.

2. Discussion on the items assigned by IRG N1827

Old Hanzi expert group discussed the feedbacks to the items listed in IRG N1827.

IRG N1827 1.a (reference list review):

The bibliography list is reviewed and completed (the result is found in Appendix).

IRG N1827 1.b (indexing method for Oracle Bone glyph without corresponding SW and UCS characters):

Japan proposed to classify the glyphs into 2 groups.

Group A: the Oracle Bone glyphs without corresponding SW character but with corresponding SW radical.

Group B: the Oracle Bone glyphs without corresponding SW character and without corresponding SW radical.

Old Hanzi expert group agreed to use some referential material(s) to index group B unanimously. Xin Jiaguwenbian (新甲骨文編, 劉釗/洪颺等編纂, 福建人民出版社, 2009, ISBN 9787211058532) would be a candidate of the material(s). To index group A, China and TCA recommended to consider a method using the number of strokes out of radical, of corresponding modern character, as a candidate of the indexing method. Japan proposed to choose one dictionary to index group A and group B. At present, Japan is not sure which dictionary is the best for this method, but Xin Jiaguwenbian would be a candidate. China and TCA commented Japan proposal is difficult for "new" Oracle Bone glyphs that cannot be found in existing dictionaries. Japan commented to the candidate indexing method given by China and TCA that the rule how to choose modern character is not fixed yet and using it for indexing is unstable. The decision of the indexing method for group A glyphs is postponed.

IRG N1827 1.c (coverage of the code points for "UCS character" column):

China and TCA proposed to use CJK Unified Ideographs including the extension A, B,

C, D. Japan requested to fix the version of SW, to stabilize the modern character corresponding to Shuowen character. China and TCA proposed to use the version by 陳昌治 (so-called "一篆一行本"), and Old Hanzi expert group accepted it unanimously.

IRG N1827 1.d (completion of the definitions for the entries required in the submission):

The definition of "Epoch/Period" is given, and "Site/Area" column is decided to be removed.

IRG N1827 2.a (review IRG N1771):

China proposed to make no names for Oracle Bone characters, but an option might be the convention used by CJK Unified Ideographs that the names are determined by their code positions. Also some difficulties were pointed by China and TCA experts; a) single character may have multiple meaning, b) some characters are not deciphered and no meaning is known, c) meaning of a character may be differently deciphered by different scholars. Japan replied that the meaning-based names are not essential.

Also China commented Japanese proposal does not cover all items in current database, because some items are dropped by the clause 3 in IRG N1771. Japan commented the clause 3 was introduced to make a stable definition of the character, so it is related with the definitions of the character and glyph for Oracle Bone script.

The discussion on the character and glyph definitions could not be finished in this meeting, thus China and TCA will submit their feedbacks before IRG#38, with written document.

IRG N1827 2.b (clarification of requirement):

Old Hanzi expert group could not understand the question by IRG. The Rapporteur gave further explanation, as follows:

> The purpose of this collection is to have it coded so that some applications/research can make use of them for their work. From user's view points, you must have some

application related requirement so that the coding would fit your requirement such as, based on what you want them coded (say, glyphs, images, or the abstract glyphs you have come up with for coding). How you want the character to be ordered so that you can index/search them easily.

I understand that in working out the collection you have some attributes, some of them are for internal use (to get the collection in place). Some may be essential in developing the coding model (so not internal to gettting the collection) which you need to distinguish and let coding people know. In other words, coding method is related to how you want your characters (glyph, glyph representatives) to be used.

Your group (or later individual contributions) need to tell IRG/WG2 this first. Otherwise, a coding model cannot be developed or may face the possibility of not being useful for any Oracle Bone development/research work.

TCA and China think this question is same with 1.b, therefore, the answer is the same with that to IRG N1827 1.b. Japan has a concern that the questions in 1.b and 2.b are different.

Japanese feedback is already submitted as IRG N1842. Japan considers the interoperability between existing dictionaries, databases, corpuses is important requirement, so IRG N1771 was designed by using existing Chinese dictionary, Yinxu Jiagu Keci Leizuan (殷墟甲骨刻辭類纂, 姚孝遂主編, 中華書局, ISBN 9787101004779, 1989), that used by several fonts and databases. Thus, Japan emphasized the importance of the mapping table between the ID in Oracle Bone database and the index number of the heading glyph in the conventional dictionary, like, Jia Gu Wen Bian (甲骨文編, 中国社会科学院考古研究所編, 科学出版社, ISBN 7101005233, 1965). Japan thinks the mapping table described in above is necessary for encoding. China and TCA agrees that the mapping table is important for users but not necessary for encoding. China and TCA consider that it is not the business of Old Hanzi expert group. Current Oracle Bone database has no information to automate the production of a mapping table, because Old Hanzi P&R had no explicit instruction to submit the information to make a

mapping table. China, TCA and Japan agreed that the mapping table cannot be made from current Oracle Bone database. Japan national body requested to note that the mapping table is necessary for encoding. TCA member body requested to note that the contributed mapping table will be welcomed.

3. Discussion on IRG N1786

China commented that TCA and China are collaborating for the font production, so China finds no problem in the font design policy.

4. Discussion on WG2 N4236

This document was submitted just before WG2#59, so China and TCA experts did not have enough time to review it. Japan explained the first question in WG2 N4236; the request of the clarification with other projects; China Character Repertoire, and other databases (e.g. CHANT). The Old Hanzi expert group is aware of their activity, however currently there is no official relationship to the group.

5. Other items

Old Hanzi expert group received the consolidated comment IRG N1835.

6. Open Issues

The comment disposition for IRG N1835
The mutual check of the comments in IRG N1787 Part 2 (the 2nd round checking)
The Old Hanzi expert group requests to close these 2 issues in IRG#38.

7. Appreciation

Old Hanzi expert group would like to express its sincere appreciation to the meeting host for the Tokyo ad Hoc meeting, Information Technology Standards Commission of Japan (ITSCJ). Old Hanzi expert group would also like to thank Ms. Toshiko Kimura for her excellent work on meeting logistics, arrangement, and hospitality.

Members attended the meeting are:

Li Guoying, Wei Lin-Mei (Selena), Zhou Xiaowen, Chen Zhuang, Dai Hong, Tatsuo Kobayashi, Masahiro Sekiguchi, Atsushi Suzuki, Toshiya Suzuki and Takao Hirase.

(end of document)

IRGN 1836Agenda

Draft Agenda for ISO/IEC JTC 1/SC 2/WG 2/IRG/Old Hanzi Ad Hoc Group Meeting, Tokyo, Japan, 2012-02-20/23

Location:　Room B3-7,
Kikai-Shinko Kaikan
3-5-8, Shiba-koen, Minato-ku
Tokyo 105-0011, Japan

Draft Agenda

1. Opening and roll call (2/20(Mon) 10:00～)
2. Approval of the agenda
3. WG2 meeting report
4. Discussion on the incoming documents
    1) Drafting the feedback to the items assigned in IRG#37 (IRG N1827)
        ① Old Hanzi coding framework
        ② Indexing method of Oracle Bone that have no corresponding SW characters
        ③ Coverage of corresponding UCS character
        ④ Completion of the definition of the entries of the submission to Oracle Bone DB
        ⑤ Clarification of bibliography information
    2) Discussion on IRG N1786 (TCA request for comment about font production)
    3) Discussion on WG2 N4236
5. Old Hanzi DB review
6. Closing (2/23(Thu) ～16:30)
    1) Approval of the meeting report
    2) Adjournment

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 IRG
N1827Appendix
Date: 2011-11-09

Title:	**Old Hanzi Principles and References (Version 3 Draft)**
Source:	**Old Hanzi Experts Group**
Status:	**Input to IRG**
Action:	
Distribution:	**IRG Members and Ideographic Experts**
Reference:	
No. of pages:	
Medium:	**Electronic**

The Oracle Bone principles and references (version 3) extracted and compiled from the following documents:

IRG N1135R, IRG N1182, IRGN1215, IRG N1267, IRG N1271 (version 2), IRG N1325, IRG N1460, IRG N1747A.

1. Format of submission

ID	Imitated Glyph	Original Glyph	Source	Period /Epoch	Area /Site	Material	SW Radical	SW Radical Number	*Corresp. Modern Char (UCS)	*Unifiable Glyph
1										
2										

2. Definitions

2.1. ID: It is the unique id that consists of one or two letter member id (G, T, K, KP, J, V, S, H, M) followed by four digit sequential number

Deleted:

ID	Imitated Script/Glyph	Original Shape/ Glyph	Source	Period/ Epoch	Ar /Si
1					
2					

Formatted ... [1]
Formatted ... [2]
Formatted ... [3]
Formatted ... [4]
Formatted ... [5]
Formatted ... [6]
Formatted ... [7]
Formatted ... [8]
Formatted ... [9]
Formatted ... [10]
Formatted ... [11]
Formatted ... [12]
Formatted ... [13]
Formatted ... [14]
Formatted ... [15]
Formatted ... [16]
Formatted ... [17]
Formatted ... [18]
Formatted ... [19]
Formatted ... [20]
Formatted ... [21]
Formatted ... [22]
Formatted ... [23]
Formatted ... [24]
Formatted ... [25]
Formatted ... [26]
Formatted ... [27]
Formatted ... [28]
Formatted ... [29]
Formatted ... [30]
Formatted ... [31]
Formatted ... [32]
Formatted ... [33]
Formatted ... [34]
Formatted ... [35]
Formatted ... [36]
Formatted ... [37]
Formatted ... [38]
Formatted ... [39]
Formatted ... [40]
Formatted ... [41]
Formatted ... [42]
Formatted ... [43]
Formatted ... [44]
Formatted ... [45]
Formatted ... [46]
Formatted ... [47]
Deleted: ¶
Deleted: s
Deleted: s

assigned by submitters.

Example: T0001 is one IRG global unique ID assigned to an Oracle Bone inscription submitted by TCA.

2.2. Imitated Glyph: The truthful trace from 'Original Shape/Glyph'
[摹寫字的定義：忠實摹寫原形的字形。]

2.3. Original glyph: The glyph selected according to the principles of Oracle Bone selection (in the item 4) in this document.
[原形的定義:根據選字原則(本文件第 4 章) 从原拓選定的字形。]

2.4. SW Radical: The glyph image of the corresponding ShuoWen Radical in Kai-style. The submitter is not required to provide the glyph image, it will be produced by the project editor based on the submitted SW Radical number.

2.5. SW Radical number : 1 – 540. The order is defined by 漢•許慎 '說文解字' (大徐本).

3. Rules

3.1. The 'Imitated Glyph' image should be of standardized dimensions given in 3.2.

3.2. The 'Imitated Glyph' image shall be in EPS format (resolution 1024×1024).

3.3. The format of bitmap images for imitated glyphs, original shapes and 540 SW Radicals are specified as follows:
128x128，Black and white bitmap

3.4. The last three columns are optional fields and they are indicated

Deleted: Script/

Deleted: picture

Deleted: picture

Deleted: is assigned

Deleted: Script/

Deleted: size and

Deleted: Script/

Formatted: Tab stops: 3.54", Left

Deleted: '

Deleted: '

with an asterisk "*". All other fields are mandatory.

3.5. The field "Corresp. Modern Char. (UCS Code)" shows the modern character similar in meaning or shape.

3.5.1. The field "Corresp. Modern Char. (UCS Code)" is filled by a single codepoint or a list of CJK Unified Ideographs separated by semicolon (;).

3.5.2. CJK Unified Ideograph (including Extension A, B, C, D) can be used for the convenience of sorting or finding a character from the database. If there is no corresponding modern character, or exists but not coded in above UCS blocks, the field must be blank. The note field should be used for the description for such cases. The UCS character corresponding to SW glyph should be determined by so-called "一篆一行本", a version of Shuowen Daxu by 陳昌治.

3.6. If the field "Corresp. Modern Char.(UCS Code)" is blank, then the "Notes" field must be filled with justifications to indicate the glyph is well-understood in meaning, for example, the meaning of the 'Imitated Glyph'. Also "Note" field can include the description of the glyph structure when the glyphic components have the corresponding modern character in UCS.

3.7. Source: The "Source" field is an important key to exclude exactly duplicated data.

The "Source" field consists of two mandatory elements and one optional element. They will be concatenated with the hyphen character '-'.

Deleted: ,

Deleted: the

Deleted: Only URO (CJK Unified Ideographs in BMP without Extension A and CJK Compatibility Ideograph) characters…

Deleted: URO

Deleted: Script /

Deleted: the

Deleted: data

Deleted: with

1. (mandatory) The 1st element is a letter indicating referenced book. The possible values are:

   (A) stands for《甲骨文合集》郭沫若主編，中華書局, ISBN 9787101016536 (13 volumes), 1978-1982

   (B) stands for《甲骨文合集補編》彭邦炯，謝濟，馬季凡編著，語文出版社，1999，ISBN 7801264967 (7 volumes), 1999

   (C) stands for《殷墟花園莊東地甲骨》中国社会科学院考古研究所，雲南人民出版社，ISBN 9787222038776, 2003

   (D) stands for《濟南市大辛莊遺址出土商代甲骨文》,《考古》2003 年 6 期,《濟南大辛莊遺址出土商代甲骨文》,《中國歷史文物》2003 年 3 期

   (E) stands for《周原甲骨文》，曹瑋，世界圖書出版公司北京公司, 2002, ISBN 9787506256650

   (F) stands for《小屯南地甲骨》考古学専刊乙種 18 号，中国社会科学研究院考古研究所，中華書局，上下冊 (1980 and 1983)

   (G) stands for《英國所藏甲骨集》李学勤，齊文心，艾蘭編著，中華書局，ISBN 9787101009569, 1985

   (H) stands for《懷特氏等所藏甲骨文集》，許進雄編，加拿大皇家安大列博物館出版, 1979, ISBN 0-88854-231-3

   (I) stands for《天理大学附属天理参考館蔵甲骨文字》天

Deleted: es

Deleted: the book

Deleted: e number

Deleted: 主編

Deleted: 庄

Deleted: 《山東濟南大辛莊甲骨》TBD

Deleted:

Deleted: 92

Deleted: TBD

Deleted: ひとものこころ

Deleted: 品

理大学天理教道友社 Vol. 1, Num. 5, 1987

(J) stands for《德瑞荷比所藏一些甲骨錄》雷煥章編著, 1997, 光啓出版社, ISBN 2-9505602-4-5, 1997

(K) stands for《瑞典斯德哥爾摩遠東古物博物館藏甲骨文字》李学勤, 齊文心, 艾蘭編著, 中華書局, ISBN 7-101-02256-1/H.154, 1999

2. (mandatory) The 2nd element is anOracle Bone number (甲骨拓片的編號) which consists of 5 digits assigned uniquely to each Oracle Bone inscription.

3. (optional) The 3rd element indicates I the side of the Oracle Bone which consists of 1 digit. The possible values are '0' for front side, '1' for back side. If an inscription is carved only on one side, this element will be omitted.

Three examples of the "Source" field are listed below.

4. A-00001 (does not have front and back side)

5. A-00001-0 (front side)

6. A-00001-1 (back side)

3.8. Period/Epoch: The "Periof/Epoch" field is the name of dynasty that the material was inscribed. For Oracle Bone, the possible value is only 商 or 周.

Deleted: 道友社, ISBN 4-8073-0254-X,

Deleted: 6

Deleted: "Several Collections of Oracular Inscriptions in Germany, Switzerland, The Netherland, Belgium" (Fr. Jean Lefeuvre), Ricci, …

Deleted: 9782950560247

Deleted:

Formatted: Superscript

Deleted: (mandatory) Oracle Bone number (甲骨拓片的編號) which consists of 5 digits assigned uniquely to each Oracle Bone.¶

Deleted: dentifier to determine

Formatted: Superscript

Formatted: Bullets and Numbering

Deleted: ¶

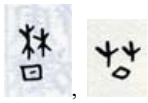
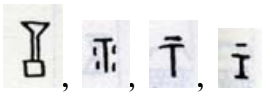
Formatted: Font color: Auto, (Asian) Chinese (Taiwan)

4. The principles of Oracle Bone selection

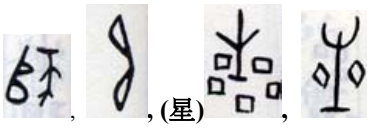
4.1. Separation principles

Two or more instances of Oracle Bone characters with the following differences will be separated.

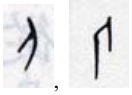
4.1.1. One or more components are different.

(莫)  , (示) 

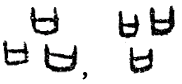
4.1.2. The number of components or lines is different.

(師)  , (星)

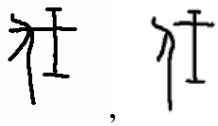
4.1.3. The direction (e.g. mirror image) of a component is different.

(人) 

4.1.4. The position of one or more components is different.

(品) 

4.1.5. Connectivity of the components is different.

(伐) 

4.2. Unification principles

Deleted: Distinction

Formatted: Font color: Auto

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman

Formatted: Left, Right: 0", Don't adjust right indent when grid is defined, Space Before: 0 pt, After: 0 pt, Line spacing: single, No bullets or numbering, Don't adjust space between Latin and Asian text, Don't adjust space between Asian text and numbers

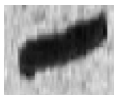
Formatted: Font: (Asian) MS-Gothic, 12 pt, Not Bold

Deleted: types of

Deleted: Whether the same set of components are connected each other or not…

Two or more instances of Oracle Bone inscriptions with the following differences will be unified unless there's any difference in the meaning:

4.2.1. The length of corresponding line is different.

(一)  , 

4.2.2. The thickness of corresponding line is different.

(王)  ,  , 

4.2.3. The size of the corresponding components different.

Deleted: each component of the same set is

4.2.4. The enclosed part is filled or not filled.

(止)  , 

5. The principles of radical classification

5.1. If an Oracle Bone glyph corresponds to a Shuowen glyph, it should be classified into Shuowen radical of the corresponding Shuowen

Deleted: is

Deleted: ing

Deleted: ed

glyph. For example,  (corresponds to ) should be classified to '明', not to '月' or '日'. In addition, Shuowen classifies some Guwen (古文) or Zhouwen (籀文) glyphs to a radical class even when these glyphs do not include Shuowen radical as their glyphic components. If the corresponding Shuowen glyph is such, the Oracle Bone glyph should be classified to the radical that the corresponding Shuowen glyph is included. For example,  (corresponds to ) is classified to '邑', not to '土' or '田'.

Original Script/Glyph	Corresp. SW Glyph	S.W. Radical
		明
		邑

5.2. The glyph should be classified into Shuowen radicals according to the Original Oracle Bone Inscriptions, e.g. "  " should be classified under radical 斤, because Shuowen radicals do not include the most significant glyphic component '單'. The next significant glyph component '斤' is used (KangxiZidian classifies '單' to the radical '口'. The classification of Oracle Bone shape

Deleted: ed

Deleted: ed

Deleted: ed

Deleted: ed

Deleted: S.W.

Deleted: es

“  ” to ‘口’ is more difficult to use than that to ‘斤’).

歸部問題：以甲骨文字形為主，如“  ”入斤部。

Example 2

Imitation Script/Glyph	Original Script/Glyph	SW Radical
		斤

Deleted: S.W.

5.3. Suppose that the shapes of the Original Oracle Bone inscriptions are different, but they share the same meaning and usage. Eventually, they have evolved into a pair of variants with two different radicals in Shuowen. According to the radical classification in Shuowen, the Oracle Bone glyphs are put under different radicals. e.g. 兀 and 元. 甲骨文異形同用，後世分為兩字，說文分見兩部，則依《說文解字》收入不同部首。如“元”、“兀”。

Deleted: T

Deleted: ,

Example 3

Imitation Script/Glyph	Original Script/Glyph	SW Radical
		一

Deleted: S.W.

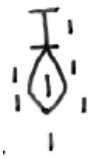
		儿

5.4. Suppose that the shape, meaning and usage of the Original Oracle Bone inscriptions are identical but new components have been added over time. If a radical can be found in Shuowen, the glyph will be put under the corresponding radical according to the Original Oracle Bone inscriptions. e.g. 汆 put under the 汆 radical and similarly 畐 in 畐 radical.

甲骨文同形同用，後世增添偏旁，《說文解字》另有部首者，則依甲骨文原形歸入相應部首。如“畐”入《說文解字》畐部，“汆”入《說文解字》汆部。

Deleted: T

Example 4

Imitation Script/Glyph	Original Script/Glyph	SW Radical
		畐
		汆

Deleted: S.W.

5.5. Suppose that the shapes of the Original Oracle Bone inscriptions are the same but they have many meanings and usages. Eventually, they

Deleted: T

have evolved into different characters. The glyph of these characters will be determined according to the shapes of the Original Oracle Bone inscriptions, and put under the corresponding radical in Shuowen. e.g. 史吏事.

甲骨文同形多用，後世分為多字，依甲骨文字形分別隸定，歸入《說文解字》相應同一個部首。如：“史”、“吏”、“事”。

Example 5

Imitation Script/Glyph	Original Script/Glyph	SW Radical
		—
		—

6. The principles of sorting the order of the glyphs of the same Oracle Bone Inscription

同字之字形排序原則

6.1. Ordering of Inscriptions

The Oracle Bone glyphs are classified into 3 groups; 1) SW-mappable glyphs, 2) SW-unmappable but with corresponded UCS character, and 3) SW-unmappable and without corresponded UCS character. They are ordered as follows:

Deleted: S.W.

Deleted: Glyph Categories

Deleted: ing

6.1.1. SW-mappable glyphs: The Oracle Bone glyphs that corresponds to ShuoWen glyphs are identified are ordered by the order of ShuoWen Jiezi Daxu version (大徐本).

6.1.2. SW-unmappable glyphs with corresponding UCS character: The Oracle Bone glyphs without corresponding SW glyph but have corresponding UCS character is available should be placed after SW-mappable glyphs. To gather similar glyphs, the glyphs sharing same corresponded UCS characters should be collected to one group.

6.1.3. SW-unmappable glyphs without corresponding UCS character: The Oracle Bone glyphs which have no corresponding SW glyphs and no UCS character should be placed after SW-unmappable glyphs with corresponding UCS character.

6.2. Ordering in Glyph Category

6.2.1. If one or more types of components or radical are different, those with smaller difference will be placed first, and those with greater difference behind.
異構字依字形差異大小排序，差異較小者置於前，差異較大者置於後。

6.2.2. Glyph variants will be placed after the typical glyph.
異寫字置於主形之後。

7. Release Process of the Database

For the record of the discussion of inclusion, deletion (because of unclear, cropped or exactly duplicated data), unification (glyphs from different

Deleted: ed

Deleted: ed

Formatted: Font: Bold

Deleted: that corresponded

Deleted: is unavailable

Deleted: ed

Deleted: ed

Deleted: that

Deleted: ed

Deleted: ed

Deleted: then

Deleted: submitted

sources(see 3.7) but cannot be separated by the separation principles), or pending should be recorded in 'Status' column of the database.

Deleted: distinguished

8. Data Format For Oracle Bone Data Exchange

For the data exchange and review work, members are going to use the data format specified as follows:

■ Images format:

8.1. use PNG storage format.

8.2. The original glyph should be scanned at 300 dpi (dots per inch).

8.3. The imitated glyph images are named [ID]+[_R] (for example, if the ID is T00001, the transcribed glyph images should be named T00001_R).

Deleted: transcribed

8.4. The original glyphs are named [ID]+[_O] (for example, if the ID is T00001, the original glyph should be named T00001_O).

8.5. Glyph determination images(Note: Not defined!) are named [ID]+[_D] (for example, if the ID is T00001, the glyph determination image should be named T00001_D).

Formatted: Highlight

8.6. The images of unifiable shapes are named [ID]+[ID of the unified glyph ID] (for example, if the Oracle Bone ID is T00001 and the unifiable shape ID is 000, the image of unifiable shape should be named T00001_000).

Deleted: Unifiable shapes

■ XML Schema:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">
  <xs:element name="OldHanZi">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="Character" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="version" type="xs:string" use="required" fixed="1.0"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="Character">
    <xs:complexType>
```

```
        <xs:sequence>
          <xs:element ref="Source"/>
          <xs:element ref="Period"/>
          <xs:element ref="Area"/>
          <xs:element ref="Material"/>
          <xs:element ref="Radical"/>
          <xs:element ref="ModernChar" minOccurs="0"/>
          <xs:element ref="Unified" minOccurs="0" maxOccurs="unbounded"/>
          <xs:element ref="Note" minOccurs="0"/>
        </xs:sequence>
        <xs:attribute name="id" use="required">
          <xs:simpleType>
            <xs:restriction base="xs:string">
              <xs:pattern value="(G|T|K|KP|J|V|S|H|M)[0-9]+"/>
            </xs:restriction>
          </xs:simpleType>
        </xs:attribute>
      </xs:complexType>
    </xs:element>
    <xs:element name="Source" type="xs:string"/>
    <xs:element name="Period" type="xs:string"/>
    <xs:element name="Area" type="xs:string"/>
    <xs:element name="Material" type="xs:string"/>
    <xs:element name="Radical">
      <xs:simpleType>
        <xs:restriction base="xs:unsignedShort">
          <xs:minInclusive value="1"/>
          <xs:maxInclusive value="540"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
    <xs:element name="ModernChar" type="xs:string"/>
    <xs:element name="Unified">
      <xs:complexType>
        <xs:attribute name="id" type="xs:string" use="required"/>
      </xs:complexType>
    </xs:element>
    <xs:element name="Note" type="xs:string"/>
</xs:schema>
```

- XML example:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<OldHanZi version="1.0">
    <Character id="T00001">
      <Source>甲骨文合集</Source>
```

```
    <Period>商</Period>
    <Area>河南安陽</Area>
    <Material>甲骨</Material>
    <Radical>001</Radical>
    <ModernChar>一</ModernChar>
    <Unified id="0000"/>
    <Unified id="0001"/>
    <Note/>
  </Character>
  <Character id="T00002">
    <Source>甲骨文合集</Source>
    <Period>商</Period>
    <Area>河南安陽</Area>
    <Material>甲骨</Material>
    <Radical> 005</Radical>
    <ModernChar>王</ModernChar>
    <Unified id="0000"/>
    <Note/>
  </Character>
</OldHanZi>
```