Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 IRG N1876
WG2N4357
N1827Appendix
Date: 2011-11-09

Title:	**Old Hanzi Principles and References (Version 3)**
Source:	**Old Hanzi Experts Group**
Status:	**Input to IRG**
Action:	
Distribution:	**IRG Members and Ideographic Experts**
Reference:	**IRGN 1827Appendix**
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	*Note
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

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

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. Only URO (CJK Unified Ideographs in BMP without Extension A and CJK Compatibility Ideograph) characters 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 URO, the field must be blank. The note field should be used for the description for such cases.

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 ‘-’.

- (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》彭邦炯主編，語文出版社, 1999, ISBN 7801264967 (7 volumes), 1999

(C) stands for《殷墟花園庄東地甲骨》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《山東濟南大辛莊甲骨》**TBD**

(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，中華書局, ISBN 9787101009569, 1992

(H) stands for《懷特氏等所藏甲骨文集》**TBD**

(I) stands for《ひとものこころ 天理大学附属天理参考館蔵品》Vol. 1, Num. 5, 道友社, ISBN 4-8073-0254-X, 1986

(J) stands for《德瑞荷比所藏一些甲骨錄》“Several Collections of Oracular Inscriptions in Germany, Switzerland, The Netherland, Belgium” (Fr. Jean Lefeuvre), Ricci, ISBN 9782950560247, 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，中華書局，ISBN 7101022561, 1999

- (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.

- (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.

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

- A-00001-0 (front side)

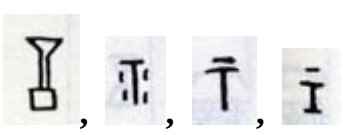
- A-00001-1 (back side)

4. The principles of Oracle Bone selection

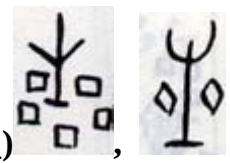
4.1. Separation principles

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

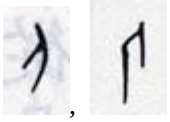
4.1.1. One or more components are different.

(莫)  ，**(示)** 

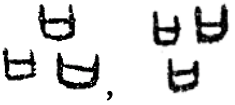
4.1.2. The number of components or lines is different.

(師)  , (星)  ,

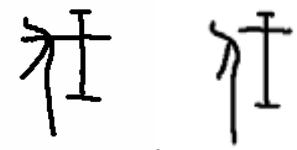
4.1.3. The direction (e.g. mirror image) of a component is different.

(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

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.

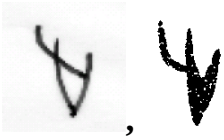
(一)  , 

4.2.2. The thickness of corresponding line is different.

(王) 

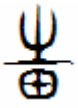
4.2.3. The size of the corresponding components different.

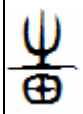
4.2.4. The enclosed part is filled or not filled.

(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 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, 

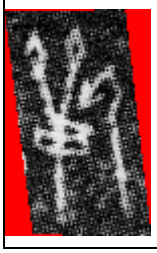
(corresponds to ) is classified to ‘邑’, not to ‘土’ or ‘田’.

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 “  ” 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

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”。

Example 3

Imitation Script/Glyph	Original Script/Glyph	SW Radical
		一
		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. 叒 put under the 叒 radical and similarly 晶 in 晶 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“晶”入《說文解字》晶部，“叒”入《說文解字》叒部。

Example 4

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

5.5. Suppose that the shapes of the Original Oracle Bone inscriptions are the same but they have many meanings and usages. Eventually, they 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
		一

		—

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:

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

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).

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).

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).

■ 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>
```