

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2 IRG N1182

Date: 2005-12-1

Title:	Report from the Old Hanzi Interest Group
Source:	Old Hanzi Interest Group
Status:	Input to IRG
Action:	
Distribution:	IRG Members and Ideographic Experts
Reference:	
No. of pages:	6
Medium:	Electronic

The Old Hanzi Interest Group has discussed the issues of the repertoire development of Old Hanzi. The group decided to obtain a common consensus between all the members before processing the realignment work separately, and to focus on the development of Oracle bone inscription repertoire first. The submission document will follow the N 1102 and N 1135, with one addition, the field “Unifiable Shapes” described below under “3. Format Update”.

The Group discussed N1165 submitted by TCA regarding an Old Hanzi Interest Group forum on old Hanzi information format and working procedure.

The group came up with the following agreements:

1. The Old Hanzi experts group plans to arrange the Old Hanzi attributes as follows:

- (1) Start with constructing an Old Hanzi script database, including these fields:
 - (1.1) Original image (figure)
 - (1.2) Glyph determination
 - (1.3) 9 Attributes of Old Hanzi inscription

- (2) Edit a wordlist: edit an Oracle Bone inscription wordlist, for IRG consideration for submit to WG2 for information.
- (3) Integrate all data into one consolidated database for easy use and make the database open to the IRG member bodies.
- (4) Glyph determination: Establish a set of common rules for glyph determination in the next IRG meeting by consulting Old Hanzi experts after the IRG#25 meeting.
- (5) The scope of mapping between Old Hanzi and modern ideographs :
 - (5.1) The mapping to modern ideographs is provided using encoded UCS characters.
 - (5.2) When there are mappings to both Simplified and Traditional characters, both will be listed.

2. The consolidated Old Hanzi database will include:

- (1).The original images of the Old Hanzi and the explanation archives.
- (2) The glyph determination and corresponding modern character.
- (3) The attribute data of collected Old Hanzi inscriptions.
(Please refer to database illustrations in IRG N 1165 and N 1168.)

3. Format Update

ID	Rep. Script/ Glyph	Original Shape/ Glyph	Source	Period/ Epoch	Area/ site	Material	SW Radical	SW Radical Number	Glyph Determ.	Corresp. Modern Char	Unifiable Shapes	Notes
1												
2												

There are two changes of the format from what was defined by IRG N1135 as follows:

1. There is a new field, “Unifiable shapes”.
2. The field “Area/Terrain” is changed to “Area/site”

4. The work plan:

- (1) Members are requested to submit the set of characters from Oracle Bone inscription categorized under the first 180 radicals of Shuowen Jiezi.
- (2) Members should submit the images and attributes of the Oracle Bone inscriptions by 8th May, 2006.

(3) Data Format For Old Hanzi Data Exchange

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

Images format:

- 3.1 use PNG storage format.
- 3.2 The original glyph should be scanned at 300 dpi (dots per inch).
- 3.3 The transcribed glyph images are named [ID]+[_R] (for example, if the ID is T00001, the transcribed glyph images should be named T00001_R).
- 3.4 The original glyphs are named [ID]+[_O] (for example, if the ID is T00001, the original glyph should be named T00001_O).
- 3.5 Glyph determination images are named [ID]+[_D] (for example, if the ID is T00001, the glyph determination image should be named T00001_D).
- 3.6 The images of unifiable shapes are named [ID]+[Unifiable shapes ID] (for example, if the Old Hanzi 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"/>
  </Character>
</OldHanZi>

```

<Note/>
</Character>
</OldHanZi>

Members attended the meeting were Dai Hong, Li Guoying, Yin Jianghong, Zhang Deshao, Shi Jianqiao, Julie S. C. Chuang, Selena Wei, Hsu Hsueh Jen, Chen Wen-Nan, ANAN Kanghong, Lu Qin, Tom Bishop, Cook Richard.

End of document