Title: Request for comment on encoding Oracle Bone Script
Source: TCA and China
Action: For consideration by JTC1/SC2/WG2

TCA and China will submit a formal proposal to WG2, asking to encode Oracle Bone Script in UCS for future WG2 meeting once the Oracle Bone Script repertoire accomplished. Before that, TCA and China have provided about 1,463 sample characters of Oracle Bone Script (SW radical 1 to 88) to WG2 for expert's discussion and feedback.

Currently proposed version is based in the result of IRG/Oracle Bone Script experts group. Experts from TCA and China recheck one by one and have removed a few characters of unclear sources. We have also invited Oracle Bone Script group expert from the Project "Chinese characters repertoire" Dr. Huang Tianshu as adviser.

1 Introduction

The Oracle Bone Script (甲骨文) is the origin of all CJK ideographs, especially modern Hanzi (漢字, i.e., Chinese characters). The Oracle Bone Script originated about 3,300 years ago, the emperors of the Shang Dynasty (商朝, about 1,558 BC to 1,046 BC) incised divination inscriptions on tortoise shells and ox scapulas, which constituted "Oracle Bone Script" and circulated to date. According to statistics, approximately at least more than 1,000,000 pieces of oracle bones carved with scripts were discovered so far, the inscriptions incised on the bones are more than one million, and a total of 4,378 characters were used. Recorded content could be divided into four categories, namely:

(a) Class and country:

    (1) slaves and civilians,

    (2) slave-owning aristocracy,

    (3) officials,

    (4) troops, penalty and prison,

    (5) wars,

    (6) territories, and

    (7) tributes.

(b) Social production:

    (1) agriculture,

    (2) fishing, hunting and animal husbandry,

    (3) handcrafts, and

    (4) commerce and transportation.

(c) Ideology and culture:

    (1) astronomy and calendar,

    (2) meteorology,

    (3) architect,

    (4) siseases,

    (5) fertility,

    (6) spirit worship,

    (7) worship,

    (8) good dreams and nightmares,

    (9) divinations, and

    (10) writings.

(d) Others.

Leave a great deal of information about the history and culture of the Shang Dynasty, which can be used for academic research worldwide.

The Oracle Bone Script is a group of fairly mature Chinese characters, applying the important character-creating and character-using methods, including pictograms, ideograms, combined ideograms, phonograms, transfer characters and phonetic loan. It is an important material for discovering the origin of Chinese characters. Viewing from a macro perspective, Oracle Bone Script is not just a heritage of Chinese culture, but some Asian countries, such as Japan, Korea and Vietnam, can also use Chinese characters to some extent. Oracle Bone Script is the root of Chinese characters; it is worthy to be treasured by countries all over the world.

The collections of the raw materials of Oracle Bone Script include the *Jiaguwen Heji* (甲骨文合集) which contains a largest amount of information, the *Jiaguwen Heji Bubian* (甲骨文合集補編) and other various collections. Following publications classified Oracle Bone Script according to "character", including *Jiaguwen Bian* (甲骨文編) written by Sun Haibo (孫海波), Xu *Jiaguwen Bian* (續甲骨文編) written by Jin Xiangheng (金祥恆), Xien Jiaguwen Bian (新甲骨文編) written by Liu Zhao's (劉釗), and Jiaguwenz Bian (甲骨文字編) written by Li Zongkun (李宗焜), which allows us to easily find different glyphs of a Oracle Bone Script character. If somebody want to know in which sentence of a book he/she can find a specific Oracle Bone Script character, he/she can look up *Yinxu Buci Zongshu* (殷墟卜辭綜類) written by Kunio Shima (島邦男) or *Yinxu Jiagu Kecleizhuan* (殷墟甲骨刻辭類纂) edited by Yao Xiaosui (姚孝遂). He/She can also find different explanations of an Oracle Bone Script character either in *Jiaguwenzjish* (甲骨文集釋) written by Li Xiaodieng (李孝定) or *Jiaguwenzi Gulien* (甲骨文字詁林) written by Yu Shengwu (于省吾). There are many other Oracle Bone Script studies and publications available, our knowledge of the Oracle Bone Script has come to a very mature stage, we can now make full use of the Oracle Bone Script in a wider range.

The Oracle Bone Script is the first well developed ancient Chinese writing, containing a considerable number of pictographic drawings as it's the oldest writing. It can help scholars to explore the significance of the earliest Chinese characters and provides a wonderful and lively glyph application for cultural and artistic development. The structure of Oracle Bone Script has not been fully finalized as it is the earliest script, thus a character often has many manifestations, such as the same character can be presented in the way of pictographs, combined ideograms, phonograms or phonetic loan. The radical structure of a Oracle Bone

Script character can be varied according to the changes of positions and number. Like Bronze Inscription (金文), Warring States Script (戰國文字) and Small Seal Script (小篆), Oracle Bone Script is different from the modern Regular Script (楷書) hence it does not belong to the same system as the modern Regular Script does. Oracle Bone Script has a lot of glyphs and cannot be mapped with the modern Regular Script of Chinese character, a glyph that can be mapped also has a unique structure of form, as a result, it cannot be treated as a font of modern Chinese characters and attach to the Regular Script, consequently, it has to be placed in an exclusive encoded area.

In summary, the Oracle Bone Script plays an essential role in the development of Chinese characters, the unique value of the Oracle Bone Script to the research and collation of ancient and modern Chinese characters lies in:

(a) The Oracle Bone Script represents the earliest stage of the development of ancient characters, it is the origin of the later Bronze Inscription, Warring State Script, Small Seal Script and Regular Script, and it is linked to the main components of various typefaces of later generations, so it is a required reference character set for the collation and textual research of Chinese characters.

(b) The glyphs and contents of the Oracle Bone Script are closely related to the history and culture of the Shang Dynasty, through the understanding of the Oracle Bone Script, people will be able to know the true facts of that ancient civilizations at their early stage.

(c) The Oracle Bone Script has been discovered for more than one hundred years, over the years, leading scholars have done a lot of in-depth research, and have achieved fruitful results. The completion of various works is quite adequate for us to establish a sound character set of the Oracle Bone Script.

(d) As the common source of the Chinese character cultural circle, the Oracle Bone Script is the common property of the Chinese cultural circle. As the oldest writing of one of the four ancient civilizations, it is also the common property of the whole world.

Since the Oracle Bone Script repertoire and its contents play an important role in the research of Chinese characters and in the history of the world civilization, it is a priority to set the international code for the Oracle Bone Script as well as to lay a good foundation for the encoding of other ancient characters.

2 Collation resources of including Oracle Bone Script in International Encoding

The collations of Oracle Bone Script character set are mainly based on the following publications:

(A) stands for《甲骨文合集》

(B) stands for《甲骨文合集補編》

(C) stands for《花園莊東地甲骨》

(D) stands for《山東濟南大辛莊甲骨》

(E) stands for《周原甲骨》

(F) stands for《小屯南地甲骨》

(G) stands for《英國所藏甲骨集》

(H) stands for《 懷特氏等所藏甲骨文集》

(I) stands for《天理大學附屬天理參考館藏甲骨文字》

(J) stands for《德瑞荷比所藏一些甲骨錄》

(K) stands for《瑞典斯德哥爾摩遠東古物博物館藏甲骨文字》

(L) Zin Jiaguwen Bian《新甲骨文編》

(M) Jiaguwen Bian 《甲骨文字編》

3 Rules for Collation

Please check the document on IRG N1271 Old Hanzi Principles and References

4 Experts of the working group

TCA: 李鋈**(Xian Li),** 鍾柏生（Bor-sheng Jung）**,**季旭昇**(Hsiu-sheng Chi),** 許學仁 **(Xue-jen Hsu), **袁國華**(Kwok-wa Yuen), **魏林梅**(Lin-mei Wei), **莊素琴 (Suh-chyin Chuang)

China:

Working Group(工作组)：趙成**(Zhao Cheng),** 李國英**(Li Guoying),** 周晓文**(Zhou Xiaowen),** 代红（**DaiHong**）,刘志基（**Liu Zhiji**）,尹江红（**Yin Jianghong**）

Adviser 咨询专家：王寧**(Wang Ning),** 黃天樹**(Huang Tiansh),** 王蘊智（Wang Yunzhi）, 赵平安（Zhao Ping 'an），董莲池（**Dong Lianchi**），赵爱学（Zhao Aixue）

Organization: Chinese Foundation for Digitization Technology, the Project of Chinese Character Repertoire, China Electronics Standardization Institute

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest** *Roadmaps***.**

A. Administrative

1. **Title:**	**Proposal on Encoding Oracle Bone Script in UCS**
2. Requester's name:	*TCA*
3. Requester type (Member body/Liaison/Individual contribution):	Liaison
4. Submission date:	*2015-10-13*
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	
(or) More information will be provided later:	*Yes*

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters):   *Yes*
       Proposed name of script: **Small Seal Script**
   b. The proposal is for addition of character(s) to an existing block: *No*
       Name of the existing block:

2. Number of characters in proposal: *2,000*

3. Proposed category (select one from below - see section 2.2 of P&P document):   *F*
   A-Contemporary   B.1-Specialized (small collection)   B.2-Specialized (large collection)
   C-Major extinct   D-Attested extinct   E-Minor extinct
   F-Archaic Hieroglyphic or Ideographic   Ideographic   G-Obscure or questionable usage symbols

4. Is a repertoire including character names provided?   *No*
   a. If YES, are the names in accordance with the "character naming guidelines"
       in Annex L of P&P document?
   b. Are the character shapes attached in a legible form suitable for review?   *Yes*

5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
       TCA
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
       Already agreed with all TCA · Experts and related organizations

6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?   *Yes, see attached PDF file*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
   of proposed characters attached?   *Yes*

7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input,
   presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?   *No*

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

---

[1] Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? _No_
    If YES explain
2. Has contact been made to members of the user community (for example: National Body,
    user groups of the script or characters, other experts, etc.)? _No_
        If YES, with whom? _Experts of Small Seal, Education, Publishing Industry and seals._
        If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example:
    size, demographics, information technology use, or publishing use) is included? _No_
    Reference: _Evidence of use of these characters were reviewed by experts of Small Seal at the time of submission for Inclusion._
4. The context of use for the proposed characters (type of use; common or rare) _Common_
    Reference: _Widely used in community of Hanzi users for Academia,   books, decorations, etc.._
5. Are the proposed characters in current use by the user community?
    If YES, where?   Reference: Education, _Publishing Industry and Stamp industry_
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
    in the BMP? _No_
        If YES, is a rationale provided?
        If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? _Yes_
8. Can any of the proposed characters be considered a presentation form of an existing
    character or character sequence? _No_
        If YES, is a rationale for its inclusion provided?
        If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either
    existing characters or other proposed characters? _No_
        If YES, is a rationale for its inclusion provided?
        If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function)
    to, or could be confused with, an existing character? _No_
        If YES, is a rationale for its inclusion provided?
        If YES, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences?
    If YES, is a rationale for such use provided? _No_
        If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
        If YES, reference:
12. Does the proposal contain characters with any special properties such as
    control function or similar semantics? _No_
        If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility characters? _No_
    If YES, are the equivalent corresponding unified ideographic characters identified?
        If YES, reference: