ISO/IEC JTC1/SC2/WG2

Universal Multiple-Octet Coded Character Set (UCS)

Secretariat: ANSI

Title: **Proposal to encode six CJK Ideographs in UCS**

Doc. Type: national body contribution

Source: Japan

Project:

Status: To be discussed at the WG2 meeting 51 in Hangzhou, China

Date: 2007-09-07

Distribution: ISO/IEC JTC1/SC2/WG2 and IRG experts

Attachments: Proposal summary form

Background and Rationale

ARIB STD-B24 Version 5.1 ("Data Coding and Transmission for Digital Broadcasting" Association of Radio Industries and Business) specifies text coding for data broadcasting, which is carried out as part of the digital broadcasting standard in Japan. It adopts 7bit and 8bit codes and their code extension technique according to ISO/IEC 2022 for text data transmission as well as ISO/IEC 10646. The primary method of data transmission is ISO/IEC 2022 based 7bit and 8bit codes and it specifies supplementary character (GAIJI) set as well as standard character sets (JIS X 0201, JIS X 0208, JIS X 0213:2004) to be available for digital broadcasting service in Japan. There are 137 CJK Ideographs defined in the GAIJI set. After full examination by Japanese national body members, six unencoded characters in UCS (including CJK Unified Ideographs in FPDAM4 and CJK Unified Ideographs Extension C in PDAM5) are found among the GAIJI set. Japan requests to encode the remaining six characters proposed in this document in the earliest opportunity in order to establish the standard transcoding for UCS based Digital TV devices because the market demand of those devices with data exchange capability according to this specification is expected to be very strong and its implementation phase has already begun in Japan.

Proposed CJK Unified Ideographs

Number of characters: two

Proposed block: CJK UNIFIED IDEOGRAPHS

Glyph	ARIB ID	Proposed Source Reference	IDS	Proposed Name
梁	47	J_ARIB-754F	▢▢氵刃木	CJK UNIFIED IDEOGRAPH-XXXX
襹	95	J_ARIB-7621	▢衤零	CJK UNIFIED IDEOGRAPH-XXXX

Proposed CJK Compatibility Ideographs

Number of characters : four

Proposed block : CJK COMPATIBILITY IDEOGRAPHS

Glyph	ARIB ID	Proposed Source Reference	Compatibility equivalence	IDS	Proposed Name
惠	39	J_ARIB-7547	U+06075 惠	▢惠丶	CJK COMPATIBILITY IDEOGRAPH-XXXX
熙	67	J_ARIB-7563	U+242EE 熙	▢▢臣己灬	CJK COMPATIBILITY IDEOGRAPH-XXXX
祓	93	J_ARIB-757D	U+07953 祓	▢衤友	CJK COMPATIBILITY IDEOGRAPH-XXXX
舘	105	J_ARIB-762B	U+08218 舘	▢舎官	CJK COMPATIBILITY IDEOGRAPH-XXXX

Table 7-11 Additional Kanji Characters, Page 73 ARIB STD-B24 Vol.1 Part 2

http://www.dibeg.org/aribstd/STD-B24/ARIB_STD-B24-1_v5.1_E1.pdf

--- end of document ---

Attachment:

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://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.

See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest *Roadmaps*.

A. Administrative

1. **Title:**	*Proposal to encode six CJK Ideographs in UCS*
2. Requester's name:	*Japan*
3. Requester type (Member body/Liaison/Individual contribution):	*Member body*
4. Submission date:	*September 7, 2007*
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	*Yes*
(or) More information will be provided later:	

B. Technical – General

1. Choose one of the following:

   a. This proposal is for a new script (set of characters): *No*

   Proposed name of script:

   b. The proposal is for addition of character(s) to an existing block: *Yes*

   Name of the existing block: *CJK Unified Ideographs and CJK Compatibility Ideographs*

2. Number of characters in proposal: *six*

3. Proposed category (select one from below - see section 2.2 of P&P document):

   A-Contemporary x   B.1-Specialized (small collection)   B.2-Specialized (large collection)

   C-Major extinct   D-Attested extinct   E-Minor extinct

   F-Archaic Hieroglyphic or Ideographic   G-Obscure or questionable usage symbols

4. Is a repertoire including character names provided? *Yes*

   a. If YES, are the names in accordance with the "character naming guidelines"

   in Annex L of P&P document? *Yes*

   b. Are the character shapes attached in a legible form suitable for review? *Yes*

5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for

   publishing the standard? *True Type, Japan NB*

   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools

---

used:

6. References:

    a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?   *Yes*

    b. Are published examples of use (such as samples from newspapers, magazines, or other sources)

    of proposed characters attached?   *No*

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 http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

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.)? | *Yes*

        If YES, with whom? | *Associations of Radio Industries and Businesses*

        If YES, available relevant documents: | *ARIB STD-B24 Version 5.1, "Data Coding and Transmission Specification for Digital Broadcasting"*

http://www.dibeg.org/aribstd/STD-B24/ARIB_STD-B24-1_v5.1_E1.pdf

3. Information on the user community for the proposed characters (for example:

    size, demographics, information technology use, or publishing use) is included? | *Yes*

    Reference: | *http://www.arib.or.jp/english/index.html*

4. The context of use for the proposed characters (type of use; common or rare) | *rare*

    Reference:

5. Are the proposed characters in current use by the user community? | *Yes*

    If YES, where?   Reference: | *The document provided for question C2*

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)? | *No*

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 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? | *No*

    If YES, is a rationale for such use provided?

        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 character(s)? *Yes*

If YES, is the equivalent corresponding unified ideographic character(s) identified? *Yes*

If YES, reference: *See main proposal.*