

**ISO/IEC JTC 1/SC2/WG2/IRG**

**Ideographic Rapporteur Group**

**Secretariat: China**

Source/Contribution Identifier: **TCA/ASCC0001**

Meeting: @/for **Singapore** Error! Reference source not found.**99**

Issue Number:

Project/Task identifier:

TITLE: **Activity Report from TCA**

Status: Member body's contribution

Keywords:

Short Description:

Proposed Conclusion/Requested Action:

**To be referenced by IRG members.**

Argument/Text of Contribution:

Contact: Mr. Zhang Zhoucai, CCID Senior Engineer, IRG Rapporteur  
Address: No. 14. 3/F, ZhiXinCun XiaoQu, Beijing 100083, China  
Tel: (8610)6234.0198, 6234.5536 Fax:(8610)6231.2924  
E-mail: joezhg@public.bta.net.cn  
Web Site:<http://www.cs.cuhk.edu.hk/~irg>

# **Activity Report from TCA**

## **1. New web site for CNS 11643 is ready (<http://www.cns11643.gov.tw>)**

To widely promote the application of CNS 11643 and to meet the user requirement for Chinese information interchange, the EDPC (Electronic Data Processing Center) of the Executive Yuan in Taiwan newly creates a “Chinese Information Interchange Code full set server” on Internet.

The main purpose of this web site is to provide Chinese bit map fonts in the whole set CNS 11643 to help user reduce the efforts of creating UDC(User Defined Characters) of their own. This web site provides a searching engine with several optional character attributes to help user easily find out any specific character they need. The searching indexes include CNS code, stroke number, Bopomofo, KangXi Radical, CangJie input method, and combinational search. Once user find out the character, he can download the bit map font of that character for the usage of interchange. The conversion tool between CNS 11643 to/from internal code, like Big-5, is also provided. This new web site is ready since October 1999 (<http://www.cns11643.gov.tw>).

## **2. Standard of sorting attributes for Chinese data processing has been drafted**

To facilitate the standardization of Chinese text sorting in Taiwan, Institute of Information Industry has collaborated scholars, government officers, and experts to draft a standard called “sorting attributes for Chinese data processing” for BSMI (Bureau of Standards, Metrology and Inspection).

The draft standard is based on the whole set of CNS 11643. Following seven attributes have been provided for each character in CNS 11643 :

- Radical
- Bopomofo
- Liu code for pronouciation
- Total stroke number
- Stroke number of radical
- Stroke number exclude radical
- Stroke order

Another two attributes include “Numeric value” and “Symbol value” have also been given to help sorting Chinese symbol characters by semantics. The draft has been submitted to BSMI, and is reviewed by the technical committee

of BSMI currently.

### **3. Chinese characters collecting works**

The Scripta Sinica is a very large-scale textual database developed by Academia Sinica, which contains a lot of ancient Chinese books and documents. Academia Sinica has made more than 10,000 UDCs for ancient Chinese books in the Scripta Sinica. Computing Center, Academia Sinica has matched 2,800 UDCs with the characters in CNS 11643, and found that 761 UDCs aren't in CNS 11643.

In addition, the CBETA (Chinese Buddhist Electronic Text Association) is developing a buddhist document database attempting to digitize all Chinese buddhist documents. Currently, 32 buddhist documents were well-done and the CBETA found about 700 characters outside the Unicode character set. And 55 buddhist documents were prepared and the CBETA found about 10,000 characters outside the Big5 character set.

In the future, Computing Center, Academia Sinica will do the following works:

1. Collect UDCs from the Scripta Sinica and from the CBETA project, and mixes them together to form a UDC set.
2. Match the UDC set with CNS 11643 and ISO 10646 BMP, CJK Extension A and B to find the duplications.
3. Remove duplicated characters from the UDC set to form a new-found character set.
4. Propose the new-found character set to be the source of CJK Extension C.

### **4. Workshop on the problems & solutions of Chinese characters for digitizing Chinese documents**

Academia Sinica hosted a workshop in which 25 scholars and experts reported the problems and/or solutions of digitizing ancient Chinese books, Buddhist documents and Taoist documents. The issues include identifying and creating UDCs, sharing information and experiences. The workshop was on June 14 – 16, and about 60 scholars and experts attended the workshop.