

CBETA Taisho Electronic Tripitaka

Huimin Bhikkhu and Aming Tu
Chinese Buddhist Electronic Text Association, Taiwan

Christian Wittern
Kyoto University, Japan

Introduction

The "Chinese Buddhist Electronic Text Association" (CBETA) has been formally established on Feb. 15, 1998 with support from the *Yinshun Foundation of North-America* and the *Chung-Hwa Institute of Buddhist Studies* to create, maintain and distribute free of charge an electronic version of the Chinese Buddhist Tripitaka.

The *CBETA Chinese Electronic Tripitaka* is based on the Taisho Tripitaka 「大正新脩大藏經」 (© Daizo Shuppansha, Inc. 大藏出版株式會社) Vols 1-55 and 85. The right to input and distribute this database has been officially granted by the copyright holder Daizo Shuppansha.

March 2001, CBETA has finished proofreading Taisho Tripitaka Volumes 1-55 & 85. The electronic text files of the Taisho Tripitaka are created from one set of XML source files. In that distribution, the following formats are included: 'App' version for online search, HTMLHelp version, HTML version and Doc Version.

Besides, the CBETA progress for year 2000 was also focused on the production of the revised edition of CBETA CD-ROM. 5000 copies were released and distributed to scholars, researchers, academic institutes, libraries and Buddhist organisations at home and abroad.

Contents of the CBETA CD-ROM

The revised CBETA CD-ROM is much more user-orientated and hence user-friendly in respect of the enhancement of the interface and function of the existing HtmlHelp version and the addition of a MS-WORD version that supports the Unicode and multi-languages platforms, as well as the equipment of the text-only Normal and App versions required for Macintosh and other retrieval system.

In addition to the scriptures in electronic form, we also include reference works that can be used with the texts. Prof. Charles A. Muller (Toyo Gakuen University) kindly gave permission to include his Dictionary of East Asian Buddhist Terms. Since this dictionary is being continually enhanced, we suggest users visit his homepage for updates. The electronic version of the "Dictionary for Buddhist Studies" 佛學大辭典 by Ding Fubao included with this CD-ROM was contributed by the Electronic Buddhadharma Society. We like to express our gratitude for this valuable contribution.

Progress of year 2000-2001

Although the primary objective of the year of 2000 was the production and release of the revised CBETA CD-ROM that included conclusion of the pre-release preparations, such as verification and revision of the Indian Canons (sutra-pitaka, vinaya-pitaka and abhidharma-pitaka) of the Taisho Edition from Vols. 1-32, other schemed operations were timely accomplished, namely input completion of the texts in Siddham script appearing in the Esoteric scriptures of Vols. 18-21 of the Taisho Edition. Proofreading and creation of the fonts thereof was to conclude by the end of year 2000. And the rest of the other volumes (Vols. 33-55 & 85) is expected to finish in the beginning of year 2001.

Expected to encounter difficulties in handling dual-line small characters appeared in the latter volumes, but should be able to finalise the first phase of proofreading of Vols. 1-55 and Vol. 85 by the March 2001 as schemed.

Research and Developing

The undertakings for this year were:

1. revising relevant application programmes, and
2. maintaining and revising data;
the specifications of which could be categorised into and comprised of
 - (1) documentation: rare characters handling, Xml files processing, various versions, such as Doc., HtmlHelp, Normal, App. and original files conversing; and
 - (2) information technology: development of relevant applications and software, CD-ROM setup programmes and programmes for multi-versions conversion, collation of various versions, and proofreading of HtmlHelp indexes.
3. conversed inputted and proofread texts into various versions, such as Xml, DOC,
4. researching and developing programmes for converting texts and post-conversion proofreading,

5. processing, encoding and web-enabling Siddham, (Romanising) Sanskrit and Japanese scripts and conversing the texts into Xml versions,
6. handling of rare characters in the Electronic Tripitaka were completed and the data was submitted to the Mojikyo Font Institute for further processing,
7. reviewing the Xml, Normal and HtmlHelp versions of Vols.
8. developing fonts for the Siddham scripts in Vols. 18-21 and an improved inputting method therefor, and
9. processing Sidham script files from Word to TXT to XML and reviewing the processed files (cxssidham).

Progresses for rare characters handling for this quarter were:

1. accomplished the processing of common rare characters
2. encoding Mojikyo rare characters
3. encoding four-corner/radical/stroke/pinyin rare characters.

Additionally,

1. completed markup of Siddham XML file, and
2. researched and developed programmes for transferring texts and post-transfer proofreading.

With respect to year 2001's Project CBETA CD-ROM, the issue of XSLT Solution has been satisfactorily dealt with. After some adjustments in the Saxon Java Source of the XSLT Processor – Saxon that could previously only be used in processing Xml files in Unicode encoding, files in Big5 XML as well as several other encodings could now be converted. Concurrently steps are taken to the testing of Perl Solution for a possible installation of Perl/Tk user interface.

In view of its dedication to and significant contribution in TEI markup of Chinese Database, CBETA was invited by the Academia Sinica to participate in the Manuscript Markup Research Team's symposium. During the deliberation, CBETA was asked to introduce and demonstrate its markup techniques in electronic Buddhist canons. CBETA was asked to return to the Academia Sinica to give a TEI training course in November at its Information Centre, and in January 2001, CBETA is scheduled to attend the PNC Conference to be held in Hong Kong in January 2001 to introduce and demonstrate its TEI markup technique.

Future plans

Preparatory works were carried out in the last quarter of 2000 for the processing of Zokuzokyo (卍續藏), such as collation and OCR testing, input related planning ... etc. And this hope to be another project processed, and which is different from the Taisho Tripitaka Project that CBETA originally produced. The later half of year 2000 marked the commencement of the preliminary test input of Zokuzokyo. Currently completed the task of indexing of Zokuzokyo, scanning of illustrations and texts and OCR conversion of the latter.

Recognising the processing of Zokuzokyo presents other distinctive issues dissimilar to those of Taisho Canons, the pre-process surveying and analysing, as well as having better utility programmes are essential and priorities on CBETA's agenda for this year.

And, undoubtedly, the Taisho Scholarly Edition will be carried continuously in the following years.