

From CBETA (Chinese Buddhist Electronic Text Association) Electronic Tripitaka Collection to BIP (Buddhist Informatics Program) and IBA (Integrated Buddhist Archives)

Huimin Bhiksu

President, Dharma Drum Buddhist College
Professor, Taipei National University of the Arts
Director, Chinese Buddhist Electronic Text Association

1. Foreword

In this age of heavy reliance on computer networking system for information exchange, the way people communicate and process data and knowledge has changed, which manifests in the unprecedented efficiency of such tasks as the acquisition, recording, organization, retrieval, display, and dissemination of newly created digitized information or electronic media. This approach is expected to substantially transform people's living environment, social structure and the development of civilization.

In history, we found religious literature had played a significant role in the development of technology applied in cultural media. For example, one of the earliest printed matters in existence is the Buddhist Diamond Sutra (yr. 868), and another example is Gutenberg Bible (yr. 1455), the first book of movable type printing in Europe. Along with the development of printing technology, the rapid increase of printed matters, and the gradual prevalence of written materials, knowledge has pervaded in the society accordingly, and which allows the civilization of human beings to progress day by day to the extent of the so-called age of 'knowledge explosion' nowadays.

Buddhism was a religion of long-established tradition, and the amount of its related materials and knowledge is as vast as the sea, which sometimes seems inaccessible. Take the publication of Buddhist scriptures for example, soon after the passing of the Buddha, the first Council was held at Rajagaha (today's Rajgir), convened by Mahakassapa the Elder, and then the second and the third councils were held at Vesālī (BC 280) and at Pātaliputra (today's Patna;

BC251) respectively. The ways and contents of the Buddhist councils reflect the backgrounds of their ages. Taking an overall view of the current century, we see the publication and translation of Buddhist scriptures being undertaken constantly, which significantly contributes to the dissemination of Buddhist knowledge and the internationalization of Buddhist studies. The situation also reflects how Buddhism has become more and more popular. However, currently, near the end of the century, due to the world's heavy reliance on computer networking system, or the Internet, there are increasing needs for digitized Buddhist scriptures or the electronic Buddhist Text in the public, and subsequent plans for digitizing and creating databases of Buddhist texts are emerging one after another in its original languages of Pali, Sanskrit, and translations of Chinese and Tibetan, etc.

Among them, the Chinese Buddhist Electronic Text Association (<http://www.cbeta.org>) was founded on Feb 15, 1998 to assume the production (including tasks of input, collation, mark-up, processing of out-of-list characters, and search system etc.) of the Chinese electronic Buddhist Text, based on Taisho Tripitaka (Daizo Shuppansha©) Vol. 1- 85¹, and Shinsan Zokuzokyo (Kokusho Kankokai ©) Vol. 1- 90, under official grant for input and distribution by the copyright holders² (as stated above). The Text is being shared on the Internet and distributed yearly in CDs. The most recent CBETA Chinese Electronic Tripitaka Collection Version 2007 has been published on Mar 10, 2007 and is being distributed in CDs (containing Taisho Tripitaka Vol. 1-55 & 85, and Shinsan Zokuzokyo Vol. 1-88).

2. Creation Procedure and Related Techniques for CBETA Chinese Electronic Tripitaka Collection

The creation procedure for the CBETA Chinese Electronic tripitaka collection contains two major stages. First, the original digital files are generated in either of the two ways, namely manual typing and optical character recognition (OCR). The texts then go through a procedure for filling in out-of-list characters and line head information, as well as collations by computer programs, to efficiently produce electronic files closer to the correct ones. Next follow manual collation and mark-up process prior to the output of the mater files of the database of CBETA electronic Buddhist Text. The operating procedure and related techniques are as follows:

2.1 Procedure for manual typing and optical character recognition

The procedure for optical character recognition (OCR): scanning in the texts → processed by

computer program for kaeriten removal →OCR→processed by computer program for incorrect strings replacement →processed by program for image analysis →electronic files containing information from original texts.

2.1.1 OCR kaeriten removal program

In generating the Taisho Tripitaka electronic files via OCR, the first problem to be solved is the kaeriten (the symbol added by Japanese scholars according to their grammar rules to facilitate the reading of Chinese Buddhist Text) appearing in the image files of the scanned texts (Figure 1). The kaeriten symbols will result in lower accuracy in OCR, and therefore need to be removed by applied kaeriten removal computer program (Figure 2).

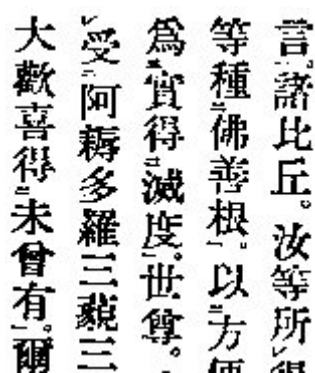


Figure 1 : image with kaeritens

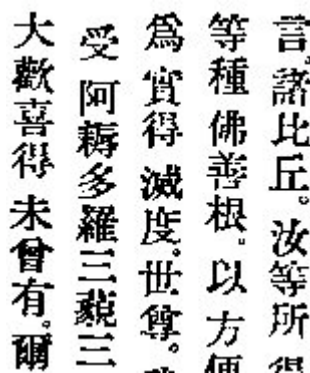


Figure 2 : image with kaeritens removed

2.1.2 Computer program for replacing incorrect strings (Figure 3) .

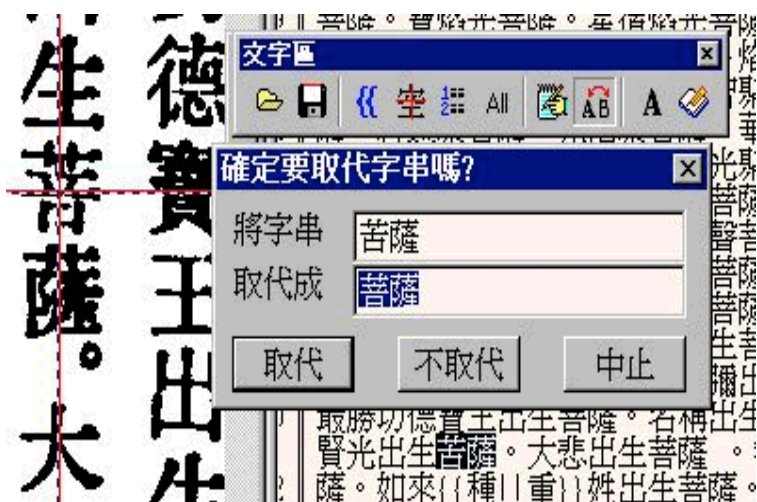


Figure 3

2.1.3 Image analysis program

The original text is recovered by analyzing the scanned image of the original Taisho Text and by adding in kaeritens, punctuation marks, and correction numerals and asterisks to the corrected text (Figure 4), using relevant application programs.

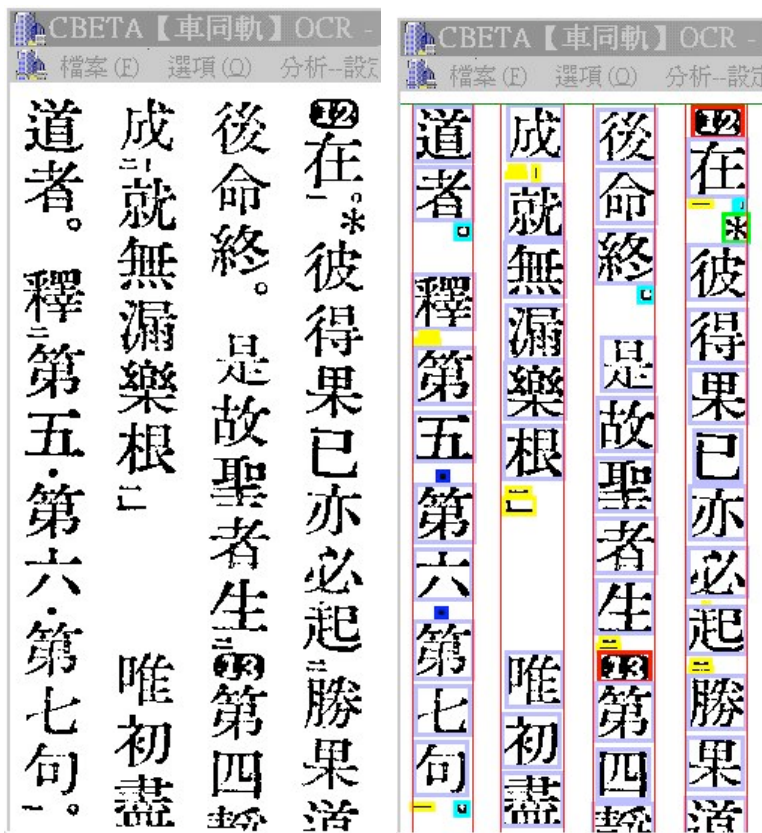


Figure 4

2.2 Process for treating out-of-list characters

Out-of-list characters (characters that are not collected in the Big5 system) are displayed by the computer in combined characters or as an image, and relevant information about these characters is also built up. As for the markings in the XML files, the customized CBETA CB encoding system is used as the standard.

Out-of-list characters in the Text of popular edition are shown with their substitutes offered by the Big5 system. For instance, 「𨮑」 is substituted for [金*本]. In an environment supporting the Unicode, the UTF8 character set can be used to display the out-of-list characters, which otherwise are displayed in combined characters.

2.3 Line head information

In order for the computer's file collation and full text searching, the texts must be presented in a format with line head information as follows:

```
T09n0278_p0395a01 ||
T09n0278_p0395a02 ||      No. 278 [No. 279]
T09n0278_p0395a03 ||大方廣佛華嚴經卷第一
T09n0278_p0395a04 ||
T09n0278_p0395a05 ||      東晉天竺三藏佛[1]駄[2a][2b]跋陀[3]羅譯
T09n0278_p0395a06 ||      世間淨眼品第一[4]之一
T09n0278_p0395a07 || 如是我聞。一時佛在摩竭提國寂滅道場。始
T09n0278_p0395a08 || 成正覺。其地金剛具足嚴淨。眾寶雜華。以為
```

The line head information takes down the location of the following text in Taisho Tripitaka, including the ordinal numbers of volume, sutra, page, column, and line, so that it not only facilitates proofreading after file collation, but also allows the searching program to precisely point to the searched results.

2.4 Collation by computer program

Seeing that manual proofreading in traditional way, undergoing four or even ten repetitions, does not always assure absolute correctness, the CBETA uses a computer-collating program (Figure 5) to compare the source files,

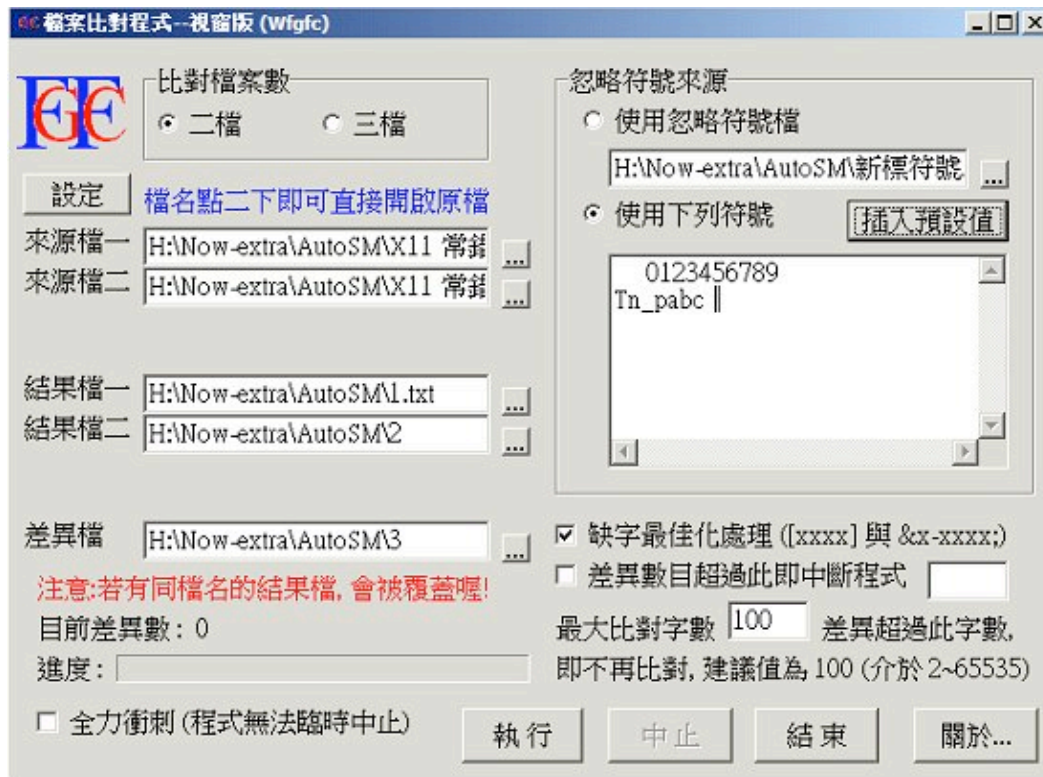


Figure 5

and to more efficiently find out the differences as follows:

```

*****
大方廣佛華嚴經卷第一
東{{普-(恙-王-心)||晉}}天竺三藏佛馱跋陀羅譯
世間淨眼品第一之一
如是我聞。一{{時||非}}佛在摩竭提國寂滅道場。始
成正覺。其地金剛具足嚴{{淨||&K2-E8C9;}}。眾寶雜華。以為
*****

```

2.5 Image-text comparison

By comparing the previously scanned image with the texts containing computer-collated results, an image-text comparing program takes up the job of proofreading. (See Figure 6.)

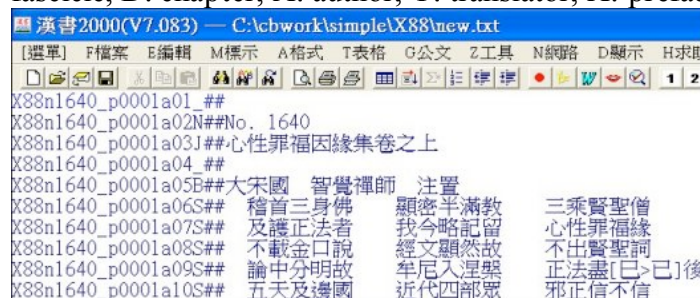


2.6 Manual proofreading

The final result out of image-text comparison by computer is prepared in hardcopy for manual proofreading carried out by volunteers.

2.7 The markup process

The markup process comprises of two parts. First, a simple markup procedure is undertaken (Figure 7) to tag with alphabetic letters such format attributes of the text as N: sutra number, J: fascicle, D: chapter, A: author, Y: translator, X: preface, S: verse, Z: mantra, W: remarks, P: paragraph, Q: other topics, P: phrase within a line.



Next, a computer program is used to transform the files with simple tags into XML documents (Figure 8) to work as the mater files of CBETA electronic tripitaka database. To meet with

different requirements, the Text database accessed through XML files can be transformed by computer programs into various formats including Normal, App, HTML, PDF, PDA, etc. for users to download them on the Internet.

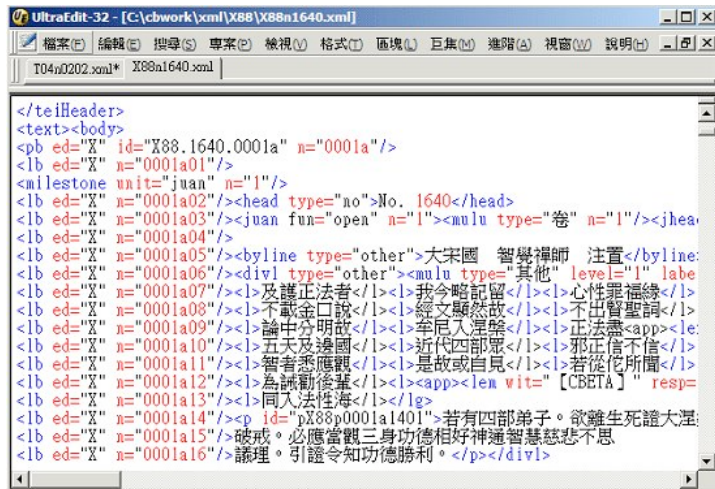


Figure 8

3. Features of CBETA Chinese Electronic Tripitaka Collection Version 2007

3.1 Contents

- A. Contains Taisho Tripitaka Vol.1-55 & 85, and Shinsan Zokuzokyo (Xuzangjing) Vol. 1-88. With 144 volumes, 3,597 categories, and 10,434 fascicles in total. Over 150,000,000 words totally in XML format.
- B. With CBReader browsing system providing various functions for browsing and making queries in the tripitaka database, and built-in operation interfaces in complex and simplified Chinese, as well as English and Japanese.
- C. Provides Buddhist Dictionaries such as Ding-Fubao Buddhist Dictionary and Nanshan Vinaya Dictionary.

3.2 Functions and features of CBReader:

- A. Multilanguage platform
CBReader is accessible with full function under Windows system in versions of complex and simplified Chinese, as well as English and Japanese. Other languages are provided

under extended options.

B. Tables of Contents

In addition to the original structures of Taisho and Xuzangjing, a CBETA table of contents combining Taisho and Xuzangjing is also provided. Users can open either the CBETA table of contents, or a single volume or category of Taisho or Xuzangjing tripitaka alone, all with multiple choices.

3.3 Search Engine

- A. CBETA has developed its own full-text search engine, supporting such operations as AND, OR, BEFORE, NEAR, and the wildcard designator, to render more precise search results.
- B. Users may check search ranges among categories, volumes, or select a certain sutra (scripture). They may also specify the searching to be taken in the queried texts, or in previous search results.
- C. The full text searching function will automatically replace non-Big5 Chinese Characters with Big5 characters.
- D. Search in the tables of contents: allows search by the title of scripture or by the topics listed in the table of contents of a certain scripture.
- E. Search for scriptures: a certain scripture can be found by specifying such information as category, group, volume number, number or title of sutra, and author or translator.
- F. Goto: This command directs the pointer of the search engine to jump to a specified location, which can be a certain book, page, column, line, or a category, volume, sutra, fascicle, or just the line header.

3.4 Citation duplication

The Duplicate Citation function will simultaneously display the texts cited from the three different versions of Shinsan Zokuzokyo. For example:

《天聖廣燈錄》「天聖廣燈錄卷第一」(CBETA, X78, no. 1553, p. 420, a13 // Z 2B:8, p. 298, a8 // R135, p. 595, a8)

Z: Zokuzokyo 《卍大日本續藏經》(卍續藏)。京都：藏經書院。

R: Reprint 《卍續藏經·新文豐影印本》。台北：新文豐。

X: Xuzangjing 《卍新纂大日本續藏經》(卍新纂續藏)。東京：國書刊行會。

Users specify citation format: Users may determine to cite the title of a sutra (a certain scripture), a chapter, or the number of a fascicle, and whether to display the correction information.

3.5 Display format

- A. Fonts : Various font styles with preferred size and color may be assigned respectively to scripture number and title, and to the verse, mantra, revision, inline annotation, correction information and so on.
- B. Users may specify display format of the body text according to their preferences: whether to feed new lines as original text, with or without the line header, whether to display correction information, the processing priority of out-of-list characters, principles for handling Sanskrit, Pali, and Sanskrit alphabet, display format for revision, and the image file size of out-of-list characters and Sanskrit alphabetic letters.
- C. Supports Unicode 3.1 standard, substantially reducing the display of combined characters.
- D. Provides 85 books, or 317 fascicles, of scriptures with modern punctuation.

3.6 External connections

- A. CBReader is connectable to external computer programs, web pages, documents, and even to any tools or resources available on the Internet. After setting up the connection with other tools, users can access external programs and web pages, or even pass parameters to those programs and web pages, while reading the scriptures on line. In other words, users are able to inquire necessary information from various tools on line real time.
- B. In addition to the embedded Ding-Fubao Buddhist Dictionary and Nanshan Vinaya Dictionary in the compact disks, also built in are current on-line dictionaries and resources websites, to provide more assistance during the browsing of the scriptures.

4. Needs of Informatics Subjects in Buddhist Curricula

4.1 Essential concepts of digitized literature

As we take a review of the writing materials in religious literature, we see they had evolved in the early ages from clay tablets (cuneiform script for example), papyrus and stones (hieroglyphs), oracle bones, pattra leaves (Buddhist scriptures), bamboo tablets, to paper in current usages. However, the new approach in the 21st century is the digitization of texts for a variety of reasons. In short, they are the following: (1) for long-term preservation; (2) hardly exhaustible and universally available; (3) able to collect great volumes of information to derive creative knowledge and synergistic effects.¹

¹ 謝清俊 “佛教資料電子化的意義” 《佛教圖書館館訊》18, 嘉義: 香光尼眾佛學院, 1999

What specific aspects regarding modern digitized media should religious education focus on? First, the authorities need to know about the multivalent characteristics (multiplayer messages) of digitized literature.² Secondly, in order for efficient management of digitized literature, they need to work out courses of standard mark-up languages.

4.2 Literature in multivalent mode (with multilayer messages)

The difference between traditional and digitized modes of literature lies in the fact that the former is monolithic and the latter is multivalent. Literature in traditional mode provides the same function as ‘books’, which is independent, non-connected, and time-consuming for retrieval, inconvenient for cross-reference, and therefore is meant to be ‘read’.

Because the basic unit of digitized texts can be as small as a character, a stroke, or even a pixel, it is much easier and speedier to re-arrange and connect these texts. Also, the digitized texts can be ‘read’ by the processor (computer) and the human mind at the same time. Therefore, as molecules with identical or different atomic valences connect and complement with each other to form other kinds of molecules, so does digitized literature possess multivalent characteristics—multiple values, meanings and appeals.

To sum it up, multivalent literature has the following characteristics:

4.2.1 Incremental Extensibility

Anywhere, at anytime, and in any form, users can all incrementally extend the reach of data as well as the layer and behavior in the database, unlike printed matters that need reprinting for inclusion of any revision and addition.

4.2.2 Structurally Distributed Management

Information in the same format is considered to be in the same ‘layer’, where users can perform various functions called ‘behaviors’. Different layers at different locations can form a structurally distributed management system over the Internet.

² Digital Library Project: University of California at Berkeley <http://elib.cs.berkeley.edu/>

"Multivalent Documents: Anytime, Anywhere, Any Type, Every Way User-Improvable Digital Documents and Systems", Thomas A. Phelps, A dissertation for Ph. D, University of California, Berkeley (1998)

"Applications of A New Document Model for Digitalization of East Asian Classical Documents", Howie Lan Instructional Technology Program University of California, Berkeley

4.2.3 Internal Complementarities

Although a particular layer is not perfect, different layers are able to complement each other. For example, in a comparison of the layer of scanned image with the layer of font reconstruction, the difference between old and new printings can easily be found. That is, the same contents show in the two different image files in a complementary way.

4.3 Digitized Literature and Standard Markup Languages

Digitized literature needs standard markup languages to provide the platform for efficient information exchange and management. In 1986, the publication of SGML (Standard Generalized Markup Language) started to provide tagging functions for extending and examining data. Then there was the issue of HTML (Hyper Text Markup Language) to meet the requirements for the World Wide Web. HTML is an implementation of SGML, which possesses such functions as assign connection, and assign format, etc. In 1998, to fully utilize the Web environment, XML (Extensible Markup Language) was developed, as a subset of SGML, omitting unnecessary parts for transmitting data on the Web, to be integrated with HTML for the display of data. XML is complementary with SGML, able to structure and describe the data on the Web and extensively support different applications.³

In the course when the Chinese Buddhist Electronic Text Association (CBETA for short, <http://ccbs.ntu.edu.tw/cbeta>) worked on the digitization of the Buddhist Tripitaka, the texts, according to Dr. Christian Wittern's report, were first marked up semi-automatically and then put in some tags, before the footnotes were processed layer-by-layer.

Current Buddhist Education has not yet provided the above-mentioned knowledge and operation environment. To cope with the new age of digitized texts, it is necessary for Buddhist educational institutes to include in their curricula courses to introduce the characteristics of and how to efficiently manage digitized texts. As for how to fulfill the prospects, let's first examine current institutes of higher education for the available programs and courses to be integrated with information science.

³ 詳參 Charles Goldfarb *The SGML Handbook* Oxford 1990. SGML。

XML (Extended Markup Language)是由W3C (World Wide Web Consortium)所制定，XML 1.0 版規格書已由W3C核准，可參 <http://www.w3.org/XML>。

5. Higher Education Subjects Integrated with Information Science

In current institutes of higher education in Taiwan, the subjects aimed to be integrated with information science may be classified as the following:

5.1 Management and Information:

Examples: departments of Information Management

5.2 Library and Information

Examples: departments and graduate institutes of Library and Information Science

5.3 Education and Information

Examples: departments of Information & Computer Education, departments and graduate institutes of Educational Media & Library Sciences

5.4 Medicine and Information

Examples: Health Informatics unit, Medical Informatics Lab (of Chung Yuan Christian University)

5.5. Communication and Information

Example: departments of information communication

5.6 Sociology and Information

Examples: graduate institute of Informatic Sociology

5.7 Language and Information

Examples: Natural Language Processing Lab of Computer Science departments and graduate institutes, Language Processing Document Analysis Lab., Chinese Speech Lab.

5.8 Geography and Information

Example: Geographic Information Research Center

5.9 Humanities Computing

Please refer to the Association for Computers and the Humanities (ACH) as introduced at the website “Humanities computing units and institutional resources”

(<http://ilex.cc.kcl.ac.uk/wlm/hc/>) offered by Willard McCarty (King's College London) and Matthew Kirschenbaum (University of Kentucky). In the list are Humanities Computing Unit of Oxford University⁴, and Centre for Computing in the Humanities of King's College London⁵ in England, and Electronic Text Center and Institute for Advanced Technology in the Humanities⁶

⁴ <http://www.oucs.ox.ac.uk/humanities>

⁵ <http://www.kcl.ac.uk/humanities/cch>

⁶ <http://www.iath.virginia.edu/>

of University of Virginia, and Computers & the Humanities Program⁷ of Brigham Young University in the US A. Also mentioned are Humanities Information Technologies Research Programme⁸ of University of Bergen in Norway, Alfa-Informatica (“Humanities Computing”)⁹ in the Netherlands, and the Humanities Computing Project of University of the Witwatersrand Johannesburg¹⁰ in the Republic South Africa.

6. Example of Informatics Program in Buddhist Education

To adapt to the informatics environment of the coming age, the Chung-Hwa Institute of Buddhist Studies worked out the Computers and the Buddhist Studies Program (CBS Program for short) in 2001, and renamed it as Buddhist Informatics Program (BIP for short) in 2002.

Students majoring in this program are required to select courses from the three fields, namely Chinese Buddhism, Indian Buddhism, and Tibetan Buddhism, as their minors. On the other hand, major students of Chinese Buddhism, Indian Buddhism, or Tibetan Buddhism may also select courses from this program as their minors.

The Buddhist Informatics Program (BIP for short) is oriented in the studies of Buddhist informatics (eg. markup languages and programming languages), applied Buddhist informatics (eg. computer and academic publishing), and Buddhist teaching technology. The three-year curriculum is organized to include fundamental courses, basic courses, advanced courses and research projects as follows:

6.1 Fundamental Courses (required courses):

100 Basic Concept of Computing:

Introduces the theory and practice of computer hardware, software and applications, such as database, operating system and network administration and so on.

101 Introduction to Humanities Computing:

Introduces the theory and practice of computer software applied in the humanities, such as computer-based instruction, research, academic publishing, and service, and other topics as databases, on-line resources, and brief history of the development of media.

103 Basic Academic Word Processing Skills:

Cultivates skills for using computer software in academic word processing (in accordance with

⁷ <http://humanities.byu.edu/Chum/department.html>

⁸ <http://www.hit.uib.no/hit/spalte1.htm>

⁹ <http://odur.let.rug.nl/alfa/Alfa-informatica/Alfa-informatica/Alfa-informatica/index.html.en>

¹⁰ <http://www.wits.ac.za/fac/arts/acl/links.htm>

the theory and format of MLA), in particular the ability to write reports and essays.

6.2 Basic Courses (grouped in three elective areas as markup languages & computer languages, electronic publishing, and instructional technology):

201 Markup Languages:

Introduces the theory and practice of all-purpose markup languages applied to the humanities, such as text encoding, SGML (Standard Generalized Markup Language), HTML (Hyper Text Markup Language), and XML (Extensible Markup Language).

202 Computers and Teaching 1:

Applies computer technology in designing, authoring, and evaluating computer-based teaching materials.

203 Programming Humanities Applications 1:

Acquires such basic programming knowledge for applications in the humanities as Perl (Practical Extraction and Report Language), string manipulation, format conversions, and data structures.

204 Computer Research Tools and Methods: as complementary elective courses for program PHA1 in the above

Acquires such basic skills for applying computer technology in linguistics and philology as KWIC (Key Words In Context) concordance, Frequencies, Collocations and KWIC, and Text-analysis program.

205 Computers and Academic Publishing:

Acquires basic skills for applying advanced computer technology in academic writing and publishing, such as academic publication and communication, desktop publishing, and electronic publishing on the Internet.

206 Introduction to Social Informatics

Emphasis on the impact brought by newly developed technology to the society, and discussion on important issues of the socialization of information.

6.3 Advanced Courses (electable in three groups as Markup Languages, Electronic Publishing, and Instructional Technology):

301 Markup Techniques:

Introduces advanced theory and practice of all-purpose markup languages applied in the humanities, such as the advanced courses of XML (XSL Pattern, XML-Data, etc.), and the

Guidelines for Electronic Text Encoding and Interchange by TEI (Text Encoding Initiative).¹¹

302 Computers and Teaching 2:

The application of advanced computer technology in teaching, such as the integration of teaching materials in the formats of digitized texts, images, sounds, and videos; hypertext and hypermedia applications; computer-based testing; network applications.

303 Programming Humanities Applications 2:

Acquires advanced programming skills for humanities applications; knowledge of integrating the modules in different programming languages; software design, and CGI (Common Gateway Interface).

304 Computer Research Tools and Methods: as complementary elective courses for program PHA2 in the above

Acquires knowledge and skill for application of computer technology in linguistics and philology, such as quantitative analysis, text data processing, databases, on-line resources, solutions of Database applied to WWW, statistics, evaluation and so on.

305 Computers and Academic Publishing 2:

Application of advanced computer technology for academic writing and publishing, such as academic publications and communication, desktop publishing, and electronic publishing on the Internet.

306 Introduction to Media informatics

6.4 Research Projects (required courses):

Students are required to complete one of the following projects as their graduate achievement.

401 Buddhist Informatics Project:

Students Apply the knowledge acquired from other courses to Buddhist informatics projects, doing research individually under professors' direction.

402 Applied Buddhist Informatics Project:

Students Apply the knowledge acquired from other courses to Buddhist informatics projects, doing research individually under professors' direction.

403 Buddhist Instructional Technology Project:

Students Apply the knowledge acquired from other courses to Buddhist informatics projects, doing research individually under professors' direction.

¹¹ Sperberg-McQueen, C. Michael and Burnard, Lou (Eds.) *Guidelines for Electronic Text Encoding and Interchange*, Chicago and Oxford, 1994.

7. IBA: An "Indra's Net" for Buddhist Studies

Draft proposal version October 1, 2007

Authors: Huimin Bhiksu, Aming Tu, Christian Wittern, William Magee, Marcus Bingenheimer

7.1 Rationale

When CBETA was founded ten years ago, few of us could have foreseen the many recent developments in information technology, especially the emerging web-technologies. Wikis (Wikipedia), blogs, Internet telephony (Skype), platforms for exchange of amateur video (YouTube), geographical information (GoogleEarth), virtual worlds (Second Life) have become virtually mainstream. At this point it is impossible to calibrate the social impact of these technologies, and in spite of GoogleEarth we are on a journey without maps or compass, led by no one in particular.

Although the trajectory is obscure, as practitioners of Digital Buddhist Studies we have to ask what these changes mean for our practice. Looking back at the last ten years it is obvious that those projects in our field which proved successful were those that could rely on long-term commitment by individuals and/or institutions. We can all remember projects that started hopefully and with promise but which, without the necessary commitment, failed to establish themselves as a permanent presence on the Web. Only a strategic, far-sighted approach makes it possible to connect scholarship and digital media effectively enough to provide continuity.

This conference is dedicated to a discussion of strategies regarding how to strengthen our practice of Buddhist Informatics. This might be an auspicious time to establish a decentralized network of institutions committed to the use of information technology for the academic study of Buddhism. We have tentatively called this projected network the *Integrated Buddhist Archives* (IBA). As a first step in its establishment, we envision a working group with the task of suggesting and drafting effective measures for connecting existing archives.

7.2 Aims

A network like IBA could provide a number of important services to academic and Buddhist communities around the world:

A. Common Portal Site & Integrated Search Interface

A growing number of databases exist which harbor large amounts of Buddhist textual materials. Often little effort is made to advertise one's collections, and so it is often difficult to know where to look for specific texts, especially for non-specialists and newcomers to the field. It seems desirable to create a common portal from which these archives and repositories could be accessed and (ideally) searched. Such a portal would also help us to heighten the profile of Buddhist Informatics within the wider community of Buddhist Studies, where projects from the Digital Humanities are often met with reservations.

B. Maintenance of Legacy Projects & Archiving

In the digital realm where objects are fragile and quickly obsolete, impermanence is certain and longevity is rare. Individual contributions to Buddhist Studies in particular often vanish with the website on which they were originally presented. Scholars move on to new projects and cease maintaining the old. This is only natural but, nevertheless, it is in our interest to offer maintenance and archival support to content providers.

An institutional network like the IBA might preserve some of those projects and keep them available once the original creators stop maintaining them.

C. Establishing Guidelines for Best Practice and Review

For many of us, the realization of a digital project is a learn-by-doing experience. For certain types of projects, however, especially those involving the digitization and archiving of texts, we, the community of practitioners, can now start to draw conclusions from our experience and share them with others. A best practice for repositories along the lines of "Guidelines for Editors of Scholarly Editions" devised by the MLA, or the "AHDS Guides to Good Practice", would help us to establish benchmarks, which also could be used by reviewers. One aim of the IBA should be to make it possible for digital projects to be included in the peer-review process.

D. Assistance with Realizing Projects

Each year IT-technologies become more complex. The minimal requirements an institution needs to set up a digital archive are competence in digitization, databasing, web-technology and project

management. Though databasing and web-technology are common skills among technicians, not all solutions are equally suited to scholarship. Exchange of personnel and workshops might assist in transferring knowledge and skills between IBA members.

7.3 Organizational Form

We propose the establishment of a working group composed of representatives from interested institutions that meets regularly (once or twice a year). Individuals might join IBA as subscribers or advisers. The working group would serve as a forum where various forms of cooperation can be explored, especially ways to connect our archives and depositories to serve their content to an ever-wider audience. The working group could also arrange workshops and disseminate information on emerging technologies that are relevant to our practice. Representatives should have enough decision making power to draft agreements and initiate projects.

We envision IBA to be a decentralized network of institutions and individuals at work on modeling Buddhist culture with the help of computers. Through regular contact, we can compare different approaches so that all projects might expand their reach. Projects involving two or more members of IBA could be initiated and maintained. Ideally, as with Indra's net, each repository can be present in - or at least be accessed from - all others.

[參考目錄]

- Erik T. Ray原著。陳建勳譯 (2001)。《XML學習手冊》。台北市: 歐萊禮(O'Reilly)。
- *Elliotte Rusty Harold*原著。陳牧群·連春雨譯 (2001)。《XML精要總覽》(*XML in a Nutshell*)。台北市: 歐萊禮(O'Reilly)。
- C M Sperberg-McQueen and Lou Burnard edit. (2002). *Guidelines for Text Encoding and Interchange* (Published for the TEI Consortium by the Humanities Computing Unit). Oxford : University of Oxford. (TEI Website : <http://www.tei-c.org>)
- David Flanagan著。陳建勳·黃吉霈編譯(2003)。《JavaScript 大全 第四版》(*JavaScript:The Definitive Guide, 4rd Edition*)。台北市: 歐萊禮(O'Reilly)。
- 《香光佛教圖書館館訊：電子佛典製作》no. 24 (2000)。
<http://www.gaya.org.tw/journal/m24/24-index.htm>

- Bhiksu, Huimin 資訊時代之宗教教育課程規劃初探---以中華佛學研究所「佛學資訊組」課程規劃草案為例。發表於1998.10輔仁大學承辦之「宗教教育研討會」
- Du, Aming 佛學與資訊的整合「中華佛學研究所佛學資訊學程」緣起與創設。2003兩岸佛教學術研究現況與教育發展研討會