

漢文電子大藏經的製作緣起與作業流程

--

以「中華電子佛典協會」為例

杜正民

佛學研究中心學報第四期

1999 年七月出版

頁 347-369

頁 347

提要

中華電子佛典協會(Chinese Buddhist Electronic Text Association, CBETA)於 1999 年 1 月 20 日在中央研究院舉辦國際「電子佛典推進協議會」(Electronic Buddhist Text Initiative, EBTI)發表會上，首度對國外團體發表其成果，得到許多與會學者專家的肯定與讚美，除對「中華電子佛典協會」的工作有正面的意義與鼓舞外，無形中也給予「中華電子佛典協會」成員一年來默默耕耘很大的獎勵。是以，謹藉此文讓各界瞭解「中華電子佛典協會」製作漢文電子佛典的緣起、目標、組織架構，以及工作進行中的各項活動，如參與相關會議、國內外的參訪、取得日本《大正新脩大藏經》的授權等。此外，並詳明「中華電子佛典協會」如何解決困難、開發程式、處理缺字、訂定作業流程等項目。並且，就

1998 年底「成果發表」的光碟內容、電子檔分類、CBETA 電子檔的殊勝處、舉例說明採用 SGML 檔案的方法、以及未來展望等相關資訊，與關心佛典電子化的先進分享，期能因此獲得更多的指正與協助。[註1](#)

成立緣起

近年來，經由眾人的努力，台灣網路上結集了不少佛教經典，也帶動佛典電子檔輸入熱潮。[註 2](#)此時的目標主要是著重於網路應用，比如將這些電子檔置放於FTP 上供人免費下載，或是透過GOPHER、WWW方便使用者瀏覽，最近更在 WWW 上提供檢索功能

(<http://ccbs.ntu.edu.tw/canon/>) 以便查詢。另個發展是將電子檔包裝設計成電子書，以使經文的呈現更加精緻。所有這些努力，莫不希望能夠透過網路，使佛典普及，讓更多人同霑法益，並利用電腦功能拓展佛典的應用範圍及閱讀方式。

此外，蕭鎮國先生於 1997 年 10 月來函表示願意提供 25 冊以CCCII 碼輸入的《大正藏》電子稿，並授權「台大佛學研究中心」進行後續之處理。經由網路上 BBS 站電子佛典討論版 (Buda-Tech)[註 3](#)的網友草擬「電子版大藏經輸入計畫」，開始進行經典輸入的計畫。同年 11 月 6 日由台大佛研中心成立 25T 小組，[註 4](#)著手進行大規模的藏經電子化籌設工作。

之後，經由恆清法師的奔波籌募所需經費，終於在「北美印順導師基

金會」與「中華佛學研究所」應允全力支持贊助此項工作後，於 1998 年 2 月 15 日假法鼓山安和分院舉辦籌備會議，當日與會的教界、學界與民間單位參與人士相當踴躍，由於這是大家期盼已久的志業，因此討論很熱絡，會議進行也很順利，因而決定於當日正式成立「中華電子佛

典協會」(Chinese Buddhist Electronic Text Association，以下簡稱CBETA)，正式開始漢文佛典電子化的作業。

主旨目標

在進行作業之前，當需先確立主旨與目標。CBETA的主旨可分為下述三點：

1. 研發佛典電子化技術，提昇佛典交流與應用
2. 利用電子媒體之特性，以利佛典保存與流通
3. 期望讓任何想要閱藏的人都有機會如願以償

依此宗旨，開始設定 CBETA 的主要目標。首先，以學術界通行的《大正新脩大藏經》為底本，初期以完成第 1-55 冊及第 85 冊藏經電子化為目標；其次，期能借助電腦校對，完成錯誤率極低之高品質電子經文；第三，作業同時希望能提出電腦缺字之有效處理方案；最後，借助電腦技術，以整合不同版本藏經的校勘。並且借助當代科技發展全文檢索工具，以提昇電子佛典之應用，同時開發單機版之使用介面，以利大眾使用，進而希望利用網路特性，將漢文佛典呈現至世界各處。[註5](#)

組織架構

確定宗旨與目標之後，即依工作需求，開始建立組織架構。CBETA 的組織分為主任委員、副主任委員、總幹事、輸入組、校對組、研發組、資訊組、網路組、發行組、財會組等。就整體架構而言，則於「中華電子佛典協會」下設「委員會」及「顧問團」兩大團體，聘請有志於佛典電子化的佛教團體，或學者專家為「委員會」會員，同時擬聘請

頁 350

諸山長老與學者專家為「顧問團」顧問。「顧問團」下設「執行顧問」一員，目前聘請中央研究院謝清俊教授為「執行顧問」，而「委員會」

下則設「常務委員」一名，代表委員會監督協會事務之進行，目前由恆清法師擔任「常務委員」。

協會事務則由「主任委員」負責整體事宜，於主任委員下擬設「副主任委員」，以協助主任委員辦理協會各項事務，目前由惠敏法師與厚觀法師分別擔任主委與副主委之職責。

行政組織則由筆者任「總幹事」，負責協調各組工作及對外行政事宜。作業系統則依作業流程，分別為經典輸入之進度及規劃的「輸入組」、成品之電腦比對及人工校對事宜的「校對組」、提出前瞻性、國際性及學術性之整體規劃及進行標記工作的「研發組」、電腦上缺字處理之「缺字組」、技術程式以及網路、單機應用程式開發的「資訊組」、將各組工作近況及成果利用網路呈現的「網路組」、成品之發行事宜及推廣活動的「發行組」，及總管財務之運用的「財會組」等，依上述工作流程之所需，而組成 CBETA 的組織架構。[註 6](#)

正式作業

確定上述之職位與責任後，CBETA 於 1998 年 3 月 1 日遷入台北市

頁 351

朱崙街慧日講堂三樓，[註 7](#) 開始購置設備與建構電腦連線等工作，三月十五日起正式以《八十華嚴》作為第一部藏經電子化經典，CBETA 電子化作業於焉開始。是以，由「輸入組」準備先行工作及提供所需的素材，由「校對組」開始進行資料比對，「缺字組」發展缺字表達方式，「資訊組」提供必須的運作程式，「研發組」提供作業方針等工作，「中華電子佛典協會」從此正式進入佛典電子化的大業。

工作進行中，「訂定規範」是相當耗費時間的，如版本格式、電腦缺字表達、組字式規範、通用字規範、電子佛典呈現方式，及大正藏內文格式與校勘輸入格式等重要的課題，都經詳細研討後再進行細目工作。

因此，經過翔實的訂定規範之後，CBETA 第一部佛典電子化的工作，於四月十二日完成第一次電腦校對，成果還算滿意，但為精益求精，故

依原計畫再繼續進行第二次及第三次校對工作。在如此嚴密的校對程序下，於五月底完成《大正藏》第十冊的校對作業，這也是 **CBETA** 初步完成的第一部電子化藏經。

國內參訪

工作進行當中，協會亦同時著手於對外的聯繫工作，於四月六日「中華電子佛典協會」主委惠敏法師與總幹事杜正民，親至台中華雨精舍拜謁印順導師並報告「中華電子佛典協會」的成立宗旨與工作目標等事宜，會中不但得到導師對此工作的肯定，同時導師也指示厚觀法師「全力支持及參與此有意義的工作」。四月二十四日仁俊長老來台，由中華佛學研究所李志夫所長，中華電子佛典協會主委惠敏法師及筆者等多人於慧日講堂五樓報告「中華電子佛典協會」的成立與工作方針、目前正在進行的工作及將來計畫等事宜，甚得長老的嘉許。進而，五月十八日於太平洋電子會議(1998 Pacific Neighborhood Consortium,

PNC) 後，由中華佛典電子協會總幹事與研發組維習安(Christian

Wittern)，假中央研究院活動中心第二會議室舉行說明會，正式對國外學者發表與介紹「中華電子佛典協會」

頁 352

的工作，參與人士皆為進行電子佛典工作多年的學者，如 Dr. Lewis

Lancaster, Dr. Tetsuya Jatsumura, Dr. John Lehman, Dr. Matthew

Ciolek, Dr. Charles Muller, 及謝清俊教授等，會中並得到他們很多寶貴

的意見與建議，CBETA 的工作方針因而得到更進一步的肯定。[註 8](#)

參與會議

CBETA 為不與當代其他漢文資料庫的進行脫節並保持聯繫，本協會工作人員也積極參與國內多種相關會議，如於剛成立時即參加的會議有：

1. 四月十八日至十九日由世新大學圖書資訊學系及台灣傳技電腦共同舉辦

的「21 世紀資訊科學與技術的展望」

2. 五月一日由中研院漢籍電子文獻協調委員會舉辦的「人文計算研討會」

3. 五月十五日至十八日在中研院舉行的 **Pacific Neighborhood Consortium 1998**

4. 五月二十一日由資策會舉辦的「中文資訊處理標準公聽會」

(ISO10646/Unicode 標準)

5. 五月二十七日由國家圖書館、台大電腦系統技術研發重點中心及資策會

共同舉辦的「台灣史料數位化研討會」

6. 六月十二日至十三日由中研院跨單位舉辦的「漢籍電子文獻資料庫建置

的回顧與前瞻研討會」等多項會議。

此外，從十月份起 CBETA 在本身的工作量達一定程度後，即積極與外界聯絡，以取得更多的技術支援與協助，茲將十月份參與的活動簡列於後，藉以瞭解 CBETA 作業情形。

87年10月1日(星期四)參加經濟部技術處與資訊工業策進會推廣服務處於台北國際會議中心舉辦的「SGML/XML 應用實務研討會--掌握企業文件數位化管理與應用」Rick Jelliffe 主講，以確定 CBETA 的 SGML/XML 標記表達方式

頁 353

87年10月2日(星期五)與台大資訊工程學系歐陽彥正教授討論 CBETA 檢索軟體等問題

87年10月6日(星期二)至中研院資訊所與謝清俊教授討論 CBETA 的作業流程與進度

87年10月7日(星期三)與 CBETA 義工茶聚，感謝他們的支持與協助

87年10月8日(星期四)與台大歐陽彥正教授及美國陳樹新教授商討資料庫與檢索問題

87年10月9日(星期五)與華康科技開發公司郭嘉生副總經理討論「漢文佛典造字與呈現」等問題

87年10月9日(星期五)至台大資訊工程學系討論 CBETA 的檢索軟體等問題

87年10月9日(星期五)與美國陳樹新教授聚餐並商量成立獎學金，招訓台大資訊相關系所學生長期參與佛典檢索與資訊工作

87年10月12日(星期一)與 Rick Jelliffe (The XML and SGML Cookbook: Recipes for Structured Information, Prentice Hall, NY 1998 一書的作者) 討論佛典 SGML 的技術與問題

87年10月15日(星期三)與蕭清鳳小姐會面感謝其支援 CBETA，並討論後續資料問題

87年10月20日(星期二)CBETA 月會：評估校對組未來作業流程，並以此流程表做為下年度作業之參考

87年10月22日(星期三)參加中研院資訊所、台大電腦系統研發重點中心、中華民國計算語言學會 SIGIR 舉辦的「資訊檢索技術的新趨勢」研討會

87年10月23日(星期五)於台大佛學研究中心接待「兩岸禪學研討會」大陸來賓，簡介並商討藏經電子化工作事項

87年10月26日(星期一)於台大佛學研究中心與大陸學者討論合作事宜

87年10月29日(星期四)於台大佛學研究中心與工作關人員討論檢索問題

87年10月30日(星期五)於 CBETA 舉行組際會議，討論標記問題

87年10月30日(星期五)於台大資訊所「網路實驗室」討論 CBETA 電子佛典檢索問題

國際參訪

除對外參與活動外，CBETA 各小組每日於網路上利用 e-mail 討論會內的事務與解決遇到的困難，這也是 CBETA 的一大特色，因為各小組皆善於運用此工具進行內部的工作研討，使得會務的進行很順利並且節省時間。因此，除依原訂進度完成訂定的目標外，為參加日本會議，特別就「大正新修大藏經原典資料庫」(SAjganikikrtaj Taiwotripitakaj, 以下簡稱 SAT) [註 9](#) 所發表《大正藏》大般若部作初步電腦校對，同時根據其差異製作

頁 354

勘誤表。

CBETA 主委惠敏法師、常委恆清法師、總幹事杜正民、顧問維習安一行四人於 1998 年 6 月 20 日至 25 日赴東京與 SAT 及「大藏出版株式會社」洽談電子佛典的合作與版權問題，旋即趕赴京都與花園大學及京都大學進行技術交流，花園大學與京都大學在在藏經電子化都有相當的經驗，如花園大學國際禪學研究所《禪基礎 ZEN BASE》光碟的負責人 Dr. Urs App 給予很多寶貴的意見，而京都大學 Dr. Tetsuya Jatsumura，則正從事於一國際性的漢字庫計畫，對 CBETA 的計畫助益頗大。而此行與 SAT 的商談，雙方皆具誠意就《大正藏》的輸入校對工作進行合作，然雙方約定必須先等 CBETA 取得日本「大藏出版株式會社」授權之後，再進行 SAT 與 CBETA 進一步合作事宜之洽談。

取得授權

因此，CBETA 與「大藏出版株式會社」的授權問題，經雙方進行多次的交流與修改契約書之後，終於在 1998 年 9 月 30 日經由郵寄方式完成簽約。「大藏出版株式會社」除授與 CBETA 使用《大正新脩大藏經》的權限外，並同意 CBETA 發行光碟等權益。「大藏出版株式會社」同意 CBETA 取得《大正藏》電子版的網路版與光碟版的發行權，實為相當優厚的條件，CBETA 的工作也因此授權而能順利進行。[註 10](#)

版權宣告

取得授權之後，特別於網頁上聲明 CBETA 資料庫為「免費提供電子佛典資料庫，以供各界作非營利性使用」。並註明 CBETA 電子資料庫則是以《大正新脩大藏經》(大藏出版株式會社(C)) 第一卷至第八十五卷為底本，且正式取得該底本版權者「大藏出版株式會社」輸入與公開之授權。

至於 CBETA 資料庫之使用說明如下：

頁 355

1. 本資料庫之使用限定於個人或團體非營利性用途。若運用於營利性目的時，必須事先取得本協會與日本「大藏出版株式會社」之同意。
2. 將本資料庫以網路、光碟或其它媒體再傳播、節錄引用及再加工發行時，

以不改變實質內容為原則，並需於成品附上本說明及版本資訊，以利使用

者查詢相關修訂資訊。

3. 若發現本資料庫中任何錯誤之處，請與本協會聯絡。
4. 若因使用本資料庫所產生之任何損害，本協會一概不負任何責任。
5. 本資料庫受相關著作權法保護。

網頁開放

為讓使用者隨時瞭解 CBETA 的進度與成果，籌劃與準備多時的

CBETA 中英文網頁(<http://ccbs.ntu.edu.tw/cbeta>)，終於在 1998 年七月初正式對外開放測試，並陸續接獲許多使用者對電子藏經輸入工作的嘉許與鼓勵信件! 進而，在七月七日中華佛學研究所舉辦的「兩岸佛學教育博覽會與中文電子佛典運用展示」中，CBETA 在師大展覽會場曾有兩場展示：一為上午的「電子佛典的運用」，另一場則為下午的「電子佛典的製作」，此兩場可說是 CBETA 經由一段時間的運作之後，正式對外的展示與發布消息，台灣各大報紙及佛教電視台即於當晚或隔天報導此項消息並刊出網址，因而獲得多方的關切與鼓舞。[註 11](#) 這些關心與讚美對默默耕耘的 CBETA 工作群都是最佳的鼓勵!

技術突破

在眾人的鼓舞下，CBETA 藏經電子化作業繼續平穩的進行著。在工作過程中，由於丹青系統的「力新國際公司」研發部副理葉齊平先生及其部門的工作人員幾個月來的協助，終於研發出「丹青 for CBETA 版」的丹青 OCR 辨識軟體；此軟體純為 CBETA 的輸入作業而設計，旨在

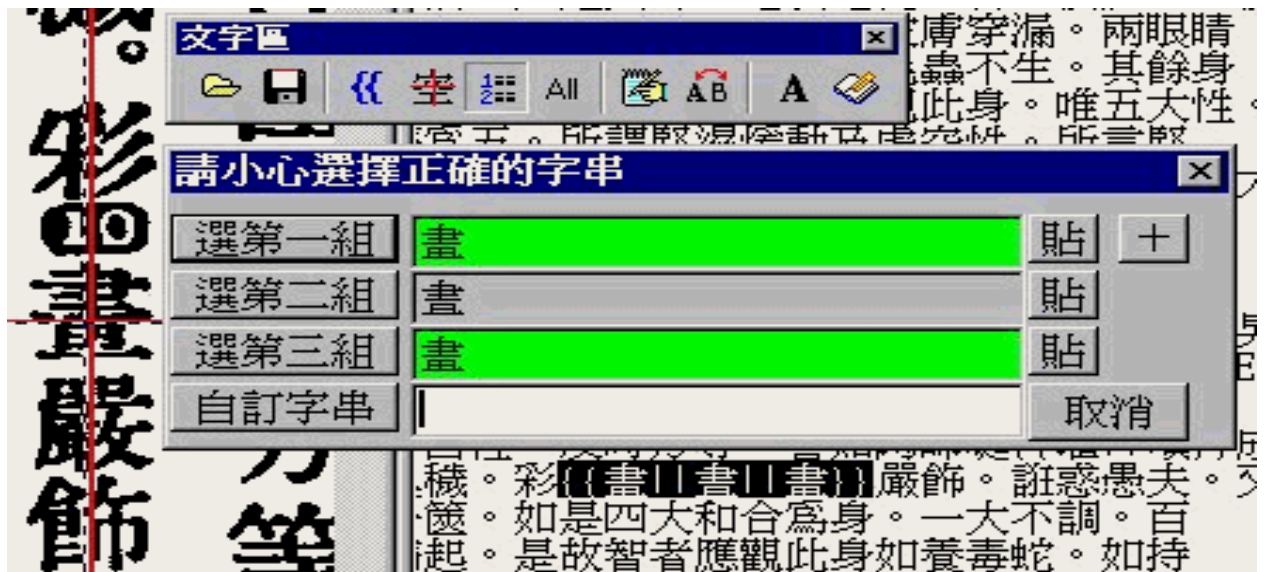
提昇漢文佛典的辨識效率，加強工作速率，及解決《大正藏》特有的問題與困難，譬如雙行小註的辨識等問題。同時也測試在不同掃描點數下的最佳辨識成果等工作，根據此次的測試，力新公司特別為 **CBETA** 提供了一些非市面銷

頁 356

售軟體所擁有的特殊新功能。

因為「力新國際公司」同仁的鼎力支持，替 **CBETA** 的輸入工作解決很多困難，使得輸入組的工作特別流暢，因而可以在第一年年底前完成《大正藏》**55** 冊所有掃描工作、轉換辨識文字、提供比對圖檔等業務。

此外，在電腦比對之後，針對結果檔，傳統是以翻書訂正，但在 **CBETA** 資訊組的努力下，開展出一個快速的「看圖校對程式」(請參照下圖)，可以一邊看大正藏掃描圖檔，一邊做文字訂正。由於它可以同時開啟一文字視窗及一圖形視窗，並且在文字視窗中快速移動到比對差異處，然後在圖形視窗中秀出相對應的圖形位置，再由校對人員選擇或載入正確的資訊，使得作業可以順暢的進行之外，對正確率與速率的提昇也有很大的助益。



缺字處理

除了技術的突破之外，電腦缺字處理也是一重要的課題，雖然國際間試圖解決缺字的方法有很多種，[註 12](#) 然而 CBETA 關於缺字的初步工作，則

頁 357

著重於現有缺字形、音、義及通用字的資料查詢及整理的工作等事項。繼而，CBETA 於「普及版」使用「一般組字式」[註 13](#) 作為表達佛典缺字的方法，則是考量讓使用者能在純文字環境下閱讀，不需另外安裝造字檔或是圖檔而設計的，這種方式除了提供了閱覽、散播上的便利性外，也不會佔用使用者對造字檔自行運用的空間等。

頁 358

因而，CBETA 沒有使用造字檔來表達缺字，而是利用系統字(Big-5) 以組字式來表達，誠如上述，主要是為顧及普遍性和使用者對造字檔的管理。此外，CBETA「普及版」中大部份的缺字是以通用字來取代，少

部份沒有適當通用字的缺字亦附有圖檔，以補「組字式」之不足。雖然，使用造字檔也曾是我們考量的方法之一，但目前則是以方便使用者操作的方式來呈現；譬如，使用者要複製經文到一般的文書處理程式上進行編輯時，可以不需先安裝造字檔就能直接複製，如此對一般使用者的操作較為方便，且在文件的散播、流通上也較方便。

當然，不同的缺字表達方法各有其利弊，因此 **CBETA** 擬建立一完整的缺字資料庫，以達成不同缺字格式的轉換，因此在光碟中 **HTMLHelp** 版的缺字處理，便同時利用了組字式和圖檔來表示。

同時，在 **CBETA** 「學術版」以 **SGML/XML (Standard Generalized MarkupLanguage/ Extensible Markup Language)** 表達的主要檔案 (Source File) 中，則以「**M 碼**」[註 14](#) 處理缺字的表達，**CBETA** 採用「**M 碼**」表示缺字的原因，一則利於國際化的使用，二則目前「**M 碼**」已具有八萬多字的缺字資料，是國際間較完整的一個字集，三者利用「**M 碼**」方便我們建立一個缺字的轉換資料庫。固然，其他的缺字表達方式也是我們考慮使用的，不過目前 **CBETA** 的缺字表達，有些缺字圖檔是取自日本「今昔文字鏡」[註 15](#) 的字型，因其乃當今最大的漢字庫；且網路上提供免費字型讓使用者下載，不增加使用者的額外負擔；而該中心亦根據 **CBETA** 缺字資料庫與其字集比對，隨時補充其尚未造的佛典缺字。且使用者可依 **CBETA** 建立的缺字資料庫，自行轉換成其他的缺字表達方式，甚或轉為將來擬採用的 **Unicode** 碼。

作業流程

在作業技術的突破與缺字困難的解決之後，經由 **CBETA** 月會多次的討論，訂定 **CBETA** 的最後成果，將以 **SGML/XML**標記語言呈現，並期

頁 359

能結合良好的界面以表達結果。**CBETA** 為達到組別間溝通管道的暢通，並經常不定期舉辦小型的「組際會議」，於會中陸續達成一些共識，譬如確定工作進行的程序：輸入組→校對組→研發組→網路組的工作流程等。[註 16](#)

作業流程確定後，接著再將 **CBETA** 作業進行的方法與步驟簡述如下：

「輸入組」的工作，主要是以掃描及辨識的方式進行，如有必要則負責打字輸入或尋找其他資源，為不浪費資源，**CBETA** 盡量不用人工鍵入的方式進行輸入工作，如能取得合作單位的認可，則可代為校對，並將所得結果做成校勘表送回原單位，以利雙方作業順利進行。[註 17](#)「輸

入組」已於 1998 年底將所需的資料準備妥當，並轉交給「校對組」進行大正藏格式化、檔案比對、看圖校對等工作，做完基本校對之後，即將檔案交給「研發組」以進行缺字處理、架構標記、學術研究、發行研發等工作，並完成 **SGML/XML** 作業流程，做成主要檔案 (Master file)。

在整個的作業過程中，校對與研發是佛典電子化最耗時的部份，如何能快速又精確的達到要求，並且能研發出符合普及與學術的要求，也一直是 **CBETA** 的重要課題。成立初期為解決缺字的問題，因而成立「缺字組」以求突破電腦缺字的困難。在整個作業的過程中，如遇任何軟硬體的需求則為「資訊組」的責任，負責研發或取得所需的工具，並由「網路組」將之準備上網，上網同時則由「發行組」籌劃出版光碟及推廣的作業。**CBETA** 工作能順利進行，乃是各組分工與合作的最佳實證。

流程圖表

有關 CBETA 的工作流程，試以流程圖表示之：

資料輸入



缺字與格式整理



頁 360

檔案比對



看圖校對及查書校正



標記處理



電子檔上網



全文檢索等應用



使用者之建議與指導



修正並更新

成果發表

以前述的作業方式進行，**CBETA1998** 年的進度在整個的摸索過程中，已經確立一完整的作業方式與流程，相信今後將可順利的進行，並將加速產品的速度。[註 18](#) 接著，於九月十五日網路組即依組際會議的結論，先行將測試版置於網路上，以「普及版」與「學術版」合併的方法上網。亦即，將缺字換成通用字，並提供缺字資料庫供人下載等方式，先行流通測試，以得知使用者的反應與需求，再行改良，以便將來正式上網之參考。

並於 1998 年 10 月 20 日 **CBETA** 月會中，全員同意於該年年底對外發表成果之後，全部同仁即全力投入「成果發表會」的籌劃事項，及有關的各項準備工作。因而，「成果發表會」成為 **CBETA** 該季末的重點。

CBETA 所欲分享的成果，可分為三大類，一、將 **CBETA** 的作業系統與方法印製成冊與有興趣於佛經電子化的單位或工作者分享，二、將 **CBETA** 作業過程所研發的程式或工具與有志於佛典電子化的工作者分享，第三則

頁 361

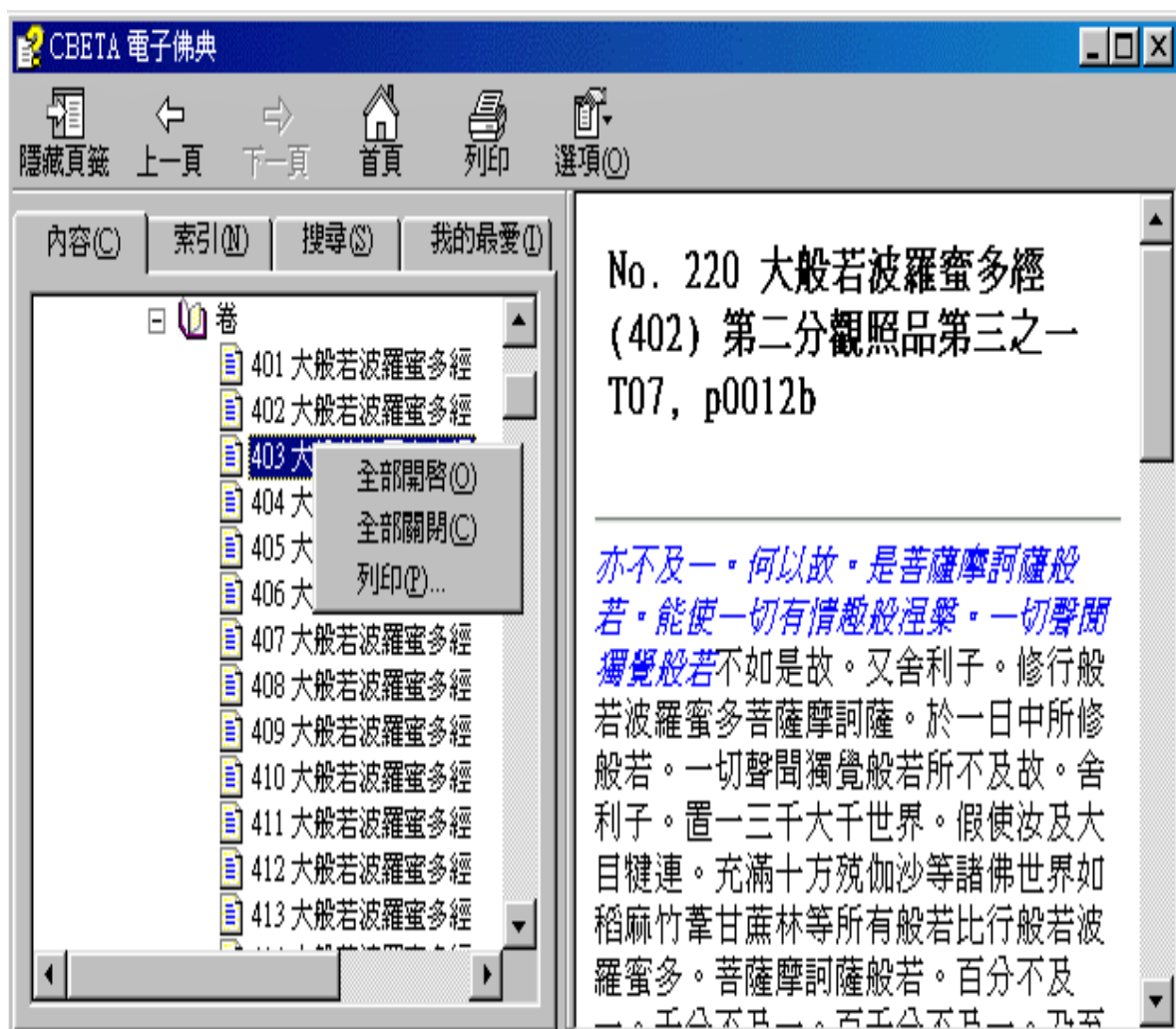
為將 **CBETA** 藏經電子檔與大眾分享。

關於藏經電子版的成果發表，在 1998 **CBETA** 電子佛典 測試版光碟及網路版中，目前有《大正藏》第五至第八冊之《般若部》、第九冊的《法華部》及第九冊下半部與第十冊的《華嚴部》，共計三部六冊。而這些成果可分為兩大類呈現，一為網路版，一為光碟版，本年度的成

果發表不但有光碟版的檢索，同時於網路上亦提供[網路檢索](#)，尚稱堪用。[註 19](#)

光碟內容

依上述說明，除了網路版本外，**CBETA** 同時也發行光碟版，以利不方便上網者使用，今再對光碟版內容作簡介，**CBETA** 電子佛典系列含有《般若部》、《法華部》、《華嚴部》，亦即《大正藏》第五冊至第十冊的普及版電子經文，並依其呈現功能分為 **Normal** 版、**App** 版、**HTML** 版及 **HTMLHelp** 版四種方式。此外還有全文檢索工具、[註 20](#) 丁福保《佛學大辭典》、[註 21](#) **CBETA** 網站光碟版、其他相關工具及線上使用說明文件等。[註 22](#) 其中 **HTMLHelp** 瀏覽器，包含有工具列、內容、索引、搜尋、我的最愛、列印、選項設定、佛學辭典、彈出式清單、及其他多種項目（請參閱下圖）。



CBETA 電子佛典是由同一套 SGML 電子主檔所產生，光碟提供下列版本，以供讀者不同的使用需求，分為一般文字檔的「普及版」；行末以句點作為結束，以利一般檢索的「App 版」；具有目錄、索引、全文檢索多功能的「HTMLHelp 版」及可以直接使用網路瀏覽器閱讀的「HTML 版」等四種不同的呈現方式。茲再對各版本說明如下：

1. 普及版

普及版為一般文字檔的格式，可利用任何支援中文 Big-5 碼的文書處理器或是瀏覽器閱讀編輯。為了達到閱讀以及流傳的方便，普及版對於電腦中無法處理的缺字，如果可以通用字代替，則以通用字代替。反之，若無可用的通用字，則以一般組字式來表示，不另設造字檔或其他的圖檔代碼。

普及版不包含《大正藏》的註釋，對於大正原版錯誤的地方，則在版本中直接更正，不另作註說明。全文依照《大正藏》的格式橫向排列，於

頁 363

每行行首註明冊數、經文號碼、頁碼、欄位及行數等訊息，方便讀者查閱。例如 T08n0221_p0001a09 則代表《大正藏》第八冊第 221 經第一頁第一欄(上欄或 a 欄) 第九行，詳見行首資訊。[註 23](#) 光碟中並含有「以卷為單位」及「以經為單位」的兩種版本，提供使用者選擇。

2. App 版

App 版的內容及格式與普及版相同，唯一的不同是為了方便部份檢索軟體能正確的檢索，因此每行行尾如果有不成句的字，則會移動到下一行的行首(亦即以句點作為行末切割處)。搬動的字數會在下一行的行首資訊後以括號顯示出，譬如 T08n0221_p0001a09(02)，其中的 (02) 代表由上面的行數搬動兩個字到本行行首。偈頌的部份因為沒有斷行的問題，故不處理。

3. HTMLHelp 版

此版是以微軟公司的 HTMLHelp 瀏覽器為主。[註 24](#) 本版本與上述的兩個版本內容相同，但是如遇有更正大正版本錯誤之處，則會以紅色顯

示。本版並以《大正藏》的欄位作為顯示基準，譬如以《大正藏》0001 頁的 a 欄單獨處理成一頁顯示。並且為了方便閱讀與檢索，因此每頁的最後二行會重現在下一頁的前面，並以藍色斜體的方式顯示。對於無通用字的缺字，

頁 364

除了使用一般組字式之外，本版提供了缺字圖檔，讀者點選組字式便可以看到該缺字圖型。[註 25](#)

本版提供了經文目錄、索引、全文檢索、線上佛學辭典，並支援網路及多媒體等功能，預計在微軟公司發行新版的 HTMLHelp 格式之後，可提供使用者透過網路直接更新經文內容的功能。而目前 HTMLHelp 版只限於中文 Windows 使用者。

4. HTML 版

HTML 版與 CBETA 的網頁資料相同

(<http://ccbs.ntu.edu.tw/cbeta>)，也與 HTMLHelp 版的格式相似，但是目前尚無全文檢索，使用者可以直接使用網路瀏覽器閱讀。但是，由於 HTML 版本是由許多的小檔案組成，如果備份到硬碟上執行，可能造成硬碟資源的嚴重浪費。如果需要在硬碟裡工作，建議使用者可以利用 HTMLHelp 版本或其他版本。倘若需要利用其中的某部份經文進行教學或其他的運用，可於經文瀏覽區選擇檢視原始檔，再選擇另存新檔。或者直接利用剪貼簿的功能，如此較不浪費硬碟資源。

5. 丁福保《佛學大辭典》電子檔

為方便使用，光碟中提供兩種格式的丁福保《佛學大辭典》的電子檔：[註 26](#) 純文字檔格式及 HTMLHelp 格式。且點選 HTMLHelp 版丁福保《佛學大辭典》，由於檔案較大，開啟時間較長。另外，目前這個 HTMLHelp 版本只能在中文 Windows 環境之下操作使用。[註 27](#)

上述為光碟所附電子檔內容，而網路

版 (<http://ccbs.ntu.edu.tw/cbeta>) 除與前述功能大同小異外，並有 FTP 可供使用者上網下載經文資料，及可隨時查閱新增資料與最新訊息等依網路特質而加的功能。

殊勝特點

當然漢文佛典電子化的工作，國際間是有一些單位在進行中，且各有其特長與殊勝處，至於 CBETA 電子佛典的特點為何？茲簡述如下：除了

頁 365

精確與嚴謹的校對是 CBETA 最基本的要求外，如何保持原來版本的格式也是一重要的考量點；另外，為了合乎電子化的特點，並不只是把書面變成電腦排版就是所謂的電子化，因此要考慮到標記(Markup) 的處理，[註 28](#) 及採取 TEI (Text Encoding Initiative) [註 29](#) 的準則，並且以 SGML/XML 的方式呈現等方式；同時，也考量人性化檢索的需求，而開始做結構與層級的索引典之考量。然而，除了國際性與本土化的考量外，學術性與普及性的同時進行也是 CBETA 很重要的考量點，也就說在學術取向的同時，也考慮到如何讓 CBETA 電子佛典能普及為一般使

用者所接受，這些考量也是 CBETA 作業進行的重點。因而以雙管齊下兩者同時進行的方式，開始作業與推廣。而上述所有成果的呈現皆是以 SGML/XML 作為主要檔案，故特於下段舉例說明 CBETA 如何以 SGML 處理藏經的內涵。

舉例說明

在 CBETA 處理《大正藏》經文的標記 (Markup) 過程中常見的問題，皆一一將之列入《CBETA 工作手冊》中，以便參考。譬如處理藏經時

【遇到不分卷該如何處理】、【遇到一個 Note 中包含另一個 Note 的處理方式】、【遇到卷終的處理】、【遇到別的版本比大正藏多字的處理】、

【遇到 Note 的內容有跨行的情形】、【Note 為該語詞的梵文時之處理】、

【遇到 Note 為"明註曰 x 南藏作 y"的處理方式】、【遇到"*"的處理方式】、

【遇到別的版本比大正藏少字的處理方式】、【遇到夾註的處理】、【遇到卷首的處理】、【遇到某個版本缺譯號的處理】等常遇到的問題與處理，皆詳明的列出 SGML 的作業方式，今舉列說明如下：[註 30](#)

頁 366

1. 【表示版本異同的符號】

大正藏：	SGML 檔案：
冊 = 策【宮】	<pre><APP n="000103"> <LEM>冊</LEM></pre>

	<pre><RDG WIT="【宮】">策</RDG> </APP></pre>
然+（圓）【明】【宮】	<pre><APP n="000107"> <LEM>&lac;</LEM> <RDG WIT="【明】【宮】">圓</RDG> </APP></pre>

說明：

表示版本異同的符號

APP=容器, 含 LEM, RDG

LEM=底本的文字在 wit 屬性記錄版本的名稱

RDG=異本的文字在 wit 屬性記錄版本的名稱

2. 【遇到別的版本比大正藏多字的處理】

大正藏：

SGML：

增
上
大
苦。

⊙(苦)⊕增⊖⊙

說明：

《大正藏》註 3 中(三)(宮)在"增"字前多一個"苦"字，SGML 的表示方式

為<APP n="049403">

<LEM>[lac]</LEM><RDG WIT="【三】【宮】">苦</RDG></APP>增。

頁 367

1991EBTI

除了 1998 年 12 月 20 日於慧日講堂舉辦的「成果發表會」[註 31](#) 外，CBETA 最近在中研院內亦曾兩場的發表會，一為 1998 年 12 月 1 日在中研院謝清俊教授的邀約與安排下，CBETA 全體同仁假中研院資訊所演講廳，與中研院各研究所有關研究人員及中研院計算中心資訊人員做交流。CBETA 的展示獲得與會人士相當的肯定，能到中研院做交流，且能獲得多數人的贊同與肯定，對 CBEAT 而言是一大鼓勵。

另一場次，則是於 1999 年元月 15 日至 21 日，假中央研究院學術活動中心與地球科學研究所會議廳，由教育部電算中心及中研院計算中心主辦，法鼓山中華佛學研究所、菩提文教基金會暨台大佛學研究中心協辦的國際「電子佛典推進協議會」(Electronic Buddhist Text Initiative, EBTI)。 [註 32](#)

會中惠敏法師 "CBETA and Taisho Tripitaka Project" 的展示與發表，是國內首次正式對國際發表「中華電子佛典協會」的成果。於會中，CBETA 的成果與作業方式，得到許多國外與會專家、學者的肯定與讚美，同時確認 CBETA 作業方式與方針的正確性。本場發表可說是台灣

佛典電子化踏上國際的重要里程碑。

會中，美國柏克萊大學 Dr. Lewis Lancaster 說，CBETA 的成立一直是他長期以來的期望，很高興現在終於得以實現；而日本全國佛典輸入組織 SAT 代表石井公成 (Dr. Josei Ishii) 等人則於會中多次表達 SAT 與 CBETA 締結姊妹會的誠意；英國牛津大學 Dr. Lou Bernard 除對 CBETA 的採用文獻編碼協定(Text Encoding Initiative, TEI) 讚美有加之外，還說 CBETA 所呈現的方式是他「想偷學的巧妙方法」(An ingenious idea I propose to steal)。 [註 33](#)

頁 368

銘謝合作

一件事業的完成，都是眾緣和合所成，CBETA 於工作進行中同時與國內外很多單位團體聯絡合作，如國內中央研究院、國家圖書館、台灣大學、佛學院所、出版單位、宗教團體、佛典工作者等。而國外則有國際電子佛典推進協議會 (EBTI)、日本 SAT 組織、美國加州柏克萊大學、日本東京大學、日本京都大學、日本花園大學及其他國外佛教團體與學術單位的支持與合作，僅此致謝。

也謹藉此文感謝諸山長老的指導，各大道場及佛教團體的支持，以及許多大德不吝提供我們經文電子檔。 [註 34](#) 同時，也感謝 Buda-Tech 討論群、電子佛典編輯小組 (EBTWG)、25T 小組、缺字小組、力新國際科技股份有限公司等團體的技術協助與指導；以及所有參與工作及協助指導我們的諸位菩薩大德，加上許多熱心的伙伴及義工們默默的耕耘與參與，佛典電子化的工作才得以順利的開展。同時，對佛教電腦資訊庫

功德會提供丁福保《佛學大辭典》電子檔，及日本「大藏出版株式會社」授權 CBETA 輸入與公開《大正新脩大藏經》，特此致謝。[註 35](#)

義工助理

最後，要感謝的是在 CBETA 幕後一群默默工作的義工。譬如，在「紙面校對」的工作中，CBETA 很榮幸能邀約到散佈於各國中、國小的教師群幫忙校對。因為職業上的訓練，他們對於錯字與錯誤的敏感度高於常人。是以，佛典經文經過嚴密的「電腦校對」後，有這群義工老師的把關，CBETA 的電子檔經文得以以較好的品質及較高的正確率問世。

此外，亦感謝多位義工幫忙各項會務工作，使得 CBETA 的工作能順利進行。

頁 369

同時，為了能讓佛典電子化的工作能永續經營，培養新生代參與佛典電子化的作業，也是 CBETA 的重點工作之一。因此，除了 CBETA 的工作群外，「研發組」亦延聘數位佛學研究所學生擔任「研究助理」，參與佛典電子檔的校勘與標記(Markup) 工作。如此，不但能增進學生的佛學與電腦知識，亦期能藉此培訓俱佛學學識與電腦技能的新一代學者，以因應二十一世紀的需求。

未來展望

截至目前為止，CBETA 除了上述的成果之外，其他的藏經電子檔，也正陸續完成中。未來，我們將朝向附有校勘資訊的「學術版」努力，CBETA 擬於學術版內附有《大正藏》的註解欄等資料，藉此提供歷代各版本漢文大藏經的原貌。並將與佛教學者合作，以完成以單部經典為

主的學術版等計畫。當然，最終則是期望能達到《CBETA 電子大藏經》「普及化與學術化、本土化與國際化兼備」的目標。

總之，期望新的成果能在近期內發行，以饗使用者。

註 1 本文是在 CBETA 主委惠敏法師、副主委厚觀法師與常委恆清法師的指導下，為 CBETA 全體工作人員的腦力結晶，由筆者匯集成文。然因時間匆促，內容難免有所疏漏，此為筆者之疏忽，敬請指正。

註 2 近年來從事的佛典電子化的工作，可分為學界、教界及民間三部分來觀察：學界有中研院謝清俊教授等多位學者專家多年來的推動佛典電子化工作，因而促成楊國屏教授等人於 1994 年策劃「電子佛典籌設會議」。教界則有佛光山於 1995 年成立「佛光山電腦大藏經」單位，以進行藏經電子化工作；及美國佛教會於 1994 年成立「佛教電腦資訊庫功德會」進行佛典電子化的工作。民間則有徐言輝先生及幾位朋友組成的「電子佛典編輯小組 (EBTWG)」，利用掃描 (SCAN) 與辨識 (OCR) 技術，以《佛教大藏經》為底本，有系統的產生電子經文檔；甚而為了深入討論佛典缺字的解決方案，於網路上成立「缺字小組」，及 Buda-Tech 討論群(詳見下註) 等網友的共同努力，以促成及帶動此熱潮。

註 3 Buda-Tech 討論群是由台大獅子吼 BBS 站及中山鹿苑佛教 BBS 站的討論版，專門進行佛典電子化相關問題的討論，並由曾國豐先生架設 Mailing List 提供討論者使用。許多佛典資料及相關技術都是本版網友努力的結果，邱大剛先生則將之整理成「佛教經系列」置於台大佛學研究中心的網站上(<http://ccbs.ntu.edu.tw/canon/>)及其他國內外網站上。有關此部份的發展過程，請參閱邱大剛，〈關於「傳統與現代佛典」〉，《佛教圖書館館訊》第 15 期，1998，頁 62-63。

註 4 所謂 25T 小組，則是由「台大佛學研究中心」主導，負責處理蕭鎮國先生所提供之 25 冊 CCCII 格式的《大正藏》經文檔，本小組即為「中華電子佛典協會」之前身。

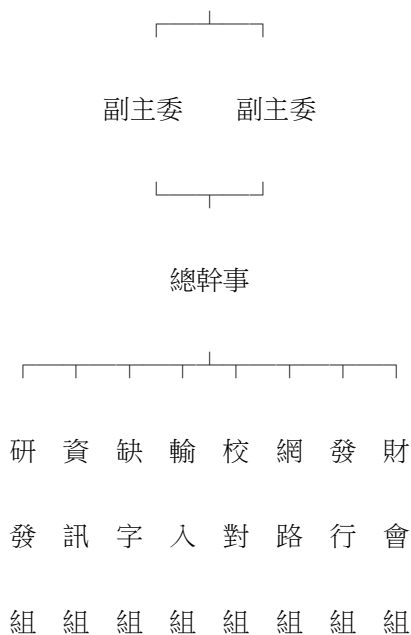
註 5 有關佛典電子化的目標，請參閱周伯戡，〈從傳統佛典到電子佛典〉，《佛教圖書館館訊》第 14 期，1998，頁 14-24。文中對電子佛典的發展有相當的期許與精闢的論點。

註 6 CBETA 組織架構圖：

中華電子佛典協會

|

委員會—常務委員—主任委員—執行顧問—顧問團



[註 7](#) 「中華電子佛典協會」聯絡資訊：

地址：台北市朱崙街 36 號(慧日講堂) 303 室
Tel:+886-2-8773-6469 Fax:+886-2-8773-6470

E-mail:cbeta@ccbs.ntu.edu.tw Homepage: <http://ccbs.ntu.edu.tw/cbeta>

[註 8](#) 其中有關謝教授的建議，請參考謝清俊，〈電子佛典中處理中文版本的方法〉1994，
URL:<http://www.sinica.edu.tw/~cdp/>；〈電子佛典的意義〉1995，
URL: <http://www.sinica.edu.tw/~cdp/>及〈網路資源對佛教之衝擊與影響〉，《佛教圖書館館訊》
第 15 期，1998，頁 6-24。他於這些論文中，對漢文佛典電子化提出很多寶貴的意見。其他各
項論點，請參閱《佛教圖書館館訊》第 15 期〈佛教資料電子化論壇〉周海文等人的文章，其中
有多篇專家學者對佛典電子化的期許與討論的文章。

[註 9](#) 日本 SAT 組織，是由日本全國各大學有志於佛典電子化的學者所組成的團體，目前正

建立一「大正新修大藏經原典資料庫」(SAjganikikrtaj Taiwotripitakaj, SAT)，詳細情形

請參考 <http://www.l.u-tokyo.ac.jp/~sat/big-5/index.html>。

[註 10](#) 謹此感謝日本 SAT 多位教授的協助，始能順利完成「中日大正藏版權簽約」的重要任務。

[註 11](#) 相關報導請參閱 CBETA 網頁 <http://ccbs.ntu.edu.tw/cbeta/step/news.html>。

[註 12](#) 有關缺字的處理可參考謝清俊，〈電子古籍中的缺字問題〉等文(謝教授有關缺字論文全文，可於 <http://www.sinica.edu.tw/~cdp/> 下載)；莊德明，〈漢字缺字處理與梵巴藏字母的輸入〉，《佛教圖書館館訊》第 14 期，1998，頁 6-13 及《佛典共用造字集的規畫》等文，URL: <http://www.sinica.edu.tw/~cdp/>；光音天，〈建立電子佛典標準交換碼---「漢字庫」的構想〉，《妙心雜誌》第 32 期，1998，頁 20-23。如能再加上維習安的著作，則可掌握佛教漢文缺字的大部分問題。Christian Wittern (維習安) 討論缺字的文章請參照：

Wittern, Christian "Chinese Character Encoding", in: *The Electronic Bodhidharma*, Nr. 3, July 1993, p. 44-47.

Wittern, Christian "The IRIZ JanjiBase", in: *The Electronic Bodhidharma*, Nr. 4, June 1995, p.58-62.

Wittern, Christian "Chinese Character Codes: An Update", in: *The Electronic Bodhidharma*, Nr. 4, June 1995, p. 63-65.

App, Urs and Wittern, Christian "A New Strategy for Dealing with Missing Chinese Characters", in *Humanities and Information Processing*, No. 10, February 1996, p. 52-59.

Wittern, Christian "Taming the Masses. A Practical Approach to the Encoding of Variant and Rare Characters in Premodern Chinese Texts.", Talk held March 7th 1997 at the Institute for Information Science of Academia Sinica in Taipei, Taiwan.

See also <http://www.gwdg.de/~cwitter/cw/taming.html>.

[註 13](#) 「一般組字式」基本規則說明：

本組字法含*/@-+?六個半形基本符號，及()兩組半形分隔符號。

今舉例說明如下：

符號	說明	範例
*	表橫向連接	明 = 日*月

/	表縱向連接	音 = 立/日
@	表包含	因 = 口@大 或 閒 = 門@月
-	表去掉某部份	青 = 請-言
-+	若前後配合，表示去掉某部份，而改以另一部份代替	閒 = 間-日+月
?	表字根特別，尚未找到足以表示者	背 = (?*匕)/月
()	為運算分隔符號	繞 = 組-且+((土/(土*土))/兀)
[]	為文字分隔符號	羅[目*侯]羅母耶輸陀羅比丘尼

為求方便，不排除採用全形注音、標點及英文符號做為組字用字根。

註 14 所謂「M 碼」就是以 M 開頭後面接著一串數字的缺字表示法，是國際上廣為採用的字表達法之一。譬如 CBETA 以 M00763 代表[條-木+火]，通用字為「條」的這個電腦缺字。

註 15 特此致謝日本「今昔文字鏡」(Mojikyo Font Center, <http://www.mojikyo.gr.jp>) 提供網上免費下載字型，以便使用者處理缺字問題。

註 16 有關 CBETA 初期的作業流程與技術突破，請參閱吳寶原，〈從實務經驗談佛典電子化初步工程之演進〉，《佛教圖書館館訊》第 14 期，1998，頁 25-31。

註 17 此外，CBETA 並於各版本的經文資料中，均附有經文相關資訊，除記載原始資料來源，提供使用者參考外，亦期望此經文資訊與經文提供者名稱，能隨經文電子檔一併流通。

註 18 上述的工作進度與產量，是依 CBETA 目前的財務能力，以平均十二位工作人員計算。依此人力估算，至 1999 年底的預計進度，期能完成藏經 1-55 冊的一半數量，亦即完成《大正藏》28 冊以上的校對。

註 19 網路版的檢索引擎是採用中正大學 GAIS 團隊所開發的元件，特此致謝。

註 20 光碟中亦提供三項檢索程式，可在使用說明的選項中，選取工具使用說明，直接選取安裝，或是將光碟上的程式複製到硬碟執行，這三項工具分別為：

- 1) CBGrep (Windows 下的文字搜尋工具)
- 2) CTGrep (Dos 下的文字搜尋工具)
- 3) Search and Replace for Windows 95/NT (shareware)

註 21 另外，光碟中提供了兩種格式的丁福保《佛學大辭典》電子檔：

1) 純文字檔格式

2) HTMLHelp 格式

[註 22](#) 譬如，有關梵文轉寫字可安裝 CBDIA.TTF 字型檔。如果安裝順利的話，將可看到類似如下的表達：Mahayana。

[註 23](#) 普及版與 App 版行首資訊規則說明

Txxnyyyyopzzzzcll 為每行經文之前的資訊，其意義如下：

T：表示大正藏（Taisho）經文。

xx：二位數，表示大正藏冊別。

n：固定不變，表示後面接經號。

yyyy：四位數，表示大正藏經號。

o：大寫之 A，B... 表示大正藏有記載之同經號之別本，

小寫之 a，b... 表示大正藏沒有記載之同經號之別本，

完全沒有同經號資料者，用下底線符號填入'_'。

p：固定不變，表示後面接頁碼。

zzzz：四位數，表示大正藏頁碼。

c：表示第幾欄，a 表示上欄，b 表示中欄，c 表示下欄，d 表示校勘欄。

ll：二位數，表示在該欄的行數。

例如：T10n0279_p0001a01，即表示大正藏第十冊 0279 經第一頁上欄第一行。

[註 24](#)HTMLHelp 為附於視窗 98 中輔助說明的瀏覽器，視窗 95 亦可以更新使用。

[註 25](#) 同註 15。

[註 26](#) 光碟所附的丁福保《佛學大辭典》電子檔，為「佛教電腦資訊庫功德會」所提供，特此致謝。

[註 27](#) 資料庫的詳細使用說明，請參閱光碟之【中華電子佛典協會資料庫基本介紹】及【使用說明】。

[註 28](#)Wittern, Christian "Minimal Markup and More - Some Requirements for Public Texts",

Conference talk held at the 3rd EBTI meeting on April 7th 1996 in Taipei, Taiwan。

Christian Wittern 相關論著，請參考 URL: <http://www.gwdg.de/~cwitter/home.htm>。

[註 29](#)Sperberg-McQueen, C. Michael and Burnard, Lou (Eds.) *Guidelines for Electronic Text*

Encoding and Interchange [TEI P3], Chicago and Oxford: Text Encoding Initiative, 1994.

[註 30](#) 因篇幅關係無法一一詳述，故僅摘錄《CBETA 工作手冊》之一小部分作為範例。

[註 31](#) 本次成果發表會，邀請國內近百名從事佛典電子化或相關工作的學者專家及宗教團體參加，並於會中發行 CBETA 手工製作的測試光碟，事後獲得很多寶貴的改進意見，特此致謝。

[註 32](#) 有關 EBTI 會議，請參閱杜正民，〈當代國際佛典電子化現況：電子佛典推進協議會 (EBTI) 簡介〉，《佛教圖書館館訊》第 15 期，1998，頁 28-39，文中對 EBTI 的起源及歷屆會議有簡潔的說明。至於最近一屆在台北舉行的 1999 EBTI 會議簡介，請參考杜正民，〈從台北看國際電子佛典實況：1999 EBTI 會議實錄〉，《現代佛教學會會訊》第二期，1999，頁 10-17。

[註 33](#) 出處: <http://users.ox.ac.uk/~lou/reports/9901taipei.htm>。

[註 34](#) 特此感謝蕭鎮國、釋厚觀、楊郁文、維習安 (Christian Wittern)、張文明、莊德明、曾國豐、陳誌雄、日本 SAT、佛教電腦資訊庫功德會、妙雲蘭若、伽耶山基金會 (香光寺)、靈山寺 (南投名間)、眾生出版社、新雨佛學社、佛典推廣小組 EBTWG (徐言輝、何宗武等人)、摩訶工作室、東坡站經文 (台大經典系列)、佛教藏 OCR 小組等提供經文電子檔以供比對。

[註 35](#) 因為協助與贊助者很多，恕無法一一詳列於此，詳細名單請參閱 CBETA 網頁

<http://ccbs.ntu.edu.tw/cbeta/friend/gongde.htm>。