

本文發表於 佛教圖書館館訊 第十四期 1998 年 6 月，轉眼就過了26年，現原文照抄，以茲紀念。

漢文電子佛經檢索軟體的實作

以《印順法師佛學著作集》光碟版為例

佛典數位化早在五、六年前就已開始引起教界的重視。數位科技因挾其無與倫比優勢：如保存久、散佈廣、檢索快、攜帶便利等，將傳統的任何媒體：木刻、石刻、印刷品、錄音帶等，都遠遠拋在後頭，所以將佛典數位化是思想前瞻的佛子很自然的想法，因為這除了有永遠保存原典的消極意義之外，更能積極地將它當成弘法的新方便。網際網路的普及，也讓大家見識到佛經上所形容的「分身無數億、法音廣宣流」的環境在網路似乎可以得到某個程度的實現。

緣此，感受到時代呼喚與歷史使命感的一些人，不約而同地投入這項工作，他們經過一段時日的摸索，逐漸發現，在到達數位化誘人的成果之前，原來竟然有這許多始料未及的問題：缺字、系統內碼、操作平台、中文檢索、標記、文件結構、多版本……，問題既複雜且繁瑣，比方說，某字該不該「造」？這就不只是技術上的問題（會遇到造字空間的限制），也涉及文字學的素養；要輸入哪個版本的佛經？原版中明顯有錯誤的地方可不可改（不改看了難過，改了增加校對的困擾，因為需要校對者的佛學素養，這就會減少可幫忙的人數，而拖延進度……）？根本的問題在於：「保存原貌」和「便利流通」。這兩個問題到目前為止還是互斥的訴求，其平衡點究竟在那裡？這情況宛如玄奘的「直譯」和羅什的「意譯」一樣，千年難論對錯。

電子佛典的工作，投入的經費龐大，開發的過程困難，需要整合的人才多樣，而願意支持的人卻少之又少（相對於其他的弘化事業），我不止一度思考，這是否值得繼續做下去？就在心生退卻之際，妙雲蘭若的法師在偶然的一次會議上，提及導師的全部著作已輸入電腦，校對也已近完工階段，問我有沒有興趣參與最後的工作：製成全文資料庫發行？我知道這不是輕鬆的工作，然憶起自己深受導師著作的啟發，所以就答允了，一旦手邊的工作結束後就開始進行。

97年10月後，開始把大部份的時間投入到導師全集的軟體製作，考慮到導師的讀者必定有翻閱原典及查佛學工具書的需求，因而想設計出同時擁有快速檢索及便利閱讀的開放式環境，讓往後的佛經資料和工具書可以很自然地納入這個系統，整合成一致的使用介面。經過長期密集的討論，我們將這個軟體環境命名為OPEN 98，以下是OPEN 98的架構和發展情形：

(一)需要一組針對佛典結構設計的開放式標記語言，作為描述文件結構及增添新功能的基礎。OML就是在這個理念下被設計出來的，我們將文件的內容、結構、擴充功能(如圖片嵌入、語言、影像等)都用統一的標記語言來描述，標記過的文件再由OPEN 98預先進行分析、建立索引檔。檢索時則可以輸出成標準的HTML(由瀏覽器解讀)或直接輸出到WORD進行編修。

(二)因為資料量相當龐大，我們需要一個效能良好的檢索核心，同時也要解決中英日梵文混合及缺字的問題。我們曾考慮過直接引進既有的技術，如中研院的CSMART、吳昇教授帶領團隊開發的GAIS系統，但是經過評估，還是決定自己開發專屬的檢索核心，除了不必負擔軟體授權費之外，更能針對OPEN 98的需要進行最細微的調整。值得一提的是，檢索程式開發的途中，遇到高手跨刀提供檢索技術和觀念，使得最新的版本(比本次發行的光碟版新，目前只有透過網路使用)檢索一億字幾乎可以在一秒之內完成，索引檔平均也只有資料檔的1/3，相較於一般市售的產品毫不遜色。

(三)成品要同時能在網路上及光碟上執行。網路雖有更新方便、管理容易、範圍廣的優點，但由於現有公共網路基礎建構趕不上使用者的成長，導致嚴重的「塞車」問題，對個人使用者而言，上網查資料意味著長時間的等待，使用意願普遍不高，因此，我們要求成品不但要能在網路執行，也要能在單機和區域網路環境運作。我們的願望是使用者只要有一部平價的個人電腦，不必連線上網，即可輕鬆地享受到我們精心製作的全文檢索軟體、字形及佛典資料庫。

(四)對於缺字這個頭痛的老問題，考量點在：佛經有必要和其他文獻做交叉參考，如以史料方面的文獻為例，假設我們的電子佛典採用一套封閉的、專屬的缺字系統，這對未來跨領域的交流非常不利。所以我們不走造字檔的老路子，而決定重新設計一套新的內碼系統，運用一些編碼的技巧，以漢語大字典五萬四千餘字為基礎，做成一套可以在網路及光碟上運行、不需任何造字檔、不需加掛任何驅動程式、並可在任何系統平台運作的「漢字庫」，本字庫業已順利運用於導師的光碟中，也獲得一致的好評。關於「漢字庫」的細節，超出本文的範圍，有機會再為文說明。

(五)雖然我們的著眼點在於電子佛典的普及，但也希望同時能滿足最嚴苛的學術要求，因此，即使OPEN 98輸出的經文已重新分段，每行字數也與原典不同，但是檢索的結果卻可以顯示原書的頁、欄及行的程度，有需要的使用者可以很快地找到原文。另外，部份的經文也將附有高麗藏及大正藏的頁碼顯示及原

書影像檔，如此，不離開螢幕也立即可享受看到原書的喜悅，這對學術研究來說可謂相當便利。

(六)OPEN 98的開放設計，讓使用者同時可以是經文的提供者，只要用OML來標記文件檔，就可以讓OPEN 98讀取並檢索，所以使用者可以自由地上傳或下載文件，創造自己專屬的操作環境。各種文件也可以很自然地融和在一起。比方說閱讀導師的著作時，可以很快地查回原典，或者查各類佛學名相辭典(目前進行中的有《佛光大辭典》、《中華佛教百科全書》、《重編一切經音義》等數種)。

看到以上所設定的目標，就知道OPEN 98的研發是一項艱鉅的工程，所幸在發展的過程中，不斷地有適時出現的助緣幫我們度過一次次的難關，而使用者的期待是支持是我們繼續下去的原動力，我們相信，無論是技術上還是非技術上的困難都是有限的，但是這一套軟體若能成功地研發出來，為教界學界帶來的利益卻是無窮：

(一)各道場只要把精力專注於資料的編輯和整理，而不必為諸多技術問題傷腦筋，這就能有效提高教界製作電子佛典的意願。

(二)使用者只要學習一套軟體和介面(如果有用瀏覽器上網經驗的人，更立即可以上手應用)，就可以同時使用多種資料庫，無形之中省下極為可觀的安裝、學習及教育成本。

(三)提供一個成本最低、功效最著的方式，讓電子佛經順利地搭上這波資訊革命的便車，以光碟的形式大量複製、留存千古，同時用網路的形式散佈於這威力絕倫、可能無限的虛擬世界。最後附上幾幀OPEN 98的執行畫面證明我們不是在畫餅，而是真正在做事。



圖一：OPEN 98的整合介面，可以看到多種資料庫，原典、今人著作、辭書都可以整合在一起。



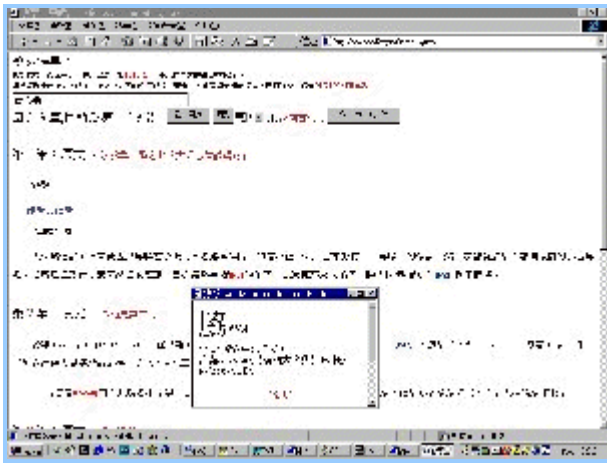
圖二：這就是98年4月8日發行的導師著作全集光碟版的起始畫面，您可以看到導師數十載的著作都收錄在這裡，每本書除了可以檢索，也可以進入閱讀模式，逐章逐頁拜讀。



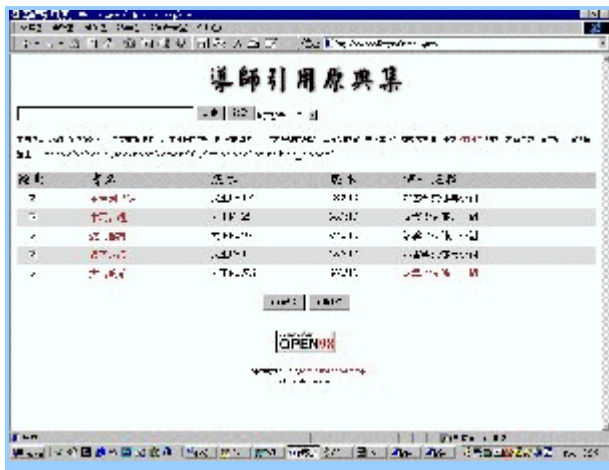
圖三：這是佛光大辭典的全文檢索結果，查詢條件是「見道之別名」，只找出一筆「正性離生」，紅色的數字{P1991-2-14}表示此句在原書1991頁第2欄第14行，藍色的[P1991-3]表示這是原書第 1991頁第3欄的開始，另外，最後一行Time taken by OPEN 98 Kernel (OPEN 98核心花的時間)顯示共花了0.048秒來查，0.014秒轉成HTML格式。比第一版的佛光大辭典(97年5月16日由佛光山寺發行)耗時8~10秒鐘，快了近200倍。

書名	冊次	頁次	冊次	冊次
...	4	1	1	...
...	4	1	1	...
...	0	4	4	...
...	2	2	2	...
...	11	2	2	...
...	40	22	22	...
...	4	2	2	...
總計	146	146	146	...

圖四：這是統計導師著作中「脅尊者」的出現頻次。OPEN 98會清楚地列出哪些書出現「脅尊者」、出現幾次，並且最後還會計總，當然，也可以按右邊的按鈕來瀏覽查詢結果。



圖五：優異的缺字顯示是OPEN 98相當值得一提的能力，不必加掛任何「驅動軟體」就可以顯示數以萬計的中文缺字，查詢則可以用通用字(如輸入「脅」可查出「月荔」，輸入「鉢」可查出「金本」，或者萬用字元「?」)。我們和中央研究院文獻處理實驗室保持相當密切的合作關係，這套目前收錄字數5萬4千餘的《漢字庫》就是受謝教授的觀念啟發，加以實現的成品，這裡要表達我們的感謝之意。



圖六：這是目前在最後修改階段、即將由妙雲蘭若發行的一片光碟之首頁：「導師引用原典集」，也是採用OPEN 98作為核心。畫面中列出的是硬碟中所能找到的五部經文。本光碟的設計考慮到無法一次出齊所有導師引用過的原典，而且資料也有修正的需要，所以特別加入了動態首頁的觀念，使用者可以自由地從網路或來函索取的方式取得原典經文，軟體可以自動偵測並動態更新您電腦上的首頁，這樣，我們提供一個開放的架構，便利大家製作新的經文，自用或送人兩相宜。

十方來、十方去，我們是深受教界法乳深恩的一群人，所以自我定位為教界的「公共研發團隊」，歡迎大家給隨時給我們批評指教，如果有任何電腦軟體方面的疑難，也請不必客氣，我們會盡力給予協助。