

J Chin Med 12(1): 1-9, 2001

# 電腦輔助擷取河間六書方劑專論資料

陳逸光

慈濟大學 醫學系

花蓮

(2000年5月18日受理, 2000年12月4日收校訂稿, 2000年12月6日接受刊載)

中醫古籍文獻整理工作, 交由電腦代勞是相當經濟實用的方法。中文全文自動擷取技術已相當成熟, 應用在中醫古籍照樣可行。本研究選取電子河間六書為測試專案, 河間六書全文資料庫共 190431 字, 本草備要藥物名詞庫共選取 203 種藥物, 專屬河間六書藥物名詞庫共選取 229 種藥物 (含前項 203 種藥物)。電腦程序運用上述之藥物資料庫擷取全文有關方劑專論之藥物組成, 結果篩選出 603 組藥物群組, 其精確率 82%, 回收率為 95%, 對輔助人工整理中醫藥古籍資料效率相當高。研究最後由人工方式核對資料, 共整理出河間方劑 512 首, 並已登錄在中醫藥古籍文獻網際網路上供使用。本研究的電腦程序、擷取流程及藥物名詞資料庫可移植到任何中醫藥古籍文獻資料上重覆使用。

關鍵詞：中醫藥典籍文獻、方劑組成自動擷取、電子中醫藥詞庫、網際網路。

## 前 言

中醫藥古籍文獻資料庫網際網站 (WWW.TCMET.COM.TW) 於民國 86 年建置完成, 現今建立在 TCMET 網站上之中醫藥古籍計有黃帝內經、金元四大家之醫學論著、景岳全書、類證治裁、備急千金方等重要中醫藥典籍供網友查閱。TCMET 網站主要由二組在 WINDOW NT4.0 作業系統下架設之網路伺服器 (Web Sever) 及微軟 SQL 資料庫伺服器 (MS SQL Database Server), 分別負責 TCMET 動態網頁之通訊及管理<sup>1</sup>。從網友使用 TCMET 記錄日誌檔 (Log file), 得知經整理的資料如景岳全書的新、古方八陣及本草學等專題資料庫深受網友歡迎。以電腦程序輔助整理中醫古籍電子文獻資料, 是本研究之重要研究課題。人工整理資料的成本遠遠超過電腦輔助整理之費用, 若能夠開發一套自動化電腦程序輔助及整理中醫藥古籍電子文獻資料且善加利用, 將可建立更多專業而效率高之中醫藥關聯性資料庫, 若要發揮網際網路資料庫功能, 這些專業表格化資料庫是十分迫切需要的。以人工方式將中醫藥古籍文獻資料編碼分類, 並重新輸入資料庫表格中是相當耗時費事而且很繁人的工作。例如 TCMET 上景岳全書方劑、藥物學之網頁及類證

---

聯絡人：陳逸光, 花蓮市中央路三段 701 號, 慈濟大學醫學系, 電話：(03)856301 轉 7207。

治裁 200 個定義之條文，當初便是由慈濟綜合醫院（花蓮總院）中醫科的醫師們協助以人工方式分類整理，最後書寫在紙本上，再由鍵盤輸入預設之資料庫表格中，相當之不容易。

自文字檔中由電腦程序自動檢索資料是當今資訊擷取 (Information Retrieval) 科技之重要研究課題，應用在中文全文檢索之技術已相當成熟<sup>2</sup>，但相對的中醫藥文獻自動擷取工作研究則較少。由於網際網路最近的熱潮近乎發燒，整理有效率的電子中醫藥文獻資料需求亦會增加，因此自動化中醫藥文獻工作倍感迫切。本研究根據中醫中藥方劑組成及全文檢索的判別特徵，進而選取適當的方法，可簡化篩選程序並提高藥物組成的擷取率，使精確率 (Precision Ratio) 達 82% 及回收率 (Recall Ratio) 達 95%。電腦自動判別之資料最後由人工方式在電腦程序中修正及整理，河間六書的方劑資料便能夠由電腦程序自動產生，且經整理之電子資料可供 TCMET 網際網路直接使用。

## 材料及方法

### 一、電腦軟體

本研究選擇 MS Visual Foxpro 5.0 作為河間六書醫籍之資料庫管理，此軟體內部是一個強大的電腦程式語言，並提供結構化查詢語 (SQL) 程序。本研究的各項查詢 (Query) 及擷取程序都是在此軟體下完成。

### 二、河間六書資料分析

素問玄機原病式、黃帝素問宣明論方、傷寒直格論方、傷寒標本心法類萃、傷寒心要及素問病機氣宜保命集世稱河間六書<sup>3</sup>，素問玄機原病式只有論而沒有方，電子河間六書資料庫在建檔之初已將每一味中藥以英文逗點分開<sup>3</sup>。中醫歷代醫籍大都按照一定之格式編寫：每一組方劑中之藥物組成皆是連續之藥物名稱，每一首方劑必定會有一方名，方劑中還包括主治、方義、製作、醫囑等內容，一般方名會放在最前而藥物組成則在方劑的中段出現。傳統醫典上方劑組成少者只有單味藥多者可以是上百味藥之組成。若能精確找出藥物組成之起始及結束，等於是完成了整首方劑的大部份工作，也正是本研究之目標，本研究中各藥物詞彙庫之辭彙集合可見於圖 1 及說明文字。

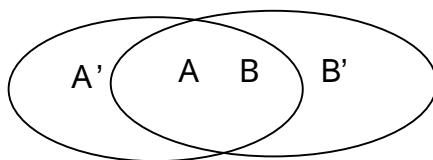


圖 1 河間六書電子文獻藥物詞彙集合

圖 1 說明：研究中選取本草備要 542 種中藥名詞<sup>4</sup>作為“種子”詞彙組成集合 A，河間六書電子文獻中所有之藥物詞彙構成集合 B， $A \cap B$  是自 B 子集中檢索出 A 子集藥物名詞之交集；有了交集後電腦程序便可進一步檢索擷取 B' (在 B 交集中被遺漏之藥物名詞)， $A \cap B$  及 B' 便是本研究所要完成的河間六書藥物組成詞彙庫。

### 三、建置河間六書藥物索引詞庫

#### 3.1 建立‘種子’索引詞庫

為標示所有在河間六書中之方劑藥物組成，本研究首先選取本草備要 542 種藥物作為全文檢的藥物詞彙。這裡稱之為‘種子’是因為：在設立檢索目標後先行選取某一特定之詞彙典，如本研究中所用之本草備要藥物名詞，即可把欲檢索之目標（河間六書中之方劑藥物組成）定位。在電腦程式設計中‘種子’觀念常被用於產先亂碼（Random Number）<sup>5</sup>，程式設計者先會給出一個數字（‘種子’），電腦才會根據此‘種子’作下一項指令而產生亂碼。種子藥物詞典在本研究中，肩負著針對主題縮小檢索範圍及定位河間六書方劑藥物組成的任務。

#### 3.2 運用種子索引詞庫擷取河間六書藥物組成

##### 3.2.1 種子索引詞庫標記文獻藥物名詞

電腦程序首先將已建檔之 542 個本草備要種子藥物辭彙，逐一在河間六書文獻中作比對，以記錄每一種藥物在全文中所出現之頻率及所在之位置，並以反轉檔（Inverted File）<sup>6</sup> 型式加以標記及儲存文獻內檢索成功之藥物名詞。

##### 3.2.2 自標記藥物名詞中篩選合理之藥物組成

藥物名詞可出現在文獻中任何地方，因此在 3.2.1 被標記之藥物名詞不一定是藥物組成，但藥物組成的特點是前後句都是藥名。因此本研究只考慮連續 2 個或以上之句子都是藥名的區域作為合理藥物組成選取範圍。古代醫家在編寫方劑論著時很少會將註解加在藥物組成中，因此藥物組成都是相鄰的藥物專有名詞，篩選程序運用此項特點，方劑專論的精確率會提高。

##### 3.2.3 篩選被遺漏之藥物名詞

上述的各項程序已足夠正確篩選出合理之藥物組成，但仍有很多在 B'子集中被遺漏之藥物。本研究再加入一個邏輯程序如下：在 3.2.2 被選取之藥物的鄰近兩個區域中，若被一個沒有被選上的句字分開，本研究將此句子的內容歸屬 B'的藥物子集（見圖 1 之說明）。然後再重複 3.2.1 所述之比對及建立反轉檔程序，河間六書方劑專論中之藥物組成被篩選出來的機會便會提高。

##### 3.2.4 建立河間六書專屬藥物索引詞庫

本草備要種子索引詞庫可以很快速地被用作定位方劑藥物組成的區域，被遺漏的藥物也可以依循上述之程序擷取補遺。接著是要建立專屬藥物索引詞庫，首先將所有在 3.2.2 及 3.2.3 被標示的藥物名詞排序並去除重覆的藥名，再由人工方式作最後確認。此階段之確認工作是在上述電腦程序篩選出來的藥物名詞表格中刪除非藥物名詞的記錄（Record），以確保所有在表格各項記錄是唯一的藥物名詞典（見附錄）。

### 四、建立及判讀藥物組成區域

#### 4.1 藥物組成起始及結束之定位切割

根據 3.2.4 之結果，再重覆 3.2.1（種子索引詞庫標記文獻藥物名詞）的程序，藥物組成起始及結束便可有效的被切割出來，但在 3.2.2 中已指出，藥物名詞可出現在文獻中任何地方，因此不是所有被切割的區域都是藥物組成，最後還得靠人工閱覽方式作擷取資料對錯的判斷。

## 4.2 人工判讀切割藥物組成的成功率

為確保每一個切割的藥物組成都被正確擷取，人工判讀成為藥物自動擷取對錯辨別的標準。本研究在切割的區域中，比對原文文獻資料並記錄選對、選錯及漏選的區域作為本研究藥物組成自動擷取回收率及精確率的評估依據。直到現在，在檢索系統中，回收率和精確率還是兩個最廣為接受的評估標準<sup>7</sup>，還可進一步計算回收率 / 精確率曲線圖（見圖 2），資料分析可信度可提高。

## 五、切割結果評估分析

本研究針對被切割定位的藥物組成區域，作為專屬藥物典自動篩選方劑藥物組成準確性的評估分析。分析的方法包括(1)精確率、(2)回收率及(3)精確回收率曲線圖，其計算方法分述如下：

### 5.1 精確率

精確率，其定義為所檢索到文章中相關文章的比例，這是一簡單易懂的概念，即檢索到的相關藥物組成區域越高，自然被歸類為較好的檢索，反之則是代表檢索出一堆不相關的‘垃圾’資料。

精確率 = 檢索所得之相關資料筆數 / 檢索所得之所有資料筆數

### 5.2 回收率

其定義是指相關文獻被檢索出之比例，對本研究而言即由人工判讀被正確切割定位的相關藥物組成區域，及沒有被切割定位出來所佔的比例，回收率越高代表被檢索程序遺漏的相關文獻越少。

回收率 = 檢索所得之相關資料筆數 / 資料庫中所有相關資料筆數

### 5.3 精確回收率曲線圖<sup>8</sup>

第三個測量的數值是精確 / 回收曲線（Precision-Recall Curve），此曲線的計算方式是將檢索出來之相關及非相關資料，計算其回收率及精確率，並將兩者之百分比在 XY-軸上展現。本研究之精確回收曲線見圖 2，其特點是精確率越大，回收率越小，可從曲線看出精確率及回收率在本研究中的相關性。

## 結 果

### 一、藥物詞庫集合結果（見圖 1）

A 集合：本草備要種子藥物辭彙集合共有 542 種中藥。

A'集合：即 A 集合減去 A∩B 集合共有 339 種中藥。

B 集合：在全文中所有的藥物組成辭彙共有 229 種藥物，見附錄（河間六書藥物辭典庫）。

B'集合：在 B 交集中被遺漏之藥物名詞，即 B 減去 A∩B 集合，共遺補 26 種藥物。

A∩B 集合：在全文及種子辭彙同時出現之專屬藥物辭彙共有 203 種藥物，以種子辭彙完成方劑藥物組成配對率達 89% (203/229)

### 二、藥物組成檢索精確率(Precision)、回收率(Recall)

河間六書的總字數共有 190431 字，被檢索出的藥物組成切割區共有 603 個屬於方劑內的藥物組成共 494 個（檢出而相關），其餘 109 個屬於非相關區域（檢出而不相關）。在文章中並未被選出之藥物組成區域有 27 個（相關而未檢出）。精確率、回收率之數據評估分析如下：

表 1 藥物組成切割精確率數據表

	相關	不相關	總數
檢出	494	109	603

表 2 藥物組成切割回收率數據表

	檢出	未檢出	總數
相關	494	27	521

回收率 = 檢索所得之相關資料筆數 / 資料庫中所有相關資料筆數

精確率 = 檢索所得之相關資料筆數 / 檢索所得之所有資料筆數

回收率 =  $494/521 = 0.95$                       精確率 =  $494/603 = 0.82$

## 二、藥物組成檢索之精確 / 回收曲線 ( Precision-Recall Curve )

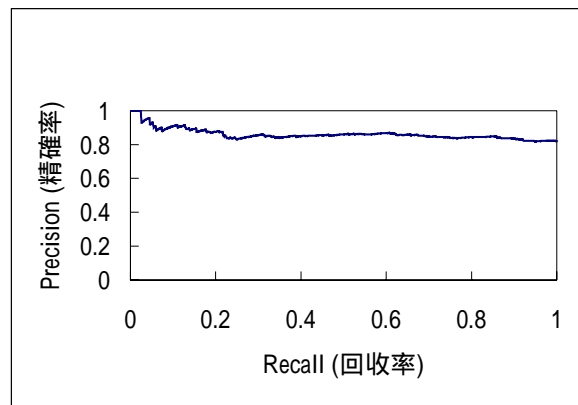


圖 2 藥物組成檢索之精確/回收曲線圖

## 討 論

中醫藥古籍文獻網際網站 (TCMET) <sup>1</sup>，自 1997 年 12 月底建立，眾多電子文獻資料經已上網，從網友的使用行為，得知使用者比較喜歡閱覽經整理分類的資料，以人工方式對全文文獻資料編碼整理是相當耗時費事的工作。因此本研究以自動化標示方劑藥物組成為研究主題，以開發一套能夠讓電腦輔助研究者不經由全文編碼便可整理方劑專論資料的程序。本研究是以逐詞索引的方式來完成藥物組成之擷取，而結果相當令人滿意，精確率 82%，回收率達 95%。高擷取率之主要原因有賴於檢索程序的設計及原始電子文獻建檔方式，兩者都是成敗的主要關鍵，歷代醫家編寫方劑專論時皆有一定格式 ( 見 3.2.2 ) 是其中一項成功主因。本研究首先選擇本草備要 542 種中藥名詞作為種子 ( 見 3.1 )，其方劑藥物組成配對率達 89% ( 見結果一 )。進一步建立文獻內專屬方劑藥物組成索引詞庫 ( 見 3.2 )，在資訊檢索領域中，這種方法常被稱為「相關回饋」 ( relevance feedback )，其施行方式是在前一階段找到的文件中，挑取重要的特徵，再回饋給系統，以期找到更多相關的資料，一些全文資料庫檢索研究中顯示「相關回饋」可提昇檢索成效達 20% <sup>9</sup>。

由於歷史年代不同中藥名詞並未能一致，因此檢索時會有誤差，如本草備要之‘白芍’，河間六書卻用‘芍藥’，因此進一步建立文獻內專屬方劑藥物組成索引典是非常有效率的作法，可以把藥物組成擷取率從 89% 提升至 98% (見結果)。分析被遺漏之藥物即 B' 集合 (見圖 1) 發現，以下列 2 種錯誤居多：(1) 一組不常見的藥物名稱被放在同一首方劑中。以河間六書中之三聖散為例，其藥物組成為烏魚骨 (海螵蛸)、燒綿灰及血餘灰 (髮)，括號內為本草備要之通用藥名，以上河間書中所載三種藥物烏魚骨、燒綿灰及血餘灰，均未曾在本研究中之藥物詞庫中出現 (見 3.2.4)。本研究的電腦方劑組成篩選程序，需要至少有一種在河間六書中出現之方劑藥物組成與本研究之藥物詞庫相吻合，這一首方劑才能夠被篩選出。像三聖散之例子，三種藥物中沒有一種藥物與藥物詞庫相吻合，因此電腦方劑組成篩選程序無法擷取而被遺漏。(2) 方劑中僅有唯一的一種中藥而且並不常見，如大豆孽散的組成，僅大豆孽一味藥，本草備要並未記載‘大豆孽’，與上述第 1 種錯誤同理，因為沒有一種藥物與藥物詞庫相吻合，所以電腦篩選程序無法擷取出這一首方劑而被遺漏。本研究在評估精確率中發現有很多非藥物組成的藥物組被錯誤擷取，原因是在中藥處方中有很多加減處方用藥，如四君子湯加陳皮、半夏名六君子湯，六君子湯減半夏名異功散，當河間六書方劑專著中討論加某某藥時，本研究之邏輯程序會誤以為是藥物組成，因為其範圍大部份仍在方劑專論中，所以對最後由人工判斷整理資料階段不會造成太多不便。

中醫藥文獻整理工作，由國人以應用軟體自行開發是理應研究發展的工作，國內有許多中文搜尋引擎作中文檢索相當成功<sup>10</sup>。可知在國內使用的外國全文期刊資料庫如 Elsevier Electronic Subscriptions 版權及使用費都是相當昂貴<sup>11</sup>，綜觀網際網路上國人自行開發的軟體也很出色。中醫中藥資訊化能夠由國人自行研發會比較經濟，而且會兼具中國文化特色。民國 82 年，由簡立峰氏在國立成功大學的一個專題講座談及中英文全文檢索有很深入之探討，他認為逐詞索引較不適用於中文，逐字索引是妥協的方法<sup>12</sup>。筆者在電子中醫藥文獻整理工作上深有同感，電子黃帝內經、金元四大家之醫學論著、景岳全書<sup>3,13</sup>等醫籍，都是以詞或句為單位，由於古文有言簡意深之特色，一行 40 個中文字 (80 Bytes) 的文章結構及上述的詞句特徵，對文字檔轉換成資料庫檔之格式相當容易，使本研究節省了很多開發的時間。中文檢索因文化背景不同不能任意自國外移植，語言學家特別指出英文詞彙的界限很容易區別，而中文則因為字與字相連而不容易確認詞彙的界限<sup>14</sup>，但中文逐字索引有其優點，因為英文有詞類變化問題如現在、過去、將來式、複數、單數等問題，反而覺得中文名詞逐字檢索比較容易。針對中文檢索的特徵，本研究選擇了本草備要中藥名詞作為特定之關鍵詞，而以關鍵詞作自動擷取的方法大致可分三類<sup>15</sup>：(1) 第一種為詞庫比對法：即利用已建立的詞庫，來比對全文資料，將文件中出現在詞庫中的片語擷取出來，本研究採此方法為主。(2) 第二種為文法剖析法：透過自然語言處理技術的文法剖析程式，剖析出文件中的名詞片語，過濾掉不適合的詞彙。(3) 第三種方法為統計分析法：透過對文件的分析，累積足夠的統計參數後，再將統計參數符合某些條件的片語擷取出來。這三種方法相當有啟發價值，本研究自動擷取方法以上述的第 1 類方法為主，藥物組成自動擷取成果令人滿意。方劑專論分類中還包括：方名、方義、製藥方法、加減方、醫囑、各醫家引述等課題，若都能夠運用前 3 項多重技巧交給電腦輔助整理，如果準確率能有七成以上，將對中醫典籍整理者減輕許多篩選判讀的負擔。

初步，本研究僅精選金元四大家之醫學論著有關河間六書部份作為電腦程序之測試，本程序可以輕易轉移至其他醫家的著作中整理出更多有效用的方劑專論。中文全文檢索國人已有眾多自行研發的系統，而中醫藥自動化全文資料擷取的工作，談論得非常少。由於電子科技日新月異，國外研究中醫中藥的機構日益眾多，加速中醫藥資料庫建造是十分迫切的課題，因此電腦輔助整理中醫藥古籍文獻若果有更多研究專題投入，將有助於中醫藥之整理及推廣。現今電腦軟、硬體、網際網路之普及、通訊費率之平價，有助於中醫古籍網際化的建構。雖然物美價廉但架設網站、整理古籍資料、系統維護、程式開發都是環環相扣，輔助電子中醫藥古籍文獻整理的目標擬定後，編寫程序以完成任務，本研究初步成效不錯，希望日後可開發更多文獻自動擷取相關技術，為中醫資訊化盡點微薄貢獻。本研究已成功的將擷取的 512 首河間六書方劑整理到 TCMET<sup>1</sup> 網際網路動態網頁資料庫中。

## 參考資料

1. 陳逸光，網際網路中醫藥典籍文獻動態資料庫研究，中醫藥雜誌 11(1):43-50，2000。
2. 林晰，中文全文檢索系統的標誌介紹，中央研究院計算中心通訊 13(9):96-98，1997。
3. 陳逸光，中醫古籍文獻資訊化研究，國際中醫藥資訊典籍學術研討會論文集（台中），pp. 195-204，1997。
4. (清)汪詡庵，本草備要，文光圖書有限公司，台北，pp.1-237，1976。
5. Yester M. Reference. In:Using Turbo Pascal. QUE Corporation, Indianapolis, p. 648, 1989.
6. Martin J. Entities and Attributes. In: Computer Database Organization. Prentice-Hall Inc., N.J., pp. 57-59, 1977.
7. 黃慕萱，檢索系統評估之發展--理論與實務，中國圖書館學會會報 59:109-126，1997。
8. Losee R. Measuring Performance. In:Text Retrieval and Filtering Analytic Models of Performance. The Kluwer International Series on Information Retrieval, pp. 78-92, 1998.
9. 曾元顯，關鍵詞自動擷取技術與相關詞回饋，中國圖書館學會會報 59:62，1997。
10. 林偉業、Harizan A、潘陳文汝，中文搜尋引擎簡介，中國圖書館學會會報 61:127-133，1998。
11. 林則孟，引進電子期刊、索引及全文資料庫的建置計劃構想與問題，中華民國科技館際合作協會通訊 4:28-31，1997。
12. 簡立峰，中英文全文檢索技術及應用簡介，國立成功大學圖書館通訊 12:1-12，1993。
13. 陳逸光、方榮瑞，中醫文獻電腦資訊化研究 - 黃帝內經，行政院衛生署中醫藥年報 9:344-364，1991。
14. 黃雲龍，中文全文資訊檢索研究架構與重要議題探討，大學圖書館 2:(3)4-26，1998。
15. 同文獻 7，pp. 60-61。

## 附錄 ( 河間六書 藥物 辭典庫 共 229 項 )

丁香	代赭石	地膽	防風	韭子	馬練花	紫苑	銅青	鬱金
人參	仙靈脾	朴硝	乳香	香附	骨碎補	菊花	鳳眼草	芎藭
大豆孽	半夏	肉豆蔻	使君子	香薷	乾地黃	菟絲子	樟柳根	芫花
大戟	玄胡索	自然銅	官桂	射干	乾葛	蛤粉	熟地黃	枳殼
大棗	玄參	艾葉	知母	桂心	乾漆	蛤蚧	穀精草	枳實
大椒	瓜蒂	血竭	羌活	桂枝	乾薑	訶子	蔓荊子	血餘灰
大黃	瓜蒌	佛耳草	金鈴子	桔梗	御米穀	陽起石	蓬莪朮	砒霜
山茱萸	甘草	何首烏	阿膠	桑白皮	梔子	雄黃	豬苓	枯萇根
山梔	甘菊	吳茱萸	阿魏	柴胡	牽牛	黃柏	橘皮	枯萇實
山藥	甘遂	杏仁	附子	桃仁	硃砂	黃耆	澤瀉	茯苓
川烏	生地	杜仲	青皮	海桐皮	細辛	黃連	獨活	茯神
川椒	生薑	沙參	青黛	海馬	羚羊角	黃精	膩粉	菖蒲
川芎	白朮	沈香	青鹽	海帶	荷葉	黃芩	蕪荑	草薺
井泉石	白芍	沒藥	前胡	海蛤	貫眾	滑石	龍骨	葶藶
升麻	白芷	牡蠣	厚朴	海藻	連翹	當歸	龍腦	葳靈仙
天門冬	白僵蠶	皂角	威靈仙	烏梅	陳皮	礞砂	龍膽	菴蓉
天南星	白檀	良薑	染槐子	烏魚骨	麥門冬	粳米	檀香	蝟皮
天麻	白礬	芒硝	枸杞子	破故紙	麥柏	腦子	縮砂	薏苡仁
巴豆	白芨	芍藥	柏子仁	神麴	麻黃	葛花	薄荷	藜蘆
巴戟	白斂	豆豉	盆硝	秦艽	寒水石	葛根	檳榔	藿香
木瓜	石斛	貝母	穿山甲	粉霜	揀桂	蜈蚣	燒錦灰	五味子
木香	石榴皮	赤小豆	紅皮	荊芥	斑蝥	蒼朮	蘆薈	
木通	石膏	赤石脂	紅花	草豆蔻	椒目	輕粉	續隨子	
牛蒡子	地骨皮	赤芍	胡桃	草烏	款冬花	辣桂	續斷	
牛黃	地榆	車前子	苦參	茵陳	犀角	遠志	麝香	
牛膝	地龍	防己	郁李仁	茴香	琥珀	酸棗仁	鱉甲	



J Chin Med 12(1): 1-9, 2001

# COMPUTER-AIDED RETRIEVAL OF PHARMACOLOGICAL INFORMATION FROM THE SIX BOOKS OF HEJIAN

Yee-Guang Chen

*Department of Medicine, Tzu Chi University,  
Hualien, Taiwan*

*(Received 18<sup>th</sup> May 2000, revised Ms 4<sup>th</sup> December 2000, accepted 6<sup>th</sup> December 2000)*

Information can be retrieved from Traditional Chinese Medicinal Literature (TCML) with the aid of the computer and considerable savings in time and money. The Chinese full text information retrieval technology is well developed and already applied to the TCML. In this study, the electronic text of *Heijian-Liu-Shu* (The Six Books of Hejian) was chosen as a pilot project. This book in form of database contents totally 190431 Chinese characters, the computer program matched 203 medicinals in the *Ben-Cao-Bei-Yao*, and finally 229 medicinals were filtered from the *Heijian-Liu-Shu* (including the above 203). The filtered medicinals database was used to extract ingredients from the *Heijian-Liu-Shu* formulas automatically. The results show that 603 ingredient groupings were indexed and the precision ratio was 82% and the recall ratio 95%. The efficiency gave us confidence in the area of computer-aided TCML information retrieval. Finally, quality control of 512 formulas was completed. These recipes have already been installed on the TCMET world-wide-web. All the computer algorithm, flows, procedures and the medicinals database which developed in this paper can be transposed to any other similar TCML information retrieval project flexibly.

**Key words:** Traditional Chinese medicinal literature, Herbal name computer-aided retrieval, Herbal name database, WWW.

---

**Correspondence to :** Yee-Guang Chen, School of Medicine, Tzu Chi University, 701,Section 3, Chung Yang Road, Hualien, Taiwan. Tel: (03)8565301 ext 7207