

以近似字詞比對方式輔助完成中醫濕證症狀標準化

李宜泓¹ 陳仁義² 葉家舟^{3,4} 林迺衡^{1*} 葉明憲^{3*}

中醫擁有眾多診斷和治療疾病的經典古籍，是一個龐大且有價值的知識寶庫。診斷疾病時，症狀是非常重要的參考因素。由於各家經典古籍記載的症狀用詞不同，且相同的症狀其定義也有差異，我們稱其是症狀尚未標準化。症狀尚未標準化是使得後人欲進行系統性的科學研究，遭遇龐大的困難之一。

本論文以電腦科學領域發展已久的完整和近似字詞比對方法，輔以資料庫的案例式推論 (Case-based reasoning) 學習機制和中醫師諮詢的三方協同合作，已整理《中醫證候學》(15)濕證門記載的大部分症狀，完成其中濕證門的濕性特徵的症狀之症狀標準化，並且產生部分的標準化症狀清單。

標準化症狀的清單將可用於製作臨床看診的問診表，用以探索疾病作用於身體的路徑，以及持續標準化各家經典古籍的症狀，使資料更為豐富。

關鍵詞：中醫，症狀識別，症狀標準化，近似字詞比對，案例式推理

台灣中醫家庭醫學雜誌 2012；7(2)：25-45

¹ 國立中正大學資訊工程學系

² 南華大學資訊管理學系

³ 佛教慈濟綜合醫院大林分院中醫科

⁴ 慈濟大學醫學系

共同通訊作者：林迺衡，國立中正大學資訊工程學系，嘉義縣民雄鄉大學路 168 號，電話：(05)2720411 ~ 33110，電子郵件信箱: naiwei@cs.ccu.edu.tw；葉明憲，佛教慈濟綜合醫院大林分院中醫科，嘉義縣大林鎮民生路二號，電話：(05)2648000-5542，電子郵件信箱: yehlinlo@gmail.com

前言

中醫證候規範化的研究是近年來中醫研究的熱點領域(1)。經由以“證候”查詢中醫期刊網醫學分部發現近年來有爆炸性的成長(如圖一。中國期刊網)。中醫證候是由中醫四診:望,聞,問,切所得之病患之症狀訊息,依據司外揣內的方法,所推演出之預測人體內在變化的理論。(3)。

中醫最核心的精華,不管是辨證,用藥,用針,按蹻導引,藥膳調理皆是依循這個理論在進行。然而中醫歷史悠久,發展過程中融合了多個學科,內容十分廣博。中醫各家各派眾多,不同派別、醫家對中醫學的認識不完全一致。中國文字豐富多彩,對於中醫學的描述也複雜多樣。這些方面都在一定程度上造成了辨證最重要的“症狀”的不一致,從而導致辨證的結果會有出入。

症狀標準化即是將多詞一義的中醫症狀同義字,用統一的症狀名詞為

代表,並針對這個名詞予以詳細的定義及分析,則能夠解決同一意義的不同症狀之形容詞所造成的困擾。而眾多的同義字其實就是一種近似字詞。

在定義了這麼多同義字前,我們也曾用人工方式做症狀的標準化(12),但是會淪於對同一義的名詞定義因工作時間長久而有前後名詞不同的情形。

所以若能夠使用電腦軟體協助解決這些情況,則將更能快速地完成症狀標準化的工作,而且不管以後是否要加入其他醫典,或是重新定義標準化後的症狀,皆能使所有的症狀資料一致。最重要的想法是未來的研究人員,能夠根據這些標準化的症狀,應用電腦的數據分析能力來探討各種中醫疾病的歸納,傳變及進行重要的中醫生理及病理機轉的研究。

張維波等人(7)和張永濤等人(10)認為醫師的直覺判斷結合資訊科技和數據分析,能使看診更為精準。但前

提是輸入的中醫診斷資料需客觀化和定量化。

張啟明、王永炎等人(9)於 2005 年時以中國高等中醫藥院校全國統編教材為基礎，收錄全部的醫案內容，利用《中醫藥學名詞》字典確立症狀的專業術語，對張維波等人的問題進行研究，提出一套建立歷代醫案數據庫的方法，實驗中應用統計辨識一個證候的常見症狀。同時依此發表了一系列感覺症狀(8)和月經類症狀(13)等十餘篇分辨疾病症狀術語的論文。這些論文讓研究者和使用者能明確地瞭解症狀的意義。鄭紅等人(2)在 2010 年提出症狀術語結合訓詁學和語言學的概念，可用以分析不同年代的醫案結構，因此可以從語言的進展中，得知中醫歷史的演變。

柯廷諭(11)則從自然語言的觀點，利用電腦軟體輔助進行中醫的症狀標準化。他指出以完整字詞比對的概念做成的標準化症狀，無法對未曾

輸入過的症狀進行標準化。同時以人工進行的症狀標準化工作，常會造成一詞多義的缺失。

本篇則是運用完整字詞比對(23、25)及近似字詞比對方法(18、21)，結合資料庫的案例式推論學習機制(22、27)，對中醫濕證的症狀進行分析。

【材料與方法】

我們使用 a.)電腦科學領域發展已久的近似字詞比對方法找到字面相似的詞彙、b.)案例式推論方法(26、27)於資料庫軟體以記錄正確和錯誤比對的詞彙供程式學習，這二種機制來完成症狀標準化。

一．完整字詞比對概念

字詞比對 (string pattern matching) 是電腦科學一直存在的問題，也是一個經典的問題(22、23、24)。完整字詞比對是檢查兩字詞是否完全相同，著名的演算法有 Boyer-Moore 以及 Knuth-Morris-Pratt 等其他方法(23、25、26)，現已存在於各項應用當中。

字詞比對的研究議題一開始以完整字詞比對為主，後續完整字詞和近似字詞比對同時都有人思考解決，直到 80-90 年代，電腦系統才陸續有快速、品質好的近似字詞比對通用解決演算法(19-21)。

二．近似字詞比對概念

假設比對時樣本(pattern) 的長度要小於或等於正文(text)，由樣本比對正文，使用萊溫斯汀差距(Levenshtein distance)衡量樣本和正文的字詞差距(22)，萊溫斯汀差距可以涵括字詞距離的新增(insertion)、刪除(deletion)和取代(substitution) (19)。每個字詞差距 t 都代表著樣本(pattern)要改變成正文(text)，最少需要 t 個編輯運算(operation)。

舉例而言，「天氣熱，很煩」是樣本，「今天天氣真的很好」是正文，用萊溫斯汀差距衡量兩字詞差異，情況如表一所示，此兩字詞的差異是 5。因為樣本要改成跟正文相

同，它必須在欲比對的第一個天字之前新增一個「今」字，第一個天字之後新增一個「天」字，「熱，」兩字須取代成「真的」，「煩」須取代成「好」，所需最少數量的編輯運算總共是 5。

三．資料庫軟體的案例式推論學習機制

俗語說：「前事不忘，後事之師。」我們使用資料庫當作儲存資料和查詢資料的媒介，是因為它的查詢速度遠較人腦回想還快。只要資料庫內有正確標準化的症狀和錯誤的字詞資料，軟體可從這些資料，知道將要標準化的文本內容(text)哪些是要被標準化的，哪些是不用被標準化的錯誤字詞。而且當資料庫的內容愈來愈完整，症狀標準化的一致性就會愈來愈可靠，涵蓋的古籍資料也會愈來愈豐富。

資料庫紀錄的正確標準化症狀，是以一種我們規範好的格式呈現，它

會對應於一群由人工判定對應正確的文本原始症狀。這些文本原始症狀會用來指示近似字詞比對程式找到近似字詞的資料後，將近似字詞的資料對應成正確的標準化症狀，我們形容這些對應關係是一張對應表，稱之為「正確標準化症狀對應表」。

資料庫紀錄的錯誤字詞資料，是以文本原始字詞資料呈現，用途是當近似字詞比對程式從文本資料(text)找到與之完全相同的字詞，則忽略此筆從文本資料找到的字詞，使之不對應於任何正確的標準化症狀，我們稱之為「錯誤字詞清單」。「錯誤字詞清單」與「正確標準化症狀對應表」的原始症狀的交集必須是空集合。

四．系統流程

我們的症狀標準化系統的流程(圖二)是先用「完整字詞(23、25、26)」比對文本資料，擷取目前資料庫已有的症狀，(圖三)再用「近似字詞(19、

21、22)」方式比對文本資料，探索目前資料庫沒有的症狀(圖四)。

第一階段：完整字詞比對，請見圖三。

第一步：讀入「正確標準化症狀對應表」。

第二步：使用此對應表的所有原始症狀，用「完整字詞」比對方法，比對文本資料，以擷取文本資料中「完全」和資料庫紀錄相符的原始症狀。

第二階段：近似字詞比對，請見圖四。

第一步：讀入「正確標準化症狀對應表」。

第二步：使用此對應表的所有原始症狀，用「近似字詞」比對方法，比對文本資料以擷取「可能」是同義異字的字詞，將其列入近似詞清單。

系統比對文本資料的方法是，依序從每筆原始症狀的每一個字，在文本內容中出現的位置定位，再分別向

前和向後讀入字詞。讀入的過程中如果字詞誤差值呈現嚴格遞增曲線 (increasing curve)，則換讀入另一方向，直到另一方向的字詞誤差值也呈現嚴格遞增曲線，就停止讀入文本字詞。然後檢查這個讀入的文本字詞，是否在字詞誤差值 t 以下，如果是就列入近似詞清單，如果不是則不列入近似詞清單。

第三步：依照資料庫記錄的「錯誤字詞資料」，刪除近似詞清單中與之「完全相符」的近似詞。

第四步：依照資料庫的「正確標準化症狀對應表」，計算近似詞清單的每一個近似詞和此對應表所有原始症狀的萊溫斯汀差距 t ，選擇 t 最小的原始症狀伴隨的標準化症狀。

第五步：對於每個近似詞，系統列出四個選項，讓使用者觀看並擇一點選：第一個是「症狀標準化正確」；第二個是「近似詞是症狀，但沒有可選的標準化症狀」；第三個是

「近似詞不是症狀，是錯誤的字詞資料」；第四個是「無法判斷，留待詢問合作醫師」。

第一個選項是讓使用者確認，該近似字詞要被標準化哪個標準化症狀，會把此筆近似字詞的症狀標準化紀錄在「正確標準化症狀對應表」，以供系統比對其他文本資料時使用；第二個選項的用意是雖然這次沒有符合的標準化症狀，讓下次近似字詞比對程式比對時，仍把該近似詞列入考慮清單；第三個選項是將該近似詞列入資料庫的「錯誤字詞清單」，當作排除清單，當下次近似字詞比對程式比對時，會不把此近似詞列入近似詞清單；第四個選項是當無法決定該近似詞歸屬哪個選項時，才使用第四個選項，留待詢問醫師或是留待下次執行近似字詞比對時找到適合的標準化症狀。

第四個選項的用意很類似第二個選項，但是相較之下，第四個是不確

定該近似詞意義才選擇，第二個是在非常確定的情況下才選擇。第一和第二個選項，與第三個選項的交集必須是空集合。

實驗結果

由於此系統使用完整字詞和近似字詞比對文本的資料時，是以資料庫的內容為基礎，所以在系統開始比對之前，須填入一些資料到「正確標準化症狀對應表」和「錯誤字詞清單」，填入資料的原則，則依前述的「資料庫軟體的案例式推理學習機制」所示。

此次欲比對的文本資料是《中醫證候學》(15)第一章第四節第一段的辨證求因區提到濕證門的濕性特徵之症狀，如表二所示。

系統自動識別症狀後的文本內容，如表三所示。井字號(#)是代表被系統識別的字詞。

系統自動識別症狀所擷取的文本症狀，如表四所示。

系統識別得到的字詞有 86 種，其中 75 種是經由近似字詞比對方式識別的，我們定義近似字詞比對方法的探索率是，近似字詞比對方法識別的字詞除以近似字詞和完整字詞比對方式識別的字詞種類總和，所以探索率是 87.21%。

討論

一. 「文本症狀」欄是症狀識別系統從輸入的文本資料識別出的字詞，系統認定這些字詞有可能是症狀，這些項目需要人工確認是不是和「文本症狀」欄的字詞相同意義。「樣本症狀」欄是已經記載在資料庫的症狀，這欄症狀被挑選的原因是系統判斷它和「文本症狀」欄的字詞最為接近。

描述接近程度的數值是呈現在「字詞誤差值」欄，有 0、1、2、3..... 等等整數，因為此實驗的系統識別濕證門文本症狀之限制條件是「尋找字詞誤差值在 2 以下的字詞，從中挑選

最接近的字詞」，所以本表格只會出現 0、1 和 2。本表格第 1 到第 11 項的項目是由完整字詞比對方法找出的文本症狀，因為這些的字詞誤差值是 0；第 12 到第 86 項是近似字詞比對方法找出的文本症狀，因為這些的字詞誤差值是 1 和 2，本系統實作上為了提高執行速度，採取分開完整字詞和近似字詞比對方法的方式處理，避免產生全部項目皆交由同一個複雜函數執行，造成執行速度緩慢的困境。

「標準症狀」欄位是人工思考「樣本症狀」欄之字詞意義，應替換成哪些標準字詞表示詞義，人工事先定義於資料庫。此表格的各欄意義如上所述。總括而言，此表格的整體意義是，系統認為「文本症狀」欄會和「樣本症狀」欄的字詞非常接近，而被標準化成與「樣本症狀」相同意義的「標準症狀」。

人工比較由近似字詞比對方法識別的字詞，從第 12 項到第 86 項的

「文本症狀」欄和「樣本症狀」欄之意義：

(1)「文本症狀」第 23, 24, 38, 72, 78 項是症狀，而且和其「樣本症狀」是同義詞。

(2)第 13, 14, 15, 16, 17, 19, 22, 27, 28, 29, 32, 35, 36, 39, 40, 41, 42, 44, 51, 57, 59, 60, 68, 69, 71, 73, 76, 79, 82, 83 項的「文本症狀」看得出是症狀字詞的一部份，可視為是症狀的一部份描述，但詞義不全。

(3)第 12, 18, 20, 33, 34, 37, 43, 49, 52, 53, 66, 67, 86 項是症狀，但和其「樣本症狀」詞義不同。

(4)第 50, 55 項是症狀，其「文本症狀」和「樣本症狀」之詞義的內容，部份相似。

(5)第 25, 26, 30, 31, 45, 46, 47, 48, 54, 56, 58, 61, 62, 63, 64, 65, 74, 75, 77, 80, 81, 84, 85 項是語助詞、描述病因之字詞，或是非症狀字詞的情況。

二．本論文的實驗環境是使用李宜泓等人(4)提出的實驗環境參數，也就是：

- (1)設定字詞誤差值在 2 以下，
- (2)「錯誤字詞清單」資料庫是空的，
- (3)使用《中醫證候學》的每一章內每一節每一證中的基本症狀，目前資料庫暫時不列入舌象和脈象。

本論文使用的近似字詞比對方法，相較於一般使用的完整字詞比對方法，能夠從文本資料中找到資料庫沒有記載的症狀，並透過人工確認和回饋內容給資料庫的機制，進而豐富資料庫記載的內容。本篇論文的實驗結果第 23, 24, 38, 72, 78 項，即說明我們有機會找到資料庫沒有記載的字詞，並且確認它們是同義詞，使得資料庫更為豐富。

三．使用完整字詞比對加上近似字詞比對的方法，輔助做症狀標準化，可以探索資料庫沒記載到的原始

症狀，進而有機會將它標記成標準化症狀。雖然使用正則表示式 (Regular expression) 也可以探索資料庫沒記載到的原始症狀，但是我們使用完整和近似字詞比對的優點是，可以在限制的字詞誤差值條件 t 下，系統能夠多一個自訂的限制條件，以限制系統擷取文本症狀的長度 n ，最多在資料庫的某筆原始症狀長度 m 的 $m+t$ 以下，也就是 $n \leq m+t$ 。而非如正則表示式一般，須要窮舉字詞中的星號和問號用法的排列組合，並且檢驗這些排列組合之下的兩字詞是否相同。使用近似字詞比對方式的好處，是正則表示式的星號和問號無法輕易辦到的。

結論

我們使用完整字詞和近似字詞比對方法，並且結合案例式推理學習機制和中醫師的協助，來輔助完成中醫症狀標準化。我們使用《中醫證候學》第一章第九節第一段的辨證求因之濕證門濕性特徵的症狀，當做此方

法的實驗材料，並且認為此方法的優點已超越正則表示式能做到的，同時會隨著資料庫內容增長，症狀識別的準確性而愈來愈高，人工介入的次數會愈來愈少，進而趨近於自動化且準確性高的症狀標準化境界。

誌謝

本論文由財團法人佛教慈濟醫院大林分院院內計畫 DTCRD-100-I-14「建構中醫診斷資料庫系列二：中醫症狀標準化之分析與應用」支持。特別感謝吳昇及施承泰提供許多字詞比對的專業指導。

參考文獻

1. 戴霞，郭偉星。中醫證候診斷標準規範化研究概況。中醫雜誌。2011;52(2):168-171。
2. 鄭紅、張啟明，“中醫症狀術語規範化研究方法探討。”山東中醫藥大學學報，Vol. 34 (1): pp. 21-22, 2010.
3. 杜青雄，黃碧萍(2010). "中醫證候量化的探討." 中華中醫藥雜誌增刊
4. 李宜泓、林迺衛、葉明憲、葉家舟、陳仁義， “以近似字詞比對方法輔助完成中醫症狀標準化”，全國計算機會議 Artificial, Fuzzy and U-learning 論文，2011 年。
5. 焦鋸、石濤，“《傷寒論》問診成就初探。”天津中醫藥，Vol. 26 (1): pp. 43-44, 2009.
6. 陳濤、陳茂華、鞏仔鵬、梁嶸、鄧李蓉、胡月琴，“基於互聯網的中醫病案 OLAP 系統的開發與應用。”安徽中醫學院學報，Vol. 29 (5): pp. 67-70, 2010.
7. 張維波，“中醫發展的數學之路。”中國醫學影像技術，Vol. 19 (204): pp. 77-80, 2003.
8. 張啟明、王義國、劉保延、王永炎，“中醫症狀中的感覺症狀。”世界科學技術，Vol. 11 (4): pp. 512-515, 2009.

- 9 . 張啟明、王永炎、張慶祥、宋觀禮, “中醫歷代醫案數據庫的建立與統計方法.” 山東中醫藥大學學報, Vol. 29 (4): pp. 298-299, 2005.
- 10 . 張永濤、梁嶸, “關於研製中醫數字問診系統的構想.” 中國醫學影像技術, Vol. 19 (204): pp. 138-139, 2003.
- 11 . 柯廷諭, “中醫症狀的結構化分析與標準化系統設計”, 國立中正大學資訊工程研究所, 碩士論文, 2011
- 12 . 林劉育, “中醫症狀的標準化初探與應用”, 國立中正大學資訊工程研究所, 碩士論文, 2011
- 13 . 李毓秋、趙宜軍、王義國、李檬、張啟明, “中醫月經類症狀單元的提取.” 遼寧中醫雜誌, Vol. 37 (1): pp. 57-58, 2010.
- 14 . 李培生, 《傷寒論》. Vol. 1. 1990: 知音出版社.
- 15 . 李洪成、李新平、李新曄, 《中醫證候學》, 2008: 中國醫藥科技出版社.
- 16 . 李兵、劉國正、符永馳、裘儉、張偉娜, “從中醫古籍數據庫建設看中醫古籍數字化.” 中國中醫藥信息雜誌, Vol. 16 (3), 2009.
- 17 . 王憶勤、李福風、湯偉昌、何建成、劉國萍等, “中醫四診信息採集與分析方法探討.” 中華中醫藥雜誌, Vol. 24 (11): pp. 1397-1405, 2009.
- 18 . Wu, S. and U. Manber, “Fast text searching: allowing errors.” Commun. ACM, Vol. 35 (10): pp. 83-91, 1992.
- 19 . Wu, S. and U. Manber. “Agrep -- a fast approximate pattern-matching tool.” in Proceedings USENIX Winter 1992 Technical Conference, pp.153-162, 1992.
- 20 . Ukkonen, “Finding approximate patterns in strings.” Journal of

- Algorithm, Vol. 6, pp.132-137, 1983.
- 21 . Myers, E. and W. Miller, “Approximate matching of regular expressions.” Bulletin of Mathematical Biology, Vol. 51 (1): pp. 5-37, 1989.
- 22 . Levenshtein, V.I., “Binary codes capable of correcting deletions, insertions and revresals.” Soviet physics, Vol. 10 (8): pp. 707-710, 1965.
- 23 . Knuth, D.E., J. James H. Morris, and V.R. Pratt, “Fast Pattern Matching in Strings.” SIAM Journal on Computing, Vol. 6 (2): pp. 323-350, 1977.
- 24 . Janet L. Kolodner, “An introduction to Case-Based Reasoning”, Artificial Intelligence Review 6(1): pp. 3-34, 1992.
- 25 . Boyer, R.S. and J.S. Moore, “A fast string searching algorithm.” Commun. ACM, Vol. 20 (10): pp. 762-772, 1977.
- 26 . Baeza-Yates, R. and G.H. Gonnet, “A new approach to text searching.” Commun. ACM, Vol. 35 (10): pp. 74-82, 1992.
- 27 . A. Aamodt and E. Plaza, “Case-based reasoning: foundational issues, Methodological Variations, and System Approaches”. AI communications. IOS press, Vol. 7: 1, pp. 39-59.

表一、萊溫斯汀差距 $t = 5$

正文	<u>今天天氣真的很好</u>
樣本	天 氣熱，很煩

表二、濕證門濕性特徵的症狀

濕證門	重著	陰濕之邪鬱滯，陽氣不得伸展，重著之感。頭重如蒙如裹，身重或疼痛，不能轉側，四肢重著沉困，或關節重著腫痛
	困倦	頭目昏蒙，肢體倦怠，神疲困頓，軟弱無力，懶言少食，痿弱不振
	腫脹	陰濕鬱滯，津氣不行，致浮腫或脹滿，足腫面浮，肢腫膚脹，關節腫脹，脘腹痞滿脹，腹脹如鼓
	蘊熱	濕鬱陽氣不能透達於外，而見蘊熱不揚，或熱甚不能自覺，或熱盛反覺舒適
	惡寒	常惡寒喜暖，寒凜，四肢冷，欲得覆被向火
	頭汗	濕邪鬱遏，陽氣內蒸，津液與濕氣上達頭面，常以頭面汗出齊頸而還，但頭汗出
	瀉利	濕盛則濡瀉，濕困脾陽，泄利或溏瀉
	小便不利	濕痺，小便不利，大便反快，
	濕流於下	濕行下受，或濕流於下，以下部之病多濕，下肢多濕，痛，痺，瘡瘍，斑疹
	聲濁	聲如從室中言，中氣之濕。
	發病緩慢	濕性黏膩，發病緩慢，病情亦多延綿，病程長，不易速解
	色脈舌特徵	濕家，身黃，白睛發黃，周身皮膚發黃，舌苔厚濁，濕潤滑膩，脈緩，濡，弦象，濕鬱，陽氣不能鼓動
	時令特徵	天時長夏之際，盛熱蒸動，濕氣流行。陰雨加劇，日晡為甚

表三、濕證門濕性特徵被識別後的文本

濕證門	重著	[77###][61###]，陽[25###]伸[74###]著之感。[12##][73##]如裹，[7##]或[16##]，[26##][79##]，[18####][38##]，或[34####][17##]
	困倦	[13##][72##]，[37###]怠，[11##]困頓，[19####]，[50##][4##]，痿[27###]
	腫脹	陰[62###]，津[28###]，至[10##][42##]滿，[52##][66##]，[3##][43##]，[33####]，[41##][21###]，[5##]如鼓
	蘊熱	[63##][45##][26##]通達於外，而見蘊[29###]，[44##][30##][22###]，[44##]盛[86###]適
	惡寒	常[9##][49##]，[76##]，四[1##]，[56###]被向火
	頭汗	濕[64###]，[45##][70##]，津液與[46###][14###]，常[15####][8##]齊頸而還，[2####]
	瀉利	濕腥則[78##]，[84##][75##]，[23##][24##]瀉
	小便不利	濕痺，[6####]，[20####]，
	濕流於下	濕[81###]，[51##]流[82##]，[83###]之[68##]濕，[40####][54###]，痺，[55####]疹
	聲濁	[32##]從[85###]，[47###]濕。
	發病緩慢	濕[71##]膩，[57##]緩慢，病情[69##]延綿，病程長，[31##]速解
	色脈舌特徵	濕家，[67##]，白[58###]，[39####][59##]，[53###]濁，濕潤滑膩，脈緩，濡，弦象，[63##]，[45##][26##][35##]
	時令特徵	[65###]夏之際，盛[36###]，[48###]行。[80###]劇，[60####]

表四. 實驗結果說明

編號	文本症狀	樣本症狀	字詞誤差值	標準症狀
1	肢冷	肢冷	0	冷的部位:四肢
2	但頭汗出	但頭汗出	0	汗出的部位:頭部
3	肢腫	肢腫	0	腫脹的部位:肢體
4	少食	少食	0	胃口狀態:不佳, 食量:很少
5	腹脹	腹脹	0	脹的部位:腹部

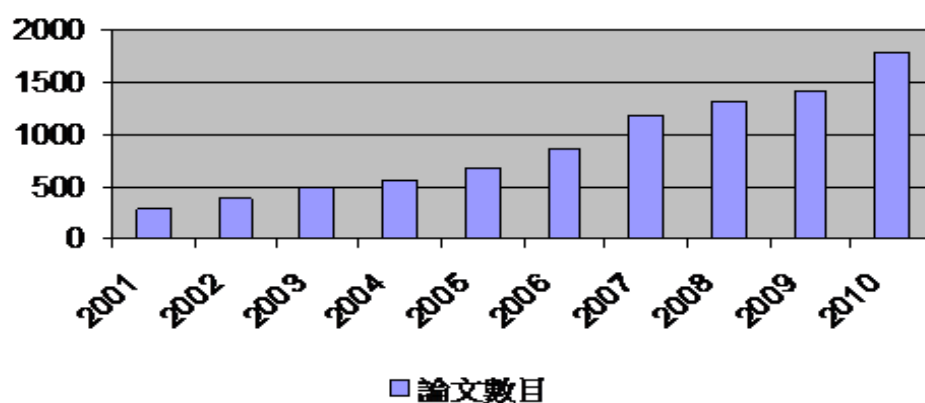
6	小便不利	小便不利	0	小便不利
7	身重	身重	0	重著的部位:全身
8	汗出	汗出	0	汗出的情形:有流汗
9	惡寒	惡寒	0	惡寒
10	浮腫	浮腫	0	腫脹的部位:全身
11	神疲	神疲	0	精神狀態:倦怠
12	頭重	頭痛	1	痛的部位:頭部
13	頭目	頭痛	1	痛的部位:頭部
14	達頭面	頭痛	2	痛的部位:頭部
15	以頭面	頭痛	2	痛的部位:頭部
16	疼痛	頭痛	1	痛的部位:頭部
17	腫痛	頭痛	1	痛的部位:頭部
18	四肢重著	四肢無力	2	肢體的力量:全身無力, 肢體力量:倦怠無力
19	軟弱無力	四肢無力	2	肢體的力量:全身無力, 肢體力量:倦怠無力
20	大便反快	大便燥結	2	大便的質:乾燥, 大便的質:乾燥而結
21	痞滿脹	痞滿燥實	2	大便的質:乾燥, 氣梗塞而痛:心下
22	能自覺	自利	2	大便的質:稀
23	泄利	自利	1	大便的質:稀
24	或溏	便溏	1	大便的質:稀
25	氣不得	利不止	2	大便的質:稀, 大便頻率:多, 疾病發作時間:很頻
26	不能	利不止	2	大便的質:稀, 大便頻率:多, 疾病發作時間:很頻
27	弱不振	利不止	2	大便的質:稀, 大便頻率:多, 疾病發作時間:很頻
28	氣不行	利不止	2	大便的質:稀, 大便頻率:多, 疾病發作時間:很頻
29	熱不揚	利不止	2	大便的質:稀, 大便頻率:多, 疾病發作時間:很頻
30	甚不	利不止	2	大便的質:稀, 大便頻率:多, 疾病發作時間:很頻

31	不易	利不止	2	大便的質:稀, 大便頻率:多, 疾病發作時間:很頻
32	聲如	聲低	1	語言聲音:低無力
33	關節腫脹	關節酸疼	2	酸痛部位:關節
34	關節重著	關節酸疼	2	酸痛部位:關節
35	鼓動	指顫動	2	抖動的部位:手指頭
36	熱蒸動	指顫動	2	抖動的部位:手指頭
37	肢體倦	體痛	2	痛的部位:身體
38	沉困	昏沉困倦	2	肢體的力量:全身無力, 肢體力量:倦怠無力, 精神狀態:昏沉
39	周身皮膚	周身無力	2	肢體的力量:全身無力
40	下肢多	肢倦	2	肢體的力量:全身無力, 肢體力量:無力
41	腕腹	腕脹	1	脹滿的部位:腕部
42	或脹	腕脹	1	脹滿的部位:腕部
43	膚脹	腕脹	1	脹滿的部位:腕部
44	或熱	潮熱	1	熱的狀態:潮熱, 熱的狀態:發熱
45	陽氣	氣喘	2	氣喘
46	濕氣上	氣喘	2	氣喘
47	中氣之	氣喘	2	氣喘
48	濕氣流	氣喘	2	氣喘
49	喜暖	腹冷喜暖	2	冷的部位:腹部, 疾病減緩因子:暖熨
50	懶言	氣短懶言	2	言語情形:懶得說話, 呼吸模式:短促
51	或濕	或悸	1	心悸
52	足腫	唇腫	1	腫大部位:唇部
53	舌苔厚	舌根痛	2	痛的部位:舌根
54	濕, 痛	舌根痛	2	痛的部位:舌根
55	瘡瘍斑	瘡瘍腫毒	2	皮膚的狀態:腫毒, 皮膚的狀態:瘡瘍
56	欲得覆	不得臥	2	睡眠的狀態:無法臥下
57	發病	發狂	1	行為動作:發狂

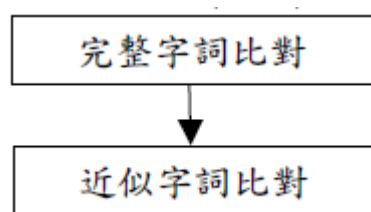
58	睛發黃	發狂	2	行為動作:發狂
59	發黃	發狂	1	行為動作:發狂
60	日晡為甚	日晡發熱	2	熱的狀態:發熱, 疾病發作時間:午後。
61	邪鬱滯	鬱怒	2	情緒:易怒然無法發作
62	濕鬱滯	鬱怒	2	情緒:易怒然無法發作
63	濕鬱	鬱怒	2	情緒:易怒然無法發作
64	邪鬱遏	鬱怒	2	情緒:易怒然無法發作
65	天時長	腹時痛	2	疾病發作時間:很頻繁, 痛的部位:腹部
66	面浮	面黃	1	面色:黃
67	身黃	面黃	1	面色:黃
68	病多	痰多	1	痰的量:多
69	亦多	痰多	1	痰的量:多
70	內蒸	骨蒸	1	熱的部位:骨頭, 熱的程度:蒸熱
71	性黏	痰黏	1	痰的質:濃稠
72	昏蒙	頭昏如蒙	2	脹悶的感覺:頭部
73	如蒙	頭昏如蒙	2	脹悶的感覺:頭部
74	展, 重	身疼重	2	重著的部位:全身, 痛的部位:全身
75	脾陽	陽萎	2	性能力:陰莖無法勃起
76	寒凜	寒熱	1	發熱情形:惡寒發熱
77	陰濕之	陰濕多汗	2	汗出的部位:生殖器周圍, 汗出的量:很多
78	濡瀉	泄瀉	1	大便的情形:泄瀉, 大便頻率:泄瀉
79	轉側	轉側不利	2	肢體動作:轉側不利
80	陰雨加	陰頭寒	2	冷的部位:龜頭
81	行下受	或下蛔	2	大便的內容物:挾有蛔蟲
82	於下	或下蛔	2	大便的內容物:挾有蛔蟲
83	以下部	或下蛔	2	大便的內容物:挾有蛔蟲
84	濕困	身酸困	2	酸痛部位:全身
85	室中言	中痞	2	氣梗塞而痛:中脘

86	反覺舒	神不舒	2	精神狀態:倦怠
----	-----	-----	---	---------

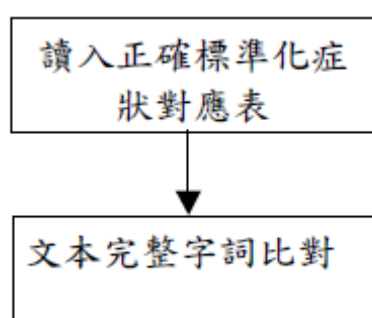
圖一 中醫證候規範的研究論文數



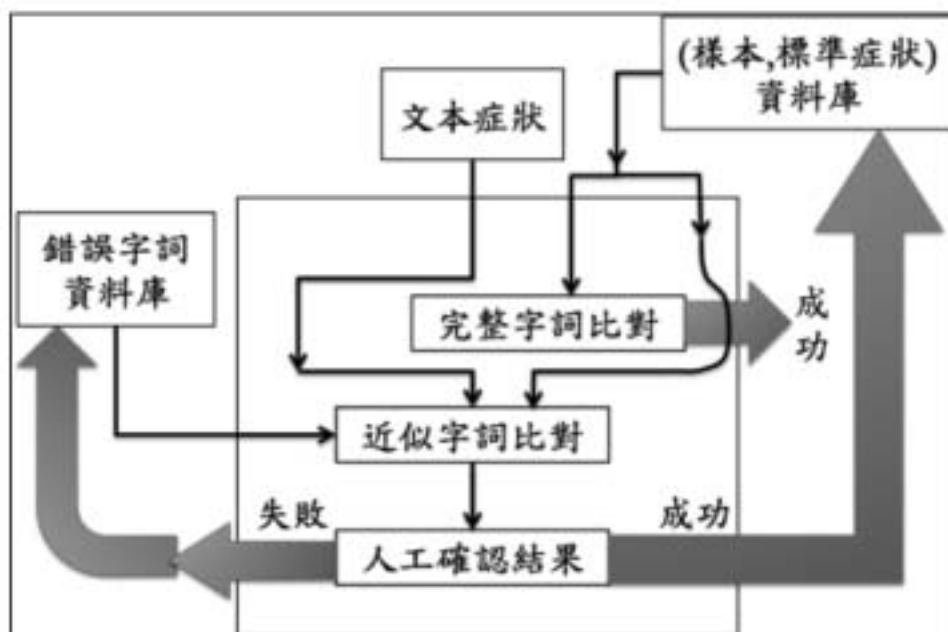
圖二、系統流程圖



圖三、完整字詞比對流程



圖四、近似字詞比對流程



Using Approximate Pattern Matching Technique to Accomplish the Damp Disease Pattern Symptoms Standardization in Traditional Chinese Medicine

Yi-Hon Lee¹, Zen-Yi Chen², Chia-Chou Yeh^{3,4}, Nai-Wei Lin^{1*}, Ming-Hsien
Yeh^{3*}

There are many classical and valuable books for diagnosis and treatments of diseases in traditional Chinese medicine. Symptoms are very important for diagnosis of diseases. However, expressions of symptoms have not been standardized in traditional Chinese medicine. The diversity of expressions of symptoms in classical books is one of the main obstacles of using information engineering to assist the research of traditional Chinese medicine.

This paper presents an approach to facilitate the standardization of symptoms in traditional Chinese medicine. This approach uses the exact pattern matching as the basis for the standardization of symptoms. It then uses the approximate pattern matching as a technique to upgrade the process of the standardization of symptoms.

We have completed a large part of the standardization of symptoms from the book of Chinese Medicine Pattern Identification, Chapter 5:Dampness Pattern. We have also performed a preliminary experiment on the effectiveness of the approximate

¹National Chung Cheng University, Department of CSIE

² Nan Hua University, Department of IM)

³ Buddhist Tzu Chi General Hospital, Da Lin Branch, Department of TCM

⁴Tzu Chi University, Department of Medicine)

*Correspondence to: Nai-Wei Lin, Department of CSIE , National Chung Cheng University, 168 University Road, Minhsiung Township, Chiayi County 62102, Taiwan, ROC; Tel:(05)2720411 ~ 33110; Email: naiwei@cs.ccu.edu.tw

Ming-Hsien Yeh, Department of TCM, Da Lin Branch of Buddhist Tzu Chi General Hospital, NO.2, Minsheng Rd., Dalin Township, Chiayi County 622, Taiwan (R.O.C.); (05)2648000-5542; Email: yehlinlo@gmail.com

pattern matching technique. The result of the experiment shows the potential benefits of the approximate pattern matching technique.

Keyword: Traditional Chinese medicine, symptom recognition, symptom standardization, approximate pattern matching, case-based reasoning

Taiwan Association of Chinese Medicine of Family journal 2012;7(2):25-45