

中醫症狀詞庫的研製

葉明憲
佛教慈濟綜合醫院
大林分院中醫科
yehlinlo@gmail.com

郭慧翔
國立中正大學
資訊工程學系
shawنشoo@hotmail.com

林迺衛
國立中正大學
資訊工程學系
naiwei@cs.ccu.edu.tw

摘要

中醫的診斷治療是以「辨證論治」的方式進行。中醫師會先根據望、聞、問、切四診收集病患的症狀及體徵資料，再根據收集到的症狀及體徵資料進行詳盡地分析，最後歸納出疾病的病因、病位、及病機，此即為「辨證」。中醫師最後根據辨證結果對病人進行適當的治療及用藥，此即為「論治」。我們希望研製一套中醫自動辨證系統，做為中醫的研究平台。然而中醫自動辨證系統的基礎在於中醫症狀標準化。

本研究建立一套中醫症狀詞庫系統，來解決中醫症狀標準化的問題。我們除了使用「症狀字詞比對」的方式進行症狀標準化，還使用「關鍵字詞比對」的方式進行症狀標準化。如果輸入的症狀已經經過標準化，並存在資料庫，系統可透過症狀字詞比對正確地查詢到其所對應的標準症狀。如果輸入的症狀尚未經過標準化，不在資料庫，系統會先將該症狀分割成一串已知的關鍵字詞，再透過關鍵字詞比對的方式，查詢這些關鍵字詞的語意。最後系統依據這些關鍵字詞的語意，查詢語意最接近的標準症狀。本研究顯示，運用語意資訊，可以明顯地提昇系統查詢標準症狀的正確率。

關鍵詞：中醫症狀詞庫、症狀標準化、語意

Abstract

The diagnosis and therapy of traditional Chinese medicine follows the approach of syndrome differentiation and treatment. For syndrome differentiation, the doctor first gathers the symptoms and signs of patients through the four examinations: inspection, listening and smelling, inquiry, and palpation. The doctor then performs a comprehensive analysis of the symptoms and signs to identify the cause, location, and mechanism of the illness. Finally, the doctor determines the treatment according to the identified syndrome. We planned to develop an automatic syndrome differentiation system as a research platform for traditional Chinese medicine. However, the foundation of an automatic syndrome differentiation system is the standardization of symptoms and signs.

This research develops a traditional Chinese medicine symptom thesaurus to solve the issue of the standardization of the symptoms and signs. In addition to string pattern matching of symptoms and signs, we also perform string pattern matching of keywords. If an inputted symptom or sign description has been standardized and is in the symptom thesaurus, the system will map this inputted symptom or sign description to its corresponding standardized symptom or sign. If an inputted symptom or sign description has not been standardized and is not in the symptom thesaurus, the system will first split the inputted symptom or sign description into a sequence of keywords. The system then performs string pattern matching of these keywords and enquires the semantics of these keywords. Finally, the system maps this inputted symptom or sign description to a standardized symptom or sign with the closest semantics. This research demonstrates that applying semantic information can significantly improve the correctness of the mapping to the standardized symptom or sign.

Keywords: traditional Chinese medicine symptom thesaurus, standardization of symptoms and signs, semantics

1. 前言

疾病在中醫被稱為「證」，所謂「證」就是由一定的「病因」作用於一定的「病位」，而發生一系列動態演變的過程[1]。

中醫先透由望、聞、問、切等四診，取得病患的身體症狀資訊，再透由分析這些症狀資訊來辨識疾病的病因及病位，這個過程稱為辨證。

我們長久以來一直希望研製一套中醫自動辨證系統，做為中醫研究的平台。然而，當我們開始著手研製中醫自動辨證系統時，首先遇到了中醫症狀尚未標準化的困難。

目前的古籍資料對於症狀的描述及其意義的解釋存在不一致的情形。我們若是想研製一套中醫自動辨證系統，假使症狀存在不一致性，電腦將會無法明確區分症狀的差異性，造成辨證的困難。因此，研製中醫自動辨證系統前，我們必須先完成症狀標準化。

有了統一的標準症狀後，中醫自動辨證系統就

能夠利用這些標準症狀的資訊，進行辯證，進而提供處方治療。

當我們著手要做症狀標準化時，遭遇到兩個主要的困難：症狀具有多詞一義及一詞多義的現象[2][3]。症狀具有多詞一義的問題為，有多個不同的症狀描述，但是它們所表達的症狀意義卻是相同的。例如，症狀「大便不利」、「大便艱澀」、和「大便不通」都是表達大便有困難的涵義，我們希望可以將這三個症狀標準化成標準症狀「大便不暢」。症狀具有一詞多義的問題為，一個症狀描述卻包含多個症狀意義。原始症狀「二便不利」表達大便有困難以及小便有困難兩種涵義，我們希望可以將這個原始症狀標準化成標準症狀「大便不暢」及「小便不暢」。

我們做症狀標準化最主要的目標，就是想讓症狀描述可以達到一詞一義以及一義一詞，使得中醫自動辨症系統不會因為症狀多詞一義以及一詞多義的問題，而在進行辯證時遇到困難。

2. 研究目的

本篇論文的研究目的就是研製一個中醫症狀詞庫系統。此系統能夠對應任意的原始症狀到標準症狀，也能夠將標準症狀對應到所有同義的原始症狀。利用此系統，我們就能夠自動進行系統化的症狀標準化，減少人工進行標準化所花費的時間，也可避免相同症狀被重複標準化成不同標準症狀的錯誤，達到症狀一詞一義以及一義一詞的結果。

3. 相關研究

在林劉育的《中醫症狀標準化與初探》研究中[4]，林劉育以人工方式將《中醫證候學》中的原始症狀以人工的方式進行症狀的標準化，並且建立一套使用症狀字詞比對的中醫症狀詞庫系統。在此系統中，只要是已經經過標準化之原始症狀，都可以提供其所對應之標準症狀。但若是其在資料庫之原始症狀，則無法提供其所對應之標準症狀，也就無法自動進行症狀標準化。

在李宜泓的《以近似字詞比對方法輔助中醫症狀標準化》研究中[5]，根據林劉育的研究基礎，他所建立的中醫症狀詞庫系統，除了使用症狀字詞比對來做症狀標準化，還增加了使用近似字詞比對的方式來做症狀標準化。除了資料庫已標準化之原始症狀，可以提供其所對應之標準症狀，若是不在資料庫之原始症狀，他的系統還會透過近似字詞比對的方法，提供建議的標準症狀，但是他的系統是利用字詞的相似程度做搜尋，因此無法提供精確的標準症狀。

4. 研究貢獻

本研究建立了一套中醫症狀詞庫系統，用來解

決中醫症狀標準化的問題。我們除了使用「症狀字詞比對」的方式進行症狀標準化，還增加了使用「關鍵字詞比對」的方式進行症狀標準化。如果原始症狀已經經過標準化，存在資料庫中，系統可透過症狀字詞比對正確地查詢到其所對應的標準症狀。如果原始症狀尚未經過標準化，不在資料庫中，系統會先將原始症狀分割成一串已知的關鍵字詞，再透過關鍵字詞比對的方式，查詢這些關鍵字詞的語意。最後系統依據這些關鍵字詞的語意，查詢語意最接近的標準症狀。本研究顯示，運用語意資訊，可以明顯地提昇系統查詢標準症狀的正確率。

5. 系統架構

本論文中所提出之系統功能分為五部分：原始症狀對應表維護、關鍵字詞對應表維護、症狀知識本體維護、標準症狀對應表維護、以及原始症狀搜尋。

資料庫儲存五個資料表：原始症狀對應表、關鍵字詞對應表、症狀知識本體、標準症狀對應表、以及未標準化原始症狀集。系統架構圖如圖 1 所示。

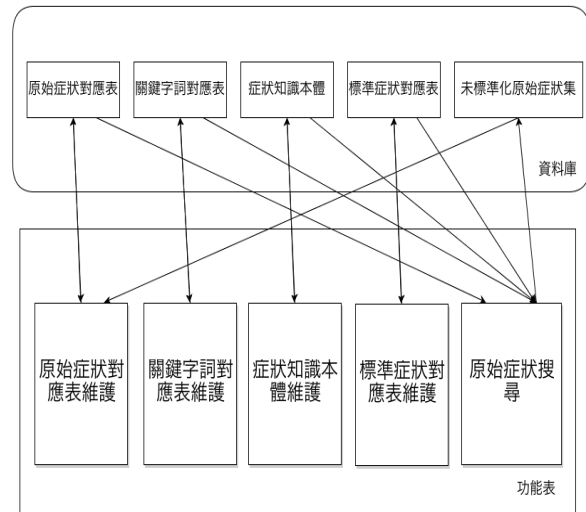


圖 1 系統架構圖

原始症狀搜尋使用原始症狀對應表作為搜尋依據，找出已標準化之原始症狀。若是在原始症狀對應表搜尋不到的症狀，會使用關鍵字詞對應表、症狀知識本體以及標準症狀對應表來做關鍵字詞搜尋，提供建議的標準症狀，這些症狀將會存到未標準化原始症狀集。系統管理者未來可以利用原始症狀對應表維護功能將未標準化原始症狀集中的症狀做標準化，並且新增到原始症狀對應表中。

原始症狀對應表記錄了原始症狀與標準症狀的對應關係。原始症狀對應表維護功能以新增、修改、刪除的方式來維護原始症狀對應表。

關鍵字詞對應表記錄了關鍵字詞與標準關鍵字詞的對應關係。關鍵字詞對應表維護功能以新增、修改、刪除的方式來維護關鍵字詞對應表。

症狀知識本體記錄了標準關鍵字詞之間的語

意關係。症狀知識本體維護功能以新增、修改、刪除的方式來維護症狀知識本體。

標準症狀對應表記錄了標準症狀與標準關鍵字詞(即階層式標準症狀)的對應關係。標準症狀對應表維護功能以新增、修改、刪除的方式來維護標準症狀對應表。

6. 系統實作

本系統的原始症狀搜尋功能主要分成兩部分：可搜尋到的原始症狀、以及未搜尋到的原始症狀，如圖 2 及圖 3 所示。



圖 2 系統介面圖



圖 3 原始症狀搜尋功能

- 可搜尋到的原始症狀
我們使用完整字詞比對[6][7][8]方式搜尋資料庫內是否已存在此原始症狀，如果有，表示始症狀已經經過標準化。
- 未搜尋到的原始症狀
如果原始症狀尚未經過標準化，不在資料

庫中，系統會先將原始症狀分割成一串已知的關鍵字詞，再透過關鍵字詞比對的方式，查詢這些關鍵字詞的語意。最後系統依據這些關鍵字詞的語意，查詢語意最接近的標準症狀。步驟如下：

I. 建立關鍵字詞對應表

我們在資料庫中建立了關鍵字詞對應表，如表 1，用來作為原始症狀斷詞的依據。關鍵字詞欄位為我們所整理出的症狀關鍵字詞，其中有包含許多同義字詞，而標準關鍵字詞欄位是把所有同義詞都統一對應到標準化的關鍵字詞。

表 1 關鍵字詞對應表一

原始關鍵字	標準關鍵字
大便	大便
不通	不暢
不利	不暢

II. 根據關鍵字詞對應表將原始症狀分割為關鍵字詞

我們需要在原始症狀中找出一些特定的關鍵字詞，作為搜尋的基準，因此我們必須對原始症狀先做斷詞，找出這些關鍵字詞。我們使用長詞優先法來對原始症狀進行斷詞。長詞優先法就是在進行斷詞時，使用資料庫中長度最長的原始關鍵字詞作比對[9]。

以原始症狀「大便不通」為例子，本系統將會斷出「大便」、「不通」。雖然在關鍵字詞對應表，如表 2 所示，中也有「大」、「便」這兩個關鍵字詞，但是因為「大便」的長度較長，因此成為斷詞的優先選擇。

我們會從原始症狀的第一個字開始對資料庫的原始關鍵字詞作比對，逐字往後做比對，直到找出最長的關鍵字詞，之後再從下一個字開始做比對，直到症狀的句尾為止。

表 2 關鍵字詞對應表二

原始關鍵字詞	標準關鍵字詞
大便	大便
不通	不暢
不利	不暢
大	大
便	大便

III. 對應原始關鍵字詞到標準關鍵字詞

在原始症狀做完斷詞後，我們再根據關鍵字詞對應表，把關鍵字詞作標準化。因為同義字即代表著相同的語意，所以我們只需用一個標準關鍵字詞就能代表這個語意，用來作為之

後搜尋的基準。

以關鍵字詞「大便」、「不通」為例，本系統會將其標準化成標準關鍵字詞「大便」、「不暢」。

IV. 根據標準關鍵字詞的語意建立症狀知識本體

因為知識本體[10]能夠表達出關鍵字詞之間的語意關係，我們希望能夠建立一個中醫症狀的知識本體，使得症狀可以用一些關鍵字詞互相的語意關係的方式來描述。

我們將關鍵字詞分為「症狀類別」、「症狀屬性」與「症狀屬性值」三種不同詞類來定義。而這三種詞類之間具有包含的關係。「症狀類別」(例:小便、大便)包含了許多不同的「症狀屬性」(例:氣味、顏色)，有些相同，有些相異，而「症狀屬性」又包含了許多不同的「症狀屬性值」(例:臭、黃)。我們只要利用少許的關鍵字詞，就能透過它們所形成的語意關係，找出所有可能的標準症狀。只要這個症狀的知識本體越完整，我們系統所能找到的標準症狀將會越精確。如圖4，為我們的症狀知識本體的一個例子。

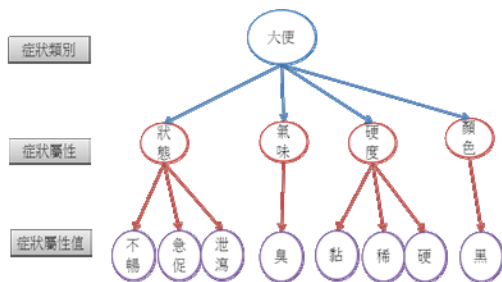


圖 4 症狀知識本體

V. 根據症狀知識本體搜尋標準關鍵字詞詞類

症狀知識本體建立完成後，我們就依據症狀知識本體裡的標準關鍵字詞定義，找出要搜尋的標準關鍵字詞的詞類。

以標準關鍵字詞「大便」、「不暢」為例，本系統將會找出它們對應的詞類定義「症狀類別」、「症狀屬性值」，如圖5。

VI. 搜尋標準關鍵字詞路徑

在找出要搜尋的標準關鍵字詞的詞類後，我們可以在症狀知識本體中，依據標準關鍵字詞的詞類之間的關係，搜尋出可能的路徑，用路徑來建立階層式標準症狀。

以標準關鍵字詞「大便」、「不暢」為例，由於它們與「狀態」這個「症狀屬性」都存在語意關係，因而形成一條路徑，建立了一個階層式標準症狀「大便:狀態:不暢」，如圖6所示。

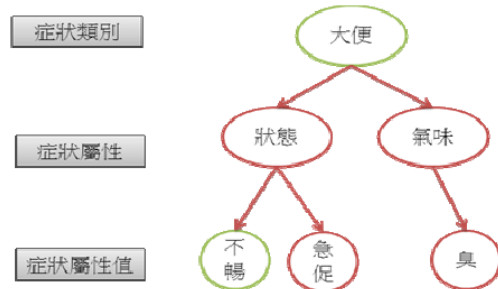


圖 5 關鍵字詞詞類對應

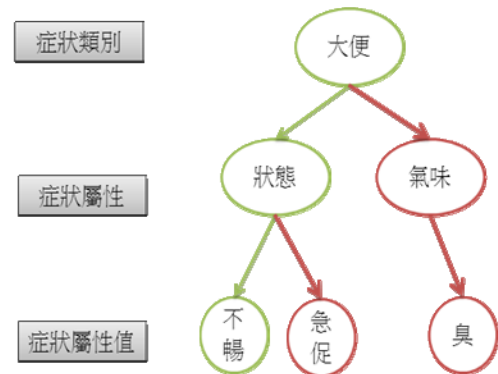


圖 6 關鍵字詞路徑

VII. 根據標準症狀對應表，對應階層式標準症狀到標準症狀

標準症狀對應表為標準症狀與階層式標準症狀的對應關係，因此我們利用此表，找出我們要搜尋的階層式標準症狀所對應的標準症狀，建議出此未標準化的原始症狀可能的標準症狀。

以階層式標準症狀「大便:狀態:不暢」為例，它根據標準症狀對應表，如表3所示，找到符合之標準症狀「大便不暢」，完成搜尋。

表 3 標準症狀對應表

標準症狀	階層式標準症狀
大便味臭	大便:氣味:臭
大便不暢	大便:狀態:不暢
大便挾血	大便:挾帶:血液
大便急促	大便:狀態:急促

以上步驟即為本系統將尚未經過標準化原始症狀「大便不通」，以關鍵字詞比對方式，提供建議的標準症狀「大便不暢」的實作方法。

本系統的標準症狀對應表維護頁面如圖 7 所示，其功能分成三個部分，新增、修改、刪除資料庫的標準症狀以及階層式標準症狀。在此頁面上方按鈕為各個症狀類別按鈕，按下某一症狀類別按鈕後，會列出所有與此症狀類別相關的標準症狀以及對應之階層式標準症狀。因為目前資料庫中的標準症狀，並無法確定包含所有的標準症狀，所以使用者可以透過新增、修改、刪除功能，對標準症狀作維護。

中醫症狀詞庫-標準症狀對應表維護

對應表匯入 對應表匯出

症狀類別 症狀屬性 症狀屬性值

新增 修改 刪除 清空

標準症狀	階層式標準症狀
<input type="radio"/> 小便不暢	排泄物小便 泌尿不暢
<input type="radio"/> 小便急症	排泄物小便 泌尿急症
<input type="radio"/> 小便不盡	排泄物小便 泌尿不盡
<input type="radio"/> 小便先禁	排泄物小便 泌尿先禁
<input type="radio"/> 小便閉塞	排泄物小便 泌尿閉塞
<input type="radio"/> 小便灼熱	排泄物小便 泌尿灼熱
<input type="radio"/> 小便無力	排泄物小便 泌尿無力
<input type="radio"/> 小便排血	排泄物小便 泌尿帶血
<input type="radio"/> 小便排膿	排泄物小便 泌尿帶膿
<input type="radio"/> 小便排砂	排泄物小便 泌尿帶砂
<input type="radio"/> 小便排膿	排泄物小便 泌尿帶膿
<input type="radio"/> 小便色清	排泄物小便 顏色清
<input type="radio"/> 小便色黃	排泄物小便 顏色黃
<input type="radio"/> 小便色綠	排泄物小便 顏色綠

圖 7 標準症狀對應表維護

本系統的關鍵字詞對應表維護頁面如圖 8 所示，其功能分成三個部分，新增、修改、刪除資料庫的關鍵字詞。在此頁面，會列出所有儲存在資料庫的關鍵字詞以及對應之標準關鍵字詞。因為目前資料庫中的關鍵字詞，無法包含所有可能的關鍵字詞，所以使用者可以透過此頁面進行關鍵字詞的變更。

中醫症狀詞庫-關鍵字詞對應表維護

對應表匯入 對應表匯出

症狀類別 症狀屬性 症狀屬性值

新增 修改 刪除 清空

關鍵字詞	標準關鍵字詞
<input type="radio"/> 大便	大便
<input type="radio"/> 小便	小便
<input type="radio"/> 尿	小便
<input type="radio"/> 小便	小便
<input type="radio"/> 瀉	小便
<input type="radio"/> 瀉	小便
<input type="radio"/> 小便	小便
<input type="radio"/> 小便	小便
<input type="radio"/> 矢氣	矢氣
<input type="radio"/> 痰	痰
<input type="radio"/> 瀉	瀉
<input type="radio"/> 瀉	瀉

圖 8 關鍵字詞對應表維護

本系統的症狀知識本體維護頁面如圖 9 所示，其功能分成三個部分，新增、修改、刪除資料庫的語意關係。在此頁面，會列出所有儲存在資料庫的語意關係。因為目前資料庫中的語意關係，沒有包含所有的語意關係，所以使用者可以透過此頁面進行語意關係的變更。

中醫症狀詞庫-症狀知識本體

對應表匯入 對應表匯出

症狀類別 症狀屬性 症狀屬性值

新增 修改 刪除 清空

Id	標準關鍵字詞	Parent Id	詞類
0 1	小便	0	object
0 2	大便	0	object
0 3	矢氣	0	object
0 4	汗	0	object
0 5	痰	0,30	object,value
0 6	瀉	0	object
0 7	瀉	0	object

圖 9 症狀知識本體維護

7. 系統評估

我們先根據《中醫證候學》，在資料庫中建立了 1487 個已標準化的原始症狀。

《中醫症狀鑑別診斷學》[11]總共包含 500 個症狀，我們選用其中 190 個相關的症狀做測試。本系統可搜尋到的症狀共有 66 個，比例為 35%，而未搜尋到的症狀共有 124 個，比例為 65%。

在未搜尋到的 124 個症狀中，透過關鍵字詞搜尋，有 75 個症狀有提供建議標準症狀，比例為 60%，有 49 個症狀無提供建議標準症狀，比例為 40%。

在有提供建議標準症狀的 75 個症狀中，有 66 個症狀建議正確的標準症狀，比例為 88%，有 9 個症狀建議錯誤的標準症狀，比例為 12%。

因此，本系統以《中醫症狀鑑別診斷學》的 190 個相關症狀做測試，提供正確標準症狀 66 個，建議正確標準症狀 66 個，共 132 個，正確率為 69%。

我們在未使用關鍵字詞搜尋時，正確率為 35%，但是利用關鍵字詞語意資訊作搜尋，我們的正確率為 69%，有效地提升了原始症狀的搜尋能力。

8. 結論

本研究建立了一套中醫症狀詞庫系統，用來解決中醫症狀標準化的問題。我們除了使用「症狀字詞比對」的方式進行症狀標準化，還增加了使用「關鍵字詞比對」的方式進行症狀標準化。如果原始症狀已經過標準化，存在資料庫中，系統可透過症狀字詞比對正確地查詢到對應的標準症狀。如果原始症狀尚未經過標準化，不在資料庫中，系統會先將原始症狀拆成一串關鍵字詞，再透過關鍵字詞比對的方式，查詢這些關鍵字詞的語意。最後系統依據這些關鍵字詞的語意，查詢語意最接近的標準症狀。本研究顯示，運用語意資訊，可以明顯地提昇系統查詢標準症狀的正確率。

參考文獻

- [1] 李洪成、李新平、李新曄：中醫證候學，中國醫藥科技出版社，大陸，2008。
- [2] 柯廷諭：中醫症狀的結構化分析與標準化系統設計，碩士論文，國立中正大學資訊工程學系暨研究所，臺灣，2011。
- [3] 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，大陸，2009；16(3)：92-93。
- [4] 林劉育：中醫症狀的標準化初探與應用，碩士論文，國立中正大學資訊工程學系暨研究所，臺灣，2011。
- [5] 李宜泓：以近似字詞比對方法輔助中醫症狀標準化，碩士論文，國立中正大學資訊工程學系暨研究所，臺灣，2012。
- [6] D. E. Knuth, J. H. Morris, V. R. Pratt, Fast pattern matching in strings. SIAM Journal on Computing, USA, 1977, 6(2): 323-350.
- [7] R. S. Boyer, J. S. Moore, A fast string searching algorithm. Communications of ACM, USA, 1977, 20(10):762-772.
- [8] R. Baeza-Yates, G. H. Gonnet, A new approach to text searching. Communications of ACM, USA, 1992, 35(10):74-82.
- [9] K.-J. Chen, S.-H. Liu, Word identification for Mandarin Chinese sentences. Proceedings of the 14th COLING, USA, 1992, 101-107.
- [10] J. F. Sowa, Semantic Networks, In Stuart C Shapiro. Encyclopedia of Artificial Intelligence, 1987.
- [11] 姚乃禮、朱建貴、高榮林：中醫症狀鑒別診斷學，人民衛生出版社，大陸，2004。