

黃帝內經貫珠詞典庫之研發與應用

陳逸光

慈濟大學 醫學系

花蓮

(2001 年 11 月 23 日受理，2001 年 12 月 27 日收校訂稿，2001 年 12 月 31 日接受刊載)

中醫經典黃帝內經之原文資料，經常被歷代醫學家引用到醫藥論著中，以闡述學術思想或醫學理論。本研究選取中醫經典資料庫 NW2001，其內容包括：黃帝內經、金元四大家著作、景岳全書及張氏醫通等中醫經典約四百萬字。所有 NW2001 典籍都交由一個文字剖析器，以 Brute Force Algorithm 切割成 N-連字串(N-Gram)詞典庫。結果發現 4-Gram 是最佳化的庫詞典在 NW2001 的精確率 (Precision) 及回收率 (Recall) 分別為 0.96 及 0.86；在中醫典籍網際網路貫珠集應用上效率很高，資料庫經過索引，網頁顯示的空等時間 (Downtime) 能控制在 8 秒內。由於內經詞句的專一性 (Specificity) 相當高，所以被萃取出來的內經 4-Gram 詞典知識庫，便可經由反轉檔 (Inverted File) 索引到 NW2001 資料庫有關中醫學術理論之出處。

關鍵詞：中醫藥典籍文獻，詞典庫，Brute Force Algorithm，N-連字串，網際網路。

DEVELOPMENT AND APPLICATION OF HUANG-DI-NEI-JING CONCORDANCE THESAURUS

Yee-Guang Chen

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

(Received 23th November 2001, revised Ms received 27th December 2001, accepted 31th December 2001)

Huang-Di-Nei-Jing, one of the Classical Traditional Chinese Medicinal Literature (CTCML), was frequently referenced by other CTML authors to write their medical principles or ideas. In this study, a CTCML database NW2001 including: *Huang-Di-Nei-Jing*, *Jin-Yuan-Si-Da-Jia*, *Jing-Yue-Quan-Shu*, *Cheng-Si-Yi-Tong* totally around four million Chinese characters was used. A word parser, developed by a computer program, can separate words into N-Grams thesaurus, which was driven by the Brute Force Algorithm. Finally, we found that 4-Grams thesaurus indexed on the NW2001 database had Precision and Recall of 0.96 and 0.86 respectively. This optimized 4-Gram thesaurus was used to develop the CTCML world-wide-web. The database was indexed and well formed that the downtime has been controlled under 8 seconds to show each homepage on the WWW. The 4-Grams data set, which extracted from *Huang-Di-Nei-Jing*, formed a knowledge-based database. By using inverted file, the 4-Gram thesaurus of *Huang-Di-Nei-Jing* could be mapped into the NW2001 database. The mapped information had a high specificity relating to medical principles of CTCML.

Key words: Traditional Chinese medicinal literature, Thesaurus, Brute Force Algorithm, N-Grams, WWW.