

行政院國家科學委員會專題研究計畫成果報告

新世代網際網路上醫學資訊系統之研發與實作(III)子計劃二:網路醫學資訊系統之
互動機制與資訊勘測之研發與實作(II)Design and Implementation of Access

Mechanism and Data Mining for a Medical Information System

計畫編號: NSC 89-2219-E-002-028

執行期限: 89年8月1日至90年7月31日

主持人: 陳銘憲教授 臺灣大學電機工程學系

一、中文摘要

新一代的網際網路技術使得遠距醫療與網路醫學資訊系統日臻成熟。本計畫為整合型計畫『新世代網際網路上醫學資訊系統之研發與實作』之一子計畫。我們研發了Web上多媒體互動機制與醫學資訊勘測之研究。首先,網路醫學資訊系統是以WWW為主體架構,開發WWW上多媒體醫學資料存取機制,以在新的多媒體資料庫系統上有效地存放各種多媒體的醫學資訊;並作影音文件同步及互動功能的模組設計;我們研發能更有效率地讓使用者端的Web瀏覽器和醫學資料庫系統連接之技術。其次,我們研發收集此醫學系統使用者行為之機制(data collection)與資訊勘測(data mining)。根據以往文獻,之前提出的資訊勘測方法,皆是分析在傳統WWW使用機制下所產生的WWW Log的資料。這樣的勘測方式不是很適合於分析網路醫學服務下的使用者行為。因此,我們在此一網路醫學系統上設計並實做出新的資訊勘測演算法以瞭解其操作模式並據以作系統效能改進。

關鍵詞: 網際網路, 資訊勘測, 網路醫學

英文摘要

The rapid advances in next generation

Internet technologies have made activities for tele-medicine and networked medical information system a reality. In this project, we conduct research and development for networked medical information systems. First, we establish an efficient Web access mechanism for medical information systems, and design multimedia data synchronization modules to allow users to efficiently access data from a networked information system through the Web interface. Next, we conduct research and develop data mining algorithms to capture the doctor and user behavior in the MedMiner (a data mining system) to fully explore the collected data. According to the previous research in data mining community, most methodology are not suitable to analyze the user behavior in tele-medicine service environment. In light of the user behavior, we can further enhance the performance of our system and improve the efficiency of multimedia communication as well as to increase user friendliness by allowing more interactive functions.

Keywords: Internet, Data mining, Tele-medicine

二、計畫緣由與目的

經過之前計畫的執行,本子計畫整合

文字、數據、圖形、影像、音訊、視訊等各種資料型式，來處理並傳送病患的基本資料、檢查（驗）報告、心電圖、各種醫學造影、心音、呼吸聲及提供互動式會診討論系統介面。在此計畫中，我們已結合網際網路之技術與網路醫學資訊系統之需要，研究發展此網路醫學資訊系統之互動機制與資訊勘測等前瞻性技術。透過總計畫之整合，將此一醫學資訊系統建構且測試於國家實驗網路上，以確實掌握新世代網際網路之技術並充分發揮此網路平台之效益。

我們設計的醫學資訊系統採用 Web 瀏覽器為主要之使用者介面，主因 WWW 的相關應用在網際網路上日益普及，我們希望能透過易於使用及用戶熟悉之介面，以提高此醫學資訊系統之普及率，從而促進醫學資訊之傳播與交流。以現有系統為例，與本計劃合作之台大醫院陳恆順醫師設立之台大遠距醫療網頁(如圖一所示)即是一個結合 WWW 與資料庫系統的醫療系統架構，病人可以藉由 Web 瀏覽器和此遠距醫療系統連接。



圖一. 台大遠距醫療網頁

首先，我們已研發出多媒體互動機制

以更有效率地將用戶端的 Web 瀏覽器和醫學資料庫系統連接。如此，廣大的網際網路使用者可以透過 Web 瀏覽器使用此一醫學資訊系統和多樣化的醫學資料。經過之前的研發後，本計畫採用三層式 (three tier) 的網路架構設計，也就是在傳統的主從式架構間建構一中間層，來有效的處理前後端的需求。在使用者端，由於採用 Web 系統是三層式的架構，因此是以瀏覽器為介面，而在伺服器端，則採用相容於 ODBC (Open DataBase Connectivity) 標準的資料庫系統來存放醫學資料。在此一架構下，我們研究相關的多媒體醫學資料存取技術，並開發出更有效率的醫學資料存取機制。我們可以整合我們的多媒體醫學資料存取技術與陳恆順醫師的醫學系統架構，在國家實驗網路架構下，提供良好之頻寬與傳輸條件，以使得同步與非同步的遠距醫療行為皆得以實現。在同步傳輸方面，我們在 Web 介面上提供即時會診討論的功能，以提供同步的遠距醫療機制。而非同步遠距醫療的部分，則是針對醫學教育的特性在 Web 介面上提供網路醫學教學的功能。我們進行系統測試與並評估實驗結果，以改進我們的系統架構來符合大眾的需求，從而提昇國家整體醫療水準。

本計畫之另一研究重點為資訊勘測 (Data Mining)，我們對醫學資訊系統做資訊勘測可獲取新的醫學資訊，另一方面也可以瞭解使用者之操作模式 (User Behavior) 並據以作系統改進。以高血壓的資訊勘測為例，我們已對醫學資料庫內的病例資料進行資訊勘測。依據高血壓的診斷之相關資料，我們可以對其中資料的關聯性 (Association) 進行研究。舉例而言，我們可以檢查心臟病最可能發生的年齡、家族病史與心臟病的關連，或是心律

異常、血管硬化與心臟病的相關性。許多醫院，包含臺大醫院，已漸對其資訊予以數位化處理，透過對醫學資訊系統內的數位化資訊作此類的資訊勘測，將可獲取許多新的資訊，以做為醫師或醫院的參考。

三、結果與討論

本子計劃今年的重點工作在持續增強此網路醫學資訊系統之功能。首先，此系統是以 WWW 為主體架構，所以開發 WWW 上多媒體醫學資料存取機制，以在新的多媒體資料庫系統上有效地存放各種多媒體的醫學資訊；並作影音文件同步及互動功能的模組設計；我們已經研發一技術能更有效率地將使用者端的 Web 瀏覽器和醫學資料庫系統連接 [9]。另一方面，我們利用最近從事 Internet telephony 之實務經驗與技術加強此一網路醫學系統之互動視訊功能 [10]。我們採用上述技術作為發展此系統的基礎，另外也針對醫學資訊系統的特殊需求研發相關功能。

在醫學資訊勘測方面，我們已具體地規畫了資訊勘測之九項步驟：(1) 決定使用目標；(2) 在資料庫中選擇適當的資料群；(3) 資料解譯和前置處理；(4) 產生所需的資料；(5) 計劃 data mining 的工作項目；(6) 發展 data mining 演算法和函式庫；(7) 對資料進行 data mining；(8) 解讀 data mining 獲得的結果；(9) 驗證所獲得的新資訊。就以往的研究來看，資訊勘測技術與其應用領域有極大之關連 (application dependent)。我們在不涉及個人研究並發展在這樣環境下的資訊勘測之技術。明確地說，我們希望獲得不同的使用群 (依其職業或部門等分類) 會使用何種系統提供的功能，以及對何類醫學資料有較密切之需求等資訊。透過使用者端的資料蒐

集，我們可以取得比在伺服器端的 data mining 更有價值的資訊。在此套醫學資訊系統上採用 data mining 的方法，也更能了解系統的使用狀況並據以作改進。我們已研究及實做出的醫學資訊搜尋器 (MedMiner) 的模組包括 Association (相關性模組) [2] 與 Traversal Patterns (路徑模組) [3] 以及 Clustering (資料叢集分析) [11]。這些資訊對於此一系統資源的規畫，如：醫學資料之分類儲存與搜尋呈現、醫學資訊系統的系統功能設計等有極大的助益。

另一方面，為提高系統的效率及提供更完整的醫學資訊服務，我們已研究評估了資訊勘測的相關技術，並且針對使用者個人行為的記錄進行了研究。為獲取充份之資料以進行使用者操作模式之資訊勘測，研發在網路上蒐集使用者行為 (data collection) 之機制亦是一研究重點。現有蒐集資料的方法是利用在伺服器端的存取記錄 (access log)。但在現有之 HTTP (Hyper Text Transfer Protocol) 協定下，Web 伺服器上收集到的存取記錄大多不能有效的界定單一使用者之存取行為。此外，今日的 WWW 系統為了有效的降低網路頻寬之需求，並加快反應時間，proxy 伺服器的設置非常普遍。大部份的 Web browser 皆透過 proxy 伺服器來存取 WWW 上的資訊，此一現象造成大部份之存取記錄皆只記錄下少數的 proxy 伺服器；如此一來，利用 Web 伺服器上所蒐集到的存取記錄來分析使用者行為是極不準確的。這種情形已被視為網路上作資訊勘測之主要困難。針對此一技術挑戰，我們已運用我們早期研發出的 APCS (Access Pattern Collection Server) [4][5] 技術來作更有效率地使用使用者端之資訊蒐集，其研究成果

已於國際會議上發表[6]。

四、結果自評

本子計畫的研究人員透過這個計畫，學習新的網路的應用技術，使用新的網路軟體套件且增加開發系統軟體的實務經驗。我們由研究發展中所得到的經驗和成果可以提供網路上相關資訊系統參考使用。我們可以預期網路醫學資訊系統之重要性必將與日俱增，而藉由電信國家型計畫之執行，我們深信這可為國家實驗網路提供同時深具前瞻技術與實用價值之成果。

五、參考文獻

- [1] J. Bai, Y.-H. Zhang, and B. Dai, "Design and development of an interactive medical teleconsultation system over the World Wide Web," *IEEE Trans. on Information Technology in Biomedicine*, Volume: 2 2, June 1998.
- [2] M.-S. Chen, J. Han, and P. S. Yu. "Data Mining: An Overview from a Database Perspective." *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6, pp. 866-883, December 1996.
- [3] M.-S. Chen, J.-S. Park, and P.S. Yu, "Efficient Data Mining for Path Traversal Patterns," *IEEE Trans. on Knowledge and Data Engineering*, vol. 10, no. 2, pp. 209-221, Mar./Apr. 1998.
- [4] C.-Y. Chang and M.-S. Chen, "On Building an Internet Gateway for Internet Telephony," Proc. of the IEEE International Conference on Multimedia and Expo, July 31 to August 2, 2000.
- [5] W. M. Detmer, and E.H. Shortliffe, "Using the Internet to improve knowledge diffusion in medicine," *Communications of the ACM*, 40, 8 (Aug. 1997), Pages 101 – 108.
- [6] I.-Y. Lin and M. -S. Chen, "On Methodology for Client-Based User Access Pattern Collection in the Web," *Proceeding of the 11th Conference on Information Networking*, Jan. 1997.
- [7] I.-Y. Lin and M.-S. Chen, "Design and Implementation of an Access Pattern Collection Server for Web Data Mining," *Proceeding of the International Symposium on Internet Technology*, April 24-26, 1998.
- [8] I.-Y. Lin, X.-M. Huang and M.-S. Chen, "Capturing User Access Patterns in the Web for Data Mining," *Proc. of the 11th IEEE International Conference Tools with Artificial Intelligence*, 1999.
- [9] Z.-R. Lin and M.-S. Chen, "Design and Performance Study of Scalable Video Storage in a Disk-Array-Based Video Server," Proc. of the IEEE International Conference on Multimedia and Expo, July 31 to August 2, 2000.
- [10] J.-S. Park, M.-S. Chen, and P.S. Yu, "Using A Hash-Based Method with Transaction Trimming for Mining Association rules," *IEEE Trans. on Knowledge and Data Engineering*, vol. 9, no. 5, pp. 813-825, Sept./Oct. 1997.
- [11] C.-H. Yun, K.-T. Chuang and M.-S. Chen, "An Efficient Clustering Algorithm for Market Basket Data Based on Small-Large Ratios," Proceedings of the 25th International Computer Software and Applications Conference (COMPSAC 2001), October 8-12, 2001.