

國科會研究計畫 91-2411-H-002-080

詞彙語意關係之自動標注—以中英平行語料為基礎

第一年期末報告

執行單位：國立臺灣大學外國語文學系

計畫主持人 高照明

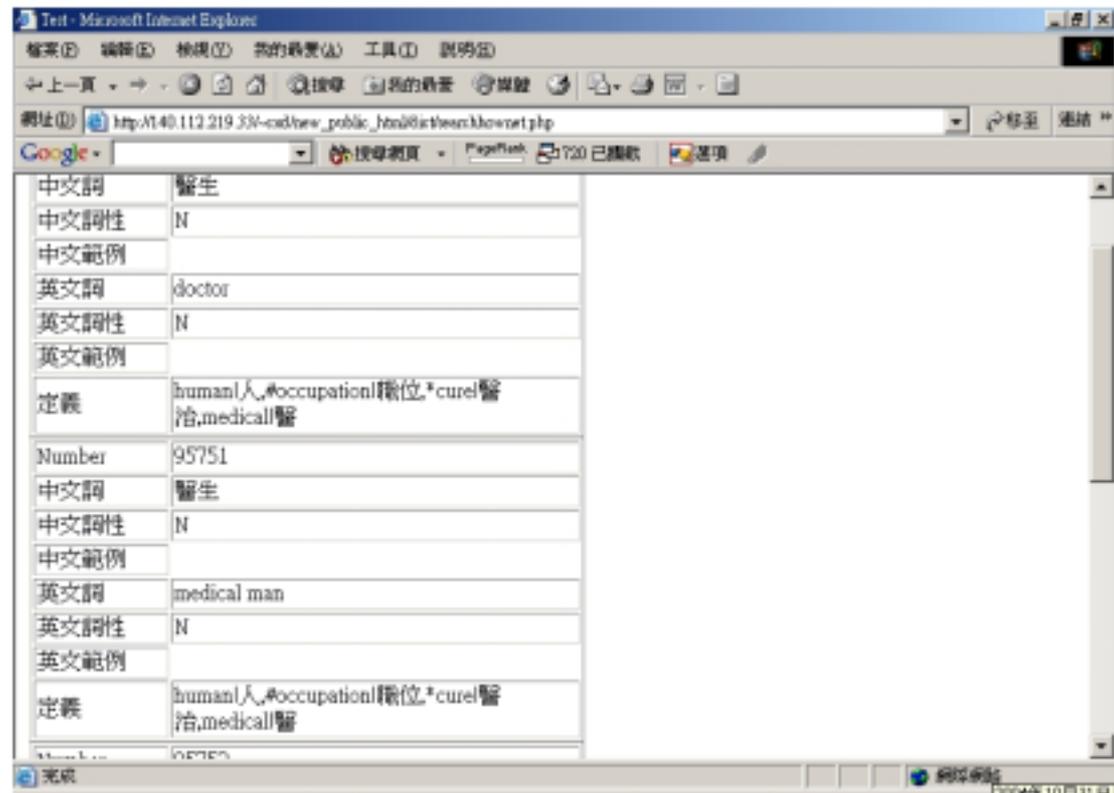
共同計畫主持人：政大資訊科學系劉昭麟

參與助理：江佳恩 黃子桓 林語君 吳昊極 戴士強 劉彥廷 張席維 余世傑

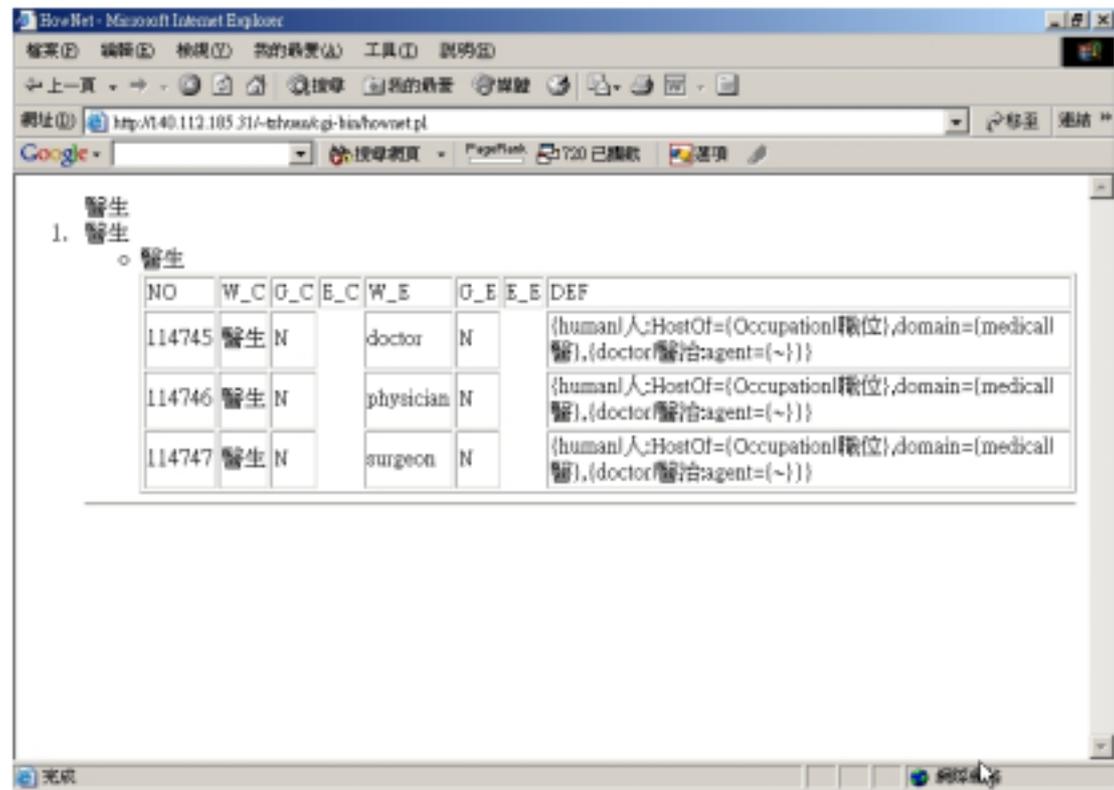
本計畫為三年計畫的第一年，主要目標是整合利用現有的中英文詞彙語意的資源為後兩年的研究奠定基礎並建立相關的資料庫、函式庫、工具程式。我們花了相當多的時間研究如何運用 Hownet 這個具備雙語詞彙語意表達模式的常識庫。Hownet2002 的語意表示方式比 Hownet 2000 更豐富清楚，特別是語意角色的部分，對自然語言理解更有幫助，但是因為結構相當複雜處理不易，所以我們針對新舊版的不同語意表達模式設計不同的應用。

我們的第一件工作是建立 Hownet 2000 及 Hownet2002 的資料庫。這兩版的基本差異可以從下列的表達模式看出來。

圖一



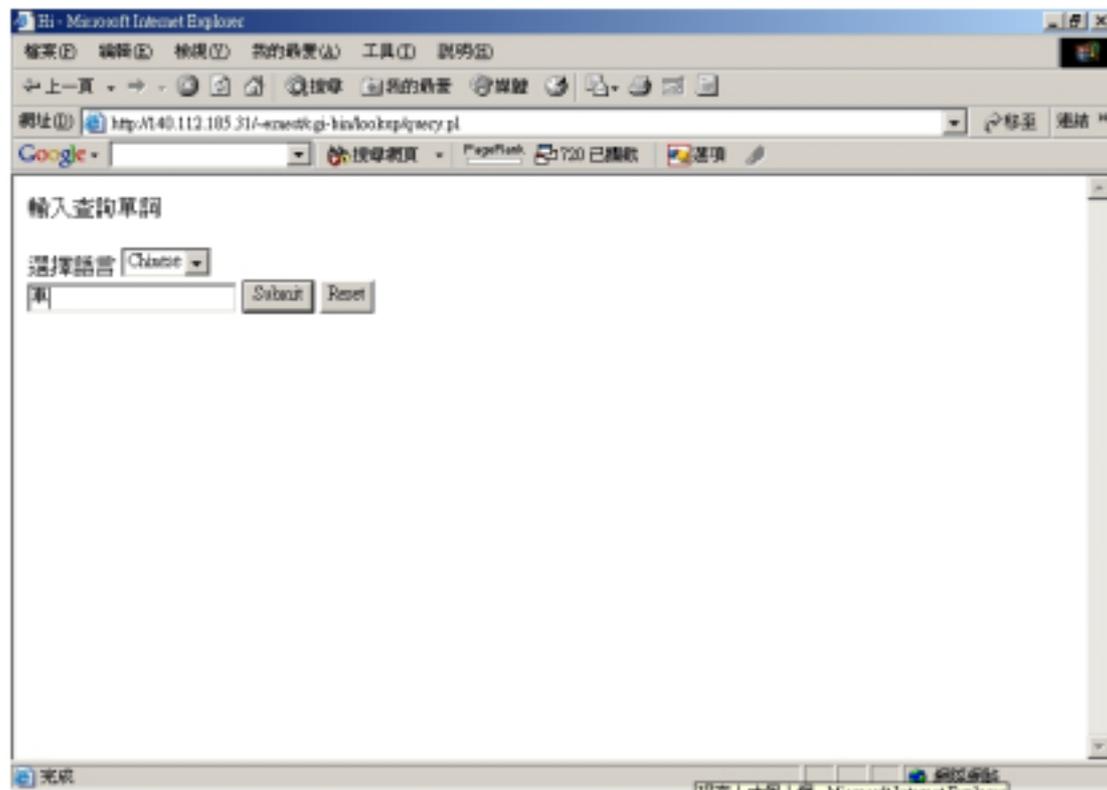
圖二



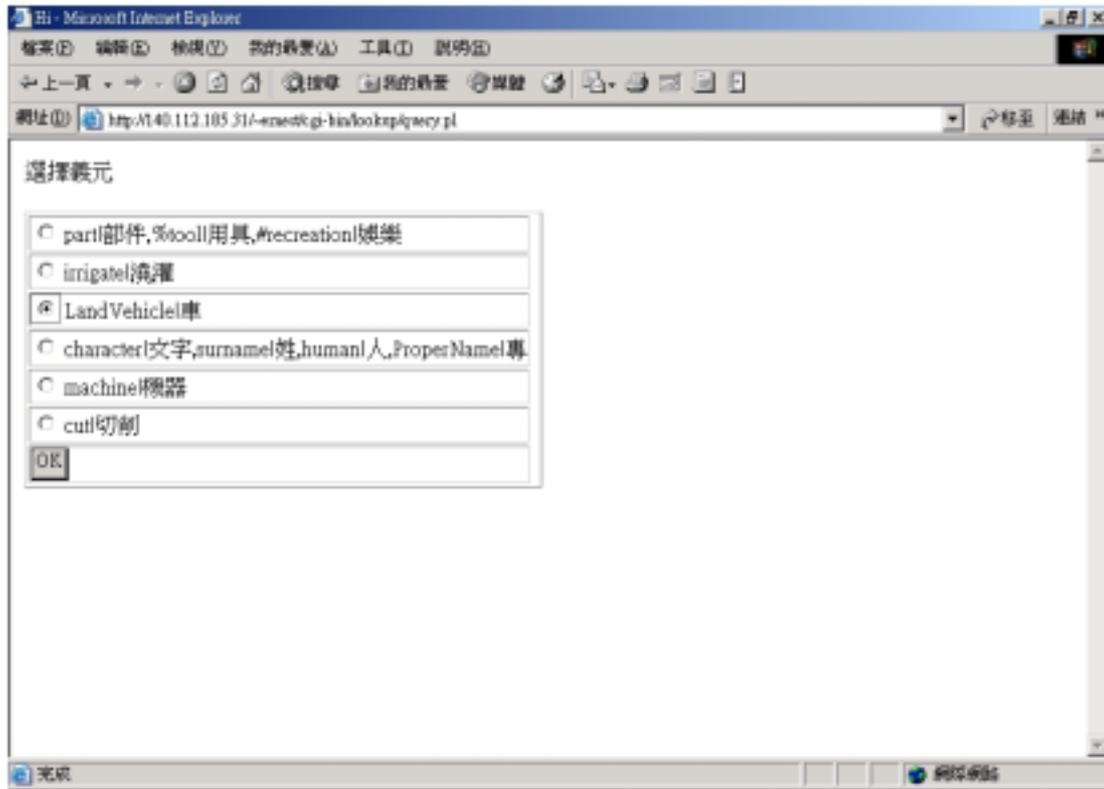
圖一舊版的 Hownet 的詞由義元表示語意，義元本身沒有內部結構，圖二新版的義元則有內部結構，且舊版的 Hownet 語意角色不清楚，例如 \* 符號可能表示在某個事件擔任 agent, experiencer，或其它的語意角色，但在新版的 Hownet 裡面哪個義元在哪個事件中擔任什麼語意角色非常明確。換言之，新版詞與詞之間的語意關係相當明確。

舊版 Hownet 的義元沒有內部結構導致對語意角色的判定較不容易，但相對的程式處理起來較簡單，檢索義元的速度也較快速。我們利用這個特色設計一個工具程式。輸入一個中文或英文詞，程式會將這個詞的所有義元列出來，使用者可以選擇其中一個或多個義元，並且可以用 AND 或 OR 來查詢包含這些義元的詞。

圖三

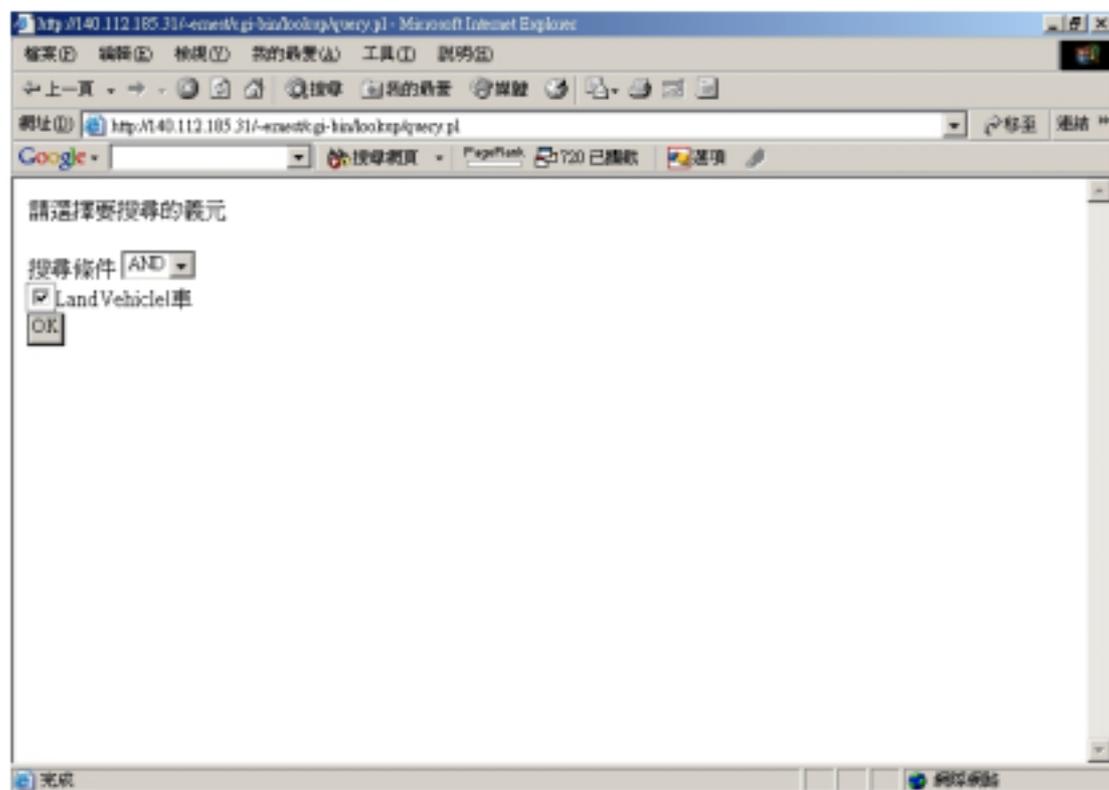


圖四

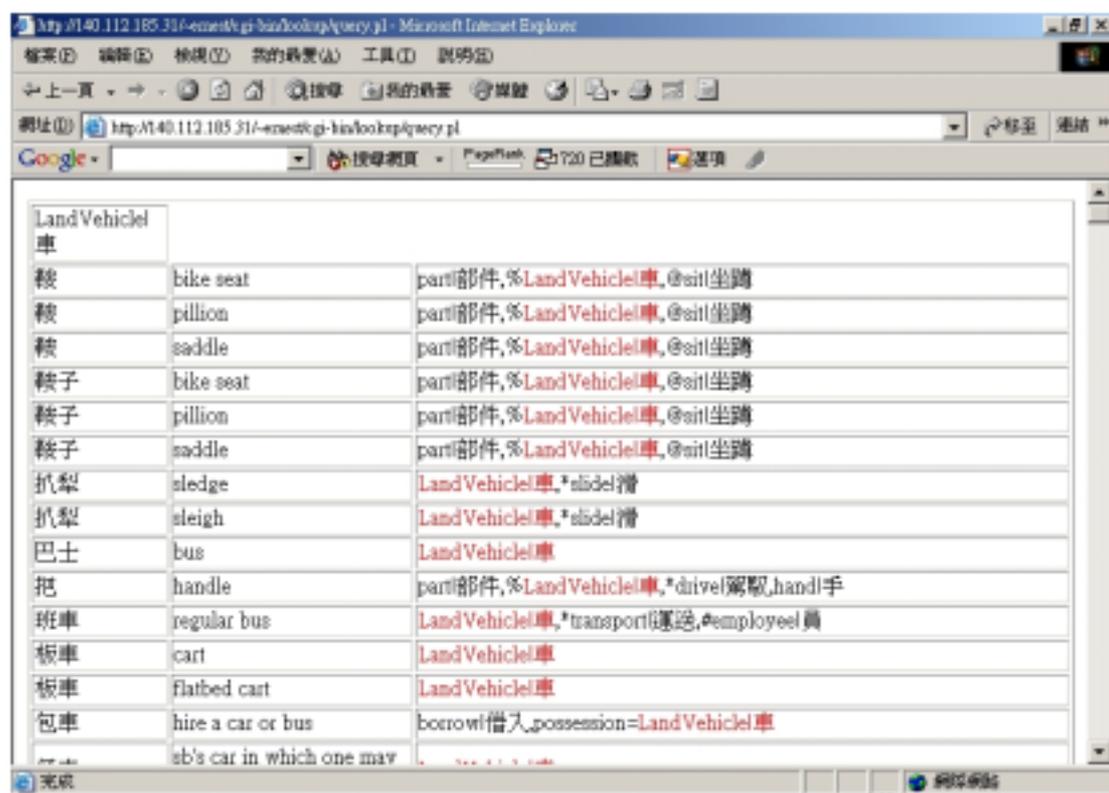


輸入車這個中文詞，程式會顯示這個詞 Hownet 義元的表示法，有交通工具，切割，人的姓等幾個不同的意義及其翻譯。使用者選擇其中一個意義，再選擇要檢索這個意義裡一個或多個義元並選擇這些義元之間的關係是 AND 或 OR，程式就會列出包含這些義元的相關中文詞。由於 Hownet 是中英雙語，我們的程式也可以輸入英文用相同的方法找出相關的英文詞及其中文翻譯。這個工具對於全自動或半自動建立中英雙語或中英跨語言的本體論(ontology)系統非常有用，例如從下面檢索的結果程式得到 LandVehicle|車下面有好幾種 LandVehicle|車，如板車，叉車，餐車，彩車，巴士，長途汽車等。而鞍、鞍子、把等詞則是車的一部份，以義元 part|部件來表示。詞彙與詞彙之間的關係透過 Hownet 的意元表達模式，原本不清楚的詞彙語意關係變得明確化。例如：餐車與吃的語意關係透過 Hownet 義元表達得非常清楚。

圖五



圖六



圖七

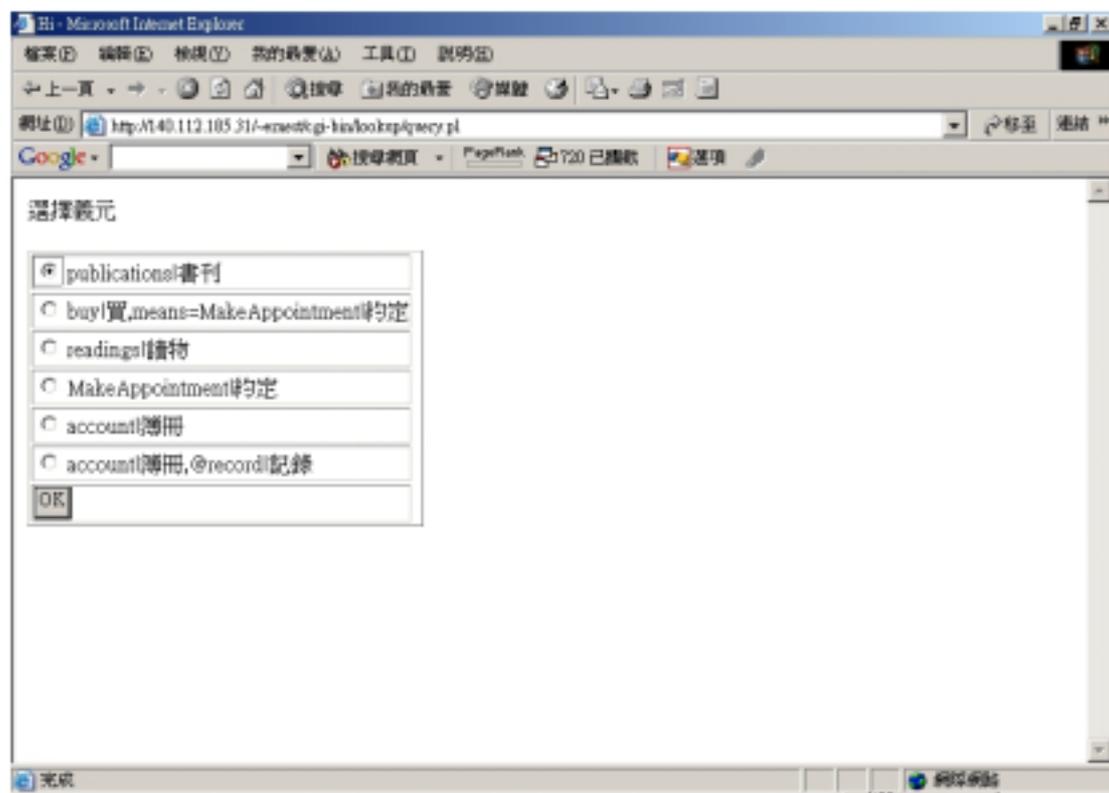
泊車	parking	put 放置,patient=LandVehicle 車
泊位	parking lot	facilities 設施,space 空間,*put 放置,#LandVehicle 車
部		Noun Unit 名量,&LandVehicle 車,&machine 機器,&publications 書刊
彩車	float	LandVehicle 車
彩車	bridal car	LandVehicle 車,#GetMarried 結婚
餐車	diner	LandVehicle 車,@eat 吃
餐車	dining car	LandVehicle 車,@eat 吃
餐車	restaurant car	LandVehicle 車,@eat 吃
叉車	forklift	LandVehicle 車,*pick 抬起
叉車	forklift truck	LandVehicle 車,*pick 抬起
錢車	forklift	LandVehicle 車,*pick 抬起
錢車	forklift truck	LandVehicle 車,*pick 抬起
錢運車	forklift	LandVehicle 車,*pick 抬起
錢運車	forklift truck	LandVehicle 車,*pick 抬起
長途汽車	coach	LandVehicle 車
長途汽車	long-distance bus	LandVehicle 車
駁車	flatcar	LandVehicle 車

圖八

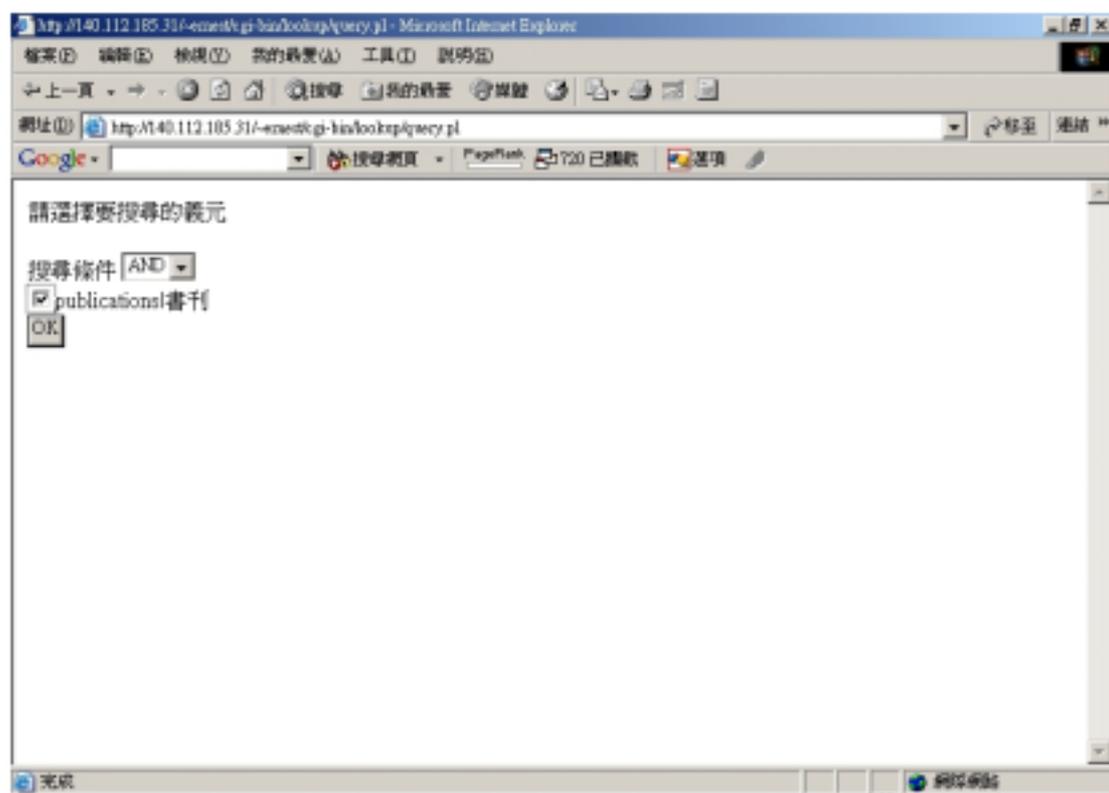
輸入查詢單詞

選擇語言

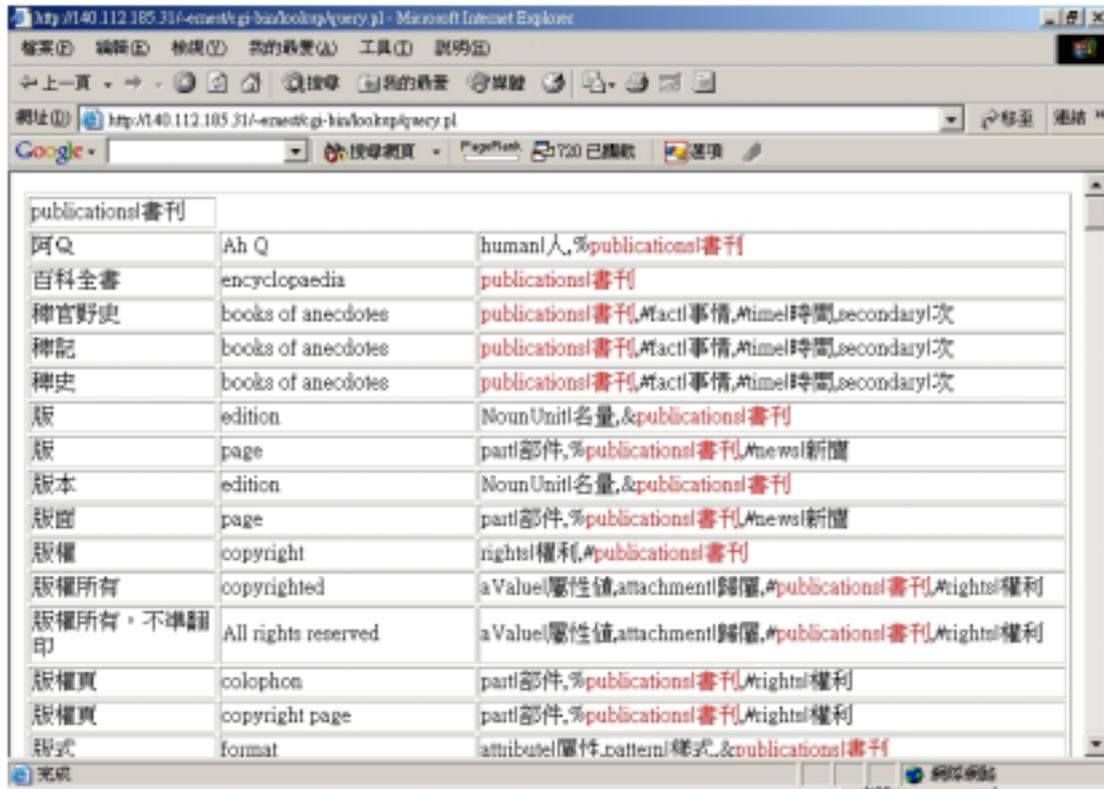
圖九



圖十

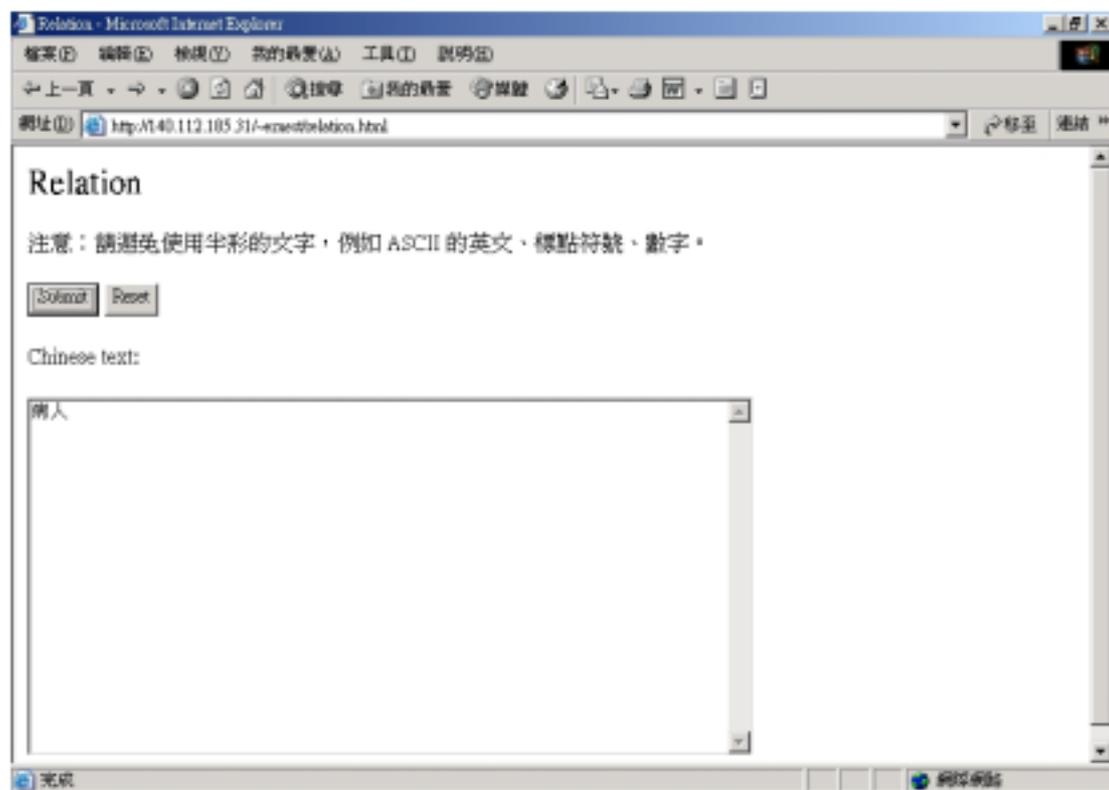


圖十一

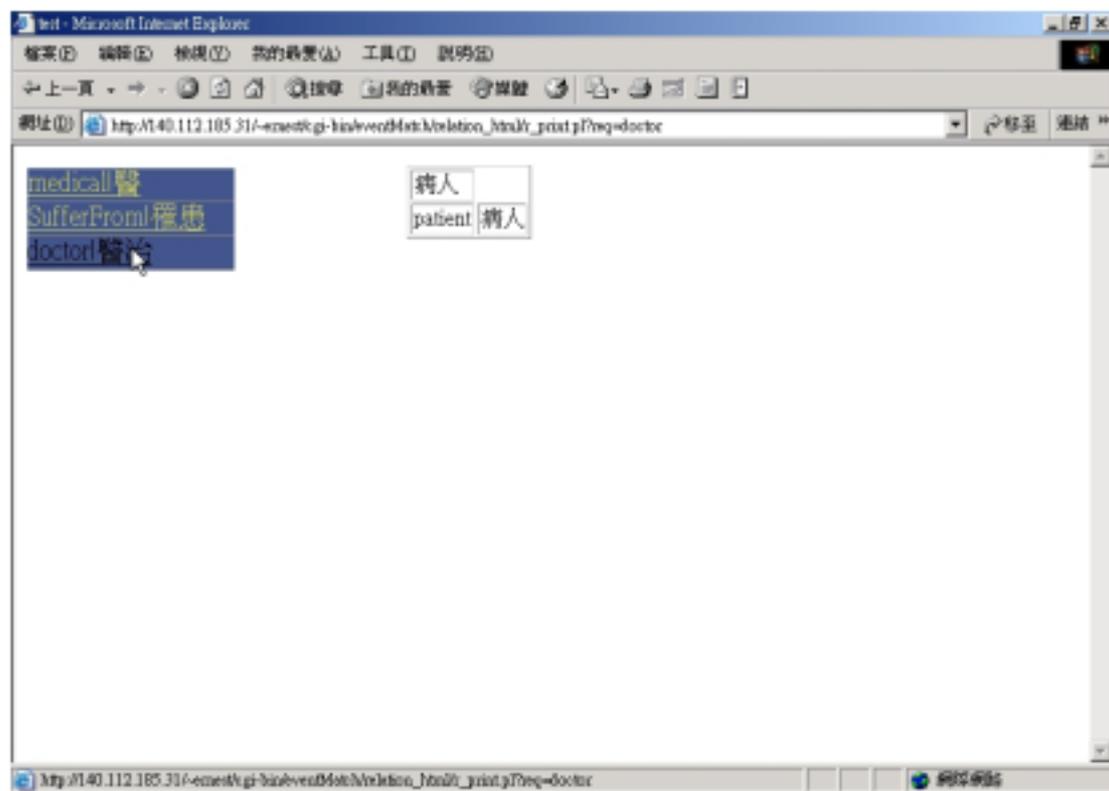


我們也利用新版 Hownet 的特性設計一個能找出義元與其事件義元之間語意角色的關係。如下圖，輸入病人程式會列出這個詞的所有義元，用滑鼠點選某一個義元就可以看到與這個義元有語意角色關係的義元。例如點選 SufferFrom|罹患這個事件義元就知道這個事件的 experiencer 是病人，而 doctor|醫治這個事件的 patient 也是病人。這個程式可以作為判定語意角色與詞彙語意的重要工具。

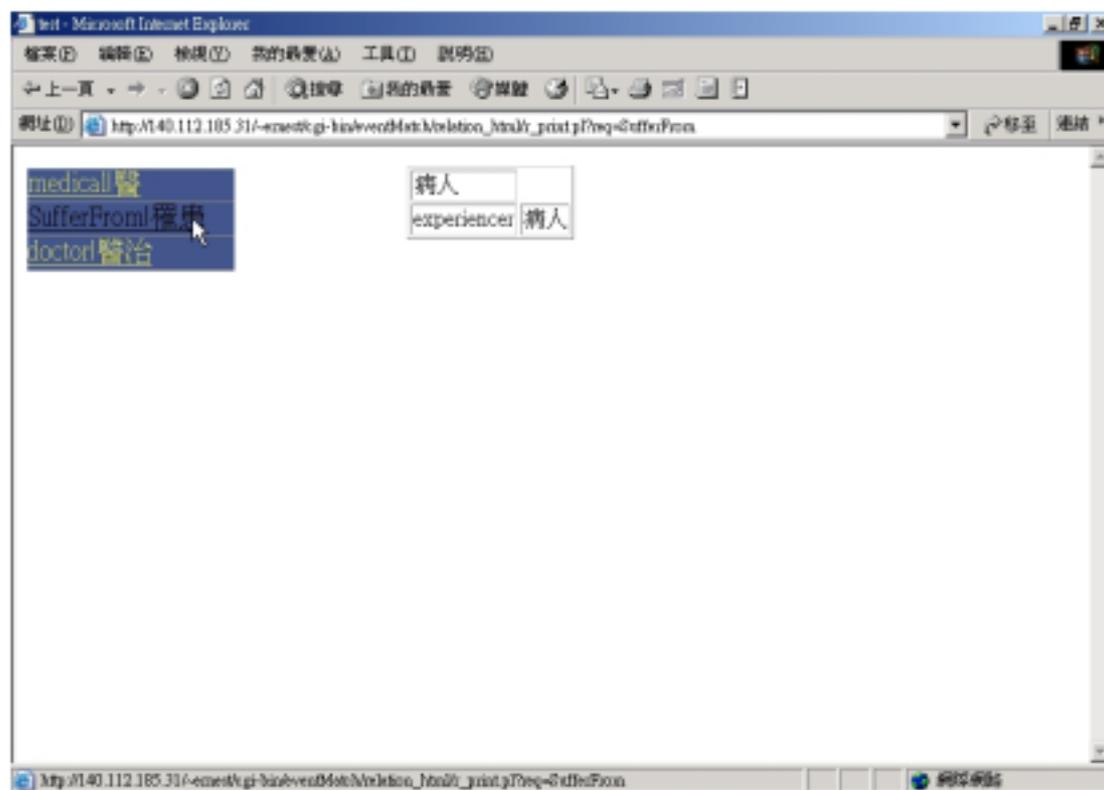
圖十二



圖十三

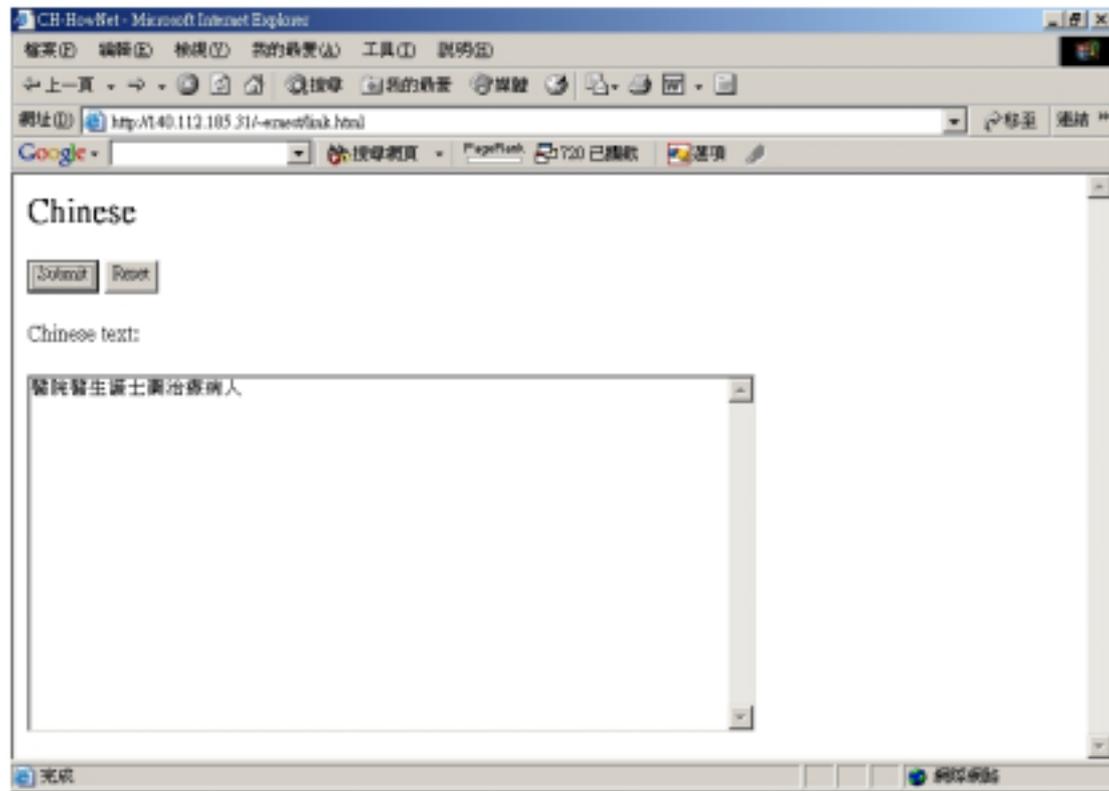


圖十四

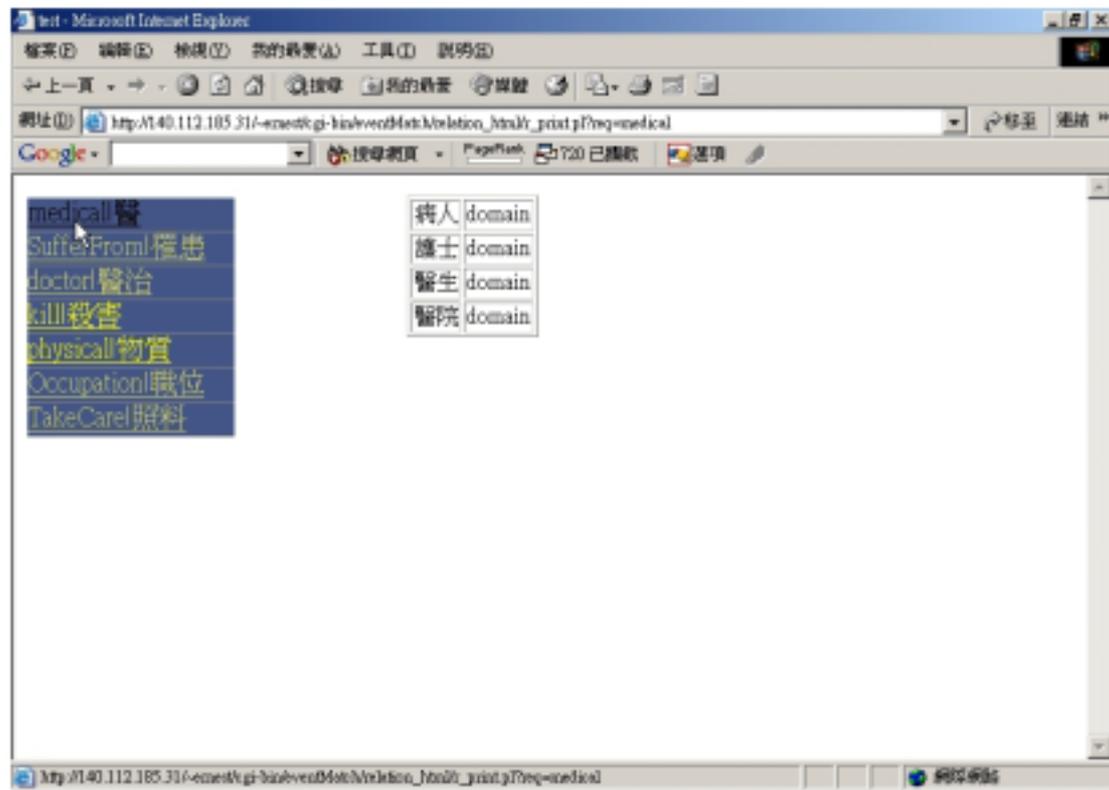


另一個類似的程式則可以找出所有在 Hownet 的義元表達法裡面有關連的任兩個詞的詞彙語意關係。例如輸入：醫院醫生護士藥治療病人，可以找出醫生護士病人醫院都是醫療領域，醫生護士是治療這個事件的 agent，醫院則是治療這個事件的 location，醫生護士都是職業等語意關係。

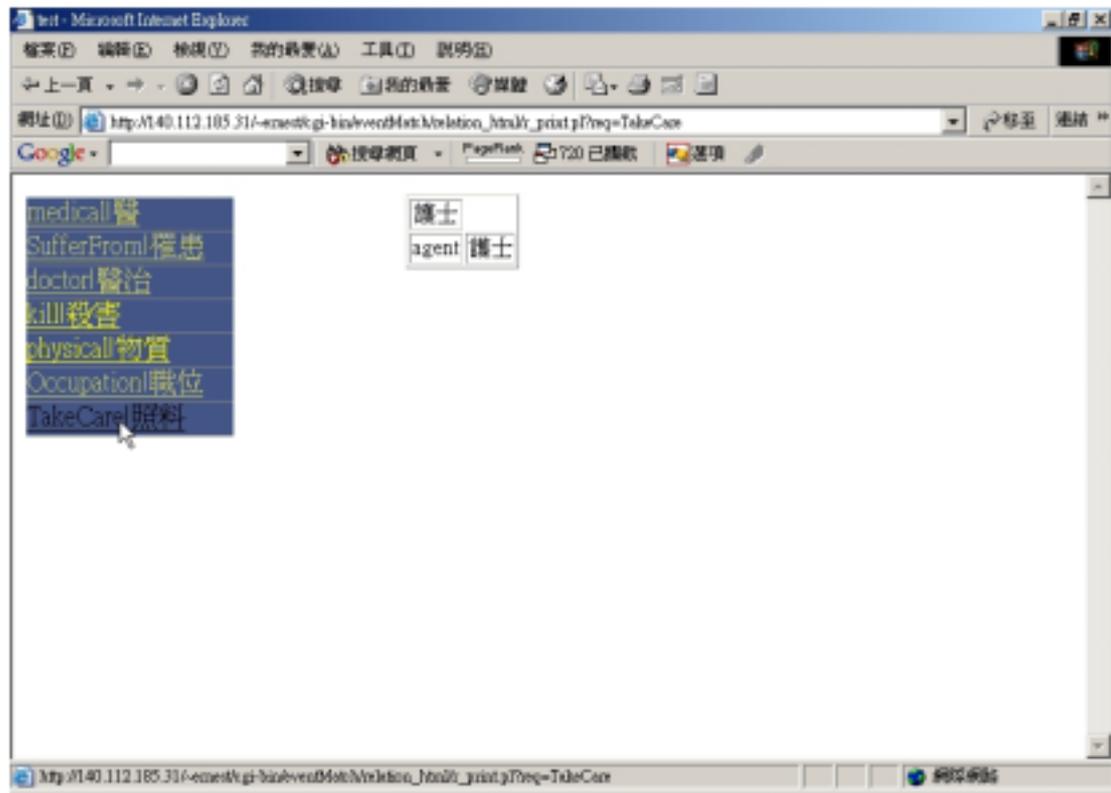
圖十五



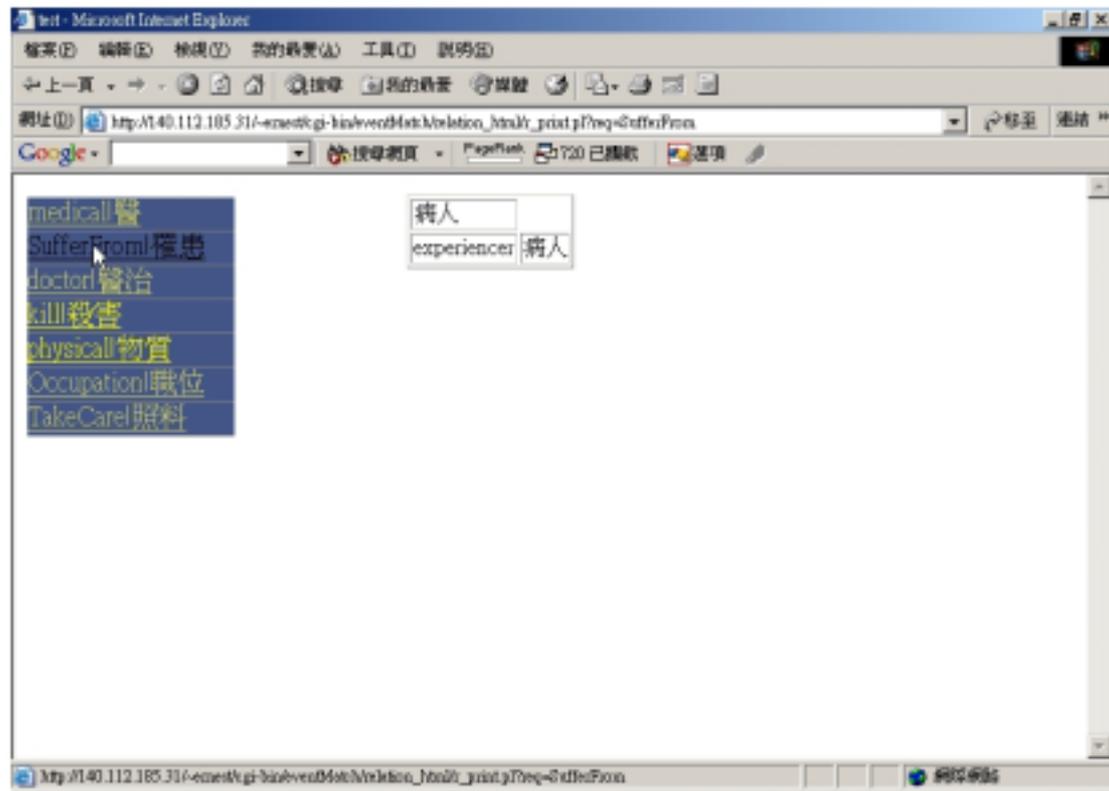
圖十六



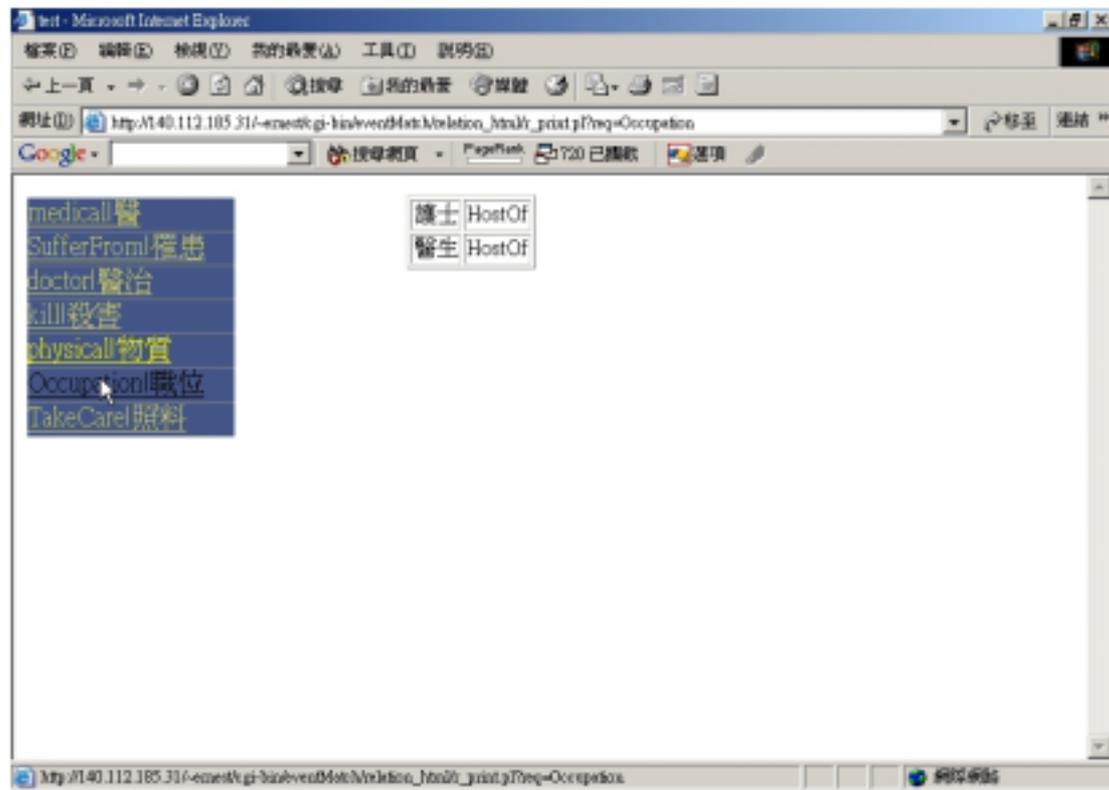
圖十七



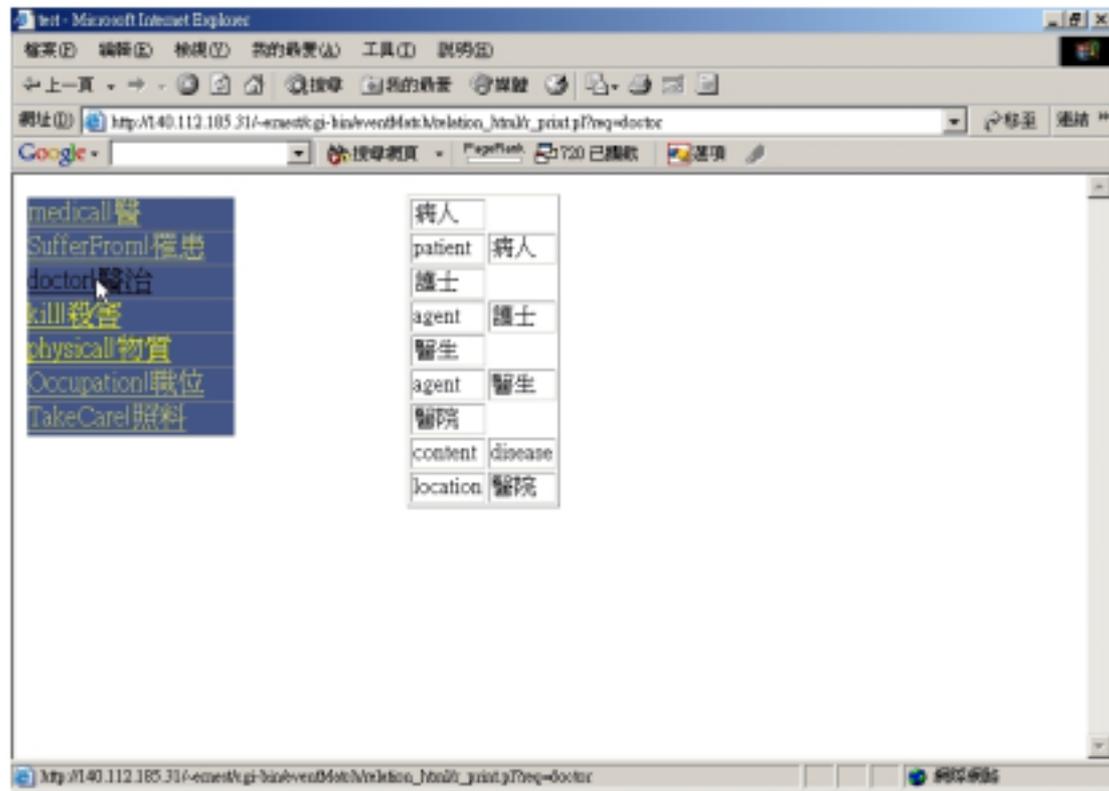
圖十八



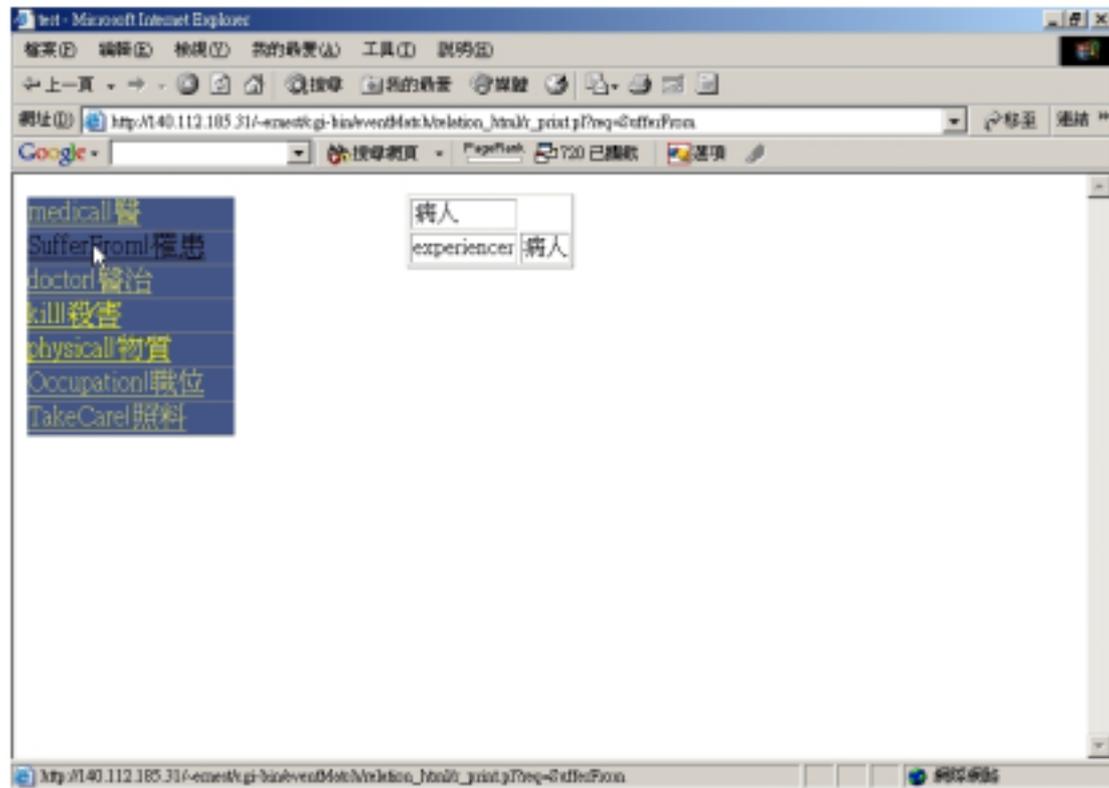
圖十九



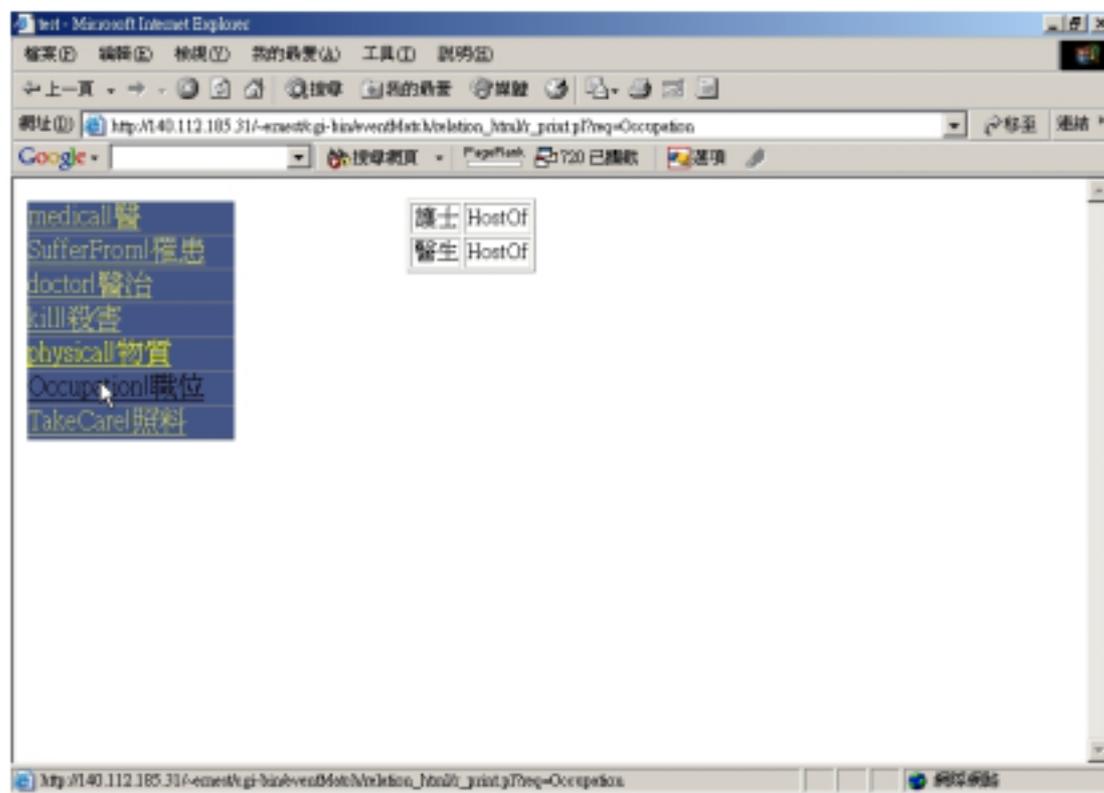
圖二十



圖二十一



圖二十二



接下來，我們將嘗試結合 Hownet 與統計的方法來解決語意歧義的問題。