

以知識概念模型為基礎之多主題對話管理系統

Ontology-Based Dialog Management for Multiple Service Integration

陳銘軍 葉瑞峰 吳宗憲

國立成功大學資訊工程學系

rusk@csie.ncku.edu.tw, p7890101@ccmail.ncku.edu.tw, chwu@csie.ncku.edu.tw

摘要

目前的對話系統，多侷限於單一功能或單一領域的應用，而在實際情況，使用者經常需要跨多個服務或多個主題。本文提出一快速且有效整合現有對話系統之方法。為了實現此一研究主題，必須先半自動地建立一個醫療概念模型作為對話系統的知識表示法，建立及整合三個不同但具有相關性之服務模組：分別為掛號諮詢模組、科別諮詢模組以及常見問答集(FAQ)模組。藉由對話管理模組依其意圖加以整合。在多主題或多服務的整合上，提出以部分樣本樹做為意圖偵測之評估方式，配合語意框架記錄並控制對話流程，並利用樣本產生系統之相對回應文句。

為了評估本文所提出的方法之可行性，由 50 個大專以上教育程度，但未參與本研究之使用者，以文字輸入的模式實際測試本系統。在效能評估上，其整體意圖偵測正確率為 86.2%、系統成功率為 77%，每筆對話的平均長度為 9.2 回合，而回答自然度則為 78.5%，足見本文所提之方法是有效可行的。

1. 緒論

語音及語言處理技術的日趨成熟，使得對話系統的實現成為可能 [1][2][3][4]。對話系統的研究，在國外方面有 MIT 的 JUPITER [5]、AT&T 的線上服務系統 [6]、Philips 的 Automatic Train Timetable Information System [7]、英國的 Nuance Automatic Banking System 以及法國的 LIMSI Arise system 旅遊資訊導覽系統 [8] 等。在國內方面，台大則提出在語意與知識之擷取 [9] 以及在分散式網路環境中對話系統代理人之架構 [10]，工研院有智慧型總機、氣象查詢系統 [11] 等多項成果，成大在對話系統的研究上也投入相當多之研究人力 [12][13][14][15]。

就對話系統言，其中一個重要課題便是如何理解人類語言，讓電腦能夠接受人類的語言和指令，直至今日仍是一個無法完全解決的問題。知識概念模型定義

除了定義了概念本身之語義定義外，同時也定義了概念與概念間的關係，並存在有推論規則可供邏輯推理，是一種具有推論能力的知識表示法[16][17]。而對話系統目前的研究大多侷限在單一功能或單一領域的應用。然而使用者之意圖，可能會牽涉到多個領域或多個服務，因此如何整合多套對話系統，進而提供更完整的資訊，是目前許多研究的主題[18]。本文提出利用意圖偵測機制，整合多種服務，此一作法不僅在整合既有系統提供解決方案，更在未來新增功能預留空間。

為驗證所提方法之正確性，我們建立一套智慧型的醫療對話查詢系統，提供三個功能模組，分別為掛號資訊諮詢模組、科別資訊諮詢模組及常見問答集資訊諮詢模組。其中

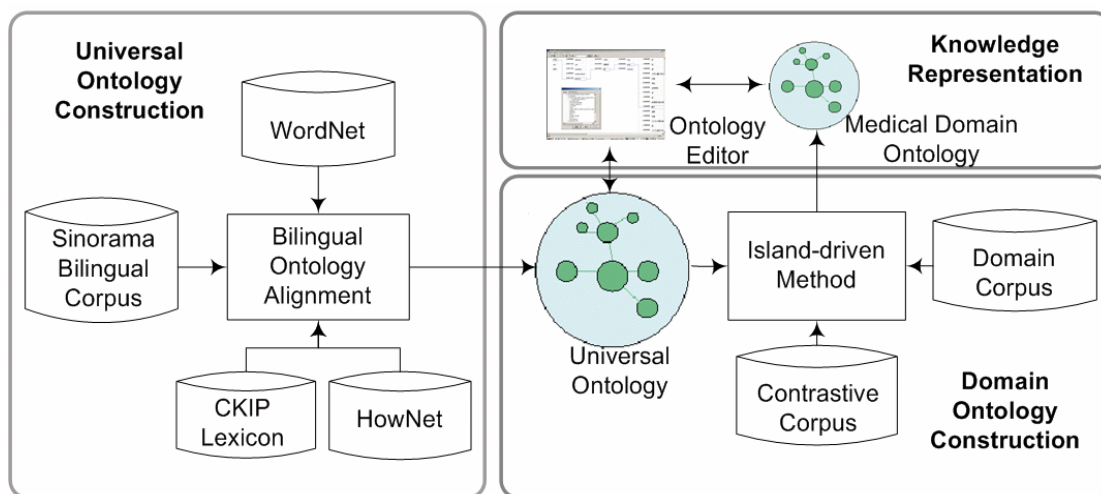
- (1) 掛號諮詢模組：乃利用自然語言處理的技術，使用者除了從網路超連結中以點選的方式進行網路掛號外，更可以符合使用者習慣之口語方式輸入。
- (2) 科別諮詢模組：此一模組利用知識概念模型中的規則來進行推論、藉以解決使用者在掛號時，可能做症狀描述但不知科別之窘境。
- (3) FAQ 諮詢模組：相關文獻的查詢，從對話過程中累積的資訊分析得到，與以單一文句查詢之 QA 系統不同。

知識表示法乃採用 WordNet 的結構為基礎，將 HowNet 的資訊整合進去，首先使用雙語語料為統計對象，基於雙語字典整合 HowNet 和 WordNet 來建立出 Universal Ontology，再利用醫療類語料和島嶼演算法[19]半自動萃取出屬於 Medical Ontology 之概念，並加入 1213 則推論規則(axioms)，形成醫療概念模型，此醫療概念模型(Medical Ontology)即為系統底層的知識庫。

在多項服務的整合上，本文提出利用部分樣本樹 PPT(Partial Pattern Tree)[20]來擷取使用者的意圖(Intention)，再由使用者的意圖搭配相關之語意框架(Semantic Frame)來控制並完成整個對話流程。

2. 知識概念模型之建構

本文提出一個兩階段建構知識概念模型之方法，如下圖一所示：



圖一、知識概念模型建構流程圖

以 WordNet 的階層架構為主，將 HowNet 整合進來，其中所用之雙語語料為光華智慧藏部分語料。建構出通用知識概念模型 (Universal Ontology)，再利用島嶼演算法將屬於領域特性的知識概念模型擷取出來成為領域知識概念模型 (Domain Ontology)。

在知識概念模型的整合上採取由下而上(Bottom-up algorithm)的策略。詞與詞或概念與概念間的上下位關係為一樹狀結構。節點之相似程度乃由終端節點開始向上傳遞累積，若子節點相似且節點本身相似程度高，則此兩個節點所代表之意義應屬於同一概念。而目前系統中的關聯整合乃建構於字網的關聯架構上，將知網的關聯分析後將之對應至字網對應的依據乃是前面所得的中、英文概念詞之對應關係加權計分，和他們彼此之間的結構上之相似度 c 如下式(1)所示：

$$\begin{aligned}
 & \Pr(\text{synset}^k | CW_i) \\
 &= \sum_{j=1}^m \Pr(\text{synset}^k, EW_j^k | CW_i) \\
 &= \sum_{j=1}^m (\Pr(\text{synset}^k | EW_j^k, CW_i) \times \Pr(EW_j^k | CW_i))
 \end{aligned} \tag{1}$$

其中

$$\Pr(\text{synset}^k | EW_j^k, CW_i) = \frac{N(\text{synset}^k, EW_j^k, CW_i)}{\sum_l N(\text{synset}_j^l, EW_j^k, CW_i)} \quad (2)$$

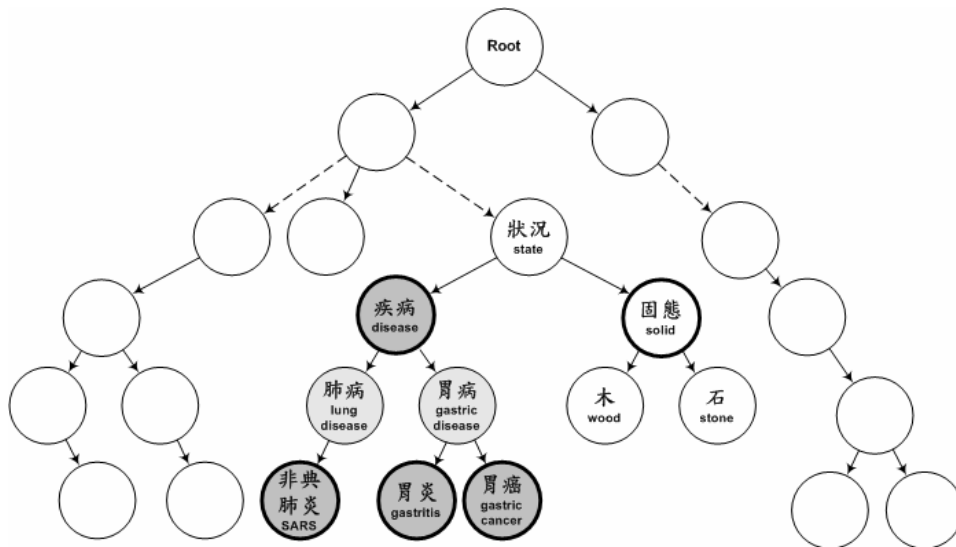
如上式所示， $N(\text{synset}_j^k, EW_j^k, CW_i)$ 為 CW_i 、 EW_j^k 和 synset_j^k 共同出現的次數。在定義於 HowNet 中之中文詞，至少有一主要特徵 $PF_i^l(CW_i)$ 和定義於 WordNet 中英文詞之上位詞中存在有一個同義詞集合， $\text{synset}_j^k(EW_j)$ ，有一致之概念時，其機率 $\Pr(EW_j | CW_i)$ 為 1，否則為 0。

$$\Pr(EW_j | CW_i) = \begin{cases} 1 & \text{if } \left(\bigcup_l PF_i^l(CW_i) \right) \cap \left(\bigcup_k \text{ancestor}(\bigcup_k \text{synset}_j^k(EW_j)) \right) \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

最後中文概念詞 CW_i 將依其機率值 $\Pr(\text{synset}^k | CW_i)$ 被整合至英文同義詞集合 synset_j^k 中，成為其中的一個元素。如此一來便完成了通用知識概念模型的建構。

接下來我們便從此通用知識概念模型中擷取出，屬於醫療領域之領域概念模型，其步驟有四：

(1) 階層線性化：知識概念模型中的階層是一樹狀的結構，如圖二所示。線性化的目的，即是將樹狀結構分解為節點之有序串列(由終端節點到根節點的路徑)所成的集合。



圖二、概念階層樹狀圖

(2) 以語料為基礎之概念抽取：語料分為目標領域語料(Target Domain Corpus)與對比領域語料(Contrastive Corpus)兩類。其中領域語料是從網路上蒐集得到的一千二百二十二篇中文醫療類 FAQ，而對比領域語料是由光華雜誌智慧藏當中選取二千一百八十篇非醫療類文章構成。並依 Tf-Idf 決定節點是否屬於目標領域，若屬於目標領域則稱此一節點為有效節點。

$$operative_node(W_i) = \begin{cases} 1, & \text{if } Tf-idf_{Domain}(W_i) > Tf-idf_{Contrastive}(W_i) \\ 0, & \text{Otherwise} \end{cases} \quad (4)$$

其中

$$Tf-idf_{Domain}(W_i) = freq_{i,Domain} \times \log \frac{N}{n_{i,Domain}}$$

$$Tf-idf_{Contrastive}(W_i) = freq_{i,Contrastive} \times \log \frac{N}{n_{i,Contrastive}}$$

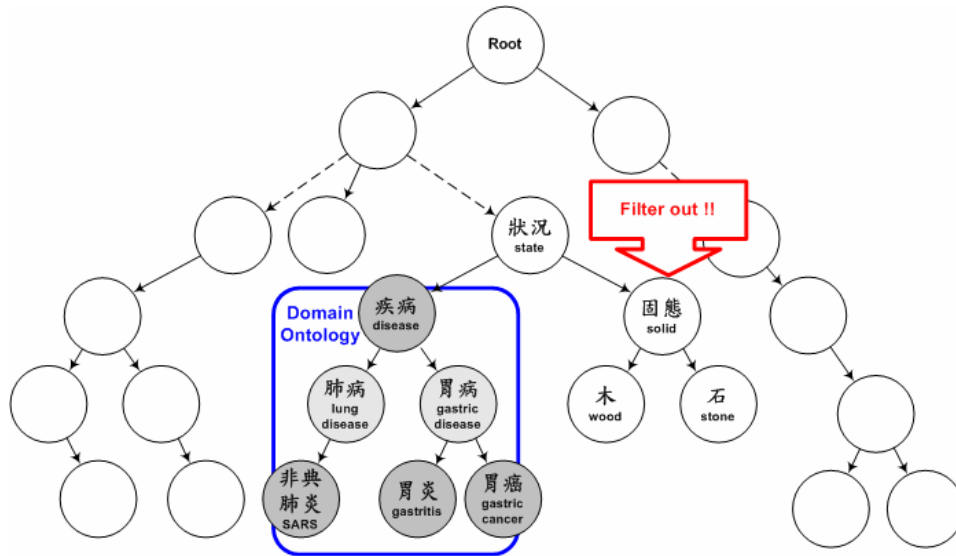
$$N = n_{i,Domain} + n_{i,Contrastive}$$

其中 W_i 為待測概念詞， $freq_{i,Domain}$ 和 $freq_{i,Contrastive}$ 分別代表 W_i 出現在目標領域語料和對比領域語料中的頻率。 $n_{i,Domain}$ 和 $n_{i,Contrastive}$ 則為目標領域語料和對比領域語料中含有 W_i 之文章數。

(3) 島嶼演算法：由於統計語料之不足，所以利用知識概念模型中的上下位關係，將一些未出現在統計語料，但屬於該領域的概念也一併萃取出來。事實上，對於一個概念而言，若是其上位關係與下位關係皆是屬於某一領域，則該概念同屬此一概念的機率是很大的。基於這樣的假設，利用島嶼演算法[19]來將潛藏於兩個領域概念之間的概念一併收錄。

(4) 合併線性成分且過濾獨立概念：經過島嶼演算法之後，可以得到一個具擴充的節點串列，原則上只要保這些節點串列合併起來，便可以建構出領域知識概念模型的雛型。但對於一些具有多個有效節點的節點串列，將予以保留並加以合併。而對於僅含有一個有效節點的節點串列將予以刪除，其原因是這些單獨存在的有效節點，多為停用詞(Stop word)或其他自然語言處理程序中所產生的干擾，如斷詞錯誤等，完成後如圖三所示。

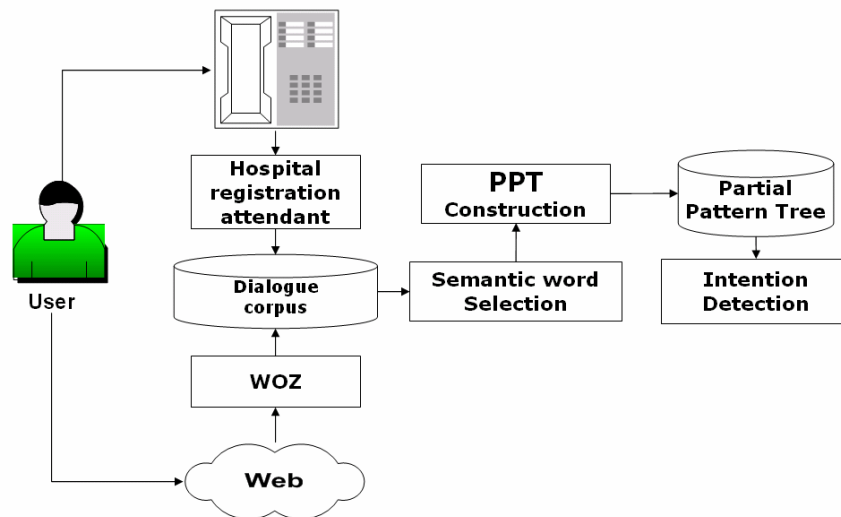
建立醫療概念模型後，依疾病、症狀及所對應的科別，加入了 1213 條規則。



圖三、領域知識概念之取得

3. 部分樣本樹之建立和意圖偵測

本文提出利用意圖來整合多項服務的方法，在對話系統的發展中經常面臨語料不足的問題，為了克服語料不足以及增強系統之包容性(Robustness)，我們使用部分樣本樹來做意圖偵測，其建構流程如圖四所示：



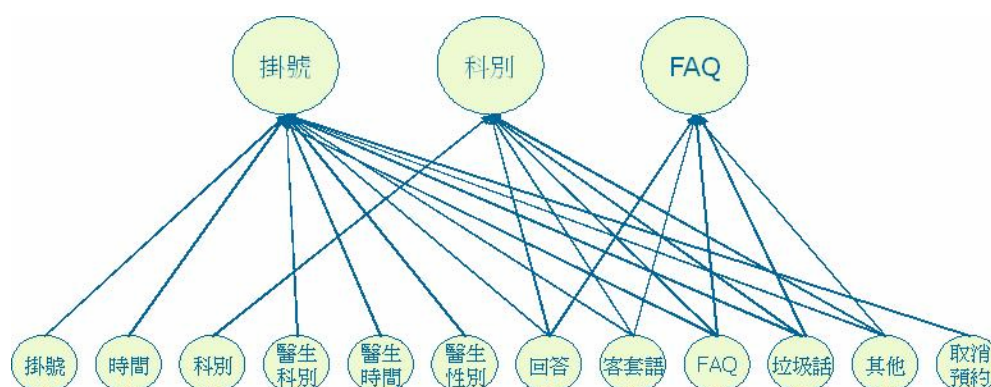
圖四、部分樣本樹和意圖偵測架構圖

以統計方法建構一套實際的對話系統時，面臨的第一個問題就是對話語料的收集，語料收集之良窳直接影響整個對話系統的建構。我們採用兩種方式收集語料，分別為 woZ 以及成大醫院實際語料。

(1) Wizard-of-Oz 方式之語料蒐集：本論文使用 ASP 和 IIS5.0 版來實做 WOZ 之語料收集平台，透過此一系統，本論文初步收集了三十四位使用者的對話共計 234(turns)之語料，用以建立系統之雛型。

(2) 成大醫院實際掛號語料：在實際語料收集部分，在 88 年一月間於成大醫院錄製之實際電話預約語音資料，再以打字的方式轉成文字檔，共有十萬多字，四千零八九筆對話資料。

在收集完語料之後，由 WOZ 所收集的語料和電話語料中選擇出 1395 句來做人工標記其意圖，做為建立部分樣本樹之訓練語料，依據系統所整合之服務，並觀察語料可歸納如下之意圖示意圖，意圖共分為 12 類，如圖五所示：



圖五、意圖示意圖

為了整合各種不同而相關的服務，必須從各種服務間找出其差異性，在此我們使用意圖(Intension)來區分服務之機制，就可從標記為各意圖的語料裡找出一組主要語意詞來代表這個意圖，簡單地說，主要語意詞即為具有鑑別性之語意概念，例如使用者想查詢有關醫生看診時間，就會提到醫生的姓名，醫生姓名便可以用來代表醫師姓名這項意圖。本文採用 LSA(Latent Semantic Analysis)的方法，其原理乃利用奇異值分解，除了對於各意圖選擇最具鑑別性的詞，作為各意圖的主要語意詞(Semantic words)。

語言的多樣性經常導致收集樣本之不充分，如使用者想要掛號可能出現的語

句可能是「我要掛號」，「我想要掛號」，「我要預約掛號」..等等，但從語料的觀察，可以發現意圖通常與主要語意詞(Semantic word)有極高之共現率，其他功能性詞彙則有可能被省略，因此本文使用部分樣本據來建立意圖辨識模組，在這裡本論文將句子是為一連串的功能性詞彙和主要語意詞的組合，可表示如下：

$$S_i = \{FP_1^i, FP_2^i, \dots, FP_{NB_i}^i, SP^i, FP_{NB_i+1}^i, \dots, FP_{NB_i+NA_i}^i\} \quad (5)$$

在式子(5)中 SP^i 代表主要語意詞，而 FP_j^i 代表第j各功能性詞彙 NB_i 和 NA_i 為在主要語意詞前和在主要語意詞之後的功能性詞彙數。根據上述定義，部分樣本句為包含主要語意詞 SP^i 的子序列，所以最常的部分樣本句即為句子本身，而最短的部分樣本句則只有主要語意詞一個詞彙，而每一個功能性詞彙都有可能被省略，所以對式子(5)定義的句子共有 $2^{NA_i+NB_i}$ 句部分樣本句。舉例說明，若有一句子為“ABC”且 A, C 為功能性詞彙而 B 為主要語意辭彙，則有四句部份樣本句“ABC”，“AB”，“BC”以及“B”。

部分樣本樹是利用語料庫中的句子，將其分解成部分樣本句後，所建立的模組，它有兩點特點，第一點為具自動學習之能力，由訓練語料所得之的部分樣本句文法，第二點為可處理贅詞及部分詞彙錯誤的情形。因此根據收集到的訓練語料首先將訓練的句子分解成所有的部分樣本句，然後以樹狀結構將所有的部分樣本句的句型資訊儲存起來即為部分樣本樹。

在實際建立部分樣本樹的過程中，每一個內部節點代表部分樣本樹上的一個獨立詞彙，因此對於每一個內部節點本論文可以表示成

$$IN_i = \{PH_i, FR_i, NS_i, Son_i\}$$

包含的參數描述如下：

PH_i :此節點在部分樣本句上所代表的辭彙

FR_i :此節點在訓練語料出現的頻率

NS_i :其下所接的內部節點個數

Son_i :記錄所有到子節點的連結

在部分樣本樹中，外部節點代表著依據部分樣本句的結尾，因此可以利用外部節點很容易的回溯找到其所代表的部分樣本句。所以在此將外部節點表示為

$$EN_i = \{PP_i, Ptr_i, IT_i\}$$

其中 PP_i :代表此外部節點所表示的部分樣本句

Ptr_i :紀錄此部分樣本句是從哪些較完整的部分樣本句中因部分功能性詞彙
被省略而來

IT_i :記錄這條路徑所代表的意圖

整個訓練過程，總共有三個步驟來建立部分樣本樹：

- (1) 將訓練的句子斷詞成為一連串的词彙序列，對於每一具訓練語料中的主要語意詞在這個步驟都將其標記唯一特殊詞彙「Semantic word」，也就是說在訓練過程中，將所有主要語意詞都看成同一個詞彙。
- (2) 將斷好詞的句子拆解成部分樣本句。
- (3) 利用接下來介紹的演算法來建立部分樣本句。

部分樣本樹建立演算法：

步驟一： Initialization

設定部分樣本樹的根節點，R

步驟二： Recursion

對所有的部分樣本句， $PP_i = \{Ph_1^i Ph_2^i \dots Ph_{N_i}^i\}$ ， N_i 為部分樣本句

PP_i 的词彙個數，執行步驟 2.1 到步驟 2.5

步驟 2.1: 根據部分樣本句的词彙順序，由根節點 R 搜尋已建利立的部分樣本樹，最後停在節點 IN_s ，也就是說由根節點 R 到停止節點 IN_s ，這條路徑符合 PP_i 的 prefix

步驟 2.2: 若此部分樣本句上所有的词彙都已被搜尋到了，也就是此部分樣本句已經存在部分樣本樹中了，跳到步驟 2.4

步驟 2.3: 對於 PP_i 上還沒被搜尋到的词彙， Ph_k^i

步驟 2.3.1：建立新的內部節點， IN_N

步驟 2.3.2：設定此 IN_N 內所代表的词彙為 Ph_k^i

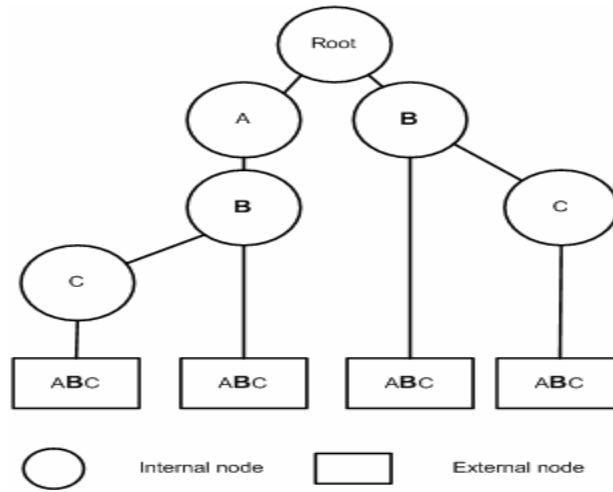
步驟 2.3.3：將 IN_N 加入 IN_s 的 Son 指標陣列中，並增加 IN_s
的 NS_s 個數

步驟 2.3.4：將 IN_s 設為 IN_N

步驟 2.4：若是此部分樣本句的外部節點不存在，則建立新的外部節點，

NE_N 並且設定 NE_N 所代表的部分樣本句為 PP_i

步驟 2.5：將在這條路徑上每一個內部節點 FP 參數加 1



圖六、建立好的部分樣本樹

圖六為訓練語句“ABC”所建立的部分樣本樹，其中“A”、“C”為功能性詞彙，“B”為主要語意詞，對上述部分樣本樹建構演算法說明，假設只考慮這句訓練語句“ABC”，首先對此句語句作前處理，先將句子“ABC”斷詞，並將主要語意詞標記為“Semantic word”代表這句子的語意，而在圖中粗體字 B 代表“Semantic word”的到結果「A, Semantic word, C」，再將斷好詞的句子拆解成所有的部分樣本句：「A, Semantic word, C」,「A, Semantic word」,「Semantic word, C」以及「Semantic word」，此時開始使用上述演算法建構部分樣本樹，第一步驟先設立好根節點 R 之後，步驟二對所有的部分樣本句一一將其加入部分樣本樹中，例如一開始先加入「A, Semantic word, C」這句，本論文利用 prefix 搜尋法，搜尋已建立好的部分樣本樹，因為一開始使沒有任何路徑存在，所以搜尋結果停留在根節點 R，此時依據未被搜尋到的詞彙順序，依序為「A」,「Semantic word」以及「C」，對每一個詞彙新增對應的內部節點，並且設定好節點間的父子關係，以及每一個內部節點的資訊，接下來因為代表此部分樣本句的外部節點尚未存在，所以本論文建立一個新的外部節點，設定其代表部分樣本句為「A, Semantic word, C」，就將第一句部分樣本句「A, Semantic word, C」加入了，而接下來當加入「A, Semantic word」這句部分樣本句時，會發現此路徑已經存

在部分樣本樹中了，因此不會增加新的內部節點，但是對每一個被搜尋過的內部節點其 FP 參數都加一，並且此時根據搜尋演算法，會停留在剛剛建立好代表詞彙為 B 的內部節點上，並且會發現其子節點並沒有外部節點，因此新增一個對應到「A, Semantic word」的外部節點，依此類推，將所有的部分樣本句加入後，就完成了步驟二，也就完成的部分樣本樹。

完成部分樣本樹的建構後，再利用部分樣本樹來偵測出意圖，在此只需對使用者的輸入與部分樣本樹中的路徑(Path)，找尋最相近似的路徑，就可以知道這句話的意圖。本論文將考慮兩個句子間的文法結構和語意相似度後，以動態規劃的方式來算出兩個句子的最佳相似分數。

假設 P_i 是使用者輸入的語句， $P_i = \{a_1, a_2, \dots, a_i\}$ ， P_T 是部分樣本樹的其中一條路徑可以被表示成 $P_T = \{b_1, b_2, \dots, b_j\}$ ，接著計算兩個句子間的結構相似度，所謂的結構相似度，就是指兩個句子之間的詞序是否相同，此定義如式子(6)

$$sim_{syn}(a_i, b_j) = \begin{cases} 0 & \text{if } a_i \neq b_j \\ 1 & \text{if } a_i = b_j \end{cases} \quad (6)$$

當在考量兩個句子之間的相似度時，除了從文法結構上來分析外，語意亦佔有重要的地位，在語意上來分析便要考慮到詞與詞之間的相關性，因此我們採用之前所建立的醫療概念模型來計算詞與詞之間的相似度。測試句與樣本樹的一條路徑中的一個詞分別為 a_i 跟 b_j 透過醫療概念模型來計算他們的相似度，在相似度的定義如式子(7)

$$sim_{sem}(a_i, b_j) = \begin{cases} 1 & \text{if } a_i = b_j \\ \left(\frac{1}{2}\right)^l & \text{if } a_i \text{ and } b_j \text{ are hypernyms} \\ 1 - \left(\frac{1}{2}\right)^n & \text{if } a_i \text{ and } b_j \text{ are synonyms} \\ 0 & \text{others} \end{cases} \quad (7)$$

在式(7)中 l 是指兩個具有上下位關係之概念在概念模型中距離多少層， n 是兩個同義詞間將其所有同意詞展開後，有多少個共同之同義詞。

在考慮到文法結構和語意的相似度後，將這兩項因素列入考慮來計算使用者的輸入句和部分樣本句之間的相似度。採用類似動態規劃的方法如式子(8)

$$\begin{aligned}
& sim_{int}(0,0) = 0 \\
& sim_{int}(i,j) = \max \begin{cases} sim_{int}(i-1,j-1) + (sim_{sem}(a_{i-1},b_{j-1}) + sim_{syn}(a_{i-1},b_{j-1})) \\ sim_{int}(i-1,j) + (sim_{sem}(a_{i-1},b_j) + sim_{syn}(a_{i-1},b_j)) \\ sim_{int}(i,j-1) + (sim_{sem}(a_i,b_{j-1}) + sim_{syn}(a_i,b_{j-1})) \end{cases} \quad (8) \\
& sim_{int}(P_i, P_j) = sim_{int}(I, J)
\end{aligned}$$

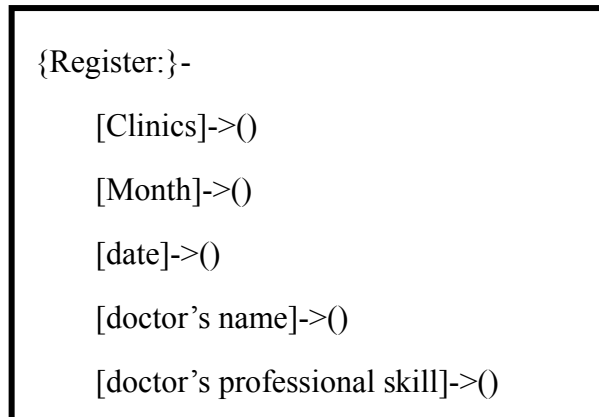
在式(8)中 $P_i = \{a_1, a_2, \dots, a_i\}$ 代表輸入的句子， P_i 是部分樣本樹的其中一條路徑可以被表示成 $P_T = \{b_1, b_2, \dots, b_j\}$ ， $sim_i(P_i, P_T)$ 代表它們之間的相似度，一開始先將其間的相似度設定為零，然後從兩個句子的第一個詞開始遞迴找尋最大相似度， $sim_{sem}(a_i, b_j)$ 是語意相似度， $sim_{syn}(a_i, b_j)$ 則是句法結構相似度。

以實例來說，系統輸入句為「我有點感冒」而樣本句有「我、有、發燒」，輸入句經過斷詞後，會得到「我、有、點、感冒」，首先第一個詞「我」對到「我」，相似度加二，第二個詞「有」對到「有」 相似度在加二變成四，第三個詞輸入句是「點」，和樣本句沒有相同的詞，因此本論文考慮「有」和「發燒」兩個詞，「點」和「有」的相似度為零，和「發燒」的相似度也為零，因此對到哪一個詞分數都一樣，因此本論文把「點」對到「有」，然後把相似度不變為四，再考慮下一個詞「感冒」，同樣的「感冒」和樣本句也沒有相同的詞，因此本論文考慮「有」和「發燒」兩個詞，「感冒」和「有」的相似度為零，和「發燒」的相似度為 1/2 因為「發燒」是「感冒」的下位詞它們之間相差一層，所以本論文就把「感冒」對到「發燒」，相似度加 1/2 變成 4 又 1/2 這就是此輸入和這句樣本句的相似度了，依此方法比對每句樣本句，找出最相近的樣本句後，就可以知道使用者這句話的意圖。

4. 對話控制模組

本節說明如何整合各項服務，及個別模組之功能說明，在前面已經解釋如何從使用者的輸入擷取出使用者的意圖，在有了使用者的意圖之後，系統便可偵測使用者是要使用那樣模組，再配合各項模組內的 Semantic Frame 產生相關的對應的回應。

在各個模組使用相對應之語意框架(Semantic Frame)相配合之來控制對話流程[23]，以實例來說明，在掛號諮詢模組中的語意框架如圖七所示：



圖七、語意框架

(1)掛號諮詢模組，主要功能是讓為協助使用者完成線上掛號，因此需要一個可提供查詢的線上資料庫，本文是以成大醫院資訊做為系統的線上掛號資料庫，提供掛號所需的各項資訊。

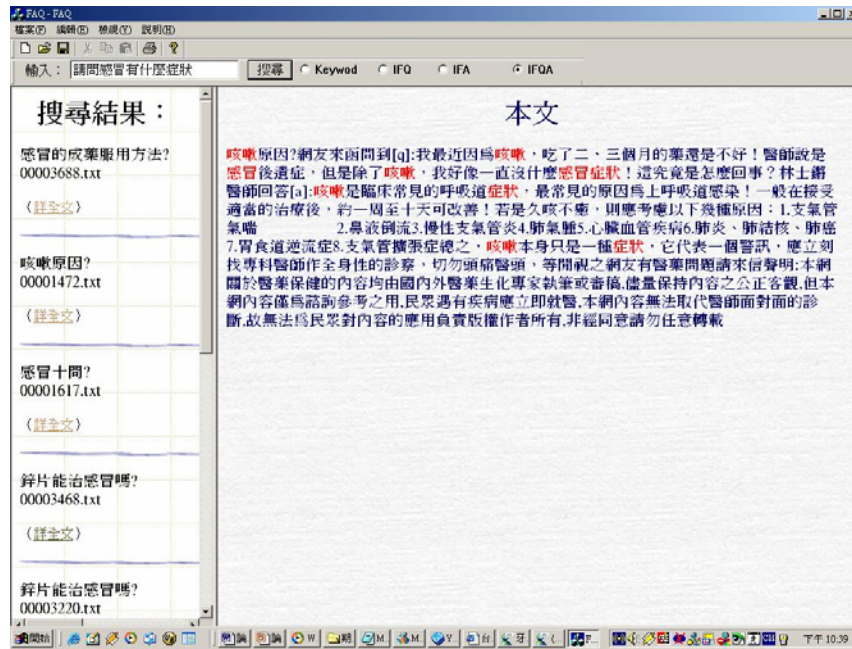
(2)此模組提供科別諮詢功能，所謂的科別諮詢功能即是引導使用者來掛號適當的科別，有病痛時，並不是每次都知該去找尋什麼樣子的醫生或掛什麼科別，本模組提供之前建立好的推論原則，來找出最適合的科別，來節省使用者的時間和醫療資源。



圖八、推論規則詳細說明

疾病發生頻率本論文總共分為 A, B, C 三個等級分別代表常發生、可能發生以及少發生，主要症狀只有一個代表此疾病最主要之症狀，次要症狀則是伴隨疾病發生的機會次之，其他症狀則代表可能但隨個人體質不同而伴隨的症狀，看診科別則是此疾病應當看哪一種科別，最後一個欄位的疾病緊急程度分為 a, b, c 三個等級，a 代表很緊急應當急診，b 代表緊急應該趕緊就醫，c 則代表普通。

(3)FAQ 諮詢模組：本論文也提供相關醫療資訊給使用者查詢，語料是從網路上收集下來，其系統界面如圖九所示：



圖九、FAQ 系統介面圖

5. 實驗與討論

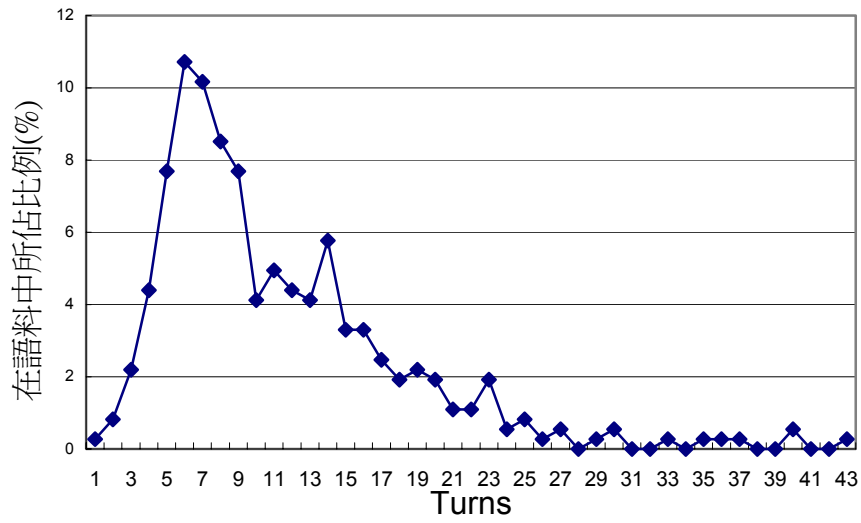
實驗所使用的機器為 Pentium IV 2G 的個人電腦 512MB RAM，開發的工具是 Microsoft Visual C++ 6.0，ASP 和 IIS5.0 版，在 Windows2000 的作業系統下進行開發與實驗。

5.1 對話語料分析

在本節將對所收集的語料進行分析，本論文的語料共分為三類，一是實際對話語料，二是 WOZ 收集來的語料，三是系統實際測試時所收集回來的語料，首先分別對各個語料對使用者一次對話的長度作分析，即是一次對話需要來回多少次(turns)，其結果如下：

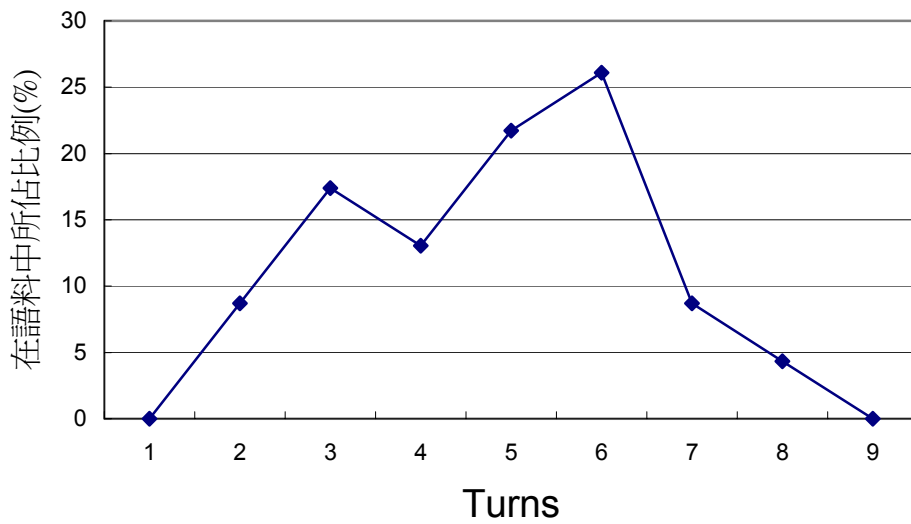
5.1.1. 電話語料

圖十中橫軸是一次對話的對話長度，縱軸是長度所佔語料的比例，在電話語料部分共收集了 4089 比對話資料(turns)、為 364 人次累計出來的，每人平均的對話次數為 11.235 次，其最長的一回對話為 47 回合，從圖十可以看出來對話長度約在 6~14 之間，平均對話長度拉長的原因是有不少特別長的對話。



圖十、電話語料長度分佈

5.1.2. WOZ 語料

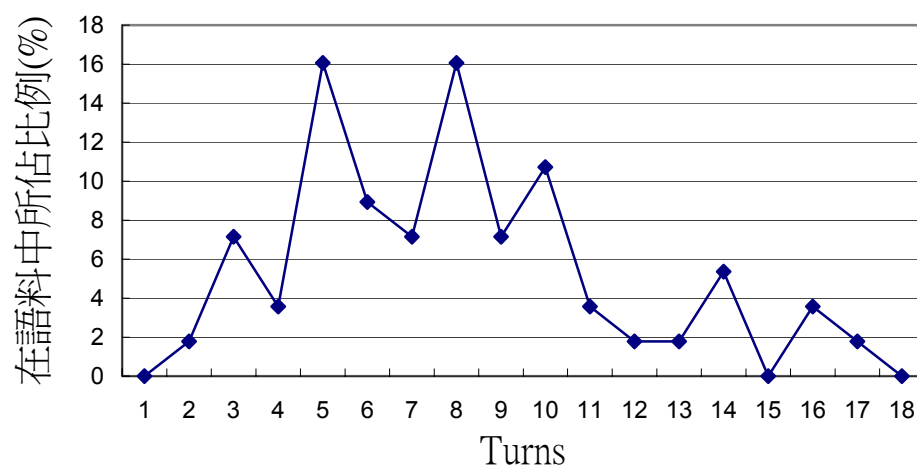


圖十一、WOZ 語料長度分佈

在 WOZ 語料部分共收集了 234 比對話資料(turns)、為 34 人次累計出來的，每人平均的對話次數為 6.882 次，其最長的一回對話為 10 回合，從圖十一可以看出來大多對話次數約在 6~8 之間。

5.1.3. 實際系統語料

在系統測試語料部分共收集了 500 筆對話資料(turns)、為 56 人次累計出來的，每人平均的對話次數為 8.928 次，其最長的一回對話為 16 回合，從圖十二可以看出來大多對話次數約在 5~8 之間，但與圖十與圖十一不同的是，圖十二有兩個高峰點分別在對話長度為 5 和 8 時，測試語料會有這樣狀況的原因是系統提供三項服務，因此使用者使用不同服務時對話的長度也就不一樣，其功能與對話長度的關係如表一所示：



圖十二、系統測試語料長度分佈

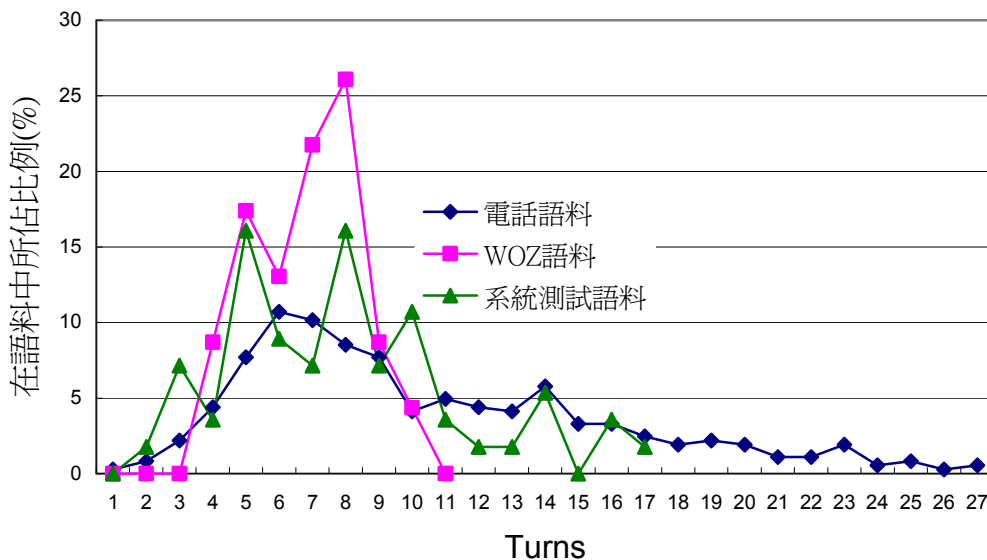
Module	Registration Module	Clinic Query Module	FAQ Module	Integrated System
Average Number of Turns	8.40	9.27	4.80	9.80

表一、各功能模組平均對話長度

從表一，我們可看出常見問答集模組的平均對話長度為 4.80，而掛號諮詢模組的平均對話長度為 8.4，因此我們可推斷圖十二中，第一個高峰是使用者使用常見問答集模組的狀況，第二個高峰則是使用其他模組或者混合的狀況。

將三種語料比較之後，我們可歸納出在電話語料的對話度最長為 11.235 次，因為使用者打電話進來後，常常有些贅詞，如：恩、喂、嘿...等等，就會形成一

次對話，在 WOZ 收集時因為目標明確且由人來回答，使用者可以快速達到目標因此平均對話長度為 6.882 次，系統測試時，因為實際對話流暢度不如人類之間的對話，因此需要較多次的嘗試後，使用者才能達成目標，平均對話長度為 8.928 次，其三種語料的分佈圖，如圖十三所示：



圖十三、對話長度分佈

4.2. 對話系統評估

在評估整個對話系統方面，我們請五十個未參與本研究之大專生來測試系統，並參考[1][24]的方法來評估整個對話系統，分別對各個模組計算對話成功率 (Task Success Rate)、平均對話長度 (Average Number of Turns) 以及答句適切度 (Contextual Appropriateness)，為了評估服務的整合，加上一個意圖偵測的正確率 (Intention detection Rate) 作為其指標，其結果如表二所示。

從表二中可得知，從對話次數來分析，在使用 FAQ 模組時所需對話次數最少，而混合使用時對話次數最多。意圖偵測部分由於掛號諮詢模組與科別建議模組意圖則較容易混淆，因此偵測正確率較低。而整體系統整合時亦然，對話成功率部分，在科別建議模組與整體系統的成功率較低，其原因為在科別建議時需要使用者提供症狀，而症狀的描述方式有非常多種如「頭痛」可以描述成「頭有一點痛」、「我左邊的太陽穴附近會痛」等等，因此導致處理複雜度提高。而整體系

Evaluation Parameters	Intention detection Rate(%)	Task Success Rate(%)	Average Number of Turns	Contextual Appropriateness(%)
Registration Module	87.3%	92%	8.40	82.0%
Clinic Query Module	84.6%	80%	9.27	75.1%
FAQ Module	92.4%	88%	4.80	85.2%
Integrated System	80.6%	68%	9.80	74.3%

表二、系統效能表

統處理時成功率下降為 68% 的原因有二：一為自然語言的混淆度而導致，二為意圖偵測錯誤，但這所佔的影響度較小。因此我們可針對每次對話收集的語料再次進行系統改進而改進整體效能，如表三即為第二次改進後的實驗結果：

由表三可明顯看出整體系統意圖偵測正確率由 80.6% 提升到 86.2%，對話長度由 9.8 次降為 9.2 次，系統成功率也由 68% 提升為 77%，這主要的效能提升，是由意圖正確率的提升以及自然語言處理能力的增強所提升。

Evaluation Parameters	Intention detection Rate(%)	Task Success Rate(%)	Average Number of Turns	Contextual Appropriateness(%)
Registration Module	92.4%	95%	7.30	85%
Clinic Query Module	90.4%	82%	8.20	79.4%
FAQ Module	94.1%	92%	4.70	88.8%
Integrated System	86.2%	77%	9.20	78.5%

表三、系統效能表 Evaluation2

6. 結論與未來展望

在本論文中，我們建立一套整合多項服務的醫療查詢對話系統，所整合的服務有掛號資訊諮詢、科別資訊諮詢以及常見問答集諮詢三項大服務。使用意圖偵測來整合服務，在意圖偵測部分則使用部分樣本樹作為判斷意圖之依據，建立部分樣本樹與幫助語言理解必須借重醫療概念模型之推論與概念的描述，透過實驗證明，系統服務成功率為 77%，充分說明了本文所提之方法是具體可行的，但仍有下列問題有待改進：

1. 在醫療概念模型抽取部分，本論文使用醫療領域的語料來找出在整個概念模型結構上屬於醫療領域的節點，但是前端的斷詞系統，並不能對句子斷出專有名詞及新詞尤其是屬於領域內之概念詞，因此造成增加雜訊以及屬於醫療領域的節點沒有被找出，因此如能在這部分統計屬於醫療領域的節點時，先做新詞偵測再斷詞，會有效提升醫療概念模型的抽取結果。
2. 對於科別建議模組部分，在症狀的描述上隨著使用者的口語化而難以讓系統理解，因此需要提出一套能漸進式的找出使用者可能症狀的方法，如此才能有效提升科別建議模組的效能。
3. 對於整個對話系統演進的部分，目前本論文的系統再做過實驗後，除了意圖偵測那一部份，其他部分的效能改進仍然需要不少人力的介入，但對於對話系統而言不斷的演進是必要的，因此如何將演進所需要介入的人力降低就是另一個重要的課題。

致謝

本研究承蒙成功大學王獻章先生提供於成大醫院實際蒐集之語料以及經濟部工研院電通所前瞻技術中心計畫經費補助特此致謝。

參考文獻

- [1] Michael F. McTEAR, "Spoken Dialogue Technology: Enabling the Conversational User Interface," ACM Computer Surveys, Vol 34, No. 1, March 2002, pp.90-169.
- [2] James F. Allen, Donna K. Byron, Myroslava Dzikovska, George Ferguson, Lucian Galescu, Amanda Stent, "Towards Conversational Human-Computer

- Interaction,” AI Magazine, 2001.
- [3] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, “Spoken Language Processing”, Prentice-Hall Inc, 2001.
 - [4] James Allen, “Natural Language Understanding”, The Benjamin/Cummings Publishing Company. 1994.
 - [5] Victor Zue, Stephanie Seneff, James Glass, Joseph Polifroni, Christine Pao, Timothy J. Hazen, Lee Hetherington, "JUPITER: A telephone-based conversational interface for weather information," IEEE Trans. on Speech and Audio Processing, vol. 8, no. 1, January 2000, pp.85-96.
 - [6] AT&T, “How may I help you?” <http://www.research.att.com/~algot/hmihy/>
 - [7] S. Bennacef and L.Lamel et al., “Dialogue in the RAILTEL Telephone-Based System,” ICSLP’96 Vol. 1. pp. 550-553
 - [8] L. Lamel, S. Rosset, J. Gauvain, S. Bennacef, M. Garnier-Rizet , and B. Prouts, “The LIMSI Arise System,” Speech Communication, 31(4), 2000, pp. 339-353.
 - [9] Kuei-Kuang Lin, Hsin-Hsi Chen,” A Semiautomatic Knowledge Extraction Model for Dialogue Management”, Master thesis of department of computer science and information engineering, National Taiwan University, 2002
 - [10] Bor-shen Lin, Hsin-min Wang, and Lin-shan Lee, "A Distributed Agent Architecture for Intelligent Multi-Domain Spoken Dialogue Systems," IEICE Trans. on Information and Systems, E84-D(9), pp. 1217-1230, Sept. 2001.
 - [11] Tung-Hui Chiang, Chung-Ming Peng, Yi-Chung Lin, Huei Ming Wang and Shih Chieh Chien, “The Design of a Mandarin Chinese Spoken Dialogue System,” in Proceedings of COTEC’98, Taipei 1998, pp. E2-5.1~E2-5.7
 - [12] Chung-Hsien Wu, Gwo-Lang Yan, and Chien-Liang Lin, “Speech act modeling in a spoken dialogue system using a fuzzy fragment-class Markov model,” Speech Communication, Vol. 38, 2002, pp183~199.
 - [13] Jhing-Fa Wang and Hsien-Chang Wang, “A Portable Auto Attendant System with Sophisticated Dialogue Structure”, Journal of Information Science Engineering, Vol.18, No.4, July. 2002, pp627-636
 - [14] Jhing-Fa Wang and Hsien-Chang Wang” Experiences of Multi-Speaker Dialogue System for Mobile Information Retrieval,” Workshop on DSP in Mobile and Vehicular Systems, April 3-4, 2003, Nagoya, Japan
 - [15] Jhing-Fa Wang and Hsien-Chang Wang, Chieh-Yi Huang, Chung-Hsien Yang” Multi-Speaker Dialogue for Mobile Information Retrieval,” ISCSLP 2002, August 23-24, 2002, Taipei
 - [16] Steffen Staab, Rudi Studer, Hans-Peter Schnurr, York Sure, “Knowledge Processes and Ontologies,” IEEE Intelligent Systems 16 (1) January February 2001 ,pp. 26-34

- [17] J. F. Sowa, Knowledge Representation: Logical, Philosophical and Computational Foundations, Brooks Cole Publishing Co.1999.
- [18] J.I, Chu-Carrol, "MIMIC" An adaptive mixed initiative spoken dialogue system for information queries," Proceedings of the 6th ACL article on Applied Language Processing, pp. 97-104
- [19] Lee, S.H., Lee H. and Kim, J.H. 1995 On-line Cursive Script Recognition using an Island-Driven Search Technique, Proceedings of the Third International Conference on Document Analysis and Recognition, Volume: 2 , 14-16 Aug. 1995. 886 -889.
- [20] Chung-Hsien Wu, Yeou-Jiunn Chen, and Cher-Yao Yang, "Error Recovery and Sentence Verification Using Statistical Partial Pattern Tree for Conversational Speech," in Proceedings of ICSLP2000, Beijing, China, 2000.
- [21] A. Miller, "WordNet: An On-Line Lexical Resource," J. Lexicography, vol.3, no.4, Dec.1990.
- [22] 董振東, 董強, "知網," <http://www.keenage.com/>
- [23] N. Fraser and G. N. Gilbert, "Simulating speech system," Computer Speech and Language 5, 1991 pp.81-99.
- [24] M. A. Walker, D. Litman, C. Kamm, and A. Abella, PARADISE: a general framework for evaluating spoken dialogue agents. In Proceedings of the 35th Annual General Meeting of the Association for Computational Linguistics, ACL/EACL (Madrid, Spain). ACL, 1997, pp.271–280.