

# 植基於自動學習之相關查詢詞推薦：應用於心血管疾病資訊檢索

## Relevant Term Suggestion Based on Automatic Learning: Application to Cardiovascular Disease Information Retrieval

邱曉莉

Hsiao-Li Chiu  
慈濟大學醫學資訊學系  
94325106@stmail.tcu.edu.tw

林紋正

Wen-Cheng Lin  
慈濟大學醫學資訊學系  
denislin@mail.tcu.edu.tw

陳郁志

Yu-Jr Chen  
佛教慈濟綜合醫院心臟內科  
chensmed@hotmail.com

王志鴻

Ji-Hung Wang  
佛教慈濟綜合醫院心臟內科  
jihung\_wang@tzu.chi.com.tw

### 摘要

當一般民眾藉由搜尋引擎在網際網路上搜尋醫療健康資訊時，礙於對醫學詞彙的認知有限，無法使用精確的查詢來描述需求，導致搜尋的結果並不理想。故本研究使用查詢詞推薦來輔助使用者建立更精確的查詢，讓使用者能找到需要的資訊。我們藉由自動學習的方法，自心血管疾病語料庫訓練出相關詞。當使用者輸入一個查詢詞後，系統會列出與此查詢詞相關的詞，讓使用者選擇有用的詞來擴展查詢。我們評估系統推薦的查詢詞與原始查詢詞的相關程度，實驗結果發現系統建議的前 10 個詞中，約有 76.7% 是與原查詢詞很相關的。此結果顯示本系統可以提供豐富的相關查詢詞供使用者選擇，以建立更好的查詢。

**關鍵字：**健康資訊檢索、查詢詞推薦、查詢擴展、機器學習

### Abstract

*Internet has become an important medium for health consumers to search for medical health information. Due to the limited knowledge of medical vocabulary, consumers usually use short and general queries which cannot reflect their information needs. To help consumers search for cardiovascular disease information, we adopt query term suggestion to assist consumers in constructing more specific queries. The suggested terms*

*are automatically learned from a corpus consisting of documents about cardiovascular diseases. Experimental results showed that in the top 10 recommended terms, about 76.7% are relevant to the original query term. It shows that our system can provide rich relevant terms to help users construct more accurate queries.*

**Keyword:** Health information retrieval, Query term suggestion, Query expansion, Machine learning

### 1、前言

現今網際網路已經成為一個重要的資訊交流的媒體，全世界每天有上千萬人利用搜尋引擎(search engine)在網際網路上找尋他們需要的資訊。由於網際網路的便捷及資料豐富，有越來越多的網路族群，如病患及其家屬，也開始利用網際網路找尋醫療健康資訊。根據 Pew Internet & American Life Project 的報告[5]，在美國約有八成的上網者曾經在網際網路上找尋醫療健康資訊。Fox[5]的調查也指出，大部份(66%)在網際網路上找尋醫療健康資訊的民眾會找尋特定疾病相關的資訊，約有一半的人(51%)會找尋疾病的治療方法，由此可見關於特定疾病的介紹是大部分民眾所需要的。在眾多疾病中，心血管疾病每年皆列國內十大死因之一，且有罹病人口遽增、年齡層下降之趨勢，因此民眾利用網際網路搜尋心血管疾病相關資訊之需求亦趨倍增。

要從網際網路上大量的資料中找到所要的資訊並不是件容易的事，因此大部分的使用者藉由搜尋引擎的協助，以便快速又方便地搜尋所需的資訊。搜尋引擎是一種資訊檢索(information retrieval)系統，其主要目的在於從大量文件中找出符合使用者需求的文件[3]。使用者在檢索文件時所用的查詢通常是由幾個詞或是句子所組成，系統便使用查詢裡的詞語與文件作比對，以找出相關性高的文件。如果尋求資訊的使用者沒有使用與文件作者相同的詞彙來描述相同的概念，很容易檢索失敗，因此資訊檢索常遭遇的困難之一就是「字彙不匹配」(vocabulary mismatch)的問題[2]。另外一個資訊檢索常見的問題是查詢太短。一般人都是使用關鍵字(keywords)到搜尋引擎裡搜尋，而且大部分的查詢都是很短的，平均不到三個字[10][11]。這些短的查詢通常是不精確的，導致查詢條件不足，使搜尋出來的結果範圍很廣，一些不是使用者要的文件也會被搜尋引擎找出來回傳給使用者。

根據Health On the Net Foundation<sup>1</sup> (HON) 在 2002 年的調查報告[6]，在網際網路上找尋醫療資訊的使用者中，有 26.33%認為資訊的正確性是最重要的，接下來是可信賴性(11.45%)及資訊的找尋(10.99%)等問題。針對搜尋工具所應具備的特性，有 45.01%的使用者認為找到相關的資訊是最重要的。由此可見民眾期望有好的工具協助他們取得所需的醫療健康資訊。

一般民眾在搜尋醫療健康資訊時，由於對醫學詞彙的認知極為有限，常常使用不精確的查詢，造成所使用之查詢詞與欲搜尋文件內之用語不相符合，或是搜尋結果所涵蓋的範圍太廣，導致搜尋結果無法滿足民眾的需求。如果有工具可以協助使用者建立好的查詢，便可以增進檢索的效能，幫助民眾找到相關的資訊[9]。故本文研究相關查詢詞推薦(query term suggestion)的方法來協助使用者建立好的查詢。我們利用自動學習的方法從語料庫中擷取出相關詞推薦予使用者選擇。當使用者在搜尋心血管疾病相關資訊時，鍵入查詢詞後，系統會自動推薦相關查詢詞給使用者選擇以建立更精確的查詢，此一新的查詢再送至搜尋引擎搜尋，如此有利使用者更快速及正確地找到所需

要的醫療健康資訊。

## 2、查詢詞推薦

有許多研究發展了多種技術去改善資訊檢索的效能，其中一個技術就是查詢擴展(query expansion)，也就是增加額外的詞彙到原來的查詢裡，以建立新的查詢[3]。查詢擴展通常都以使用者提供的查詢為基礎，就是當原始查詢的檢索效能不好時，可以追加與原始查詢相關的詞彙來改善。如果原始查詢因字彙不匹配的問題而無法找到相關文件時，系統可以將原始查詢詞彙的同義詞，也就是描述相同概念的其他詞彙，加入原始查詢，以便提高與文件內詞彙匹配的機會。而如果原始查詢太短不夠精確，也可以藉由查詢擴展，將原查詢詞彙的相關詞加入查詢，增加查詢條件，以縮限搜尋結果的範圍。

擴展詞的來源，主要分成二大類：一是從文件中自動擷取出來，一是從人工建立的語言知識庫中獲得。從文件中擷取出擴展詞的方法通常是利用機器學習(machine learning)的方法，在文件集中找出詞與詞或是詞與查詢語句的相關程度，再根據相關程度來決定要用哪些詞來擴展查詢。擴展詞的另一個來源是人工建立的語言知識庫，包括字典、同義詞辭典、知識本體(ontology)等。這些知識庫是根據專家的知識建立而成，與從文件集中自動擷取出來的知識相比，正確性是相當的高。不過使用人工建立的知識庫也有一些缺點，例如建立知識庫費時費力、知識庫涵蓋的範圍可能不夠廣泛、更新較慢等。

以使用者的角度來看，查詢擴展技術可分為半自動與自動的方式。自動化的查詢擴展技術是不需依賴使用者的介入，由系統自行找出新的查詢詞彙加入原始查詢中以形成新的查詢。而半自動(或是互動式)的查詢擴展技術則是要求使用者介入，協助系統作查詢擴展。其中一種常用的方法是使用者相關回饋(user relevance feedback)[3]，首先系統使用原始查詢進行檢索，然後使用者對檢索所得的文件進行相關性判斷後再回饋給系統，以便從檢索出的文件中找出相關詞彙，以修飾原始查詢，再次進行檢索。而使用者與系統的互動，將持續到使用者滿意其檢索結果後才結束。另一種互動式的查詢擴展方法是相關詞推薦，由

<sup>1</sup> <http://www.hon.ch/>

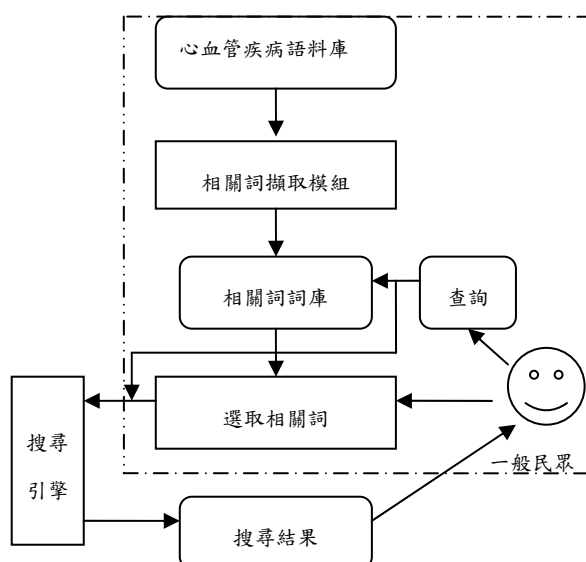


圖 1：相關查詢詞推薦系統架構圖

系統找出與原始查詢相關的詞彙，推薦給使用者選擇，以作為加入原始查詢的擴展詞。半自動查詢擴展的主要優勢為減輕使用者重新組織查詢語句的負荷，但其缺點為使用者在得到良好檢索結果之前，必須花時間與精神來與系統互動[8]。有研究指出使用互動式的查詢擴展來找出相關詞去擴展原來的查詢可以改進資訊檢索系統的效能[7][9]。

本研究使用相關詞推薦的方法來協助使用者建立好的查詢。相關詞的來源則是用自動學習的方法從語料庫中擷取而得。查詢詞推薦系統的架構如圖 1 所示。首先建置心血管疾病語料庫，然後從中訓練出相關詞，並將所擷取出之相關詞儲存於相關詞詞庫內。當使用者輸入一個中文查詢，經系統比對相關詞詞庫後便列出與原始查詢相關的詞彙，建議給使用者選擇有用的詞來擴展查詢，以建立更精確的查詢，最後再將新的查詢送至搜尋引擎去搜尋。

### 3、自動學習相關詞

#### 3.1、建立心血管疾病語料庫

我們蒐集與心血管疾病相關的醫學報導、網路醫療問答、醫院衛教資訊等類型的文章，然後由專科醫護人員確認為正確的心血管疾病相關資訊共 240 篇文章。另外由花蓮慈濟醫院心臟內科主治醫師協助建立各種心血管疾病的簡介，內容大致包含有疾病名稱、定義、

```

<?xml version="1.0" encoding="Big5">
<DOC>
<SOURCE>花蓮慈濟醫院心臟內科</SOURCE>
<CLASS>心血管病知識</CLASS>
<AUTHOR>心臟內科主治醫師 陳維志醫師</AUTHOR>
<AFFILIATION>花蓮慈濟醫院心臟內科</AFFILIATION>
<DATE>2006/10</DATE>
<DISEASE>
<NAME>心律不整(Arrhythmia)</NAME>
<DEFINITION>
什麼是心律不整?
心臟的跳動是規律的，並隨著身體狀況(活動量、情緒)加快或減慢。通常每分鐘心臟規則跳動60到100次。
心臟的規律和規則跳動是靠心臟的電氣傳導系統控制。如果心臟的電氣傳導系統出現問題，就可能造成心跳不正常。當心跳過慢、過快
</DEFINITION>
<CAUSES>
心律不整的原因有哪些?
造成心律不整的原因很多，有冠状動脈心臟病、內分泌疾病如甲狀腺亢進、神經系統疾病、肺部疾病、電解質異常、精神疾病、藥物戒藥
等、可樂、咖啡等提神飲料中含有咖啡因或興奮劑影響自主神經系統，並影響電氣傳導系統進而引起心律不整。
</CAUSES>
<FACTORS>
以下狀況有可能引發心律不整
1.缺血性心臟病
2.心臟瓣膜疾病
3.電解質之不平衡
4.另外有一些藥物及飲料也可能產生心律不整
</FACTORS>
<SYMPTOMS>
心律不整有哪些症狀?
心律不整可以至無症狀或感覺心跳不規則、心悸、呼吸不順、胸悶、頭暈腳輕、四肢無力、昏厥，嚴重者甚至直接發生猝死。
</SYMPTOMS>
<EXAMINATION>
心律不整的理學檢查
醫生通常會先聽診。
</EXAMINATION>
<TESTS>
心律不整的檢查有哪些?
1.心電圖
2.二十四小時心電圖
3.超音波
4.電氣生理
5.心導管
</TESTS>
<COMPLICATIONS>
心律不整的合併症有哪些?
心律不整如果太慢或太快可引起心輸出量不足而造成生命危險，嚴重可幾分鐘內死亡。
</COMPLICATIONS>
<TREATMENT>
<PROCESSING>
心律不整要如何治療?
輕微的心律不整可能不需治療而不需服藥。有些心律不整只需發作時服藥。
嚴重的心律不整則可能需急救，以電擊或其他方法挽救生命。
</PROCESSING>
<PREVENTION>
如何預防心律不整?
生活壓力和有些食品會影響自主神經系統並改變心律，因此平時儘量少喝含咖啡因或興奮劑飲料如茶、可樂、咖啡。
盡量勿吸煙、勿喝含酒飲料。保持生活正常。嚴重的心律不整會導致猝死，遇有心律不整的現象時，應儘早就醫。
</PREVENTION>
</DISEASE>
</DOC>

```

圖 2：「心律不整」的 XML 文件

疾病致因、危險因子、症狀、理學檢查、檢驗報告、合併症、治療方式、預防方法、護理指導等，還有與心血管疾病相關的醫學知識如常見檢查介紹、常見症狀介紹等資訊共約有 24 篇文章。醫師所撰寫之心血管疾病相關知識，其內容以一般民眾容易閱讀及了解為主，可以提供給一般民眾參考。心血管疾病語料庫便由上述的文章所組成。

為了便利文件處理，及讓這些文件未來能重複使用，我們將文件標記為 XML(Extensible Markup Language)格式。我們定義了一組標籤集(tag set)，各個標籤(tag)分別對應到上述的各類資訊。另外還有一些標籤標示文件的描述資料(metadata)，包括來源出處、文件類別、作者、單位、日期等。以心律不整為例，圖 2 為標記後的 XML 文件。

#### 3.2、自動訓練相關詞

建置好心血管疾病語料庫之後，我們利用自動學習的方法，從心血管疾病相關的文件中訓練出相關詞。機

器學習可以依預先設定的關係、學習演算法、物件特徵設定、文件集類型等因素，從文件集中找出不同類型的相關詞。例如利用共現資訊(co-occurrence information)，我們可以從醫學相關的文件中找出常常一起出現的詞彙。

首先我們用中文斷詞系統[1]將訓練文件(即心血管疾病語料庫之文件)斷詞，把各個中文詞區隔開，以利後續處理。由於此斷詞系統並非針對醫學領域而設計，所以有許多醫學專業詞彙被斷錯，我們便手動將這些錯誤更正。斷詞之後接著我們使用相互訊息(mutual information)[4]來計算詞與詞之間的相關性。在資訊理論中，相互訊息是用來計算已知一個變數的情形下，關於另一個變數的不確定性所減少的量，也就是量度兩變數間的普遍相關性。在統計式自然語言處理的研究中，相互訊息常被用來計算詞彙與詞彙之間的相關性。如果二個詞之間的相互訊息很高，我們就認為這二個詞之間有相當程度的關連性，也就是互相可視為相關詞。如果二個詞之間的相互訊息趨近於零分，表示這二個詞之間沒有什麼關係，是互相獨立的。在本研究中，二個詞 $w_1$ 與 $w_2$ 之間相關性的大小是用下列相互訊息的公式來計算，所擷取出來的相關詞會存入相關詞詞庫。

$$MI = P(w_1, w_2) \times \log \frac{P(w_1, w_2)}{P(w_1)P(w_2)} \quad (1)$$

其中 $P(w_1, w_2)$ 為 $w_1$ 及 $w_2$ 同時出現的機率， $P(w_1)$ 及 $P(w_2)$ 分別為 $w_1$ 與 $w_2$ 的機率。

### 3.3、相關查詢詞推薦

當使用者輸入一個查詢詞後，系統會找出與此查詢詞的相互訊息為正分的詞，從分數最高的開始依序列出，讓使用者可以選擇有相關的詞來擴展查詢，如圖3及圖4所示。

使用者在選擇認為相關的查詢詞後，便可以送到搜尋引擎去搜尋相關資料。例如在圖4中，使用者選擇了“電燒術”、“治療”、“電氣燒灼術”三個詞來擴充查詢，然後新建立的查詢被送到Google去搜尋，找出與使用電氣燒灼術治療心律不整有關的網

頁，如圖5所示。



圖 3：系統的網頁介面

#### 查詢結果

原始查詢詞：心律不整

- |                                         |                                           |                                |
|-----------------------------------------|-------------------------------------------|--------------------------------|
| <input checked="" type="checkbox"/> 電燒術 | <input type="checkbox"/> 心跳               | <input type="checkbox"/> 嚴重    |
| <input checked="" type="checkbox"/> 治療  | <input type="checkbox"/> 傳導               | <input type="checkbox"/> 心電圖   |
| <input type="checkbox"/> 猝死             | <input type="checkbox"/> 心悸               | <input type="checkbox"/> 休克    |
| <input type="checkbox"/> 節律             | <input type="checkbox"/> 合併症              | <input type="checkbox"/> 導管    |
| <input type="checkbox"/> 記錄器            | <input type="checkbox"/> 致命性              | <input type="checkbox"/> 電極    |
| <input type="checkbox"/> 電燒             | <input type="checkbox"/> 成功               | <input type="checkbox"/> 二十四小時 |
| <input type="checkbox"/> 心搏過速           | <input checked="" type="checkbox"/> 電氣燒灼術 | <input type="checkbox"/> 不正常   |
| <input type="checkbox"/> 發作             | <input type="checkbox"/> 根治               | <input type="checkbox"/> 電流    |
| <input type="checkbox"/> 心搏             | <input type="checkbox"/> 檢查               | <input type="checkbox"/> 病人    |
| <input type="checkbox"/> 昏厥             | <input type="checkbox"/> 意識               | <input type="checkbox"/> 引起    |

至HMSE查詢

至Google查詢

重新查詢

圖 4：“心律不整”的推薦查詢詞



圖 5：擴展後之查詢送至搜尋引擎搜尋的結果

<sup>2</sup> <http://www.google.com.tw/>

#### 4、實驗結果

本研究之主要目的是希望一般民眾透過相關查詢詞推薦系統，選擇與原始查詢相關的詞，來建立更精確的查詢，如此可以協助使用者搜尋到真正需要的資訊。如果系統推薦很多與原始查詢相關的詞，使用者就越有可能建立出好的查詢。相反地，如果系統所推薦的都是不相關的詞，則使用者就很難利用這些推薦詞來建立查詢。因此系統所推薦的詞是否與原始查詢相關是系統效能的一個重要指標。我們設計了一個實驗來評估系統所推薦的查詢詞與原查詢之間的相關程度。

根據行政院衛生署國民健康局網站<sup>3</sup>統計的熱門查詢及向心血管疾病病患及其家屬調查的結果，我們使用 12 個民眾最常用的與心血管疾病相關的查詢詞來測試系統。這些查詢詞分別為高血壓、冠心病、心血管疾病、心肌梗塞、心律不整、胸痛、心導管、肥胖、糖尿病、戒煙、中風、高膽固醇。我們請三位花蓮慈濟醫院專業護理人員針對每一個測試查詢詞判斷系統所推薦的前 30 個詞是否與原查詢詞很相關、部分相關、或是不相關。每一個推薦詞的相關程度由三位護理人員中的多數意見決定。如果三位專業護理人員所評定的相關程度皆不相同，無法決定系統推薦的查詢詞與原查詢詞的相關程度，則由第四位護理人員針對該推薦詞進行評估，以決定相關度。例如系統針對查詢詞“心血管疾病”所推薦的一個詞是“機率”，由三位護理人員評定的相關度分別為很相關、部分相關、不相關，因此無法決定相關程度，故由第四位護理人員針對此項推薦詞評定與原查詢詞的相關度，評定結果為很相關，因此“機率”與“心血管疾病”的相關程度是很相關。

我們計算系統推薦的前 10、20、30 個詞中，相關詞所佔的比率有多少。我們採用兩種標準來認定相關詞，一種是嚴格的標準，只有被認定為很相關的推薦詞才視為相關詞；另一種是比較寬鬆的標準，只要是很相關及部分相關的推薦詞都算是相關詞。實驗結果如表 1 及表 2 所示，表 1 為在所有推薦詞中很相關的詞所佔的比率，表 2 為在所有推薦詞中很相關及部份

表 1：系統評估結果－很相關的詞所佔的比率

|        | document | sentence | word   |
|--------|----------|----------|--------|
| Top 10 | 53.33%   | 76.67%   | 60.00% |
| Top 20 | 53.33%   | 66.25%   | 59.52% |
| Top 30 | 47.78%   | 62.50%   | 59.56% |

表 2：系統評估結果－很相關及部份相關的詞所佔的比率

|        | document | sentence | word   |
|--------|----------|----------|--------|
| Top 10 | 75.83%   | 87.50%   | 69.17% |
| Top 20 | 73.75%   | 81.67%   | 70.23% |
| Top 30 | 68.89%   | 78.89%   | 70.51% |

相關的詞所佔的比率。在計算相互訊息時，我們分三個範圍計算共現資訊：同一篇文章(document)，同一個句子(sentence)，前後 1 個字(word)。我們比較不同共現範圍對系統效能的影響，表 1 及表 2 中各欄分別列出了不同共現範圍的評估結果。

實驗結果發現，將共現資訊的計算範圍設定在一個句子之內有最好的效果。通常一個句子就是描述一件事情或想法，句子內的詞大都會跟所描述的事情相關，所以這些詞彼此之間的關連性應該很高。只看很相關的詞的部份，系統建議的前 10 個詞中，大約有 7 個詞是與原查詢詞很相關的；前 20 個詞中，大約有 13 個詞是與原查詢詞很相關的；前 30 個詞中，則約有 18 個詞是與原查詢詞很相關的。若把很相關及部份相關的詞都算是相關詞，系統建議的前 10 個詞中，大約有 8 個詞是與原查詢詞相關的；前 20 個詞中，大約有 16 個詞是與原查詢詞相關的；前 30 個詞中，則約有 23 個詞是與原查詢詞相關的。

若以嚴格的標準判定相關度(表 1)，共現資訊計算範圍設定在同一篇文章內所訓練而得的推薦詞的效能表現較差。以寬鬆的標準來看(表 2)，以文章為共現範圍的表現就比較好。這可能是因為在同一篇文章中出現的句子很多，雖然都跟一個主題有關，不過可能是描述不同的事情或想法，所以不同句子的詞之間的關連性會比較低。至於前後各一個字的共現範圍則是太窄，使得擷取出的相關詞不夠多。在 12 個測試詞中，有 6 個詞的推薦詞只有十來個。另外目前訓練語料庫

<sup>3</sup> <http://www.bhp.doh.gov.tw/BHP/do/moreConditionList>

內的文章不到 300 篇，還是太少，會影響相互訊息的準確性，進而影響了系統的效能。未來我們將持續收集更多的訓練資料，以獲得更好的訓練結果。

## 5、結論

在本研究中我們使用查詢詞推薦的方法來幫助一般民眾在網際網路上找尋心血管疾病的相關資料。我們用自動學習的方式從訓練資料中抽取出相關詞，推薦給使用者選擇，以建立良好的查詢。實驗結果顯示將共現資訊的計算範圍設定在一個句子之內有最好的效果。系統建議的前 30 個詞中約有 62.5% 是與原始查詢詞很相關的，若把部份相關的詞一併考慮，則有 78.9% 的詞是與原查詢詞相關。

由於目前的訓練資料較少，疾病介紹的部分只是眾多心血管疾病的一部分，未來我們會持續收集更多的訓練資料，以獲得更好的訓練結果。另外，除了由語料庫中訓練出相關詞，我們也要研究由人工建立的語言知識庫中，例如 UMLS(Unified Medical Language System)<sup>4</sup>，擷取出相關詞。我們的方法目前雖然只應用於心血管疾病，但只要訓練資料夠多夠完整，未來可以擴大應用在其他專科或是整個醫學領域，來幫助使用者找到更符合其需求的醫療健康資訊。

## 誌謝

本研究承蒙國科會計畫 NSC 94-2218-E-320-006 及 NSC 95-2218-E-320-006 補助支持，特此誌謝。

## 參考文獻

- [1] 朱怡霖 (2002). 中文斷詞及專有名詞辨識之研究. 國立台灣大學資訊工程研究所碩士論文.
- [2] 陳光華, 莊雅蓁 (2001). “資訊檢索之中文詞彙擴展”, 資訊傳播與圖書館學, 8(1), 59-75.
- [3] Baeza-Yates, R. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison Wesley, 1999.
- [4] Church, K.W. and Hanks, P. (1990). “Word Association Norms, Mutual Information, and Lexicography.” *Computational Linguistics*, 16(1),

22-29.

- [5] Fox, S. (2005). “Health Information Online: Eight in ten internet users have looked for health information online, with increased interest in diet, fitness, drugs, health insurance, experimental treatments, and particular doctors and hospitals.” [On-line] Pew Internet & American Life Project, May 17, 2005.  
[http://www.pewinternet.org/pdfs/PIP\\_Healthtopics\\_May05.pdf](http://www.pewinternet.org/pdfs/PIP_Healthtopics_May05.pdf)
- [6] Health On the Net Foundation (2002). “Excerpt of the 8th HON's Survey of Health and Medical Internet Users.” [On-line]  
[http://www.hon.ch/Survey/8th\\_HON\\_results.html](http://www.hon.ch/Survey/8th_HON_results.html)
- [7] Kakade, V. and Sharangpani, M. (2004). “Improving the Precision of Web Search for Medical Domain using Automatic Query Expansion.” [On-line]  
<http://www.stanford.edu/class/archive/cs/cs276a/cs276a.1032/projects/reports/vkakade-smadhura.pdf>
- [8] Li, J.-F., Guo, M.-Z., and Tian, S.-H. (2005). “A New Approach to Query Expansion.” In *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics (ICMLC2005)*, 2302-2306.
- [9] Qing, T., Robert, M., Ngo, L., and Dibble, E. (2006). “Assisting Consumer Health Information Retrieval with Query Recommendations.” *Journal of the American Medical Informatics Association*, 13, 80-90.
- [10] Silverstein, C., Marais, H., Henzinger, M., and Moricz, M. (1999). “Analysis of a Very Large Web Search Engine Query Log.” *ACM SIGIR Forum* 33(1), 6-12.
- [11] Tseng, Y.-H. (1998). “Solving Vocabulary Problems with Interactive Query Expansion.” *Journal of Library Information Science*, 24(1), 1-18.

<sup>4</sup> <http://www.nlm.nih.gov/research/umls/umlsmain.html>