

第二章

相關研究

本研究的相關研究，主要分為兩部份。第一部份是針對網路搜尋查詢修正的相關研究，第二部份則是介紹隱含回饋的相關研究。

2.1 網路搜尋的查詢修正

網路搜尋的查詢修正是資訊擷取領域近幾年來的熱門議題。Bruno M. Fonseca 等曾提出利用查詢的**概念**(Concept)去做**查詢擴展**(Query Expansion)[8]。其主要作法是利用搜尋引擎取的查詢日誌，從中探勘使用者查詢關鍵字之間的**關聯法則**(Association Rule)。利用關聯法則判斷關鍵字之間的上下義關係，以建立由關鍵字所組成的**關聯圖**(Relation Graph)，如圖 2.1 所示。由關聯圖中，所找出的最小迴圈(Minimum Cycle)就對應到一個概念。因此，利用概念與查詢關鍵字間的關係，就可達到查詢擴展的目的。

因此，這篇論文與本研究究竟有何不同呢？Fonseca 這一篇論文，主要方法為探討概念與查詢關鍵字的關係，達到查詢擴展，也同樣的使用了過去使用者的資訊，但卻僅僅利用查詢關鍵字來做為資訊來源之主要考量，而忽略了點選網頁之重要性。而本研究則是雙重考慮了查詢關鍵字與點選網頁兩大資訊來找出與目前使用者有相似資訊需求的以往經驗使用者，進而分析這些經驗使用者的查詢過程，來達到查詢修正。

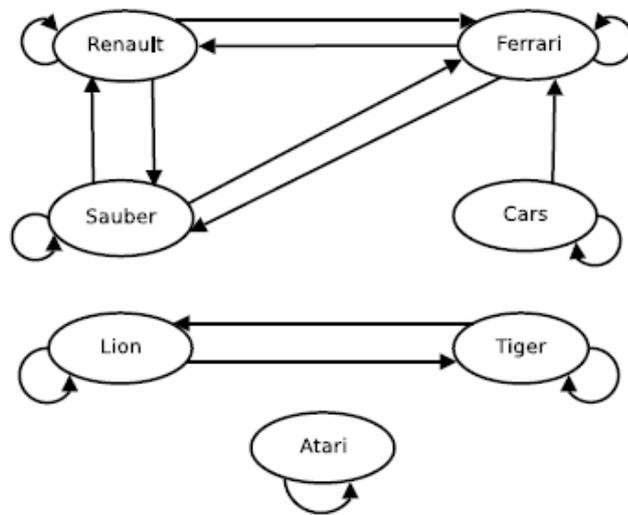


圖 2.1： 關鍵字“jaguar”的查詢關鍵字關聯圖。

同樣使用查詢日誌作查詢擴展的還有[6]。其主要的重點在於作者認為使用者點選過的相關網頁中，其所包含的關鍵字項目(Keyword Term)是與使用者的關鍵字查詢有相當程度關係的。所以，作者首先針對查詢日誌中，找出每一個查詢項目(Query Term)連到其相關網頁的機率值(此處的相關定義為使用者搜尋時點選過的都算)，再從這些相關網頁中的每個關鍵字項目去與關鍵字查詢計算相似度，則相似度最高的 n 個關鍵字項目與關鍵字查詢做 AND 成為新的關鍵字查詢，來達到查詢擴展。因此，這篇論文雖然利用長期情境，但卻沒有利用到短期情境的線索。

[17]這篇論文中，主要是去統計由查詢日誌所萃取出來的特性，藉著貝氏網路模型(Bayesian Network Model)去計算出使用者下一個查詢類型的機率，以產生修正後的查詢。作者去統計兩種查詢日誌的特性，一個是 Query Refinement Class，這是指使用者下一個查詢跟這一個查詢的關係，是 Generalization，還是 Specialization，諸如此類的。另一個是 Informational Goal，這是指使用者查詢的主要目標，共有 15 種。統計後的資訊，利用貝氏網路模型去計算機率值，進而做查詢修正。

[3]這篇論文認為以往查詢擴展的方法均是自動化地幫助使用者重新系統地闡述(formulate)使用者的資訊需求，而相關回饋的方法則是著重於讓使用者給予回饋，去釐清模擬兩可的資訊需求。因此，本篇問文主要的想法就是結合概念化的觀念到查詢擴展的技術上，使更精確的求出使用者的資訊需求。本篇論文利用分群(Cluster)的技術，將搜尋引擎對使用者初始查詢所回傳的結果分成數群。將每群的概念萃取出來後，再利用傳統查詢擴展的技術[23，12]去更進一步的了解使用者的資訊需求。

以上的相關研究均是近幾年對於網路搜尋查詢修正的研究，而利用**本體論**(Ontology)作查詢修正的研究也日益眾多。以下將做個概略介紹。

[20]這篇論文中，Hugo Liu 等作者認為目前需要讓使用者給予關鍵字來做查詢的搜尋引擎是不適合每位使用者的。例如：初接觸搜尋引擎的使用者，他們不知道如何給予適合他們資訊需求的關鍵字。因此作者認為理想中的搜尋引擎是可以讓使用者給予表示其目標的自然語言查詢，而也可以根據使用者給的自然語言查詢，來找出使用者的真正資訊需求。所以，作者提出利用 Open Mind Common Sense (OMCS)[27]這種常識(commmonsense)的本體論，以及使用者一開始所選擇的使用者目標，去將使用者的自然語言查詢推導成符合使用者資訊需求的關鍵字查詢，以這樣的方法作查詢修正。

另外，在[34]這一篇論文中，主要也是利用一種模糊本體論(Fuzzy Ontology)做查詢修正。基本上這篇論文所用的模糊本體論也是一種描述關鍵字與關鍵字之間語義關係的本體論。而這篇論文學習本體論的方法是去找出資料集中每篇文件的關鍵字，若任兩個關鍵字曾一起出現在同一篇文件中，就會去比較這兩個關鍵字，那一個出現的文件數較少，少者為多者的下義，反之亦然，以這樣的方法去學習所謂的模糊本體論。而查詢修正的動作則是回傳上下義的關鍵字給使用者去做選擇。

而 Nenad Stojanovic 這位作者曾作了一系列利用本體論達到查詢修正的論文，不過他的研究並不是單純使用在傳統網路搜尋上。其主要的研究目的是為了實現語意網 (Semantic Web) 的精神。例如 [10] 這篇論文中，針對網頁或論文的集合，利用本體論將 Database 裡的內容轉成具有語意含義的知識庫，再針對使用者在網路介面所下的一般查詢也將其轉換成有條件限制的本體論基礎查詢 (Ontology based query)。使用者的查詢有可能會有很多不同的意義。因此，系統根據使用者下的查詢，第一步先去分析其查詢的 Ambiguity，也就是利用數學公式去分析使用者的查詢所代表的不同意義。接著，依據第一步的結果，利用知識庫產生修正資料。最後，觀察哪幾種意義的比重較高，較符合使用者的需求作排名，以此作查詢修正。

另外，[29] 這一篇主要是針對使用者想要查詢本體論中的資訊，去做查詢修正。例如有一個研究本體論，使用者想找出在某處服務且是某種研究領域的所有教授，系統針對這些有限制條件的查詢去做查詢修正，因此是跟傳統網頁搜尋比較不同。其解法也是先去確定使用者查詢的 Ambiguity，再搭配上使用者搜尋的行為，以藉此找出最符合使用者需求的資訊。

2.2 隱含回饋

近年來，隱含回饋已經引起很多研究者的興趣 [2, 22]。以下則介紹一些有關於隱含回饋的相關研究。在 [2] 這篇論文中，Xuehua Shen 等作者利用機率模型 (Probabilistic Model) 的方法，將使用者先前的查詢與先前點選過的網頁結合到目前使用者的查詢，再與網頁作 Kullback-Leibler (KL) Divergence Method [36]，來增進萃取的正確性。

而在 [22] 這一篇論文中，作者利用一種稱為查詢鏈 (Query Chains) 的資訊，去學習一些新的排名函式 (Ranking Functions)。而何謂查詢鏈呢？就是使用者搜尋資訊時，與搜尋引擎

互動的所有動作，包括所給予的關鍵字查詢、點選過的網頁以及使用者通常會注意前幾篇網頁等等之類的隱含回饋資訊。作者同時也提出 6 種型態的隱含回饋策略(Implicit Feedback Strategies)，分別探討使用者給予的關鍵字查詢以及點選過網頁之間的關係。根據隱含回饋策略所判斷出的結果，決定回傳網頁的 Preference，最後利用 Ranking SVM 來學習排名函式。

而上述兩篇論文雖然都利用隱含回饋中的短期情境資訊，但卻都沒有利用到長期情境的資訊，這也是與本研究最大不同之處。