

以模糊理論與中介資料建構模糊查詢處理系統 ——以房屋選購系統為例

劉俞志

元智大學資訊管理系

林志麟

元智大學資訊管理系

吳炎崑

萬能科技大學工業管理系

摘要

模糊查詢是指在查詢條件中，含有模糊項的語意變數，能表達自然語言與人類思考方式。本研究旨在提出一模糊查詢的處理架構。主要作法是藉由房屋選購之使用者查詢需求與符合資料的回饋，加上模糊理論來建構模糊查詢條件語句的中介資料，使得模糊項依據查詢特質做不同的適切轉換，進而將模糊查詢的條件轉換為傳統查詢條件。此外，也將模糊的加權觀念應用在轉換機制上，使其能將模糊的查詢條件與加權值，轉為一合理的明確範圍查詢。並將查詢結果根據其與條件的符合度加以排序，方便使用者能快速找到較符合需求的資料。本文除了以例子詳述查詢條件的轉換，亦說明如何運用企業單位實際交易資料做轉換，透過定期更新資料的內容，讓模糊語句轉譯成清晰查詢時更能反映資料的現況。最後並實作一模糊查詢轉換器，將之整合到目前現有的資料庫系統。

關鍵字：模糊查詢、模糊數、房屋選購、近似關係。



Applying Fuzzy Theory and Meta Data for Fuzzy Query System: A Case Study of Housing Purchase

Julie Yu-Chih Liu

Department of Information Management, Yuan Ze University

Jun-Lin Lin

Department of Information Management, Yuan Ze University

Yan-Kuen Wu

Department of Industrial Management, Vanung University

Abstract

A fuzzy query is a query whose predicate contains a fuzzy term to represent the natural language and thought of humans. This work presents a novel framework for processing fuzzy queries. The proposed approach is to construct the metadata for the predicate in the fuzzy query by applying the fuzzy theory and the feedback resulting from comparing the user demands and the user query results on a house purchase. The fuzzy queries may be translated into the corresponding crisp queries according to the metadata with respect to fuzzy terms in the fuzzy queries. Fuzzy weighting is applied to the translation mechanism in order to rationalize the range value resulting from the translation. The query results are listed in order of matching level, thus allowing the user to find the required data conveniently. The above approach is illustrated by examples. This work also illustrates the maintenance of metadata periodically based on an enterprise's transaction data. Finally, a translator is implemented for fuzzy queried and incorporated it into a traditional database system.

Key words: Fuzzy Query, Fuzzy Number, Housing Purchase, Similarity Relation



壹、緒論

傳統的 DBMS 系統能依據使用者的查詢條件有效地在明確資料中，取得相符的資料，對於不確定性或不精確的口語化查詢，則欠缺處理的能力。然而，現實中的需求充滿著不明確性與可容忍性，舉例來說，欲購買房屋之使用者可能想要“大安區附近、中等價位、坪數接近 35 坪的房子”，其中“附近”、“中等”與“接近”均為不明確詞，而原始需求為“位於大安區、700~750 萬、35~40 坪的房子”的使用者，有時會容忍“大安區、690 萬、33 坪的房子”，但傳統的查詢系統只能列出完全吻合查詢的資料。此外，一般查詢系統僅能回饋針對特定屬性排序的資訊，無法以查詢結果與查詢語句的符合程度做為排序的依據，例如：與查詢條件完全符合的資料，其符合度為 100%，應該排在結果的最前面，80%的次之，以利使用者有效地篩選資料。

為了提供口語化的模糊查詢需求，本文應用模糊理論於房屋選購之查詢系統上。目前，將模糊查詢用於資料庫的作法可分三類：第一是採用傳統關聯式資料庫的引擎，並利用模糊理論轉換模糊查詢為清晰查詢語言；第二類是建立模糊關聯式資料庫，擴展資料儲存型態，使系統能直接處理模糊資料；第三類為混合式。本研究之作法方向亦屬為第一類，主要考量目前的資料庫管理系統並無有效支援模糊資料的儲存結構，即便擴展資料的儲存型態，亦無法透過現有的引擎，得到最佳化存取。反之，若直接轉換查詢語句，而不更動企業現有的資料庫系統，除了可減少資料庫重新建置的成本，亦可快速運用於現階段系統。

本研究主要的目的在，提出一查詢轉置器(Translator)的建置架構與作法，並藉由查詢需求與符合資料的回饋，建置模糊查詢語句的轉換表，使查詢語句中的模糊條件轉換為清晰條件時，能因時因地做適切的修正。而查詢所得的結果，除依據使用者下達之容忍度予以適度的放寬篩選條件，並利用模糊理論與權重觀念加以計算或排序，找出更符合使用者的結果。

貳、模糊集合的基本概念

模糊理論可將人的主觀思考或判斷的過程模型化，並以模糊集合(Fuzzy set)來做定量性處理的方法。模糊集合的觀念是由傳統的明確集合(crisp set)所延伸(Zimmermann 1987)，主要的不同在於，明確集合僅以“是”或“否”來表示屬性值是否屬於某個集合，但在模糊集合中則是利用模糊的隸屬關係和範圍來確定屬性值所屬的集合。底下針對模糊查詢轉換所需用到之模糊集合概念做一簡單介紹。

定義 1.1 (Zadeh 1965)：若 U 為 Universe of discourse，則在 U 中的模糊集合 A ，是指利用隸屬函數(Membership function) μ_A ，來說明 U 中元素屬於 A 的程度

$\mu_A : U \rightarrow [0,1]$ 。若 U 為有限集合，即 $U = \{u_1, u_2, \dots, u_n\}$ ，則 $A = \sum_{i=1}^n \mu_A(u_i) / u_i$ ，而 A 的 α -cut A_α ，以及 A 的 α -level-cut $A_{\alpha\text{-level}}$ 分別定義如下：

$$A_\alpha = \{u_i \mid \mu_A(u_i) \geq \alpha\}, \quad \alpha \in [0,1] \quad (1)$$

$$A_{\alpha\text{-level}} = \sum_{i=1}^n u_{A_{\alpha\text{-level}}}(u_i) / u_i \quad (2)$$

其中，若 $\mu_A(u_i) > \alpha$ 則 $\mu_{A_{\alpha\text{-level}}}(u_i) = \mu_A(u_i)$ ；反之，若 $\mu_A(u_i) \leq \alpha$ ，則 $\mu_{A_{\alpha\text{-level}}}(u_i) = 0$ 。

模糊集合具有傳統集合的基本運算性質(Yager et. al. 1987)，例如：包含(Inclusion)、相等(Equality)、補集(Complement)、交集(Intersection)與聯集(Union)等等，在此僅介紹本文中會用到的兩種運算：交集與聯集。首先，令 A 、 B 為在字集合 U 中的兩個模糊集合，而其隸屬函數分別以 μ_A 與 μ_B 表之，其中 $\mu_A : U \rightarrow [0,1]$ ， $\mu_B : U \rightarrow [0,1]$ ，且 $A = \{(u_i, \mu_A(u_i)) \mid u_i \in U\}$ ， $B = \{(u_i, \mu_B(u_i)) \mid u_i \in U\}$ ，則 A 和 B 的標準(standard)運算*如下：

定義 1.2 (Kaufmann & Gupta 1991)：模糊集合 A 交集(Intersection) B ， $A \cap B$ ，其標準交集定義為：

$$\mu_{A \cap B}(u_i) = \min(\mu_A(u_i), \mu_B(u_i)) \quad \forall u_i \in U \quad (3)$$

定義 1.3 (Kaufmann & Gupta 1991)：模糊集合 A 聯集(Union) B ， $A \cup B$ ，其標準聯集定義為：

$$\mu_{A \cup B}(u_i) = \max(\mu_A(u_i), \mu_B(u_i)) \quad \forall u_i \in U \quad (4)$$

一般文獻通稱模糊交集(Fuzzy Intersection)為 t-norms，而稱模糊聯集(Fuzzy Union)為 s-norms，文獻(Gupta 1991；Schweizer & Sklar 1963；Yanger 1980；Frank 1979；Yn & Bien 1994)中都曾提出其他 t-norms 及 s-norms 的方式來定義兩模糊集合的交集與聯集。

定義 1.4 (Gupta 1991)：模糊集合 A 交集 B ，以代數乘積(Algebraic Product)定義為：

$$\mu_{A \cap B}(u_i) = A(u_i) \cdot B(u_i) \quad \forall u_i \in U \quad (5)$$

定義 1.5 (Gupta 1991)：模糊集合 A 聯集 B ，以代數加法(Algebraic Sum)定義為：

$$\mu_{A \cup B}(u_i) = A(u_i) + B(u_i) - A(u_i) \cdot B(u_i) \quad \forall u_i \in U \quad (6)$$

隸屬函數(Membership functions)有多種：三角形、梯形、L-R、高斯分佈(又稱指數函數)、 π 函數、S 函數與 Z 函數等等。隸屬函數的訂定會因屬性而異，可以藉由統計法與專家歸納獲得，底下僅介紹本文所需之各型態模糊數的隸屬函數定義。

所謂模糊數(Fuzzy numbers)是一個定義在 Universal set $U = R$ (實數)上的模糊集合 A (參考定義 1.1)，其特質請參考文獻(Buckles & Petry 1984)。

定義 1.6 (Kaufmann & Gupta 1991)：梯形(Trapezoidal)模糊數，其隸屬函數 μ_A 定義如式子(7)，其表示式可簡化為 (a, b, c, d) ，可視為三角形模糊數 (a, b, d) 的延伸。

* 在模糊集合論上，「標準運算」並不是唯一的運算，只是其中較常用的種定義而已。



$$\mu_A(x) = \begin{cases} 0, & x < a. \\ \frac{x-a}{b-a}, & a \leq x < b. \\ 1, & b \leq x \leq c. \\ \frac{d-x}{d-c}, & c < x \leq d. \\ 0, & x > d. \end{cases} \quad (7)$$

定義 1.7 (Kaufmann & Gupta 1991)：令 F_L ， F_R 為 Reference functions，若一模糊數在區間 $[m_{1A}, m_{2A}]$ 之隸屬度為 1，則稱此數為平坦 L-R 模糊數，符號記為 $(m_{1A}, m_{2A}, u_A, v_A)$ ，如圖 1，其隸屬函數為： $\forall x \in R$,

$$\mu_A(x) = \begin{cases} F_L\left(\frac{x-m_{1A}}{u_A}\right), & x \in (-\infty, m_{1A}). \\ 1, & x \in [m_{1A}, m_{2A}]. \\ F_R\left(\frac{x-m_{2A}}{v_A}\right), & x \in (m_{2A}, \infty). \end{cases} \quad (8)$$

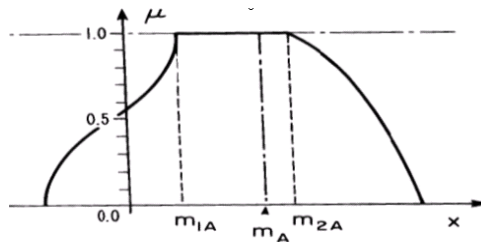


圖 1：平坦 L-R 模糊數

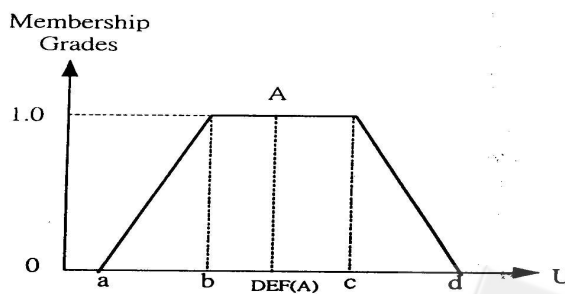


圖 2：梯形模糊數去模糊化值



模糊數的算術則涉及到多個區間表示，定義如下：

定義 1.8 (Kaufmann & Gupta 1991)：模糊數的四則運算：

令 $A = (a_1, b_1, c_1, d_1)$ 及 $B = (a_2, b_2, c_2, d_2)$ 為兩梯形模糊數：

$$\text{模糊數加法： } A + B = (a_1 + a_2, b_1 + b_2, c_1 + c_2, d_1 + d_2) \quad (9)$$

$$\text{模糊數減法： } A - B = (a_1 - d_2, b_1 - c_2, c_1 - b_2, d_1 - a_2) \quad (10)$$

$$\text{模糊數乘法： } A \cdot B = (a_1 \cdot a_2, b_1 \cdot b_2, c_1 \cdot c_2, d_1 \cdot d_2) \quad (11)$$

$$\text{模糊數除法： } A / B = (a_1 / d_2, b_1 / c_2, c_1 / b_2, d_1 / a_2) \quad (12)$$

兩模糊集合間的接近程度計算，可透過 Buckles 與 Petry(1983)的相似關係架構或者是 Shenoï 和 Melton(1989)所提出的近似關係為基礎的模式運用。

定義 1.9 (Buckles & Petry 1983)：相似關係(Similarity relations)為一映射 $S_K : D_K \times D_K \rightarrow [0,1]$ ，

D_K 為第 K 個屬性值域，滿足

1. 反身性(Reflexive)： $S_K(x, x) = 1$.

2. 對稱性(Symmetric)： $S_K(x, y) = S_K(y, x)$.

3. 遞移性(Transitive)： $S_K(x, z) \geq \text{Max}_{y \in D_K} \{ \text{Min} \{ S_K(x, y), S_K(y, z) \} \}$, $\forall x, y, z \in D_K$.

定義 1.10 (Shenoï 和 Melton 1989)：滿足相似關係中的反身性與對稱性，則稱為近似關係(Proximity relations)。

模糊系統中的模糊資料需作去模糊(Defuzzication)的處理，才具備有明確的特性，而去模糊化就是將模糊性資料轉換成明確性資料的一種方法。去模糊化的方法並無固定之方式，需視問題及資料的特性而定，同理，在模糊數的運用上，有時需要將模糊數與清晰值(crisp number)彼此做轉換，例如比較一模糊數與清晰值的大小，其操作方法定義如下：

定義 1.11(Chen 1995 ; Kaufmann & Gupta 1988)：令 A 為一梯形模糊數， $A = (a, b, c, d)$

且 $0 \leq a \leq b \leq c \leq d \leq 1$ ，此模糊數 A 之去模糊化值，以(DEF(A))表之(參圖 2)，

可計算如下：

$$DEF(A) = (a + b + c + d) / 4 \quad (13)$$

Zadeh(1975)在 1975 年以自然語言中出現的字或語句做為語意變數值，又稱為模糊項(Fuzzy Term)。例如設 Age 為語意變數，則變數值為 young、old、……等。而修飾詞(Modification)(Hou & Chen 1995)被定義為加諸於語意上的模糊關係，用以修飾模糊項，例如對模糊項的修飾詞 very young、more or less old，等等。模糊項可分三種(Chen & Jong 1997)，一、簡單(Simply)模糊項，例如“高”、“普通”、“矮”；二、混合(Composite)模糊項：對簡單模糊項加入一些修飾詞，例如“很高”、“稍微高”；三、複合(Compound)模糊項：使用邏輯運算子連接數個模糊項，例如“高並且很壯”。

含有修飾詞 very 或 more or less 之模糊項，可分別用 Concentration rules 與 Dilation rules 修正其隸屬函數。請參考文獻(Giarratano & Riley 1989)。另外，含有 AND 的複合模糊項，可用模糊集合之交集運算(如定義 1.2 或 1.4)表達其隸屬函數，而含有 OR 的複合模糊項，則是用模糊集合之聯集運算(如定義 1.3 或 1.5)。

參、相關文獻

一、模糊查詢系統

模糊查詢是指在查詢條件中，含有模糊項(Fuzzy term)的語意變數，達到更接近自然語言與人類思考方式的查詢表達。而查詢所存取的資料庫可分別為傳統資料庫或模糊資料庫。1982年 Buckles 與 Petry(1982)提出模糊關聯式資料模式，其中屬性值可為集合值，屬性值域(Domain)的元素可為模糊項，並以相似性關係(Similarity relations)來描述模糊項之間的相似度。之後，更進一步擴充屬性值至模糊數(Fuzzy numbers)(Buckles & Petry 1984)。此外，Petry 也把模糊資料庫內的模糊資料分為四種型態(Petry 1996)：清晰值(precise values)、區間值(interval values)、模糊值(fuzzy values)與缺值(null values)，並提出不同模糊資料屬性運算處理的方法。

在模糊資料庫的模糊查詢，有學者以模糊數積分值排序為理論基礎(徐瑞豐 1995)，將屬性積分值與查詢條件之積分值的接近程度作為查詢依據，不同型態的模糊屬性值可藉由模糊數積分計算加以比較，並以平均化運算子(Averaging Operators)取代 Max 與 Min 運算子造成的資訊流失。但由於計算過程較複雜，而且每次更新資料亦必需同時計算其屬性模糊數的積分值，較適用於以查詢為主且異動較少的系統。

另一種作法是藉由模糊理論將模糊查詢運用於傳統資料庫，包括：Kacprzyk 與 Zadrony(2001)、Chen 等人(1986)、Chen 和 Jong (1997)以及江芳彬(1993)的作法，本研究亦屬此類。Kacprzyk 與 Zadrony 的作法是讓使用者以 FQUERY 下達查詢，系統則藉由轉換器將 FQUERY 轉換為一般 SQL，提交資料庫處理，而此轉換器是以 ADD-ON 的方式外掛於關聯式資料庫(ACCESS)中。Chen 等人的作法在利用 Threshold 設定值，將模糊查詢轉換為傳統關聯式資料庫可接受的 SQL 查詢，並藉由過濾資料來避免計算不必要的隸屬度，增進系統效能。此外，江芳彬(1993)的作法則是先將資料庫中儲存的明確資料值模糊化，來完成模糊查詢。然而其計算隸屬度時，需對資料庫內所有記錄做運算，所以隨著查詢條件的複雜化，計算時間耗費越大。

混合式的作法(Chen & Chen 2000)則依據各類查詢與資料庫差異採用不同處理，如下：

1. 查詢及資料庫內存放資料的屬性值都是模糊型態，則使用相似函數(Similarity Function)來測量兩者間的相似度。
2. 查詢及資料庫內存放資料的屬性值都是清晰值，則使用相似矩陣(Similarity Matrices)來測量兩者間的相似度。
3. 清晰查詢相對於資料庫內屬性值是模糊型態，則運用語意變數的隸屬函數將清晰查詢轉換為模糊查詢。
4. 模糊查詢相對於資料庫內屬性值是清晰值，則同樣運用語意變數的隸屬函數將資料庫內的清晰資料轉換為模糊型態。

對此，劉俞志和黃堯銘(2002)則提出一套轉換機制，將模糊資料庫裡各種不同表示法的模糊資料，轉換表示成分割區歸屬程度表示法，再採用近似語意相等(Proximity Semantic Equivalence, PSE)，做為相似程度衡量的核心機制，達到以單一比較方式即能整合處理多種不同表示方法的模糊資料型態的功能。此外，Lin 將權重應用於模糊查詢中(Lin & Chen 2000)，允許使用者藉由模糊數或明確值的權重設定，得到更符合需求的結果，並且查詢所得的每一筆記錄均附有一個符合(Match)程度值，供使用者參考。

二、房地產買賣考慮因素

房地產購買之考慮因素已被廣泛討論，本文根據近期文獻整理這些因素於表 1，並以此作為查詢系統之介面設計的參考依據。

表 1：相關文獻房地產買賣考慮因素整理

文獻名稱	考量因素
何智明(1998), 消費者住屋購買行為關鍵因素之研究-以台北都會區為例	1.生活便利性, 2.購屋能力, 3.賣方推廣促銷策略, 4.經濟與金融發展, 5.建商信譽與房屋品質, 6.樓層與使用實坪, 7.國際、政治與治安, 8.房屋開發管制與防洪措施, 9.消防與醫療, 10.減少交易糾紛與增強售後服務, 11.貸款條件, 12.座向、區域與陽宅地理, 13.空屋資訊與購屋安全。
王琬宜(2000), 地理資訊系統在不動產查詢與分析上之應用	1. 基本屬性因素： (1)行政區, (2)坪數, (3)價格, (4)屋齡, (5)房、廳、衛, (6)類型。 2. 空間環境屬性因素： (1)文教設施：幼稚園、國小、國中、圖書館。 (2)服務設施：市場、郵局、警察局、消防隊、醫院及衛生所。 (3)交通設施：公車站、捷運站、停車場、主要道路。 (4)運動遊憩設施：公園。 (5)鄰避設施：加油站、變電所、垃圾掩埋場、垃圾焚化廠、納骨塔、公墓。
郭嘉昌(2000), 新竹科學工業園區員工對住宅環境偏好之分析	1. 內部環境屬性變數： (1)住屋外型設計與色調, (2)住屋空間大小, (3)房屋狀況(如漏水、龜裂、脫漆等), (4)安全性與衛生條件, (5)使用之建材類別、等級, (6)採光及通風性, (7)樓層的選擇。 2. 外部環境屬性變數： (1)停車空間, (2)便利性(上班、購物、就學等), (3)附設遊樂場、游泳池、健身房等, (4)學區, (5)住屋安全管理, (6)環境的治安狀況, (7)住屋的視野景觀, (8)交通便利, (9)公設配置比例低。 3. 需求特性： 購買原因、住宅型態、居住面積、住宅格局、住宅價格。

翁志賢(2000),消費者購屋行為之探討—品牌權益,品牌態度與購屋評估準則對購屋意圖之影響	<ol style="list-style-type: none"> 1. 房屋屬性：房屋的坪數、土地的坪數、房屋的格局、公設比、建材好壞、防震結構、透天厝或公寓。 2. 交通運輸：距市中心的距離、距大眾運輸的距離、距主要道路的距離、緊臨道路的寬度、停車便利性。 3. 公共設施(周圍環境)：離小孩學校的距離、附近有無公園、附近有無市場、與文化中心或美術館的距離。 4. 購屋負擔能力：訂金及簽約金的多寡、房屋總價、貸款利率、可貸款的金額。
蘇振昌(1999),消費者的網路資訊搜尋對購屋行為之影響	地段、交通、屋齡、坪數、樓別、生活機能、景觀、外觀設計、空間規劃、公共設施、建材設備、生活管理、安全管理、清潔維護。
鄭淵聰(1998),中老年人對老人住宅購買行為及行銷策略之初探性研究-以台中都會區為例	<ol style="list-style-type: none"> 1. 住宅本身屬性：住宅外觀、建材品質、格局、採光通風、私密性、附屬設施、坪數及房廳數。 2. 住宅周遭環境因素：交通便利性、停車設施、公共環境品質、生活便利性、安全性。 3. 其他因素：價格、付款優惠條件、建設公司可靠性。
何志南(2000),住宅消費模式之建構與分析	<ol style="list-style-type: none"> 1. 居室舒適性：樓地板面積、空間大小、房間數、住宅設置、衛浴室、樓層數、屋齡、住宅型態、住宅使用狀態等。 2. 設施可及性：至市中心距離、至學校距離、至醫院距離、至高速公路距離、至鐵路距離等。 3. 鄰里環境：臨避設施、社區和諧、穩定性、安全、寧適性、種族、學校品質、商店服務、鄰里信譽、空氣品質、污染等。

表 1 考量之因素中，有些因素屬於非靜態、較難取得或較難評估與描述之屬性，例如：鄰避設施、社區環境、建材好壞、公共設施、建設公司之可靠性。有些則屬於主觀性或流於繁瑣細節，例如：採光、安全、格局，無法以簡單文字表達之特質。因此，本文僅選擇下列屬性做為系統之考量，包括：區域、坪數、總價、屋齡、樓層、朝向、類型、用途。其中隱含使用者的已知與主導權，例如：使用者對區域的選擇，暗示了使用者知道並且接受該區域之環境、交通狀況，因而不需再去查詢該區域的鄰避設施、社區環境或交通狀況等資料。主要原因，除了考量屬性的資料獲取性、可靠性，也考量其有效性。

肆、研究內容

轉為模糊查詢轉為清晰查詢的方法共分七個步驟，步驟一規劃查詢轉置器的架構與處理流程，步驟二訂定查詢轉換所需的中介資料（或稱轉換表），步驟三根據模糊條件的組成內容分類，並列出各類條件的轉換方法，步驟四加入權重處理方式，步驟五計算查詢結果的隸屬度，步驟六考慮效能改善的三種作法，最後根據查詢介面選擇轉換方式。

一、模糊查詢轉置器架構

本研究主要是利用一轉置器架構，將使用者所下的各式模糊查詢，轉換成清晰條件查詢，使其能直接由 DBMS 之最佳化器(Optimizer)，產生較佳之查詢計劃(Query Plan)，得到快速有效的查詢結果。

轉置架構如圖 3，其中介於使用者與資料之間的架構，稱為轉置器，轉置器會根據條件子句的模糊性質，做不同的處理。若條件值為清晰資料，如“房屋的朝向要在南方”，則可直接轉換為清晰查詢子句。需要進一步處理的模糊查詢條件，如“房屋的價格要在低價位”，則將條件中的語意變數與模糊項，依轉換表中的對應資料，轉換模糊條件成清晰條件。隨後將所有處理過的清晰查詢條件，交予 DBMS 由房屋特性資料庫中取出符合條件的記錄。取出之記錄經由計算其與查詢條件的符合度，做再次的篩選與排序，最後產生依符合度高低所排列的查詢結果。此外，轉換表之資料是根據企業實際交易資料制定且定期更新，因此具有適時適地性。

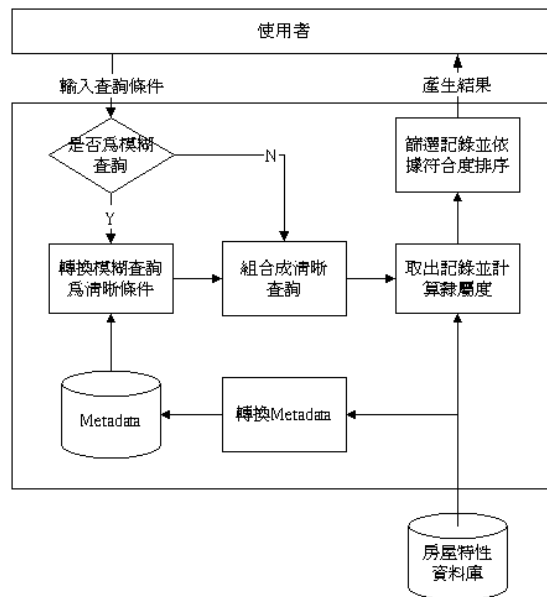


圖 3：查詢執行架構與流程

二、訂定模糊查詢的中介資料

系統在訂定模糊資料的中介資料（或稱轉換表）時，應根據實際需求與可忍受差異程度，來設定屬性值的近似程度。以信義房屋為例，本研究取近五年(1997~2001)於台北市的需求與成交資料間的差異做統計。依其資料來源將之區分為(1)需求條件間

的近似程度，(2)需求條件與成交結果的平均差異(容忍度)，(3)成交資料用於語意變數模糊項的資料來源三種。

(1) 需求條件間的近似程度：

以區域為例，經資料統計後發現，填寫需求區域為士林區的同時會填寫天母區的機率是 0.8，並將其視為兩者間的近似關係，表 2 即以士林區為例(已排除非鄰近的行政區)，列出其與鄰近行政區彼此的近似程度。

表 2：士林區與鄰近行政區間的近似程度(資料來源：信義房屋)

區域	近似程度(士林區)
士林區	1
天母區	0.8
中山區	0.6
內湖區	0.5

表 3：客戶購屋前需求與實際成交結果差異(資料來源：信義房屋)

成交／需求	下限	上限
價格	29%	26%
坪數	23%	25%
樓層	46%	81%
屋齡	N.A.	83%

(2) 需求條件與成交結果的平均差異(容忍度)：

統計成交資料與尚未購屋前填寫之需求資料的平均差異做為需求容忍度。以坪數為例，可發現如實際成交時所購得的房屋坪數低於原始需求所填之坪數者(稱之下限)，其平均差異為 23%，相對的，成交坪數高於原始需求所填之坪數者(稱之上限)，其平均差異為 25%，其餘相關屬性如表 3，列出其平均差距的正負百分比。

其中價格與屋齡的下限應無容忍度，因為在其他屬性條件不變下，此兩者數字越低表示越趨向於優勢-價廉物美；但考慮使用者可能在一定的預算價格內想查詢出其他屬性條件最好的組合，故仍保留價格的下限容忍度而刪除屋齡的下限平均差異。

(3) 成交資料用於語意變數模糊項的資料來源

研究中所使用的語意變數與模糊數見表 4，其模糊項的中介資料，同樣是以信義房屋近五年實際成交的物件資料依價格高低排序，低於 30%為低價位；30%-40%則介於低價位與一般價位之間；40%-60%為一般價位；60%-70%介於一般價位與高價位之間；高於 70%為高價位，得出結果如表 5(代表台北市)，此資料並可以月為單位重新計算。

表 4：使用之語意變數與模糊項

語意變數	模糊項
價格	高價位、一般價位、低價位.
坪數	大坪數、中等坪數、小坪數.
樓層	高、中、低.
屋齡	新成屋、中古屋、老房子.

表 5：“價格”的模糊項及隸屬函數

模糊項	Membership function
高價位	$\begin{cases} 0, & p < 1590 \text{萬.} \\ \frac{p-1590}{1660-1590}, & 1590 \text{萬} \leq p < 1660 \text{萬.} \\ 1, & p \geq 1660 \text{萬.} \end{cases}$
一般價位	$\begin{cases} 0, & p < 850 \text{萬.} \\ \frac{p-850}{1080-850}, & 850 \text{萬} \leq p < 1080 \text{萬.} \\ 1, & 1080 \text{萬} \leq p < 1590 \text{萬.} \\ 1 - \frac{p-1590}{1660-1590}, & 1590 \text{萬} \leq p < 1660 \text{萬} \\ 0, & p \geq 1660 \text{萬} \end{cases}$
低價位	$\begin{cases} 1, & p < 850 \text{萬.} \\ 1 - \frac{p-850}{1080-850}, & 850 \text{萬} \leq p < 1080 \text{萬.} \\ 0, & p \geq 1080 \text{萬.} \end{cases}$

三、模糊條件之分類與轉換

將模糊條件根據之前提出作法(左心強等人 2002)，分為三種狀況，以不同方式轉換成清晰條件：

- 狀況一：查詢條件為修飾詞+屬性值間具近似關係的清晰值，例如：區域最好在士林區。其處理方式是運用近似關係(定義 1.10)定出的中介資料(例如表 2)產生隸屬函數，並以修飾詞(Modification)所代表的門檻值(參表 6)代入公式(1)處理。
- 狀況二：查詢條件為修飾詞+區間值(Range Value)所組成，例如：坪數最好在 30-40 坪。其處理方式是運用 L-R 模糊數(定義 1.7)，以表 3 的中介資料產生隸屬函數，同上，以修飾詞所代表的門檻值代入公式(1)。
- 狀況三：查詢條件為修飾詞+語意變數的模糊項所組成，例如：價位最好在低價位。其處理方式是以修飾詞(Modification)所代表的門檻值代入語意變數模糊項的隸屬函數(例如表 5)，並代入公式(1)處理。

例 4.1：找出符合查詢條件“行政區最好在士林區”的房屋資料。(狀況一)

- (1) 見表 2， $U(\text{士林區}) = 1/\text{士林區} + 0.8/\text{天母區} + 0.6/\text{中山區} + 0.5/\text{內湖區}$ 。
- (2) 參考表 6 得修飾詞“最好在”的門檻值為 0.7。
- (3) 使用公式(1)，並以 0.7 代入得區域條件值天母區，士林區。
- (4) 轉換為 SQL 查詢：`select * from 房屋資料表 where 區域 in ('天母區', '士林區')`。

例 4.2：找出符合條件“坪數最好在 30-40 坪”的房屋資料。(狀況二)

- (1) 根據定義 1.7，利用表 3 產生隸屬函數： $\forall x \in R$,

$$\mu_A(x) = \begin{cases} 0, & x < 30 \times (1 - 0.23). \\ F_L\left(\frac{x - 30 \times (1 - 0.23)}{30 \times 0.23}\right), & x \in (30 \times (1 - 0.23), 30). \\ 1, & x \in [30, 40]. \\ F_R\left(\frac{40 \times (1 + 0.25) - x}{40 \times 0.25}\right), & x \in (40, 40 \times (1 + 0.25)). \\ 0, & x > 40 \times (1 + 0.25). \end{cases}$$

- (2) 參考表 6 得修飾詞“最好在”的門檻值為 0.7。
- (3) 以 0.7 代入隸屬函數並四捨五入求得坪數條件值 28-43 坪。
- (4) 轉為 SQL：`select * from 房屋資料表 where 坪數 between 28 and 43`。

例 4.3：找出符合“價位最好在低價位”的房屋，門檻值為 0.7。(狀況三)

- (1) 根據表 5，找出低價位的隸屬函數。
- (2) 參考表 6 得修飾詞“最好在”的門檻值為 0.7。
- (3) 以 0.7 代入隸屬函數，得 $(1080 - 850) * 0.3 + 850 = 919$ (萬)。
- (4) 轉為 SQL：`select * from 房屋資料表 where 價格 <= 919`。

表 6：修飾詞代表的門檻值

修飾詞	門檻值
完全在	1
最好在	0.7
大概在	0.5

表 7：語言的權重值和其對應之模糊數

模糊項	Trapezoidal Fuzzy Numbers
高	(0.7113, 0.7589, 0.8065, 0.8541)
中	(0.4257, 0.4733, 0.5209, 0.5685)
低	(0.0476, 0.0925, 0.1401, 0.1877)

四、權重處理

在使用者所下達的模糊查詢中，可能針對不同屬性給予不同的權重值，因此，為了增加查詢的彈性，本系統增加了多屬性權重的考量。權重處理採(Lin & Chen 2000)中權重所使用的語意詞和其所對應的梯形模糊數的定義，其中以語意變數的模糊項來決定各查詢項目的權重值(如表 7)。

介紹完權重所代表的意義後，接著說明權重的計算方法，令兩梯形模糊數 W_1 及 W_2 分別為兩屬性 A_1 與 A_2 之權重， $W_1 = (a_1, b_1, c_1, d_1)$ 且 $W_2 = (a_2, b_2, c_2, d_2)$ ，Lin 和 Chen (2000)提出權重正規化的計算方式，以 \bar{W}_1 及 \bar{W}_2 ，分別代表 W_1 與 W_2 之正規化權重，並參考公式(9)至(12)：

$$\bar{W}_i = W_i / (W_1 + W_2) = \left(\frac{a_i}{d_1 + d_2}, \frac{b_i}{c_1 + c_2}, \frac{c_i}{b_1 + b_2}, \frac{d_i}{a_1 + a_2} \right) \text{ 其中 } i=1,2 \quad (14)$$

例 4.4：假設某一使用者對物件的價格及坪數分別為 $W_{\text{價格}} = (0.4, 0.5, 0.6, 0.7)$ 及 $W_{\text{坪數}} = 0.9$ ，對於明確的權重值，依(Lin & Chen 2000 ; Chen & Chen 2000)將 0.9 擴充(0.9, 0.9, 0.9, 0.9)，則其正規化權重值，分別計算如下：

$$\begin{aligned} \bar{W}_{\text{價格}} &= \frac{(0.4, 0.5, 0.6, 0.7)}{(0.4, 0.5, 0.6, 0.7) + (0.9, 0.9, 0.9, 0.9)} = (0.4, 0.5, 0.6, 0.7) / (1.3, 1.4, 1.5, 1.6) \\ &= (0.25, 0.33, 0.43, 0.54) \end{aligned}$$

$$\bar{W}_{\text{坪數}} = \frac{(0.9, 0.9, 0.9, 0.9)}{(0.4, 0.5, 0.6, 0.7) + (0.9, 0.9, 0.9, 0.9)} = (0.56, 0.60, 0.64, 0.69)$$

五、計算隸屬度

隸屬度之計算，根據下列幾種狀況有不同處理方式：

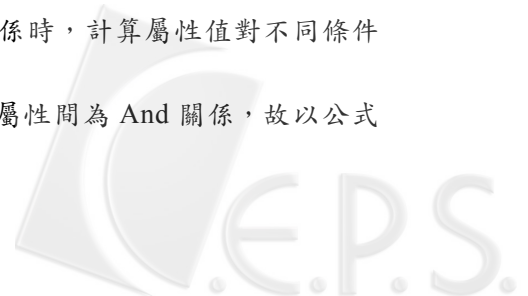
狀況四：將查詢結果所得的記錄之屬性值與條件值做比較，並運用近似關係(定義 1.9)定出的中介資料(例如表 2)代回隸屬函數做處理。

狀況五：將查詢結果所得的記錄之屬性值與條件值做比較，並運用 L-R 模糊數(定義 1.7)以表 3 的中介資料代回隸屬函數做處理。

狀況六：將查詢結果所得的記錄之屬性值與條件值做比較，並以語意變數模糊項的隸屬函數(例如表 5)代入做處理。

狀況七：如該屬性的查詢條件為多個且彼此間為 OR 關係時，計算屬性值對不同條件值之隸屬度後，再以公式(4)聯集運算來處理。

狀況八：計算完個別屬性值對條件值的隸屬度後，若各屬性間為 And 關係，故以公式(3)交集運算做處理。



狀況九：計算完個別屬性值對條件值的隸屬度後，若各屬性有不同權重值，先以公式(14)計算出各屬性值的正規化權重值，再以公式(13)解模糊化得相對權重值，將結果與個別所屬的屬性值隸屬度相乘後加總，並與 1 求 minimum(因隸屬度最大值為 1)，以求得此一記錄的隸屬度。

例 4.5：原查詢條件為“房屋的行政區在士林區”，查詢所得的其中一筆記錄其行政區屬性值為“天母區”，則隸屬度之計算如下：(狀況四)

(1) 根據表 2， $U(\text{士林區}) = 1/\text{士林區} + 0.8/\text{天母區} + 0.6/\text{中山區} + 0.5/\text{內湖}$

(2) 得該記錄行政區屬性值的隸屬度為 0.8。

例 4.6：原查詢條件為“房屋的坪數在 30-40 坪”，查詢所得的其中一筆記錄其坪數屬性值為“30 坪”，則隸屬度之計算如下：(狀況五)

(1) 根據定義 1.5，利用表 3 產生隸屬函數： $\forall x \in R$

$$\mu_A(x) = \begin{cases} 0, & x < 30 \times (1 - 0.23). \\ F_L\left(\frac{x - 30 \times (1 - 0.23)}{30 \times 0.23}\right), & x \in (30 \times (1 - 0.23), 30). \\ 1, & x \in [30, 40]. \\ F_R\left(\frac{40 \times (1 + 0.25) - x}{40 \times 0.25}\right), & x \in (40, 40 \times (1 + 0.25)). \\ 0, & x > 40 \times (1 + 0.25). \end{cases}$$

(2) 得該記錄坪數屬性值的隸屬度為 1。

例 4.7：原查詢條件為“房屋的價位在“低價位”，查詢所得的其中一筆記錄其價格屬性值為“919 萬”，求其隸屬度。(狀況六)

(1) 根據表 5，找出低價位的隸屬函數。

(2) 以 919 萬代入隸屬函數，得 $1 - (919 - 850) / (1080 - 850) = 0.7$ 。

例 4.8：原查詢條件為“房屋的類型在“公寓”或“電梯華廈”，查詢所得的其中一筆記錄其類型屬性值為“公寓”，求其隸屬度。(狀況七)

(1) 根據表 8， $U(\text{類型}) = 1/\text{公寓} + 0.8/\text{電梯華廈}$ 。

(2) 代入公式(4)求 $\text{Max}(1, 0.8) = 1$ 。

表 8：房屋類型的近似程度

	透天	公寓	電梯華廈	別墅	套房
透天	1	0.5	0.5	0.8	0.3
公寓	0.5	1	0.8	0.5	0.7
電梯華廈	0.5	0.8	1	0.5	0.7
別墅	0.8	0.5	0.5	1	0.3
套房	0.3	0.7	0.7	0.3	1

例 4.9：原查詢條件為“房屋的行政區在士林區，坪數在 30-40 坪”，查詢所得的其中一筆記錄其行政區屬性值為“天母區”，坪數屬性值為“30 坪”，求其隸屬度。

(狀況八)

(1) 由例 4.4 及例 4.5 可得行政區屬性值的隸屬度為 0.8，坪數屬性值的隸屬度為 1。

(2) 代入公式(3)求 $\text{Min}(0.8,1)=0.8$ 。

例 4.10：原查詢條件為“房屋的行政區在士林區，坪數在 30-40 坪”，查詢所得的其中一筆記錄其行政區屬性值為“天母區”，坪數屬性值為“30 坪”，權重值分別為行政區：中，坪數：高，求其隸屬度。(狀況九)

(1) 由例 4.4 及例 4.5 可得行政區屬性值的隸屬度為 0.8，坪數屬性值的隸屬度為 1。

(2) 其正規化權重：

$$\bar{W}_{\text{區域}} = (0.3, 0.36, 0.44, 0.5), \quad \bar{W}_{\text{坪數}} = (0.5, 0.57, 0.69, 0.75)$$

(3) 以公式(13)解模糊化得相對權重：

$$\bar{W}_{\text{區域}} = 0.4, \quad \bar{W}_{\text{坪數}} = 0.6$$

(4) 相乘再加總各屬性值隸屬度與權重， $\text{Min}(0.4*0.8+0.6*1, 1)=0.95$ 。

六、效能改善

根據前述的設計方法，轉置器的效能改善要考慮底下三個問題，解決方式陳述於後：

問題一：邏輯連接詞 and/or 以 min/max 為運算子時，易造成查詢不合理的結果(Gupta 1991 ; Yanger 1980)。

問題二：在 4.3 節中的轉換作法是將模糊條件中各屬性修飾詞所代表的門檻值轉換為清晰條件，若使用者另設一查詢門檻值，該如何轉換使系統得到較佳效能。

問題三：加入權重處理時，不適用於 4.3 節中轉換成清晰條件的作法。

針對問題一，以例 4.9 來說，令該例所提的記錄為 A，假設另有一筆記錄 B 同樣位於天母區，其行政區屬性的隸屬度為 0.8，但其坪數為 28.6 坪，求坪數屬性的隸屬度也為 0.8，經 Min 運算後兩筆記錄 A、B 的隸屬度相同，但這樣的結果並不理想。但若以定義 1.4 代數乘積做運算子，則結果記錄 A 的隸屬度為 $1*0.8=0.8$ ，記錄 B 的隸屬度為 $0.8*0.8=0.64$ ，此運算值比較符合所預期的結果。代數乘積及代數加法運算子比 Max 以及 Min 適用的原因是(Zimmermann 1996)：代數乘積得到一個小於或等於最小值之運算結果，而代數加法則大於或等於最大值，此性質為負向補償性質(negatively compensatory)。故本文 And 與 Or 條件計算方式將以代數乘積及代數加法做為運算子。

針對問題二，先將查詢所設的門檻值與修飾詞所代表的門檻值比較，取其大者作門檻值，再根據 4.3 節之分類做法，轉換模糊條件為清晰條件。

針對問題三，加入權重處理時，雖有部份記錄會有未達門檻值的屬性，但因加入權重調整後，該記錄仍可達到門檻值。為避免逐筆檢視所有記錄的繁複計算，而影響查詢的效率，本文中藉由反推出屬性的最小隸屬度，來決定查詢條件的放寬程度。

考慮查詢門檻值為 0.9，價格的相對權重為 0.3，坪數的相對權重為 0.7，則要求得可滿足查詢門檻值的記錄，其價格最小隸屬度推算如下：

1. 首先令價格及坪數之隸屬度分別為 x, y ，因須符合查詢條件的門檻值 0.9，在個別乘上屬性的相對權重值 0.3 及 0.7 後，可得到下面的式子：

$$0.3x + 0.7y \geq 0.9 \quad (x \geq 0, y \geq 0)$$

2. 令 $y=1$ ，可得 $x=0.67$ ，即要找出滿足查詢門檻值為 0.9 的記錄，其價格隸屬度 x 須要大於或等於 0.67。
3. 進一步將隸屬度 $x \geq 0.67$ 代回價格的隸屬函數反推，就可得到放寬後的價格查詢條件值(因為相當於價格的門檻值由 0.9 降至 0.67)。

如本例改為： $0.3x + 0.7y \geq 0.6$ ($x \geq 0, y \geq 0$)，做法仍令 $y=1$ ，即便價格隸屬度 $x=0$ ，該記錄的整體隸屬度還是會大於查詢門檻值。也就是說，當坪數的權重值 0.7 大於查詢門檻值 0.6 時，不論價格隸屬度 x 為多少，記錄的整體隸屬度仍有可能大於或等於查詢隸屬度，故查詢時可不轉換價格屬性的查詢條件。

接下來推導查詢中有三個屬性的狀況：假設今有一查詢，其門檻值為 $T(0 \leq T \leq 1)$ ，三個屬性 X, Y, Z 的相對權重值分別為 W_x, W_y, W_z ($0 \leq W_x, W_y, W_z \leq 1$ 且 $W_x + W_y + W_z = 1$)，則要求得可滿足查詢條件門檻值 T 的記錄。首先令 X, Y, Z 之隸屬度分別為 x, y, z ($0 \leq x, y, z \leq 1$)，在 3 個屬性的隸屬度個別乘上其相對權重值後，可得到下面的式子：

$$W_x x + W_y y + W_z z \geq T \quad (15)$$

即查詢所得的結果，需要滿足(15)的條件才算成立。

接著說明如何決定是否將個別屬性轉換為查詢條件的判斷式的推導過程，以屬性 X 為例，如果 $W_y + W_z \geq T$ ，則即使隸屬度 $x=0$ ，式子(15)仍有機會成立，判斷式變為：

$$W_y y + W_z z \geq T$$

表示縱使 X 的隸屬度 $x=0$ ，只要屬性 Y 或 Z 的個別權重值大於或等於 T ，記錄的整體隸屬度仍有可能大於或等於查詢隸屬度，故查詢時可不轉換該屬性的查詢條件。又因為相對權重值 $W_x + W_y + W_z = 1$ ，故：

$$\begin{aligned} (1 - W_x) &= W_y + W_z \geq W_y y + W_z z \geq T \quad (\because 0 \leq y, z \leq 1) \\ \therefore (1 - W_x) &\geq T \end{aligned}$$

即： $(1 - \text{屬性相對權重值}) \geq \text{查詢門檻值}$ (16)

使用判斷式(16)的好處，是將不同屬性權重值代入(16)時，不須再加總一次其他屬性的相對權重值，在屬性有增減時，也不需改變此判斷式。

如果將查詢條件中的設定值代入判斷式(16)後，無法滿足 $(1 - \text{屬性相對權重值}) \geq \text{查詢門檻值}$ 的條件，則下一步是要求得隸屬度 x 的最小值以反推出該屬性的查詢條

件，將 $W_y y + W_z z = 1 - W_x$ 代入(15)中：可得

$W_x x + (1 - W_x) = T$ ，所以 $x = (T - 1 + W_x) / W_x$ ，即：

$$\text{最小隸屬度} = (\text{查詢門檻值} - 1 + \text{屬性相對權重}) / (\text{屬性相對權重}) \quad (17)$$

由於(16)，(17)兩式僅和個別屬性的最小隸屬度有關，故可適用於查詢條件中含有多個不同屬性權重值的情況(不一定剛好是3個)。

問題三之情形，若先求出最小隸屬度，則可將模糊查詢轉換成唯一(unique)的清晰查詢條件；反之，則需透過兩種方式，取得查詢結果。一是：逐筆對每一記錄(之各屬性隸屬度)之權重總和是否跨過門檻值，如此不僅造成資料庫 I/O 的成本亦產生逐筆檢視的計算成本。若令 R 為查詢之表格，其資料筆數(cardinality)以 $|R|$ 表之，而令利用轉換後之清晰查詢所取得之資料筆數為 $|Q|$ ，則查詢選擇率(selectivity) $s = |Q| / |R|$ ，那麼在資料的檢視成本可改善 $(1-s) \times 100\%$ 。在 I/O 方面的改善則依查詢所得資料之分佈而定，若大多集中於部分資料頁，改善較大，最高亦不超過 $(1-s) \times 100\%$ 。第二種方式是：將模糊條件轉換為各種不同屬性隸屬度組合之清晰條件，但此法則明顯造成查詢的負擔。然而，以式子(17)求可行解必定產生前置計算(pre-computation)的成本，當資料筆數 (cardinality) 極小，(例如，數次的 I/O，便可擷取完畢)，就不值得耗費執行此前置處理的成本。

七、查詢型態之分類與轉換

轉換的方式需考慮不同查詢型態，查詢型態是依使用者介面設計而定，不同的查詢型態會有不同的轉換流程，本文中將查詢型態區分為查詢型態一(修飾詞法)和查詢型態二(權重法)，分別說明如下。

查詢型態一(修飾詞法)：(以圖 4 為例)

- I. 取出使用者查詢語句中模糊項及修飾詞相對應的中介資料。
- II. 先將查詢所設的門檻值與修飾詞所代表的門檻值比較，取其大者作門檻值，再根據前述做法，轉換模糊條件為清晰查詢的三個狀況分別處理。
- III. 將步驟 II 結果組合為標準 SQL 指令對資料庫進行查詢。
- IV. 因查詢條件間為 And 或 Or 的關係，以公式(5)、(6)計算所有查詢結果記錄的符合度，並依符合度遞減程度排序。

查詢型態二(權重法)：(以圖 5 為例)

- I. 取出使用者查詢語句中模糊項及權重值相對應的中介資料。
- II. 以公式(14)計算各屬性值的相對權重值，再以(13)解模糊化。
- III. 如以(15)判斷(1-該屬性相對權重值後)大於查詢所設門檻值，則該屬性不需做查詢條件轉換；反之則以(17)為門檻值並轉換為清晰查詢。此一步驟主要在先行篩選掉達不到門檻值的資料，以減少步驟IV中需要計算隸屬度的資料量。
- IV. 計算所有查詢結果記錄的符合度，再次依門檻值做篩選，最後依隸屬度遞減程度排序。

伍、系統實作

一、系統環境

本系統以 Web 介面做開發，方便日後移植為企業 Intranet 系統或擴充開放為電子選購系統，並提供房屋之外觀展示圖片及便利親和的操作。系統之伺服器端所採用的軟、硬體，列於表 9，客戶端則採支援 IE5.5 以上瀏覽器的作業系統。

表 9：系統伺服器軟硬體配備

軟體		硬體	
作業系統	Windows2000	CPU	P III 1G
程式撰寫	ASP+VBScript	RAM	512M
應用軟體	IIS5.0	硬碟	20GB
資料庫	SQL2000		

二、系統介面與實例

系統使用的資料包含用以制定中介資料的查詢與成交資料及房屋資料表內容均源於信義房屋的房屋成交資料。查詢變數(列於表 10)是參酌表 1 各文獻中所歸納之影響因素，且考量實際資料可否取得及系統執行的效率，從中選擇較重要且適於記錄處理之項目而得。

系統實作部份以兩種不同查詢型態的介面—修飾詞法及權重法加以說明，其中僅以投射房屋資料表的部份屬性包括：區域、坪數、類型及總價，並選擇部份 tuples 為說明範例，如表 11 為房屋資料表的部份。

表 10：本文使用的查詢變數一覽表

查詢項目	說明
區域	以三碼的郵遞區號表示，如大安區、信義區
價位	買方所希望的房屋價格
坪數	買方所希望的房屋坪數
用途	共四項：透天厝、公寓、大樓、別墅
類型	共七項：住宅、店面、辦公、廠房、套房、車位、土地
屋齡	買方所希望的屋齡
樓層	買方所希望的樓層
朝向	分東、南、西、北、東北、東南、西北、西南

表 11：房屋資料表的部份記錄

編號	區域	坪數	類型	總價
A0001	士林	38.68	電梯華廈	250
A0002	士林	19.73	公 寓	310
A0003	士林	31.92	電梯華廈	525
A0004	士林	35.32	電梯華廈	1200
B0001	中山	15.57	套 房	350
B0002	中山	24.09	公 寓	690
B0003	中山	39.38	電梯華廈	1700
B0004	中山	43.93	透 天	1180
C0001	天母	30.00	電梯華廈	900
C0002	天母	60.00	別 墅	2500
C0003	天母	30.00	電梯華廈	670
C0004	天母	24.06	公 寓	733

請輸入查詢條件

區 域	完全在 ▾ 中正區 ▾
總 價	最好在 ▾ ==> <input checked="" type="radio"/> 低價位 ▾ <input type="radio"/> 300 ▾ 至 600 ▾ 萬
坪 數	最好在 ▾ ==> <input checked="" type="radio"/> 大坪數 ▾ <input type="radio"/> 30 ▾ 至 35 ▾ 坪
屋 齡	大概在 ▾ ==> <input checked="" type="radio"/> 新成屋 ▾ <input type="radio"/> 5 ▾ 至 以下 ▾ 年
樓 層	大概在 ▾ ==> <input checked="" type="radio"/> 不拘 ▾ <input type="radio"/> 15 ▾ 至 以上 ▾ 樓
朝 向	完全是 ▾ 南 ▾ 方
用 途	完全是 ▾ <input checked="" type="checkbox"/> 住宅 <input type="checkbox"/> 辦公 <input type="checkbox"/> 土地 <input type="checkbox"/> 車位 <input type="checkbox"/> 廠房 <input type="checkbox"/> 倉庫
類 型	大概是 ▾ <input type="checkbox"/> 透天 <input checked="" type="checkbox"/> 公寓 <input checked="" type="checkbox"/> 電梯華廈 <input type="checkbox"/> 別墅 <input type="checkbox"/> 套房
符合率	<input type="radio"/> 100% <input checked="" type="radio"/> 90% <input type="radio"/> 80% <input type="radio"/> 70% <input type="radio"/> 60% <input type="radio"/> 50% <input type="radio"/> 40% <input type="radio"/> 30% <input type="radio"/> 20% <input type="radio"/> 10%

圖 4：查詢型態一(修飾詞法)的操作介面

例 5.1：查詢「區域最好在士林區，坪數最好在 30~40 坪，類型完全在公寓或電梯華廈，價位完全在低價位，符合率為 80%。」的房屋。

- (1) 查詢門檻值：因查詢條件所設定的符合率為 80%，故其值為 0.8。
- (2) 查表 6 得區域修飾詞“最好在”的門檻值為 0.7，以查詢與區域修飾詞的門檻值求 MAX 得 0.8，再查表 2 並使用公式(1)，得區域條件值天母區，士林區。
- (3) 代入公式(8)並查表 3，產生坪數(30~40 坪)的隸屬函數： $\forall x \in R$

$$\mu_A(x) = \begin{cases} 0, & x < 30 \times (1 - 0.23). \\ F_L\left(\frac{x - 30 \times (1 - 0.23)}{30 \times 0.23}\right), & x \in (30 \times (1 - 0.23), 30). \\ 1, & x \in [30, 40]. \\ F_R\left(\frac{40 \times (1 + 0.25) - x}{40 \times 0.25}\right), & x \in (40, 40 \times (1 + 0.25)). \\ 0, & x > 40 \times (1 + 0.25). \end{cases}$$

- (4) 查表 6 得坪數修飾詞“最好在”的門檻值為 0.7，以查詢與坪數修飾詞的門檻值求 MAX 得 0.8，代入坪數(30~40 坪)的隸屬函數並四捨五入得坪數條件值 28~43 坪。
- (5) 查表 6 得修飾詞“完全在”的門檻值為 1，以查詢與區域修飾詞的門檻值求 MAX 得 1，再查表 8 並使用公式(1)，得類型條件值公寓，電梯華廈。
- (6) 查表 6 得價格修飾詞“完全在”的門檻值為 1，以查詢與價格修飾詞的門檻值求 MAX 得 1，再查表 5 並代入低價位隸屬函數，得價位條件值為小於 850 萬。
- (7) 轉換成對應的 SQL 查詢 select * from 房屋資料表 where 區域 in ('天母區', '士林區') and 價格 between 28 and 43 and 類型 in ('公寓', '電梯華廈') and 價格 <= 850。
- (8) 查出記錄 A0001、A0003、C0003 計算其個別屬性隸屬度：
 區域：查表 2 得 A0001、A0003 隸屬度為 1，C0003 為 0.8。
 坪數：代入坪數(30~40 坪)的隸屬函數，得三筆記錄隸屬度皆為 1。
 類型：因類型條件間為 Or 關係，查表 8 得三筆記錄隸屬度皆為 U(類型) = 0.7/公寓 + 1/電梯華廈，並代入公式(5)求 S-norm，得三筆記錄隸屬度皆為 1。
 價格：查表 5 得三筆記錄隸屬度皆為 1。
- (9) 因各條件間為 And 關係，代入公式(4)求得各記錄 t 之符合度 $\text{Mat}(t)$ ：
 $\text{Mat}(t) = \text{區域隸屬度} * \text{坪數隸屬度} * \text{類型隸屬度} * \text{價格隸屬度}$ 。
 $\text{Mat}(A0001) = 1 * 1 * 1 * 1 = 1$ ，同理， $\text{Mat}(A0003) = 1 * 1 * 1 * 1 = 1$ ，
 $\text{Mat}(C0003) = 1 * 1 * 1 * 0.8 = 0.8$ 。
- (10) 排序後結果列示如下：

編號	區域	坪數	類型	價格	符合度
A0001	士林	38.68	電梯華廈	250	1
A0003	士林	31.92	電梯華廈	525	1
C0003	天母	30.00	電梯華廈	670	0.8



請輸入查詢條件							
區域	中正區		總價	低價位 300 至 600 萬			
坪數	大坪數 30 至 35 坪		屋齡	新成屋 5 至 以下 年			
樓層	不拘 15 至 以上 樓		朝向	南方			
用途	<input checked="" type="checkbox"/> 住宅 <input type="checkbox"/> 辦公 <input type="checkbox"/> 土地 <input type="checkbox"/> 車位 <input type="checkbox"/> 廠房 <input type="checkbox"/> 倉庫						
類型	<input type="checkbox"/> 透天 <input checked="" type="checkbox"/> 公寓 <input checked="" type="checkbox"/> 電梯華廈 <input type="checkbox"/> 別墅 <input type="checkbox"/> 套房						
請輸入權重限制							
區域	高	總價	高	坪數	高	屋齡	中
樓層	中	朝向	低	用途	中	類型	低
符合率	<input type="radio"/> 100% <input checked="" type="radio"/> 90% <input type="radio"/> 80% <input type="radio"/> 70% <input type="radio"/> 60% <input type="radio"/> 50% <input type="radio"/> 40% <input type="radio"/> 30% <input type="radio"/> 20% <input type="radio"/> 10%						

圖 5：查詢型態二(權重法)的操作介面

例 5.2：查詢「區域：士林區；坪數：30~40 坪；類型：公寓或電梯華廈；價位：低價位；符合率為 90%；權重分別為區域：中、坪數：中、類型：高、價位：高」的房屋。

1. 代入公式 (8) 並查表 3，產生坪數 (30~40 坪) 的隸屬函數：

$\forall x \in R:$

$$\mu_A(x) = \begin{cases} 0, & x < 30 \times (1 - 0.23). \\ F_L\left(\frac{x - 30 \times (1 - 0.23)}{30 \times 0.23}\right), & x \in (30 \times (1 - 0.23), 30). \\ 1, & x \in [30, 40]. \\ F_R\left(\frac{40 \times (1 + 0.25) - x}{40 \times 0.25}\right), & x \in (40, 40 \times (1 + 0.25)). \\ 0, & x > 40 \times (1 + 0.25). \end{cases}$$

2. 由表 7 可查得權重值：

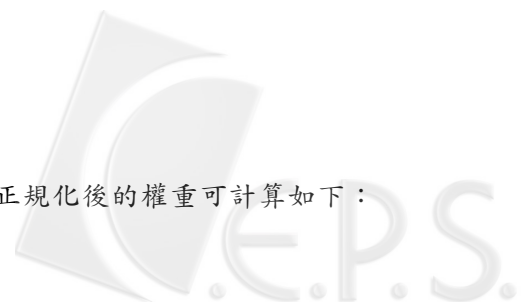
$$W_{\text{區域}} = (0.4257, 0.4733, 0.5209, 0.5685)$$

$$W_{\text{坪數}} = (0.4257, 0.4733, 0.5209, 0.5685)$$

$$W_{\text{類型}} = (0.7113, 0.7589, 0.8065, 0.8541)$$

$$W_{\text{價位}} = (0.7113, 0.7589, 0.8065, 0.8541)$$

令 $\Psi = W_{\text{區域}} + W_{\text{坪數}} + W_{\text{類型}} + W_{\text{價位}}$ ，正規化後的權重可計算如下：



$$\bar{W}_{\text{區域}} = W_{\text{區域}} / \Psi = (0.1496, 0.1782, 0.2114, 0.25)$$

$$\bar{W}_{\text{坪數}} = W_{\text{坪數}} / \Psi = (0.1496, 0.1782, 0.2114, 0.25)$$

$$\bar{W}_{\text{類型}} = W_{\text{類型}} / \Psi = (0.25, 0.2859, 0.3273, 0.3756)$$

$$\bar{W}_{\text{價位}} = W_{\text{價位}} / \Psi = (0.25, 0.2859, 0.3273, 0.3756)$$

依公式(13)解模糊得相對權重：

$$DEF(\bar{W}_{\text{區域}}) = 0.1973, DEF(\bar{W}_{\text{坪數}}) = 0.1973, DEF(\bar{W}_{\text{類型}}) = 0.3097$$

$$DEF(\bar{W}_{\text{價位}}) = 0.3097$$

3. 檢查是否需轉換為清晰查詢條件：

門檻值：因查詢條件所設定的符合率為 90%，故門檻值為 0.9。

區域：因 $(1 - \text{區域權重值}) = (1 - 0.1973) = 0.8027 < 0.9$ (門檻值)，代入公式(17)，得最小隸屬度為 $(0.9 - 1 + 0.1973) / 0.1973 = 0.49$ ，查表 2 並使用公式(1)得到士林區、天母區、中山區、內湖區。

坪數：同理，採公式(17)，得最小隸屬度 $(0.9 - 1 + 0.1973) / 0.1973 = 0.49$ ，代入坪數(30~40 坪)的隸屬函數得條件 26-45 坪。

類型：同理，採公式(17)，得最小隸屬度 $(0.9 - 1 + 0.3097) / 0.3097 = 0.68$ ，查表 8 並使用公式(1)，得到條件值公寓、電梯華廈、套房。

價格：同理，採公式(17)，得最小隸屬度 $(0.9 - 1 + 0.3097) / 0.3097 = 0.68$ ，查表 5 並代入低價位隸屬函數，得其條件值為小於 924 萬。

4. 轉換成 SQL 代入得 `select * from House where 區域 in ('天母區','士林區') and 坪數 between 26 and 45 and 類型 in ('公寓','電梯華廈','套房') and 價格 < 924`。

5. 查出記錄 A0001、A0003、C0001、C0003 計算其個別屬性隸屬度。

區域：查表 2 得 A0001、A0003 隸屬度為 1，C0001、C0003 為 0.8。

坪數：代入先前產生的隸屬函數，得四筆記錄隸屬度皆為 1。

類型：因類型條件間為 OR 關係，查表 8 得四筆記錄隸屬度皆為 $U(\text{類型}) = 0.7/\text{公寓} + 1/\text{電梯華廈}$ ，並代入公式(5)求 S-norm，得四筆記錄隸屬度皆為 1。

價格：查表 5 得 A0001、A0003、C0003 隸屬度為 1，C0001 為 0.22。

6. 求得各記錄符合度

$$\text{Mat}(A0001) = \min\{1 * 0.1973 + 1 * 0.1973 + 1 * 3097 + 1 * 0.3097, 1\} = 1.$$

$$\text{同理, Mat}(A0003) = 1, \text{Mat}(C0001) = 0.73, \text{Mat}(C0003) = 0.97$$

7. 以查詢門檻值篩選掉記錄 C0001 後排序，結果列示如下：

編號	區域	坪數	類型	價格	符合度
A0001	士林	38.68	電梯華廈	250	1
A0003	士林	31.92	電梯華廈	525	1
C0003	天母	30.00	電梯華廈	670	0.97

陸、結論

本文提出一模糊查詢轉置器之架構與處理流程，主要利用查詢需求與符合資料的回饋以及模糊理論，建構模糊查詢條件語句的中介資料，使得模糊項會依據查詢特質做不同的適切轉換。換句話說，就是運用企業實際交易資料來訂定隸屬函數及模糊項，透過資料的定期更新，讓模糊語句與清晰條件的轉換較其他文獻更能反映現況資料。此法迥異於(張振芳 1991)利用房屋的基本屬性、品質指標(銷售人員自訂)及使用者對指標偏好程度為準則，後者易流於主觀。

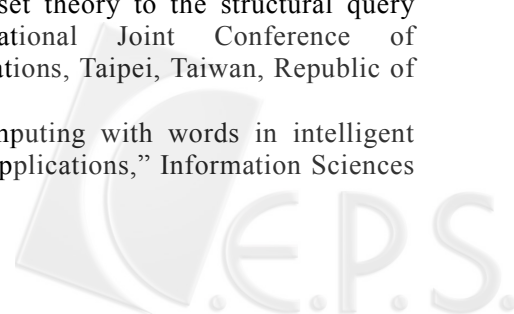
本文藉由查詢範例闡述轉置器所運用的各項模糊理論及技術，並且提供兩種口語化查詢型態，其查詢結果能根據使用者下達之容忍度放寬篩選條件，也能根據其與條件的符合度加以排序，並以代數乘積與代數加法取代傳統以標準交集及標準聯集運算的作法，提昇了資料搜尋的效率以及符合度的準確性。而轉換後所得的 SQL 標準語法亦可利用 DBMS 之最佳化器(Optimizer)，產生較佳之查詢計劃，得到快速有效的結果。

目前針對房屋選擇問題的文獻多屬於行為面的分析，較少著力於方法的應用。現有之購屋輔助系統大多整合專家系統，藉由專家知識與經驗法則，分析房地產購買之適當性(張振煌 1995)，此法能讓使用者能適切提出所需條件之各屬性值，但缺乏探討消費者本身之期望。另外，尚有以 GIS 輔助房屋的選購，主要構想是整合與公共行政相關之資訊及相關之空間資料，利用地理資訊系統動態展現鄰里環境，以輔助使用者之購屋查詢，而其考慮之公共設備僅試行在路燈管理的部分(江芳彬 1993)。此兩種作法主要均在輔助購屋者下決策，決策後的屬性值不一定為明確值，若能整合本文提出之模糊查詢系統，則能同時達到輔助決策與有效查詢之效果。

參考文獻

1. 王琬宜，2000，地理資訊系統在不動產查詢與分析上之應用，政治大學地政學系碩士論文。
2. 左心強、黃堯銘、劉俞志，2002，『應用模糊理論於房屋選購查詢系統』，2002 數位生活與網際網路科技研討會，六月。
3. 江芳彬，1993，關聯式資料庫中模糊查詢整合模式之建立-以房屋仲介資訊為例，台灣工業技術學院碩士論文。
4. 何志南，2000，住宅消費模式之建構與分析，成功大學碩士論文。

5. 何智明, 1998, 消費者住屋購買行為關鍵因素之研究-以台北都會區為例, 中興大學碩士論文。
6. 徐瑞豐, 1995, 模糊查詢與模糊關連式資料庫整合模式-以觀光旅遊全球資訊網路系統為例, 台灣科技大學碩士論文。
7. 翁志賢, 2000, 消費者購屋行為之探討-品牌權益, 品牌態度與購屋評估準則對購屋意圖之影響, 成功大學碩士論文。
8. 張振芳, 1991, 房地產仲介業略資訊系統中之買賣配對專家系統的建構研究, 東海大學企業管理研究所碩士論文。
9. 張振煌, 1995, 消費者購屋決策專家系統, 政治大學地政系碩士論文。
10. 郭嘉昌, 2000, 新竹科學工業園區員工對住宅環境偏好之分析, 中華大學碩士論文。
11. 劉俞志、黃堯銘, 2002, 『運用近似語意相等在模糊資料庫上的模糊查詢』, 第13屆國際資訊管理學術研討會, 五月, pp.447-454。
12. 鄭淵聰, 1998, 中老年人對老人住宅購買行為及行銷策略之初探性研究-以台中都會區為例, 雲林科技大學碩士論文。
13. 蘇振昌, 1999, 消費者的網路資訊搜尋對購屋行為之影響, 輔仁大學碩士論文。
14. Buckles, B. P. and Petry, F. E., "A fuzzy representation of data for relational databases," *Fuzzy Sets and Systems* 7, 1982, pp:213-226.
15. Buckles, B. P. and Petry, F. E., "Extending the fuzzy databases with fuzzy numbers," *Information Sciences* 34, 1984, pp:145-155.
16. Buckles, B. P. and Petry, F. E., "Information-theoretical characterization of fuzzy relational databases," *IEEE Trans. Systems. Man. Cybernet.*, SMC (13:1) 1983, pp:74-77.
17. Chen, S. M. and Jong, W. T., "Fuzzy Query Translation for Relational Database Systems," *IEEE Trans. Systems. Man. Cybernet.*, Part B:CYBERNETICS (27:4) 1997, pp:714-721.
18. Chen, S. M., "Using fuzzy reasoning techniques for fault diagnosis of the J-85 jet engines," *Proceedings of the Third National Conference on Science and Technology of National Defense*, Taoyuan, Taiwan, Republic of China, 1995, pp:216-219.
19. Chen, S. M., Ke, J. S., and Chang, J. F., "Techniques of fuzzy query translation for database systems", in *Proc. 1986 Int. Computer Symp*, Tainan, Taiwan, R. O. C, Dec.1986, pp: 1281-1290.
20. Chen, Yu-Chuan and Chen, Shyi-Ming, "Techniques of Fuzzy Query Processing for Fuzzy Database Systems," *Proceedings of the Fifth Conference on Artificial Intelligence and Applications*, 2000, pp:361-368.
21. Frank, M. J., "On the simultaneous associativity of $F(x, y)$ and $x+y-F(x, y)$," *A Equations Mathematicae* (19:2) 1979, pp:194-226.
22. Giarratano, J. and Riley, G., "Expert Systems:Principles and Programming," Boston, MA:PWS-Kent, 1989.
23. Gupta, J. Oi, "Theory of T-norms and fuzzy inference methods," *Fuzzy Sets and Systems* (40:3) 1991, pp:431-450.
24. Hou, Y. C. and Chen, C. M., "Apply the fuzzy set theory to the structural query language of database," *Proceeding International Joint Conference of CFSA/IFIS/SOFT'95 on Fuzzy Theory and Applications*, Taipei, Taiwan, Republic of China, 1995, pp:107-113.
25. Kacprzyk, Janusz and Zadrozny, Slawomir, "Computing with words in intelligent database querying: standalone and Internet-based applications," *Information Sciences* 134, 2001, pp:71-109.



26. Kaufmann, A. and Gupta, M. M., "Fuzz Mathematical Models in Engineering and Management Science," Amsterdama: North-Holland, 1988.
27. Kaufmann, A. and Gupta, M. M., "Introduction to fuzzy arithmetic," Van Nostrand Reinhold, New York, 1991.
28. Lin, Yun-Shyang and Chen, Shyi-Ming, "Using Automatic Clustering Techniques for Fuzzy Query Processing in Relational Database Systems," The Eleventh National Conference on Information Management, 2000, pp:1185-1194.
29. Petry, F. E., "Fuzzy Databases: Principles and Applications," Kluwer Academic Publishers, Boston, 1996.
30. Schweizer, B. and Sklar, A., "Associative functions and abstract semigroups," Publ. Math. Debrecen (10) 1963, pp:69-81.
31. Sheno, Sujeet and Melton, Austin, "Proximity relations in the fuzzy relational database model," Fuzzy Sets and Systems (31:3) 1989, pp:285-296.
32. Yager, R. R., Ovchinnikov, S., Tong, R. M., and Nguyen, H. T., "Fuzzy Sets and Applications : Selected Papers by L. A. Zadeh," Wiley-Interscience, 1987, pp:46-79.
33. Yager, R., "On a general class of fuzzy connectives," Fuzzy Sets and Systems (4:3) 1980, pp:235-242.
34. Yn, W. and Bien, Z., "Design of fuzzy logic controller with inconsistent rule base," Journal of Intelligent & Fuzzy Systems (2:2) 1994, pp:147-159.
35. Zadeh, L. A., "Fuzzy sets," Information and Control (8) 1965, pp:338-353.
36. Zadeh, L. A., "The concept of a linguistic variable and its application to approximate reasoning," Inf. Sci. (8:13) 1975, pp:199-249.
37. Zimmermann, H. J., "Fuzzy Set Theory and Its Applications," Kluwer Academic Publishers, Dordrecht, 1996.
38. Zimmermann, H. J., "Fuzzy Sets, Decision Making, and Expert Systems," Kluwer Academic Publishers, 1987.

