

一個輔助投資知識管理之 K 線知識本體

李俊宏

中州技術學院資訊工程系

leonlee@dragon.ccut.edu.tw

洪偉智

中州技術學院工程技術研究所

welton14@yahoo.com.tw

摘要

本論文提出一個以 K 線理論為基礎之知識本體來輔助投資知識之儲存與管理。我們首先將財金時間序列轉換成 K 線的表示法。接著利用模糊語詞來表示原本不精確且模糊的 K 線。然後使用者便能夠利用決策樹等理論挖掘出有興趣的 K 線型態。進而將這些型態儲存起來，以利日後進行投資決策使用。讓使用者能根據所分類的知識本體在不同的情況下進行投資知識的再使用，取得更好的投資決策建議。我們也建立了一個系統來驗證所提出的想法。依據實驗結果，本方法可以與其他的財金時間序列方法一起並用來增加使用者進行投資決策時的正確性。

關鍵詞：K 線理論、財金時間序列、模糊理論、知識本體

一、簡介

在人類社會中有許多領域的參與者需要經年累月的累積經驗方能成為該領域的專家。在專家系統的研究中[1]，最難的部份便是如何有系統地將這些專家的知識擷取出來，利用特殊的表示法，編寫到電腦中，進而利用電腦自動化計算推理的特性重複將這些專家的知識作最大的利用。

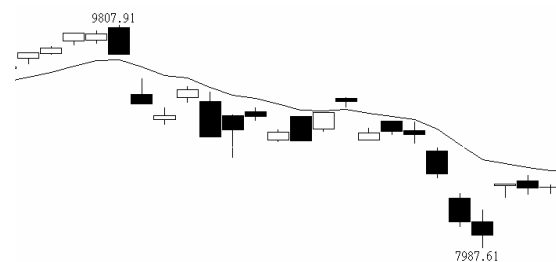
如股市的投資決策便相當須要專家的知識來輔助。長久以來股票等金融交易的走勢預測一直是許多人研究的題目。因為這不僅僅是一種學術上研究，同時它還包含了許多商業上的價值。然而具時間序列性質的股票資料，由於其數字化以及連續的性質，須要特別的處理與轉換方能產生好的資料挖掘、型態辨識或財金預測的結果[2]。而且股市專家並不只單一的考慮股票數字的變化，而且同時會利用過往的經驗來輔助他們的判斷。如利用一些技術指標，配合大盤的或類股的走勢資訊來進行個股變化的預測。

在過去有許多處理時間序列的方法在超過

百篇的論文中被提出來。如：群聚、分類、時間序列切割等。在實現了成打的時間序列分析方法後，Keogh 和 Kasetty 主張針對不同的時間序列使用不同的處理方式是相當合理的[3]。事實上，在現實的環境中，同樣的數據在不同的環境下或由不同的人解釋也會有不同的結果或解讀。

時間序列資料的前置處理是與資料的領域和知識相關的工作。我們相信對於每一個時間序列的分析與處理並沒有一個最有效率的方法，然而小心地針對資料進行前置處理相信必然可以降低預測演算法的複雜度和增加準確度。

源自於日本江戶時代米市交易之 K 線圖 (Candlestick Chart)，是目前各類交易市場中最常被廣泛使用來描述市場交易行為的一種圖表之一[4]。其特點是從單一條 K 線就可以清楚視覺化地表示出每個交易日的開盤價、收盤價格、最高價、與最低價之間的關係。同時也可以觀察出本日的交易趨勢是漲或是跌，行情波動的劇烈與否。這些要素充分的反應了市場之供需層面以及投資者的心理層面。並且若將連續的 K 線呈現出來，便可以形成具有時間關聯性之 K 線圖表。並且一些連續 K 線的組合可以形成一個 K 線樣式 (Candlestick Pattern)。如圖一所示為台灣股市大盤的 K 線走勢圖。其中黑色與白色之條狀圖為交易單日之 K 線，中間之線條為每五日之平均收盤價格所連起來之平均線。



圖一. 台灣股市大盤的 K 線走勢圖

有經驗的投資者可以藉由觀察時間序列之 K 線圖的變化尋找出特定的樣式，而輕易的獲悉在一段時間之中，整體交易行情的趨勢是否持續，景氣是否有反轉的跡象，以及是否出現買賣訊號。這也是 K 線圖目前廣泛的被使用於各類交易市場之原因[5]。

一般介紹 K 線理論的書籍，通常所呈現的

是一些在時間序列上之 K 線樣式圖形，每一根 K 線代表了當日的漲跌態勢，再由一小段時間序列之 K 線組成一個 K 線樣式，再針對這一些 K 線樣式圖形做一番投資策略的解釋，用以教導投資理論的初學者或是交易員如何去分析 K 線樣式，並且去理解這一些 K 線樣式所代表的意涵，以達成幫助他們做出一個比較容易獲利且不容易受到損失的投資決策。然而，K 線理論的初學者面對這一些沒有任何數據可以做為學習參考的 K 線樣式圖，而且描述 K 線圖形語意的也僅僅是一些比較模糊的字句；例如：開高、開低、長上影、中紅 K 線 等等，可預見的這個學習過程將不會十分容易，初學者必需花費漫長的時間去學習，反覆的練習與分析，才足以建立比較可靠的決策經驗。

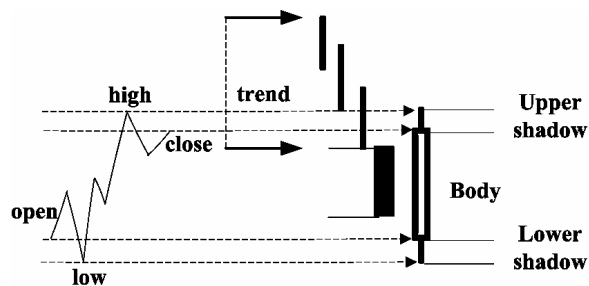
在[6]中，作者嘗試著以模糊邏輯的方式，將這一些存在於時間序列中之不明顯的經驗法則，轉換成一般人比較容易理解之語意，即為模糊邏輯之語言變數 (linguistic variable)。其好處是此方法不以複雜的數學模型來預測未來市場行情是漲或是跌，而是以 K 線理論中的 K 線樣式之時間序列做為未來市場行情判讀之參考資料來源。由於有 K 線理論的經驗知識做為輔助，使得整體的預測過程可以得已明瞭。同時因為模糊邏輯之語意變數的作用，改善整個 K 線理論的學習過程並且讓使用者可以互相分享彼此之見解。本論文跟隨之前的研究，以知識本體的理論來加強對模糊 K 線的定義，以期可以讓電腦系統也可以使用過往的投資經驗。

本論文的各個章節如以下所述：第二節簡介相關的背景知識、第三節介紹 K 線的知識本體、第四節描述系統之實現、第五節則為本論文研究的總結、最後一節是本研究之參考文獻。

二、相關背景知識

(一) K 線理論

K 線的功能，主要是在說明交易價格與時間之間的變化關係。如圖二所示。



圖二. K 線的轉換與 K 線樣式

一條 K 線基本上可以分成三個部分，K 線實體 (body)、K 線的上影線 (upper shadow) 與下

影線 (lower shadow)。使用者可以利用 K 線來描述某一段時間內的數字變化。如交易時間長度是以天做為時間單位，那 K 線線體將會由當日的開盤價 (open) 與收盤價 (close) 之價格差所形成，上影線與下影線的部分則由當日的最高價 (high) 與最低價 (low) 與相對應之開收盤之價格差所形成。K 線線體的顏色會因為開盤與收盤的價格差而有所不同。當開盤價小於收盤價則以白色的線體之 K 線表示該交易日是屬於一個趨勢向上的行情，反之則使用黑色線體來表示該交易日行情趨勢向下。

單一根 K 線可以呈現出當次交易的交易行為。不過對於投資者而言，這些有限的資訊量要提供做為決策建議與評估仍嫌不足，所以資深的投資專家會對於時間軸上連續多日之 K 線加以分析、觀察，也就是觀察的對象由單一 K 線擴展為 K 線樣式。K 線樣式可以由一根或是多根的 K 線所組成。假設圖二左方的線型為單一股票價格一日的走勢，市場在一個向下趨勢中，在一根黑色 K 線結束後，隔日的開盤價開低，中間價格振盪走高，最後的收盤價大於前一天的收盤價。此一 K 線樣式在 K 線理論中被稱為多方吞噬線，是一個價格將反轉的訊號。投資者可以藉由這個訊號做出買進的決策。

(二) 模糊 K 線理論

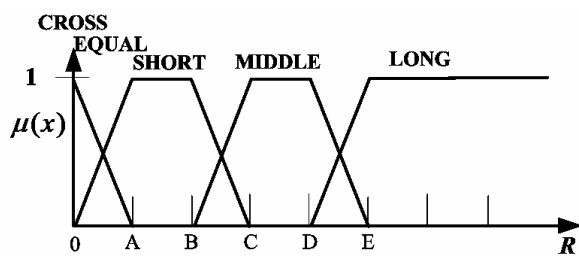
藉由觀察 K 線樣式的變化，確實可以做為投資者決策的重要評估要件。但是到底上下影線、實心線要多長或者是這些 K 線怎樣的組合才有最大的影響力，以及如何從一堆的數據資料中找出這些 K 線樣式對人類而言是一件耗力費時且艱難的工作。因此我們提出了一個以模糊語詞來描述 K 線以及 K 線樣式的方法，將找尋 K 線樣式的工作交由電腦來自動化地執行，以提供投資人投資決策參考。

以下為將數據轉成模糊 K 線前的前置處理公式：

$$\begin{aligned} L_{upper} &= [high - \max(open, close)] / open \\ L_{lower} &= [\min(open, close) - low] / open \\ L_{body} &= [\max(open, close) - \min(open, close)] / open \end{aligned} \quad (1)$$

其中“L”代表上下影線或實心線的長度；“open”，“close”，“high”，and “low”代表特定時間內的數據變化；“max”和“min”為在“open”，“close”二者之間取得最大值或最小值的函數。

接下來我們定義了四個模糊語句來描述上下影線和實心線的長度：EQUAL、SHORT、MIDDLE、LONG。其模糊歸屬函數如圖三所示。



圖三. K 線長度之模糊歸屬函數

其中 EQUAL 使用右線性模糊歸屬函數來定義。方程式如下所示：

$$\text{right_linear}(x : a, b) = \begin{cases} 1 & x < a \\ (b-x)/(b-a) & a \leq x \leq b \\ 0 & x > b \end{cases} \quad (2)$$

而 LONG 則由左線性模糊歸屬函數來定義。

$$\text{left_linear}(x : a, b) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \leq x \leq b \\ 1 & x > b \end{cases} \quad (3)$$

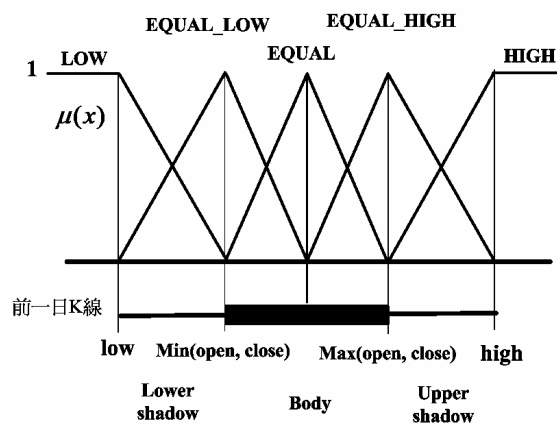
而 SHORT、MIDDLE 則由梯型模糊歸屬函數來定義。方程式如下所示：

$$\text{trapezoid}(x : a, b, c, d) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \leq x < b \\ 1 & b \leq x < c \\ (d-x)/(d-c) & c \leq x < d \\ 0 & x \geq d \end{cases} \quad (4)$$

我們利用不同參數的改變來描述不同使用者對模糊詞句的定義。此外實心線的顏色也是一個重要的 K 線性質，以下的規則實現實心線的定義：

- If open - close > 0 then the body color is BLACK.
- If open - close < 0 then the body color is WHITE.
- If open - close = 0 then the body color is CROSS.

要模型化一個 K 線樣式只有定義 K 線是不足的，我們尚須定義連續二條 K 線之間的關係。二條 K 線的關係與當日的開盤價位、收盤價位、當日開收盤價位與前日的收盤價位相比較之差距有關。因此我們用來描述 K 線關係的模糊詞句為：OPEN LOW, OPEN EQUAL_LOW, OPEN EQUAL, OPEN EQUAL_HIGH, and OPEN HIGH, CLOSE LOW, CLOSE EQUAL_LOW, CLOSE EQUAL, CLOSE EQUAL_HIGH, 和 CLOSE HIGH。其模糊歸屬函數如圖四所示。



圖四. K 線關係之模糊歸屬函數

我們使用右、左線性函數來模型化 LOW 和 HIGH。而其中 EQUAL_LOW、EQUAL、EQUAL_HIGH 由以下之三角模糊函數來定義：

$$\text{triangle}(x : a, b, c) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \leq x \leq b \\ (c-x)/(c-b) & b < x \leq c \\ 0 & x > c \end{cases} \quad (5)$$

為了使 K 線模糊化的結果更能夠接近實際的狀況，我們對於這些描述 K 線的模糊集合再定義了一些修飾詞如 VERY, ABOVE, BELOW 等。這部份我們使用 FuzzyToolKit [7] 的協助來實現。舉例來說明，當有一個明確值輸入於 X 軸且對應到的範圍是在本系統之中是屬於 MIDDLE 的模糊集合範圍，若是有加上 ABOVE 這個關鍵字來修飾這一個 MIDDLE 模糊集合，則 MIDDLE 模糊集合範圍將是會從原先定義的集合範圍擴展至 X 軸的 Max。若是 MIDDLE 的修飾詞修改成為 BELOW，那經過修飾過後的模糊集合範圍將會由原本的範圍往左延伸至 X 軸座標等於 0 的位址。

(三) 知識本體

Ontology 一開始起源於哲學，後來 Ontology 這個詞漸漸的用於其它的領域，所以對於每個領域對於 Ontology 的定義都有一點不同。在 [8] 中對 Ontology 的定義為對某個領域內所用到的概念給予明確且正式的敘述，並且描述了概念與概念之間的關係。[9] 的作者認為 Ontology 中定義了分享資訊所需要的共同字彙，包含給予字彙清楚的定義以及關係的描述，進而讓機器可解讀這些概念與關係。經由概念被清楚的定義與描述，當雙方在談論事情時，都能知道所要表達的東西為何，也可讓機器了解人所想要表達的想法，甚至是讓機器與機器相互了解。在 [8] 中列出了幾項為何要發展大家 Ontology 的理由：

- 分享人於人或智慧型代理人 (Agent) [10] 之間對於資訊架構的共同協議：就是說大

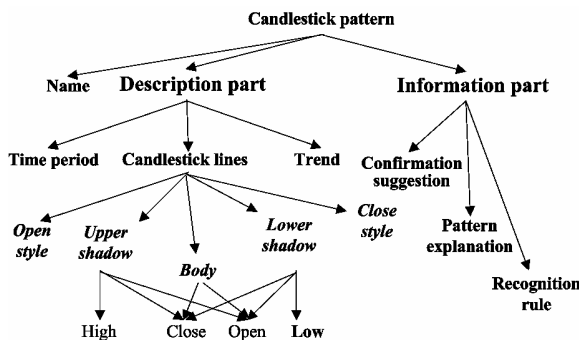
家可以用一個共同的標準來發佈與分享資訊。例如：現在有三家提供購物的網站，大家都用同樣格式來發佈的資訊。當我們要買東西，透過代理人很容易就可以用同一套方法把大家的資訊都收集起來。

- 領域知識的再利用：要對某個領域的認識用Ontology給表示出來，不是件很容易的事，因為對於不同的領域的知識表達，都需要不同的領域專家。若研究學者或團體已發展了此領域的Ontology，則可以我們直接使用已存在的Ontology。當我們需要建立更大的範圍的Ontology時，則可整合現存較小範圍的Ontology。
- 將領域知識與運作知識分開：當我們為某個領域開發了某個應用程式，若我們將領域知識與運作知識分開，我們只要替換掉領域知識的部份，用同一種運作的方法，就可以開發出適合另一領域的應用程式了。這也是Ontology常被用到的地方。

簡單來說，我們可以用 Ontology 來表達對於某一領域的知識。Ontology 通常由 Class、Slot 和 Relation、Instance 所組成。Class(也時候稱為 Concept)為 Ontology 中最主要的部分，可以為來描述領域中的一種想法。例如交通工具、火車、飛機或 K 線等，這些都是 Class。Property (也稱為 Attribute、Property、Role)是用來描述 Class 的特性或屬性如上下影線，實心線等 而Relation是用來說明Class 與Class 之間的關係。Instance 稱為實例，為概念或類別的一個例子，實例將會推有類別的特性與屬性還有與其它類別的關聯。例如：黃豆交易市場的 K 線和台股大盤的 K 線都是 K 線走勢圖的實例，但是交易的標的物不同，價格走法也不同。

三、K 線的知識本體

在本論文中，我們利用模糊理論以及知識本體，定義出一個 K 線理論的知識，而寫好的電腦程式便能利用這個定義好的知識本體來記錄對投資人有意義的 K 線樣式。所定義之知識本體簡圖如圖五所示：



圖五. K 線樣式之知識本體簡圖

一個 K 線樣式有三個屬性：Name 為特定 K 線樣式的名稱，Description Part 描述總成 K 線樣式的 K 線，Information Part 則記錄此一 K 線樣式的人為建議。而其中 K 線之上下影線、實心線，開收盤的方式則如第三節中所述。一個 K 線樣式的例子如表一所示。

表一. K 線樣式實例

Pattern name: 多頭吞噬	Pattern information part
Pattern description part	Confirmation suggest: 須確認本樣式是有效。
Followed trend: Down trend	Confirmation information: 隔日之開盤價不能低於前一日的收盤價
K 線 Candle line0 Open style: VERY OPEN_LOW Close style: VERY CLOSE_HIGH Upper shadow: null Body: ABOVE MMIDDLE Body color: WHITE Lower shadow: null Candle line1 Open style: BELOW OPEN_HIGH Close style: BELOW CLOSE_HIGH Upper shadow: null Body: ABOVE SHORT Body color: BLACK Lower shadow: null	Recognition rule: 1. 走勢在一個下降趨勢中。當日開低。 2. 第二天的實心線整個蓋過前一天的實心線 .. Pattern explanation: 在下降趨勢中，為一黑一紅組合稱之多頭吞噬。多頭吞噬與空頭吞噬意義大於貫穿與烏雲罩頂，但首先需決定大盤的趨勢停損可設在多頭吞噬排列最低價
Interested time period: DAY	

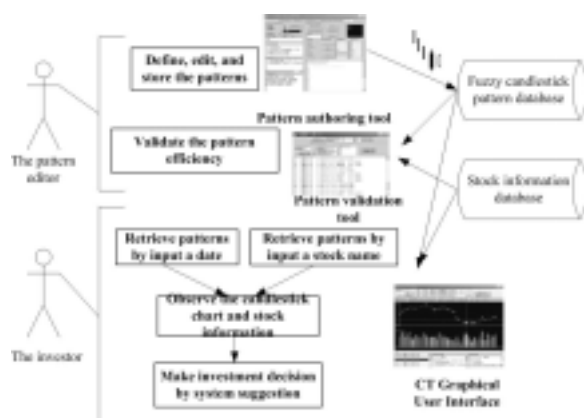
在 K 線樣式的描述部份可以使用決策樹等演算法來由時間序列的資料中找出相關的 K 線樣式，再驗證被找出來的 K 線樣式的有效性，最後由投資者自行為這些有投資意義的 K 線樣式加上描述部份。由系統所辨識出來的 K 線樣式資料不但是能夠解釋其意義且對於使用者是非常有意義之參考資訊。藉助電腦的高效能運算與快速資料處理的優點，來縮短投資者或是交易員在尋求投資標的物的過程，進而成為交易員或是投資者的一個決策輔助工具。交易員及投資決策者藉由辨識出來的 K 線樣式，可以很清楚的理解目前市場的趨勢與並利用作為風險的評估。如此逐步的累積投資的知識和經驗，相信對投資者在進行投資決策時會有相當大的助益，並且可以與他人分享投資的經驗。

四、系統實現

我們並且以模糊K線樣式理論為核心技术，實做一模糊K線理論學習系統。我們期待此系統能夠針對台灣證券交易市場的資料加以收集分析，主動的去分析這一些資訊，再把所辨識出來的K 線樣式以圖式方式標示出來交給投資人進行編輯進而建立系統內建的K線理論知識庫。如此系統便能對於K線樣式規則做解釋。

這個系統一方面使得初學者可以比較容易

理解K線理論的樣式與規則型態。使用者也可以透過一K線樣式產生器來編輯出屬於個人獨特的投資見解K線樣式規則，並且交由系統去驗證此K線樣式的合理性。圖六為此系統之運作流程。



圖六. 模糊K線理論學習系統運作流程

圖七與圖八則分別為系統之使用者介面與K線之使用者編輯工具。



圖七. 系統之使用者介面



圖八. K線編輯工具

我們同時以台積電 (TSMC2330) [11] 為例進行系統的測試。我們以 1997-10-23 至 2002-12-25 的台積電股價資料當輸入，讓系統

自動找出相關的 K 線樣式。共有 1342 個樣式被轉換出來。其中系統共找出了 79 個有價值的 K 線樣式。我們再以這些樣式為知識庫，對 2003 年的台積電股票進行投資。

表二為使用樣式投資的過程表。我們定義了如下的差異函數 v 來驗證樣式的有效性。

$$v = \frac{close_{t+5} - close_t}{close_t} \times 100 \quad (6)$$

其中 $close_t$ 為樣式出現當日的收盤價，而 $close_{t+5}$ 為樣式出現後五日的收盤價。

表二. 台積電預測結果

買進日期	差異
2003-12-19	3.3
2003-11-21	-3.1
2003-11-20	-0.7
2003-11-04	-0.7
2003-10-01	5.3
2003-09-30	6.8
2003-09-17	-4.9
2003-09-12	3.6
2003-08-15	4.9
2003-08-13	5.9
2003-07-28	3.4
2003-07-22	2.7
2003-07-02	3.3
2003-06-17	-0.9
2003-05-30	9.3
2003-03-13	8.6
Total variance: 46.8	

由以上的結果可知，電腦自動擷取的 K 線的確有其一定的效力。我們雖然無法保證使用 K 線樣式的確可以讓投資人獲利，然而我們相信如果使用者善用這些樣式可以作出更正確的投資決策。

五、結論

K 線理論是一些投資專家累積了長時間的經驗法則，投資者可藉由 K 線理論中所記載之 K 線樣式對於市場之 K 線樣式加以觀察，以分析投資時機是否來臨並做出投資決策。但是往往要掌握其精神與技巧卻要花上一段的時間去學習如何分辨這一些特殊 K 線樣式。藉由本論文所提出的方法，可以有效的將金融時間序列的轉換成為一個比較容易令人理解的語意，進行建立一個支援 K 線理論記錄之知識本體。使用者可利用系統自行找出有意義之 K 線樣式，也可以將 K 線理論之過去的經驗法則透過模糊邏輯的方式轉

換成為知識定義與規則並且記錄於知識庫之中。如此可以縮短了過去對於經驗法則的學習時間,存在於知識庫中之相關經驗也可以達成與他人分享之目的。

六、參考文獻

- [1] G. Joseph and R. Gray, "Expert Systems – Principles and Programming 3rd", New York, PWS Publishing Company, 1998.
- [2] K. Mehmed, *Data Mining: Concepts, Models, Methods, and Algorithm*, John Wiley & Sons, 2003, pp. 19-38.
- [3] E. Keogh and S. Kasetty, "On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration," in *2002 proc. SIGKDD conf.*, pp. 102-111.
- [4] G. L. Morris, *Candlestick Charting Explained: Timeless Techniques for Trading Stocks and Futures*, 2nd. ed., McGraw-Hill Trade, pp. 8-139, 1995.
- [5] J. C. P. Shieh, *Contemporary Investments—Analysis And Management*, Taipei, Taiwan, Best-Wise, pp. 317-431, 1998.
- [6] C.H.L Lee, A. Liu, and Wen-Sung Chen, "Pattern Discovery of Fuzzy Time Series for Financial Prediction," *IEEE Transaction on Knowledge and Data Engineering*, vol 18, no. 5, 2006, pp. 613-625.
- [7] NRC Fuzzy Toolkit. Available: http://www.iit.nrc.ca/IR_public/fuzzy/fuzzyJDocs/index.html
- [8] N. F. Noy and D. L. McGuinness, "Ontology Development 101: A guide to Creating Your First Ontology," Technical Report KSL-01-05, Stanford Medical Informatics, Stanford University, 2001.
- [9] B. Chandrasekaran, J.R. Josephson, and V.R. Benjamins, "What Are Ontologies, and Why Do We Need Them," *IEEE Intelligent Systems*, vol. 14, 1999, pp. 20-26.
- [10] C-S Lee and C-Y Pan, "An Intelligent Fuzzy Agent for Meeting Scheduling Decision Support System," (SCI) *Fuzzy Sets and Systems*, vol. 142, no. 3, 2004, pp. 467-488.
- [11] <http://www.tse.com.tw/ch/index.php>