

# 基於知識本體與模糊語意於RSS搜尋之研究

黃鼎涵、楊豐兆

E-mail: 9707368@mail.dyu.edu.tw

## 摘要

在目前Web 2.0網路盛行的時代，RSS(Really Simple Syndication)已經被採用為訂閱與發佈網站消息的重要媒介，網站建置者可以藉由RSS來發佈資訊給使用者，使用者也可以挑選他想要接收的RSS類別，這種知識分享與使用者互動之設計正好符合Web2.0的精神。由於每天在網路上所發佈資訊過多，一旦當使用者訂閱越多的RSS，只要RSS檔每更新一次，就會湧入大量的資訊。隨著時間過的越久，其資訊量越是驚人，使用者如何在這麼多的資訊中，有系統的過濾並且找出自己所需要的資訊，或是想從資訊中進行延伸閱讀的動作，來了解事件的前後關係，甚至可以從搜尋的過程中發現新的知識。本研究採取Web的搜尋介面，以現有網路新聞分類類別先建立出知識本體，依據使用者所輸入搜尋的文句意義與RSS檔的標題欄位做模糊語意的分析和關聯，再將根據語意模糊出來的結果到建立的知識本體中做解析和映對，最後呈現出與使用者語意最相近的資訊呈現出來。本研究成果包括：第一部份是建立起以新聞為主的知識本體，第二部份是運用模糊理論以及與語意網來建構搜尋平台，第三部份則是結合搜尋平台以及知識本體完成整個介面。

關鍵詞：知識本體;模糊語意;RSS

## 目錄

內容目錄 中文摘要	iii	英文摘要	iii
iv 致謝辭		v 內容目錄	v
vi 表目錄		viii 圖目錄	viii
ix 第一章 緒論	1	第一節 研究背景	1
1 第二節 研究動機	2	第三節 研究目的	2
3 第四節 研究範圍與限制	5	第五節 研究流程	5
5 第六節 論文架構	8	第二章 文獻探討	8
9 第一節 模糊理論	9	第二節 模糊搜尋	9
12 第三節 RSS	16	第四節 語意網	16
16 第五節 知識本體	18	第六節 TOVE	18
Ontology工程方法	22	第三章 系統需求分析	24
系統目標	24	第一節 系統目標	24
系統核心設計	40	第二節 使用者需求分析	33
的建置	41	第一節 新聞知識本體	40
50 第五節 同義詞與相關詞	49	第二節 知識本體的建置	40
50 第五章 系統實作與效能評估	55	第三節 模糊語意搜尋的建置	49
55 第一節 系統開發工具與環境	55	第四節 模糊語意搜尋的效能	55
55 第二節 系統介面與功能	55	第五節 系統效能評估	55
60 第六章 結論與未來研究發展	66	第一節 研究貢獻	66
60 第二節 未來研究方向	67	第二節 參考文獻	68
67 表目錄	68	表目錄	68
表 5-1 語意搜尋效率測試結果	61	表 5-2 字串搜尋效率測試結果	61
63 表 5-3 測試結果比較表	64	圖目錄	64
7 圖 2-1 三角形歸屬函數	11	圖 1-1 研究流程圖	11
12 圖 2-2 梯形歸屬函數	12	圖 2-1 三角形歸屬函數	11
17 圖 2-3 鐘型歸屬函數	12	圖 2-2 梯形歸屬函數	12
25 圖 2-4 RDF Graph	17	圖 2-3 鐘型歸屬函數	12
25 圖 2-5 TOVE本體論工程	22	圖 2-4 RDF Graph	17
28 圖 3-1 系統架構圖	25	圖 2-5 TOVE本體論工程	22
28 圖 3-2 離散型模糊集合	28	圖 3-1 系統架構圖	25
30 圖 3-3 新聞與類別模糊之關係矩陣	28	圖 3-2 離散型模糊集合	28
35 圖 3-4 新聞類別的模糊語意網路	29	圖 3-3 新聞與類別模糊之關係矩陣	28
37 圖 3-5 語意搜尋架構圖	30	圖 3-4 新聞類別的模糊語意網路	29
42 圖 3-6 新聞分享與RSS發佈的使用案例圖	34	圖 3-5 語意搜尋架構圖	30
46 圖 3-7 模糊語意搜尋的使用案例圖	35	圖 3-6 新聞分享與RSS發佈的使用案例圖	34
47 圖 3-8 網際網路使用者與搜尋系統的循序圖	35	圖 3-7 模糊語意搜尋的使用案例圖	35
37 圖 3-9 新聞RSS搜尋的活動圖	37	圖 3-8 網際網路使用者與搜尋系統的循序圖	35
42 圖 3-10 系統部署圖	38	圖 3-9 新聞RSS搜尋的活動圖	37
45 圖 4-1 TOVE Ontology工程方法概念流程圖	42	圖 3-10 系統部署圖	38
46 圖 4-2 新聞分類四大類別關係圖	45	圖 4-1 TOVE Ontology工程方法概念流程圖	42
46 圖 4-3 新聞分類類別與子類別階層圖	46	圖 4-2 新聞分類四大類別關係圖	45
47 圖 4-4 國際區域類別與子類別階層圖	46	圖 4-3 新聞分類類別與子類別階層圖	46
47 圖 4-5 新聞來源與子類別階層圖	47	圖 4-4 國際區域類別與子類別階層圖	46
48 圖 4-6 新聞知識本體架構圖	48	圖 4-5 新聞來源與子類別階層圖	47
		圖 4-6 新聞知識本體架構圖	48
		圖 4-7 相關詞樹狀結構	48

50 圖 4-8 模糊推論 . . . . .	51 圖 4-9 CKIP定義的Tagging Tree . . . . .
52 圖 4-10 詞彙之間POS Similarity模糊集合 . . . . .	53 圖 4-11 詞彙之間TV Similarity模糊集合 . . . . .
54 圖 5-1 系統功能架構 . . . . .	56 圖 5-2 會員登錄前首頁 . . . . .
57 圖 5-3 會員登錄後首頁 . . . . .	57 圖 5-4 新增RSS Feed . . . . .
58 圖 5-5 刪除RSS Feed . . . . .	58 圖 5-6 搜尋RSS新聞的介面 . . . . .
圖 5-7 模糊搜尋結果範例 . . . . .	60

## 參考文獻

- 一、中文部份 李建興(2003), 基於Ontology架構之CMMI Level 2度量分析Web Service研究與建構, 私立長榮大學經營管理研究所未出版之碩士論文。 吳柏林(1994), 模糊統計分析:問卷調查研究之新方向, 國立政治大學研究通訊, 2, 65-80。 吳柏林, 曾能芳(1998), 模糊迴歸參數估計及在景氣對策信號之分析應用, 中國統計學報, 36(4), 399-420。 張益華(2005), 基於知識本體的語意檢索系統之研究-以學校公文及法規為例, 大葉大學資訊管理系未出版之碩士論文。 戚玉樑(2005), 以本體知識為基礎的知識庫建制程式及其應用, 中原大學資訊管理所未出版之碩士論文。 鐘正男(2004), 以知識本體為基礎的語意查詢系統之研究-以圖書館為例, 大葉大學資訊管理系未出版之碩士論文。 林建宏(2006), 正規化概念分析建構電腦病毒特徵之知識本體, 雲林科技大學資訊管理學系未出版之碩士論文。 戚玉樑, 林建良(2004), 使用OWL-QL開發領域本體知識庫之知識提取, 2004 電子商務與數位生活研討會。 胡訓誠(2003), 應用本體論設計ISO文件管理資, 國立高雄第一科技大學資訊管理學系未出版之碩士論文。
- 二、英文部份 Akshay, J., Tim, F., & Sergei, N. (2006). Text understanding agents and the semantic web. Proceedings of the 39th Hawaii international conference on system sciences, 4-6. Alexander, M. (2001). Ontology learning for the semantic web. Kluwer Academic Publishers, 16(2), 72-79. Asuncion, G., & Oscar, C. (2002). Ontology languages for the semantic web. IEEE intelligent systems, 17(5), 54-60. Baldwin, J. F., & Zhou, S. Q. (1984). A fuzzy relational inference language. Fuzzy sets and systems, 14(22), 155-174. Baumgart, A. S., Knapp, H., Suetterlin, P., & Schader, M. (2007). A profile-based peer-to-peer RSS information distribution. IEEE intelligent systems, 5-7. Berners, L. T., & Fischetti, M. (1998). Weaving the web: The original design and ultimate destiny of the World Wide Web by its inventor (1st ed.). San Francisco: Harper Business. Brickley, D., & Guha, R. V. (2004). RDF vocabulary description language 1.0: RDF schema [Online]. Available: <http://www.w3c.org/TR/rdf-schema/> [2004, February 10]. Buckles, B. P., & Petry, F. E. (1982). A fuzzy representation of data for relational databases. Fuzzy sets and systems, 5, 213-226. Celik, D., & Elci, A. (2005). Searching semantic web services: An intelligent agent approach using semantic enhancement of client input term(s) and matchmaking step. Computer Software and Applications Conference, 2, 916-922. Chen, G. Q., Vandenbulcke, J., & Kerre, E. E. (1992). A general treatment of data redundancy in a fuzzy relational data model. Journal of the American society of information science, 43, 304-311. Dan, C., Frank, V. H., Ian, H., Deborah, M., Peter, F., Patel-Schneider, & Lynn, A. S. (2001). DAML+OIL (March 2001) reference description [Online]. Available: <http://www.w3.org/TR/daml+oil-reference> [2001, December 18]. Dan, C., Frank, V. H., Ian, H., Deborah, M., Peter, F., Patel-Schneider, & Lynn, A. S. (2001). Annotated DAML+OIL ontology markup [Online]. Available: <http://www.w3.org/TR/daml+oil-walkthru/> [2001, December 18]. Dubois, D., & Prade, H. (1983). Ranking fuzzy number in the setting of possibility theory. Information Science, 30(2), 183-224. Feff, H. (2004). OWL web ontology language use cases and requirements [Online]. Available: <http://www.w3c.org/TR/webont-rrg> [2004, February 10]. Frank, M., & Eric, M. (2004). RDF primer [Online]. Available: <http://www.w3.org/TR/REC-rdf-syntax/> [2004, February 10]. Graham, K., & Jeremy, C. (2004). Resource description framework (RDF): Concepts and abstract syntax [Online]. Available: <http://www.w3.org/TR/rdf-concepts/> [2004, February 10]. Gruninger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies. Unpublished master's thesis, Department of Industrial Engineering University of Toronto. Canada, M S A. Guttman, R. H., Moukas, A. G., & Maes, P. (1999). Agents that buy and sell. Communications of the ACM., 42(3), 81-91. Guu, S. M., Pang, C. T., & Liu, J. Y. C. (2002). On the proximately semantic measure of fuzzy data in extended possibility-based fuzzy relational databases. Working paper of department of information management. Yuan-Ze University. Han, J., & Kamber, M. (2001). Data mining: Concepts and techniques. New York: Morgan Kaufmann. Hesketh, B., Pryor, R., & Gleitzman, M., & Hesketh, T. (1988). Practical applications and psychometric evaluation of a computerized fuzzy graphic rating scale. Fuzzy sets in psychology, New York: North-Holland, 425-454. Keeble, R. J., & Macredie, R. D. (2000). Assistant agents for the world wide web intelligent interface design challenges. Inter-acting with Computers, 12, 357-381. Kilir, G. J., & Folger, T. A. (1988). Fuzzy sets, uncertainty and information. NJ: Prentice-Hall. Kilir, G. J., & Yuan, B. (1995). Fuzzy sets and fuzzy logic theory and application. NJ: Prentice-Hall. Lee, C. S., Chen, C. P., Chen, H. J., & Kuo, Y. H. (2002). A fuzzy classification agent for personal e-News service. International journal of fuzzy systems, 4(4), 849-856. Lee, C. S., Pan, C. Y., & Chang, M. J. (2002). A fuzzy decision agent for meeting scheduling supported system. International conference on fuzzy systems and knowledge discovery, Singapore, 2002. Lin, C. T., & George, L. C. S. (1991). Neural-network-based fuzzy logic control and decision system. IEEE transactions on computers, 40(12), 1320-1336. Manton, K. G., Woodbury, M. A., & Tolley, H. D. (1994). Statistical applications using fuzzy sets. New York: John Wiley and sons, Inc. Patrick, H. (2004). RDF semantics [Online]. Available: <http://www.w3c.org/TR/rdf-mt/> [2004, February 10]. Protege'3.3.1(2007), [Online]. Available: <http://protege.stanford.edu/> [2007, September 20]. Roger, H., L., Cecil, E., H., & Veda, C., S. (2001). A smart web query method for semantic retrieval of web data. Data & knowledge engineering, 38(1), 63-84. RSS(2007), [Online]. Available: <http://www.rssboard.org/> [2007, September 20]. Smith, M. K., Welty, C., & Mc Guinness, D. L. (2004). OWL web ontology language guide [Online]. Available: <http://www.w3c.org/TR/owl-guide/> [2004, February 10]. Stefan, D., Prasenjit, M., & Sergey, M.

(2000). Framework for the semantic web. *IEEE internet computing*, 4(6), 68-73. Uschold, M., King, M., Moralee, S., & Zorgios, Y. (1998). The enterprise ontology. *The knowledge engineering review*, 13(1), 31-89. Zadeh, L. A. (1972). A fuzzy set theoretical interpretation of hedges. *Journal of Cybernetics*, 2(2), 4-34. Zadeh, L. A. (1995). Discussion: Probability theory and fuzzy logic are complementary rather than competitive. *Technometrics*, 37(3), 271-276. Zhongzhi, S., He, H., Jiewen, L., Fen, L., & Haijun, Z. (2006). Agent-based grid computing. In *applied mathematical modeling*, 30(7), 629-640. Zimmermann, H. J. (1991). *Fuzzy set theory and its application*. London: Kluwer academic publishers.