

A Hierarchical Approach to Assess Keyword Dependencies in Fuzzy Keyword Ontologies

Christer Carlsson
IAMSR, Åbo Akademi University
Joukahaisenkatu 3-5A
FIN-20520 Åbo, Finland
Email: christer.carlsson@abo.fi

Robert Fullér
IAMSR, Åbo Akademi University
Joukahaisenkatu 3-5A
FIN-20520 Åbo, Finland
Email: rfuller@abo.fi

Mario Fedrizzi
DISA, Faculty of Economics
University of Trento
via Inama 5, 38122 Trento, Italy
Email: mario.fedrizzi@unitn.it

Abstract—The Knowledge Mobilization project (KNOWMOBILE) has been a joint effort by Institute for Advanced Management Systems Research, Åbo Akademi University and VTT Technical Research Centre of Finland. Its goal was to better “mobilize” knowledge stored in heterogeneous databases for users with various backgrounds, geographical locations and situations. The working hypothesis of the project was that fuzzy mathematics combined with domain-specific data models, in other words, fuzzy ontologies, would help manage the uncertainty in finding information that matches the users needs. In this way, KNOWMOBILE places itself in the domain of knowledge management. In this paper we describe an industrial demonstration of fuzzy ontologies in information retrieval in the paper industry where problem solving reports are annotated with keywords and then stored in a database for later use. Furthermore, using Bellmann-Zadeh’s principle to fuzzy decision-making we will show a method for identifying keyword dependencies in the keyword taxonomic tree.

Keywords—Fuzzy ontology; keyword ontology; fuzzy coverage

I. INTRODUCTION

Fuzzy ontologies have been proposed as a solution for addressing semantic meaning in an uncertain and inconsistent world. Since many concepts that are needed for intelligent systems lack well defined boundaries, or precisely defined criteria of membership, we need fuzzy logic to deal with notions of vagueness and imprecision. As with fuzzy logic, reasoning is approximate rather than precise. The aim is to avoid the theoretic pitfalls of monolithic ontologies, facilitate interoperability between different and independent ontologies [1], and provide flexible information retrieval capabilities [2]–[5]. In the KNOWMOBILE project, we developed a concept of a tool for searching plant knowledge with a search engine based on a fuzzy ontology. The usage scenario for the tool was that a process expert, dealing with a problem in the process chemistry of a paper machine, wishes to find past problem solving cases of a similar setting in order to find possible solutions to a current issue. This setting is a universal one: pieces of knowledge, called “nuggets”, are written and stored by companies on different domains in the form of incident reports.

In the KNOWMOBILE project we have focused on the chemistry of the “wet end” in order to limit the work effort needed to construct the domain ontology and concentrate on a subject on which domain expertise and actual data were

available. Nuggets are documents that can contain all kinds of raw data or multimedia extracted from different information systems. An expert author annotates the nuggets with suitable keywords, and it is these keywords that the search is then based on. In addition to providing exact results to queries, the tool uses a fuzzy domain ontology to extend the query to related keywords (see Figure 1). As a result, the search results include nuggets that may not necessarily deal with exactly the same process equipment, variable, function or chemical, but nuggets that may still provide valuable insight to solving the problem at hand. The terms defined in the ontology should be familiar to the users of the system, e.g. for plant operators, process developers and chemical suppliers. On the other hand, the terminology should be based on sound principles of conceptual modeling and ontology development. Furthermore, the goal should be to have common concepts and plant models with equipment manufacturers and engineering contractors that provide the initial information and often participate in modifications and upgrade projects. This creates a link to the ongoing development of engineering data models for various industrial areas. Fuzzy keyword ontologies should thus make use of relevant product and plant modeling standards and more general upper ontologies. The key research question of the KNOWMOBILE project is [6]: How to build fuzzy ontologies for the process industry domain to enhance knowledge retrieval?

II. FUZZY ONTOLOGY FOR PROCESS INDUSTRY

Our demonstration works with both engineering and operational knowledge of an industrial plant and we used the following taxonomic system,

- Top layer: general concepts (based on international standards).
- Middle layer: vocabulary defined and shared by business partners to share knowledge of, e.g. the type and structure of process equipment. This layer extends the top layer with domain-specific keywords.
- Bottom layer: custom, company-specific concepts, e.g. specific products and component types, or even individual process plants.

The system of keywords represents concepts, properties, relationships, axioms, and reasoning schemes relevant for the

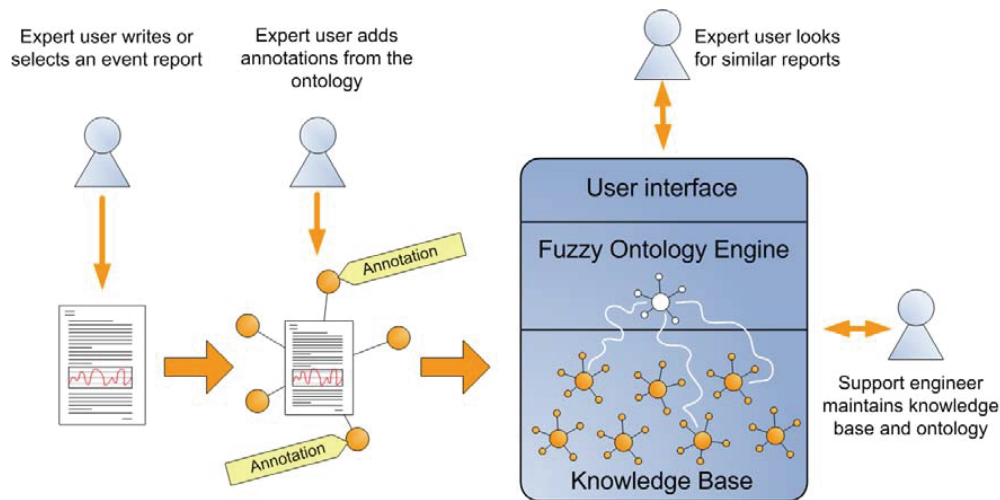


Fig. 1. System concept - a knowledge base of event reports [6].

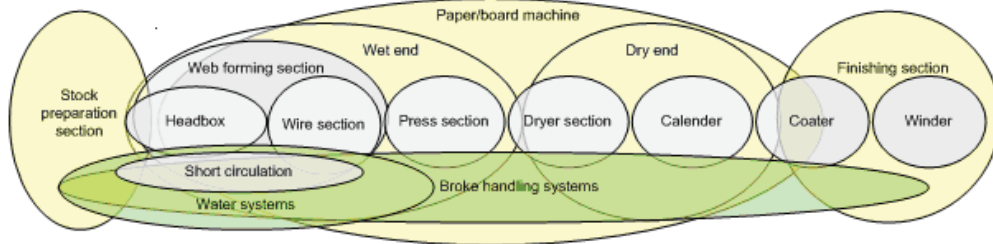


Fig. 2. System example - a fuzzy decomposition of a paper making line. [6].

application area. On the basis of various upper ontologies and industrial data models we identified that the following keyword categories are needed to characterize event reports:

- Systems: types of real-world components of a process plant, e.g. machines, buildings, software and people. (see Figure 2)
- Functions: phenomena and activities carried out at an industrial plant.
- Variables: properties and state variables of various entities, e.g. temperature.
- Events: types of interesting periods of plant life described in event reports, e.g. test runs or equipment failures.
- Materials: raw materials, products, consumables etc. handled in a process plant.

Our basic approach to conceptualize our application is shown in the informal Unified Modeling Language (UML) class diagram below (Figure 3).

Event reports describe events that are related to various entities of a process plant, e.g. to equipment, processing functions and materials. Nothing is assumed about the internal structure of event reports. Instead, they are characterized by an expert with keywords selected from a fuzzy ontology. The expert can select the keywords from five categories: event,

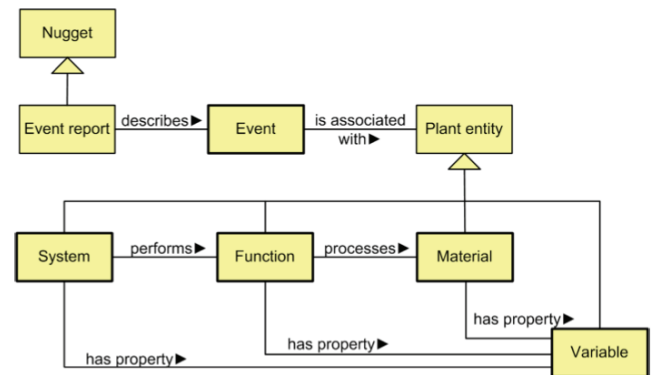


Fig. 3. Overall domain concepts [6].

system, function, material and variable. All keywords represent an entity type and can have subtypes and smaller parts. In the KNOWMOBILE demonstration tool keywords are used to characterize event reports and other nuggets stored in a database.

Classification (is-a) and decomposition (part-of) can be found in most ontologies and data models. They are important in the industrial context as well. So, the keywords in each cat-

For developing a software tool the ontology should be expressed and stored in a more formal way. The basic approach for representing a fuzzy ontology is illustrated in Figure 6 with a combination of UML class and instance diagrams.

From all kinds of nuggets, the demonstration only looks at event reports that usually describe problematic situations. Each report, like the specific instance "Report #1" in the diagram, is related to types of systems, plant functions, events, etc. Only the reference "Report #1" to the keyword "Holes" is shown in Figure 6. As indicated by the "instance of" associations, all keywords seen by the users are individuals, i.e. instances of a subclass of "Keyword category". Fuzzy dependencies between keyword instances are described by a few fundamental relationship types like is-a (specialization), part-of and, as an option, instantiation. In the first version we focus on "specializations", i.e. fuzzy classification of keywords. The degree of overlapping (or inclusion) of the sets represented by the keywords is described by linguistic labels, i.e. natural language words like "moderate" or "significant". So, the instance named "Specialization #1" in Figure 6 tells us that "Holes" is "to a large extent" understood as a subclass of "Quality problem but only represents a minor part of its scope". In addition, the keyword "Holes" may also specialize other problem types. In the KNOWMOBILE project we have modelled dependencies with one single dependency relationship. For simplicity it is symmetric and has a strength value between zero (independence) and one (full dependence).

IV. A MINIMAX APPROACH TO ASSESS KEYWORD DEPENDENCIES

Now we show a method for approximating keyword dependencies in the keyword ontology. This method uses Bellman-Zadeh's principle to fuzzy decision making [8]. An event type fuzzy taxonomy is shown in Figure 7. For example, consider the second column of event classification matrix "Problem". All "Technical problems" are "Problems" and they represent around 80% of all possible problems. That is "Technical problem" covers "Problem" with degree 0.8. Similarly, "Human errors" are "Problems" and they represent around 30% of all possible problems. That is "Human error" covers "Problem" with degree 0.3. Following Bordogna and Pasi ([9], page 804) we will assume that if A and B are two keywords in the keyword taxonomic tree then

$$\text{coverage}(A, B) = \text{fuzzy inclusion}(A, B). \quad (1)$$

For example, "System fault = {Device fault, Design flaw}" that is, "System fault" is a union of these two events. Furthermore, "Function failure = {Design flaw, Drift, Oscillation}" that is "Function failure" is the union of these three events. "Design flaw" covers "System fault" with degree 0.6 and at the same time "Design flaw" covers "Function failure" with degree 0.4 (see Figure 7). Moreover, "System fault" and "Function failure" do not have any more component in common. We compute the degree of dependency between "System fault"

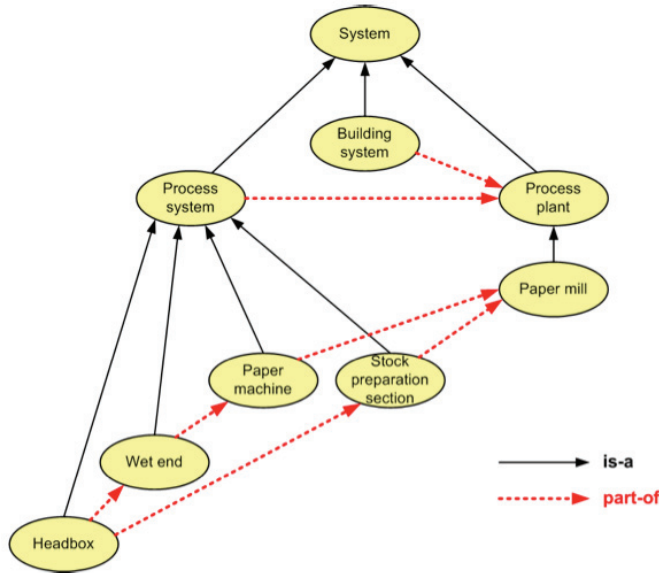


Fig. 4. Simplified event classification matrix [7].

egory are linked by is-a and part-of relationships as illustrated in Figure 4. Furthermore, the ontology should model functional and other kinds of dependencies between keywords in various keyword categories. As an example, systems can be or are used for some purposes, i.e. they play various roles in carrying out one or more functions. This creates a link between the keywords "wire section (a part of paper machine)" and "formation (a quality measure of the produced paper)". Modeling classifications, decompositions and various dependencies leads to a situation where we have a taxonomy tree for each keyword category and a set of part-of relationships trees describing the decomposition to various domain entities (Figure 5). In addition, there are dependency relationships linking keywords to each other.

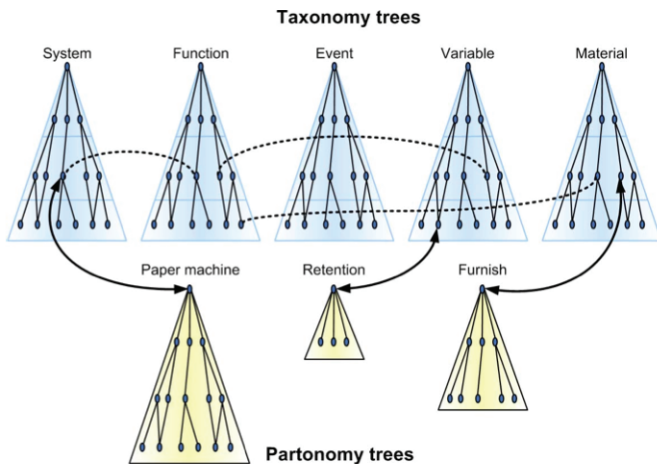


Fig. 5. The fuzzy ontology defines classification, decomposition and miscellaneous dependency relationships between keywords [7].

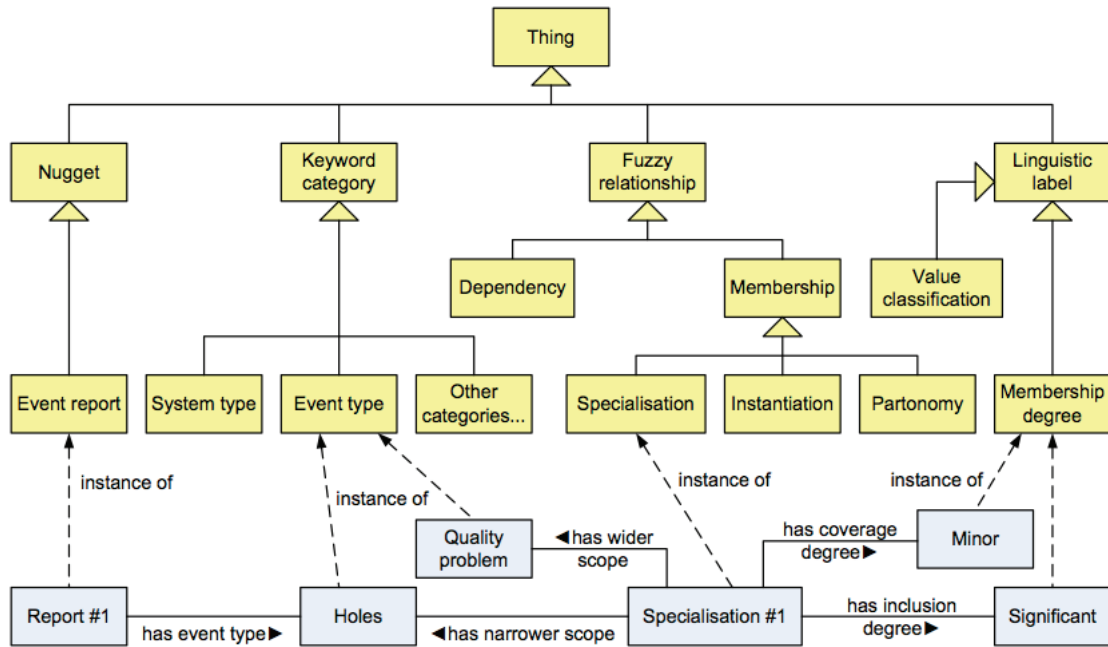


Fig. 6. Representing fuzzy keyword ontology with object classes and their instances [7].

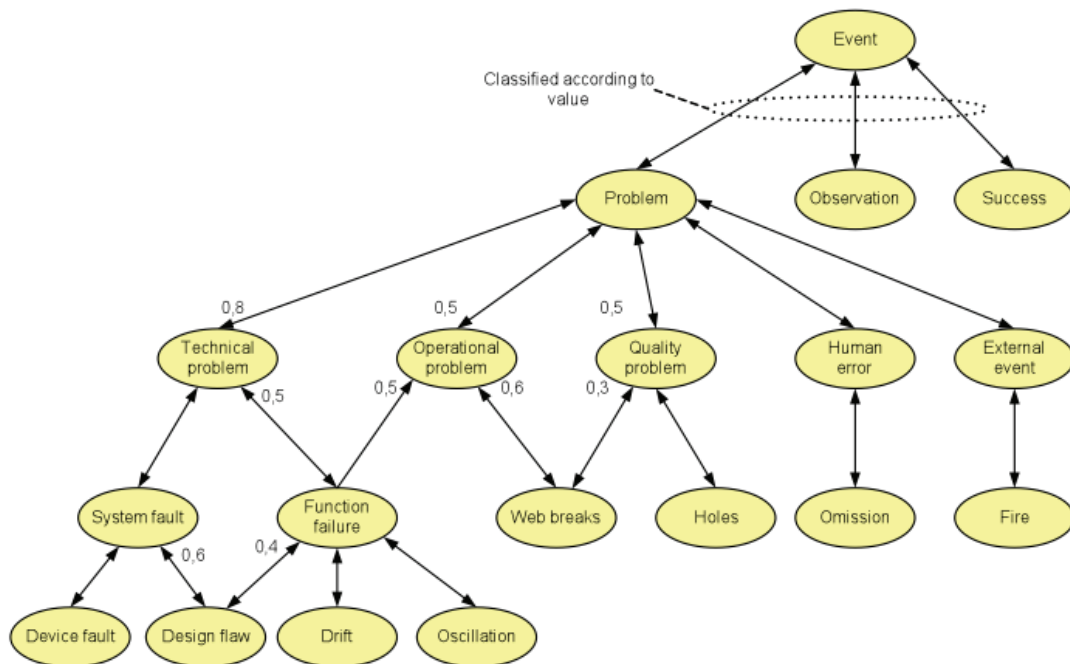


Fig. 7. A fragment of the event type fuzzy taxonomy [7].

(SF) and "Function failure" (FF) as their joint coverage by "Design flaw" (DF)

$$\text{dependency}(SF, FF) = \min\{\text{coverage}(DF, SF), \text{coverage}(DF, FF)\},$$

that is,

$$\text{dependency}(SF, FF) = \min\{0.6, 0.4\} = 0.4 \quad (2)$$

It is easy to see in Figure 7 that keywords "Function failure" and "Fire" are independent since they do not have any component in common. In this case we have,

$$\text{dependency}(\text{"Function failure"}, \text{"Fire"}) = 0. \quad (3)$$

Zero means independence, one means full dependence, and values between zero and one denote intermediate degrees of dependency between keywords. It can happen that two keywords have more than one joint component. Then we apply Bellman-Zadeh's principle (max-min approach) to fuzzy decision making to measure their dependency. For example, suppose that "System fault" and "Function failure" were to have two joint components, where the first one is "Design flaw" and the second one "Fluctuation" that has a coverage values 0.7 and 0.5, respectively. Then we measure the degree of dependency between "System fault" and "Function failure" according to Bellman-Zadeh's principle to fuzzy decision making as

$$\begin{aligned} \text{dependency}\{SF, FF\} &= \max\{\min\{0.6, 0.4\}, \min\{0.7, 0.5\}\} \\ &= \max\{0.4, 0.5\} = 0.5. \end{aligned}$$

Supposing that all the coverage degrees are given by experts. Then we can summarize our algorithm as follows: *Compute the degrees of dependency between keywords on the immediate upper level using the max-min approach. Then repeat this procedure until the top layer.* For example, consider keywords "Technical problem" (TP) and "Operational problem" (OP). Then we find (Figure 7),

$$\begin{aligned} \text{dependency}\{TP, OP\} &= \min\{\text{coverage}(FF, TP), \\ &\quad \text{coverage}(FF, OP)\} = 0.5. \end{aligned}$$

One can further improve this model by introducing degrees of inclusion and coverage between concepts as suggested by Holi and Hyvönen [10], and Holi [11].

V. DEMO ARCHITECTURE AND IMPLEMENTATION

The component-based demo architecture has been implemented by VTT Technical Research Centre of Finland, using the Protégé ontology editor to maintain the fuzzy ontology in OWL format (as shown in Figure 8). The GUI component (Graphical User Interface) guides the user in specifying the information query, and presents the results. Tools for browsing and evaluating the fuzzy reasoner component directly were also provided. A database adapter is used to access report data, which in this case was stored locally in XML files. Similarly, an ontology adapter is used to provide access to the fuzzy ontology, in this case stored in OWL files.

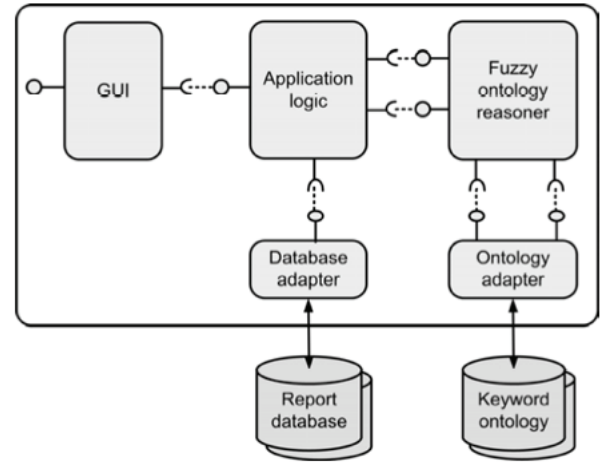


Fig. 8. Component-based demo application architecture [7].

The adapters help hide the different interfaces and protocols of different data sources (e.g. SQL, HTTP) and provide transparent access via an agreed interface. The fuzzy ontology reasoner component is used to process ontology-based information. Its main function in the demo is to extend a list of query keywords to a list of their closest neighbors in terms of fuzzy ontology relationships. For maintenance and evaluation purposes, the component interface also provides methods for directly accessing the ontology concepts and relationships. Finally, the application logic component binds all the functionality together by taking the query, using the reasoner component to extend it, passing the extended query to the report database and then combining and ordering the results for the GUI. The fuzzy ontology with fuzzy concepts, relations, and instances was defined using Protégé version 3.4. The developed ontology was exported for the reasoning software as a standard OWL file. Protégé is a widely used tool for developing ontologies. As such it provides some advantages, for example a forms-based interface for editing the basic classes and adding the individual keywords, reports, as well as their relationships. Protégé does not have a built-in support for modeling uncertainties, however in late 2010 Straccia [12] introduced new plug-in, called SoftFacts, which is an ontology mediated top-k information retrieval system over relational databases.

VI. SUMMARY

We have examined the possible use of fuzzy ontologies in the retrieval of stored reports of industrial plant knowledge. Since no well-established definitions or mature tools exist, we have adopted a practical approach - defining those constructs that we deemed necessary in order to adequately capture domain semantics, and implementing a demonstration application in which to test those constructs against actual industrial data. The demo has been evaluated with the help of industry experts. Although the extended query tool seems to be promising, there has been a great deal of worry about the amount of

work needed for maintaining the fuzzy ontology relationship weights.

REFERENCES

- [1] V. Cross, "Fuzzy semantic distance measures between ontological concepts," IEEE Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS 2004), Banff, AB, Canada, pp. 635-640, June, 2004.
- [2] G. Acampora and V. Loia, "Fuzzy control interoperability and scalability for adaptive domotic framework," IEEE Trans. Ind. Inf., vol. 1, pp. 97-111, May 2005.
- [3] G. Acampora, M. Gaeta, V. Loia, A.V. Vasilakos, "Interoperable and adaptive fuzzy services for ambient intelligence applications," ACM Trans. Auton. Adap. vol. 5, pp. 8:1-8:26, May 2010.
- [4] C.-S. Lee, Z.-W. Jian and L.-K. Huang, "A fuzzy ontology and its application to news summarization," IEEE Trans. Syst. Man Cybern. Part B Cybern., vol. 35, pp. 859-880, October 2005
- [5] C.-S. Lee, M.-H. Wang, G. Acampora, C.-Y. Hsu and H. Hagra, "Diet assessment based on type-2 fuzzy ontology and fuzzy markup language," Int. J. Intell. Syst., vol. 25, pp. 1187-1216, October 2010.
- [6] J. Hirvonen, T. Tommila, A. Pakonen, C. Carlsson, M. Fedrizzi, R. Fullér, "Fuzzy keyword ontology for annotating and searching event reports," International Conference on Knowledge Engineering and Ontology Development (KEOD 2010), Valencia, Spain, October, 2010.
- [7] T. Tommila, J. Hirvonen, A. Pakonen, "Fuzzy ontologies for retrieval of industrial knowledge - a case study," VTT Working Papers, no. 153, November 2010.
- [8] R. E. Bellman and L. A. Zadeh, "Decision-making in a fuzzy environment," Manage. Sci., Ser. B, vol. 17, pp. 141-164, December 1970.
- [9] G. Bordogna, G. Pasi "A flexible multi criteria information filtering model," Soft Comput., vol. 14, pp. 799-809, June 2010.
- [10] M. Holi and E. Hyvönen, "Modeling degrees of overlap in semantic web ontologies," ISWC Workshop Uncertainty Reasoning for the Semantic Web, Galway, Ireland, Nov. 2005.
- [11] M. Holi, Crisp, fuzzy and probabilistic faceted semantic search, Doctoral dissertation, Aalto University School of Science and Technology, Helsinki, Finland, 2010.
- [12] U. Straccia, "SoftFacts: A top-k retrieval engine for ontology mediated access to relational databases," 2010 IEEE International Conference on Systems Man and Cybernetics, pp. 4115-4122, October 2010.