

Research on Automatic Fuzzy Ontology Generation for World Wide Web

GJCST Classification
H.5.3, H.5.4

K.Srikanth¹, J. Anitha², Ch. Heymaraju³, A.Harish⁴, B.Sasikanth⁵

Abstract-Ontology is an effective conceptualism commonly used for the World Wide Web. Fuzzy logic can be incorporated to ontology to represent uncertainty information. To tackle this problem, this paper proposes the FOGA (Fuzzy Ontology Generation framework) for automatic generation of fuzzy ontology on uncertainty information. The FOGA framework comprises the following components: Fuzzy Formal Concept Analysis, Concept Hierarchy Generation, and Fuzzy Ontology Generation. We also discuss approximating reasoning for incremental enrichment of the ontology with new upcoming data. Finally, a fuzzy-based technique for integrating other attributes of database to the ontology is proposed. **Index Terms**-Intelligent Web services and World Wide Web, ontology design, uncertainty, "fuzzy," probabilistic, knowledge representation formalisms and methods, concept learning.

Keywords - Ontology Generation, Fuzzy Ontology.

I. INTRODUCTION

Ontology is a conceptualization of a domain into a human understandable, machine-readable format consisting of entities, attributes, relationships, and axioms [1]. It is used as a standard knowledge representation for the World Wide Web [2]. However, the conceptual formalism supported by typical ontology may not be sufficient to represent uncertainty information commonly found in many application domains due to the lack of clear-cut boundaries between concepts of the domains. For example, a document can be very relevant, relevant, or irrelevant to a research area. In addition, keywords extracted from scientific publications can be used to infer the corresponding research areas. However, it is inappropriate to treat all keywords equally as some keywords may be more significant than others. To tackle this type of problems, one possible solution is to incorporate fuzzy logic [3] into ontology to handle uncertainty data. Traditionally, fuzzy ontology is generated and used in text retrieval [4] and search engines [5], in which membership values are used to evaluate the similarities between the concepts in a concept hierarchy. However, manual generation of fuzzy ontology from a predefined concept hierarchy is a difficult and tedious task that often requires expert interpretation. So, automatic generation of concept hierarchy and fuzzy ontology from uncertainty data of a domain is highly desirable. In this paper, we propose a framework known as FOGA (Fuzzy Ontology Generation framework) that can automatically generate a fuzzy ontology from uncertainty data based on

Formal Concept Analysis (FCA) [6] theory. The generated fuzzy ontology is mapped to a World Wide representation in OWL (Web Ontology Language) [7]. The rest of this paper is organized as follows: Section 2 discusses related work on ontology generation and FCA. Section 3 gives some basic definitions and operators of the fuzzy theory. The FOGA framework is presented in Section 4. Section 5 discusses the approximating reasoning technique to incrementally furnish the generated ontology with new instance. The problem of integrating extra attributes in database to the ontology is given in Section 6. Performance evaluation of the proposed FOGA framework is given in Section 7. Finally, Section 8 concludes the paper.

II. RELATED WORK

2.1 Ontology Generation

Although editing tools [8], [9] have been developed to help users to create and edit ontology, it is a troublesome task to manually derive ontology from data. Typically, ontology can be generated from various data types such as textual data [10]. Compared to other types of data, ontology generation from textual data has attracted the most attention. Among techniques used for processing textual data, clustering is one of the most effective techniques for ontology learning. Conceptual clustering techniques such as COBWEB and CLASSIT are powerful clustering techniques that can conceptualize clusters for ontology generation. We have created DESK to fill this gap, using the Programming By Demonstration paradigm [2], [5], [6] to allow the page modification by a non-expert user. In programming by demonstration the system infers procedural information from examples of what the user wants to achieve. The programming by demonstration paradigm has an intrinsic ambiguity because general information has to be derived from particular cases provided by the user. To solve such ambiguity some strategies have been used, such as monitoring all user interaction (vs. watching only the initial and the final state), using multiple examples (e.g. negative examples), or interactively asking the user to help or decide. The extraction of structured information, like the difference and context model used by DESK, from a semi-structured document (HTML code) is very similar to the way wrappers operate [8]. Wrappers provide a uniform access to the information stored in heterogeneous repositories like data bases, files and so forth.

About^{1,2,3,4&5} -GIT, GITAM University, Visakhapatnam
¹srikanthwizard@gmail.com, ²anithanv28@gmail.com, ³heymanraju@gmail.com, ⁴saihari08@gmail.com, ⁵battula.sasi@gmail.com



III. FUZZY THEORY

In this section, we review some fundamental knowledge of fuzzy theory [3].

Definition 1 (Fuzzy Set). A fuzzy set A on a domain U , is defined by a membership function μ from U to $[0,1]$, i.e., each item in A has a membership value given by μ . We denote $\varphi(s)$ as a fuzzy set generated from a traditional set of items S . Each item in S has a membership value in $[0, 1]$. S can also be called as a crisp set.

Definition 2 (Fuzzy Relation). A fuzzy set A on a domain $G \times M$, where G and M are two crisp sets is a fuzzy relation on G, M .

Definition 3 (Fuzzy Sets Intersection). The intersection of fuzzy sets A and B , denoted as $A \cap B$, is defined by $\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$.

Definition 4 (Fuzzy Sets Union). The intersection of fuzzy sets A and B , denoted as $A \cup B$, is defined by $\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$.

Definition 5 (Fuzzy Set Cardinality). Let S_f be a fuzzy set on the domain U . The cardinality of S_f is defined as $|S_f| = \sum_{x \in U} \mu(x)$

Where $\mu(x)$ is the membership of x in S_f .

Definition 6 (Fuzzy Sets Similarity). The similarity between two fuzzy sets A and B is defined as

$$E(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Definition 7 (Fuzzy Sets Subsethood). The subsethood of a fuzzy set A of a conceptual cluster B is calculated as

$$\text{Subsethood}(A,B) = \frac{|A \cap B|}{|B|}$$

Definition 8 (Fuzzy Set Max-min Composition). Let $P(X,Y)$ be a fuzzy relation on X, Y and $P(Y,Z)$ be a fuzzy relation on Y, Z . The max-min composition of $P(X,Y)$ and $Q(Y,Z)$, $P \cdot Q$, is defined by:

$$\mu_{P \cdot Q}(X,Y) = \max \min(\mu_P(X,Y), \mu_Q(Y,Z)), \varphi_x \in X, y \in Y.$$

The max-min composition indicates the strength of relation between the element of X and Z .

IV. THE FOGA FRAMEWORK

Fig. 1 shows the proposed FOGA (Fuzzy Ontology Generation framework), which consists of the following components.

4.1 Fuzzy Formal Concept Analysis

The Fuzzy Formal Concept Analysis incorporates fuzzy logic into Formal Concept Analysis to represent vague Information.

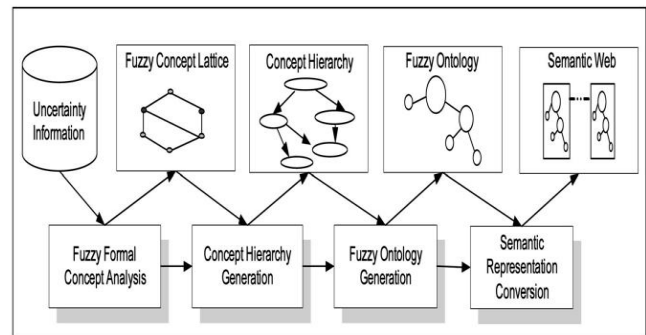


Fig 1. The FOGA framework.

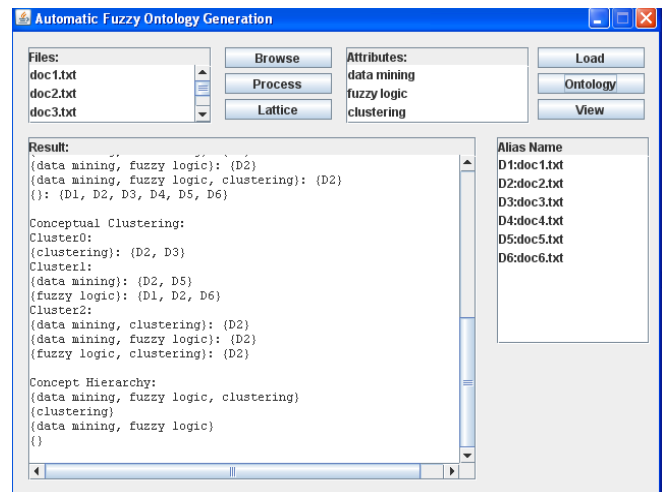


Fig.2.Ontology Generation Frame Work

TABLE 1

A Cross-Table of a Fuzzy Formal Context

	Data Mining	Clustering	Fuzzy Logic
D1	0.8	0.12	0.61
D2	0.9	0.85	0.13
D3	0.1	0.14	0.87

TABLE 2

Fuzzy Formal Context in Table 1 with an α cut $\alpha = 0.5$

	Data Mining	Clustering	Fuzzy Logic
D1	0.8	-	0.61
D2	0.9	0.85	-
D3	-	-	0.87

V. DESK AS AN AUTHORIZING TOOL

After the proposition extraction step, we have a set of propositions as fuzzy rules. The next step is to use the generated rules for reasoning new data. For example, assume that we have a fuzzy rule IF x is A THEN y is B , where A and B are fuzzy sets. Then, if we have a new proposition x is A , we need to find what conclusion we can get about y .

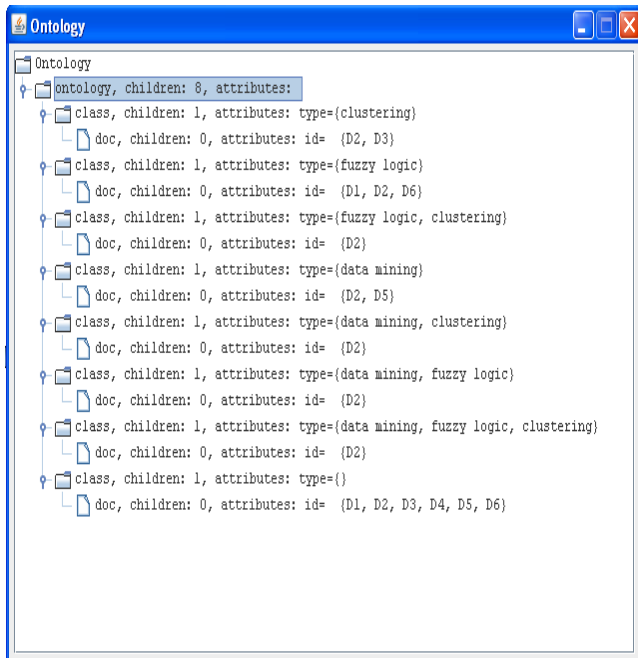


Fig.3. Ontology Tree Structure

TABLE 3

A Fuzzy Formal Context Having Cross Relation with the Fuzzy Formal Context in Table 2

	D1	D2	D3
Author1	1.0	-	1.0
Author2	0.5	1.0	-

The cross relation represents an intercontext relation that probably occurs between the fuzzy formal contexts when the set of objects of a context is regarded as the set of attributes of an other contexts. For example, the context represented by the cross table shown in Table 3 has cross relation with the context in Table 2, while the documents are used as attributes of the authors. The membership value of 1.0 implies that the author is the first author of the document, while 0.5 implies that the author is the second author.

VI. PERFORMANCE EVALUATION

6.1 Generating Ontology from Citation Database To evaluate the proposed FOGA framework for ontology

generation, we have collected a set of 1,400 scientific documents on the research area Information Retrieval published in 1987-1997 from the Institute for Scientific Information's (ISI) Web site [52]. The downloaded documents are preprocessed to extract related information such as the title, authors, citation keywords, and other citation information. The extracted information is then stored as a citation database. First, we construct a fuzzy formal context $K_f = \{G, M, I\}$, with G as the set of documents and M as the set of citation keywords. The membership value of a document D on a citation keyword CK in K_f is computed as $\mu(d, C_k) = n1/n2$, where $n1$ is the number of documents that cite D and contain CK and $n2$ is the number of documents that cite D . This formula is based on the premise that the more frequent a keyword occurs in the citing paper, the more important the keyword is in the cited paper.

VII. EVALUATION USING RECALL, PRECISION, AND F-MEASURE

We have classified manually the documents downloaded from ISI into classes based on their research themes. These classes are used as a benchmark to evaluate the clustering results in terms of recall, precision, and F-measure. As discussed earlier, we extract citation keywords of documents as their attributes. Since these attributes are descriptors for the generated clusters, if more keywords are extracted and used, the more meaningful the cluster descriptors are constructed. To verify this, we vary the number of keywords N extracted from documents from 2 to 10, and the similarity threshold T_s from 0.2 to 0.9 when performing conceptual clustering. The measured precision, recall and F-measure are presented in Table 7, respectively.

TABLE 7
Performance Results Using Recall Measurement

	$T_s=0.2$	$T_s=0.3$	$T_s=0.4$	$T_s=0.5$	$T_s=0.6$	$T_s=0.7$	$T_s=0.8$	$T_s=0.9$
N=2	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.98
N=3	0.99	0.99	0.99	0.99	0.98	0.98	0.97	0.97
N=4	0.98	0.98	0.97	0.97	0.94	0.95	0.94	0.94
N=5	0.89	0.87	0.87	0.88	0.87	0.89	0.89	0.89
N=6	0.8	0.81	0.83	0.83	0.83	0.85	0.85	0.85
N=7	0.81	0.8	0.82	0.82	0.83	0.84	0.86	0.86
N=8	0.79	0.79	0.81	0.82	0.82	0.84	0.85	0.85
N=9	0.76	0.77	0.8	0.8	0.81	0.83	0.84	0.84
N=10	0.73	0.75	0.78	0.78	0.79	0.81	0.83	0.83

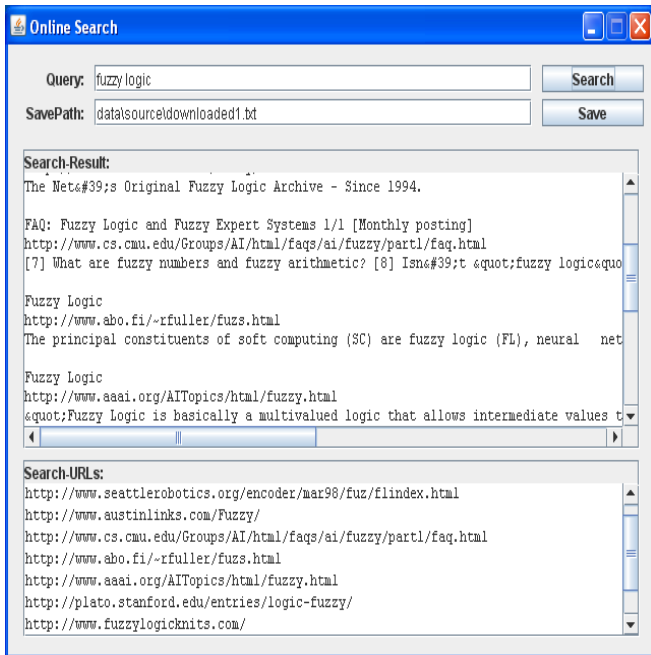


Fig 4. Performance Result Using World

Wide Web Mining Techniques

VIII) CONCLUSIONS

In this paper, we have proposed the FOGA framework for fuzzy ontology generation on uncertainty information. FOGA consists of the following steps: Fuzzy FormalConcept Analysis, Fuzzy Conceptual Clustering, FuzzyOntology Generation, and World Wide Web Representation Conversion. In addition, we have also proposed an approximating reasoning technique that allows the generated fuzzy ontology to be incrementally furnished with new instances. Finally, we have also proposed a technique to integrate extra attributes in a database to the ontology. Our authoring tool provides automatic support for the customization of dynamic web documents based on comparing the pages generated by the system with a modified version provided by the end-user. DESK is based on PEGASUS, a system used to represent the World Wide Web information structured by models that allow a clear separation between contents and presentation. DESK uses domain information stored in PEGASUS and presentation models for finding the context of changes made by user. Our authoring tool also determines whether the user is enabled to do these modifications depending on a user model. With DESK the user only needs to take care of editing HTML pages using any standard HTML editing tool such as PageMaker or Netscape Composer.

IX) REFERENCES

- 1) N. Guarino and P. Giaretta, Ontologies and Knowledge Bases: Towards a Terminological Clarification. Toward Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing. Amsterdam: IOS Press, 1995.

- 2) Castells, P., Szekely, P.: Presentation Models by Example. En: Duke, D.J., Puerta A. (eds.): Design, Specification and Verification of Interactive Systems '99. Springer-Verlag, 1999, pp. 100-116.
- 3) L.A. Zadeh, Fuzzy Logic and Approximate Reasoning," Synthese, vol. 30, pp. 407-428, 1975.
- 4) Castells, P. Macías, J.A.: An Adaptive Hypermedia Presentation Modeling System for Custom Knowledge Representations. Proceedings of the World Conference on the WWW and Internet
- 5) D.H. Widyantoro and J. Yen, A Fuzzy Ontology-Based Abstract Search Engine and Its User Studies," Proc. 10th IEEE Int'l Conf. Fuzzy Systems, pp. 1291-1294, 2001.
- 6) B. Ganter and R. Wille, Formal Concept Analysis: Mathematical Foundations. Springer, 1999. [7] W3C, Web Ontology Language Overview,"http://www.w3.org/TR/owl-features/, 2006.
- 7) OntoWeb deliverable 1.3, http://www.ontoweb.org/, 2006.
- 8) A. Gomez-Perez, O. Corcho, and M. Fernandez-Lopez, Ontological Engineering: With Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web (Advanced Information and Knowledge Processing). Springer, 2004.
- 9) Macías, J. A., Castells, P.: Adaptive Hypermedia Presentation Modeling for Domain Ontologies. To appear in proceedings of 10th International Conference on Human-Computer Interaction (HCII '2001). New Orleans, Louisiana, 2001.
- 10) Murray, T.: Authoring Knowledge Based Tutors: Tools for Content, Instructional Strategy, Student Model, and Interface Design. Journal of the Learning Sciences 7, 1, 1998, 5-64.
- 11) Muslea, I.: Extraction Patterns for Information Extraction Tasks: A Survey. Proceedings of AAAI Workshop on Machine Learning for Information Extraction. Orlando, Florida, July 1999.
- 12) Vassileva, J.: Dynamic Course Generation on the WWW. Actas 8th World Conference on Artificial Intelligence in Education (AIED'97). Kobe, Japon, 1997, 498-505.
- 13) Weber, G. and Specht, M.: User modeling and Adaptive Navigation Support in WWW-based Tutoring Systems. Proceedings 6th International Conference on User Modeling (UM97). Sardinia, Italy, 1997.