Research Article

An Intelligent Information Retrieval Approach Based on Two Degrees of Uncertainty Fuzzy Ontology

Maryam Hourali and Gholam Ali Montazer

IT Engineering Department, School of Engineering, Tarbiat Modares University, P.O. Box 14115-179, Tehran, Iran

Correspondence should be addressed to Gholam Ali Montazer, montazer@modares.ac.ir

In spite of the voluminous studies in the field of intelligent retrieval systems, effective retrieving of information has been remained an important unsolved problem. Implementations of different conceptual knowledge in the information retrieval process such as ontology have been considered as a solution to enhance the quality of results. Furthermore, the conceptual formalism supported by typical ontology may not be sufficient to represent uncertainty information due to the lack of clear-cut boundaries between concepts of the domains. To tackle this type of problems, one possible solution is to insert fuzzy logic into ontology construction process. In this article, a novel approach for fuzzy ontology generation with two uncertainty degrees is proposed. Hence, by implementing linguistic variables, uncertainty level in domain's concepts (Software Maintenance Engineering (SME) domain) has been modeled, and ontology relations have been modeled by fuzzy theory consequently. Then, we combined these uncertain models and proposed a new ontology with two degrees of uncertainty both in concept expression and relation expression. The generated fuzzy ontology was implemented for expansion of initial user's queries in SME domain. Experimental results showed that the proposed model has better overall retrieval performance comparing to keyword-based or crisp ontology-based retrieval systems.

1. Introduction

The process of searching specific information among a large number of information items is known as information retrieval (IR). Users of IR Systems expect to find the most relevant items to a certain query. The computing parameters such as recall and precision are used for effectiveness appraisal of these systems [1]. Generally, an information retrieval system does not present an ideal behavior. Users often receive large result sets, and they have to spend a considerable time to find these items which are really relevant to their initial queries. Indeed, this kind of searching information will neglect relevant documents that do not contain the index terms which are specified in the user's queries. Working with specific domain knowledge, the mentioned problem can be tackled by incorporating a knowledge base such as an ontology which builds the relationships between index terms and existing information retrieval systems [2].

One of the motivations of the semantic web is the implementation of ontologies to overcome the limitations of keyword-based search [3]. Ontology is a conceptualization of a domain into a human understandable, machine-readable format consisting of entities, attributes, relationship, and axioms [4]. It is used as a standard knowledge representation for the semantic web [5]. An ontology-based information retrieval system is constructed on specified domain knowledge. Applying knowledge structures for filtering and searching the user's relevant needed information is the objective of inserting an ontology to information retrieval systems. If the searched information is covered under the concept of the user's knowledge domain, using ontology will increase the probability of relevancy [6]. However, the conceptual formalism supported by a typical ontology may not be sufficient to represent uncertainty, commonly found in many application domains due to the lack of clear-cut boundaries between concepts of the domains. One possible solution to confront the uncertain and vague information is inserting the fuzzy logic to ontology constructing process [7]. Fuzzy set theory, among computational intelligence

techniques, is a promising approach to improve the effectiveness of information retrieval systems [8]. It deals with uncertainty that may be present in document and query representations as well as in their relationships. It has already been used for indexing, clustering, and recommendation [2]. The fuzzy ontology is based on modification of an existing crisp ontology. The modification process is entirely incremental and the conversion process to a fuzzy ontology adds membership values to the currently existing relations, and may also add new entries to the ontology [9].

Several researchers have considered the usage of fuzzy ontologies to model the uncertainty in the domain knowledge [10], where Lee et al. proposed a fuzzy ontology for news summarization. In their paper, the fuzzy inference mechanism generated the membership degrees for each fuzzy concept of the fuzzy ontology [11]. They also proposed a novel type-2 fuzzy ontology and applied it to diet assessment by combining the type-2 fuzzy sets and the ontology model to propose a type-2 fuzzy ontology. Moreover, they used the type-2 fuzzy ontology to diet assessment domain to propose a fuzzy diet assessment agent for people with the average levels of physical activity [12]. A Type-2 Fuzzy Ontology (T2FO) is a knowledge representation model for describing the domain knowledge with uncertainty. It is an extension of the domain ontology and contains six layers, including a domain layer, a category layer, a fuzzy concept layer, a fuzzy variable layer, a fuzzy set layer, and a Type-2 fuzzy set (T2FS) layer. The concepts and relations of the T2FO are constructed by fuzzy variables, fuzzy sets, and T2FSs [12]. In a similar work, Lee et al. developed a type-2 fuzzy ontology and used it for personal diabetic-diet recommendation [13]. Quan et al. presented an automatic fuzzy ontology generation for semantic help-desk support for supporting customer services utilizing the semantic web technologies. It has been focused on an automatic generation approach, known as fuzzy formal concept analysis (FFCA) for fuzzy machine service ontology that can also deal with uncertainty data [14].

One application of fuzzy ontologies is in query expansion task. The main aim of query expansion is to add new meaningful terms to the initial query. Query expansion technology can improve the efficiency of the search engine by adding other terms which are closely related to the original query terms and disambiguate the user query [9]. Bahri et al. addressed fuzzy ontology implementation and query answering on databases. They propose a language to define fuzzy ontology schema and to query fuzzy ontology databases and an inferential engine to infer fuzzy concept instances and their membership degrees [15]. Pan et al. proposed a framework of fuzzy query languages for fuzzy ontologies and present ed query answering algorithms for query languages over fuzzy DL-Lite ontologies [16].

In this paper, a new method for ontology generation is proposed based on fuzzy theory with two degrees of uncertainty. Considering two uncertainty degrees in concept expression and relation expression and combining uncertain models to generate a new fuzzy ontology is the main contribution of this work. In this model, linguistic variables are used and membership degree of concepts to a certain domain and similarly membership degree of relations to concepts in SME domain are modeled as well. Then, these uncertain models are combined and a new ontology with two uncertainty degrees, both in concept expression and relation expression, is proposed. Finally, performance appraisal of information retrieval system based on proposed query expansion algorithm is measured in SME domain by comparing the lack of ontology, the crisp ontology, and the fuzzy ontology situations.

The rest of this paper is organized as follows: Section 2 reviews related work to the subject of paper. In Section 3, basic concepts of fuzzy set theory are explained. Fuzzification of the modification process ontology is depicted in Section 4. In Section 6, information retrieval system performance appraisal is discussed, and finally Section 7 concludes the paper.

2. Related Work

Fuzzy logic systems (FLSs) have been credited with providing an adequate methodology for designing robust systems that are able to deliver a satisfactory performance when contending with the uncertainty, noise, and imprecision attributed to real-world environments and applications [10]. As a result, FLSs have been used in wide range of applications including fuzzy ontologies.

Because the aim of this paper is proposing an information retrieval approach based on fuzzy ontology, in this section some fuzzy information retrieval models which are used as a tool for improving the retrieval performance are presented.

Parry has implemented fuzzy ontology for information retrieval which is focused on medical documents retrieval. This ontology has fuzzy values in its relations [17]. Zhai et al. presented a fuzzy ontology for semantic information retrieval in e-commerce domain, and a semantic query expansion method is used for this purpose. Their framework includes three parts: concepts, properties of concepts, and values of properties in which property value can be either standard data types or linguistic values of fuzzy concepts [18]. They also implemented fuzzy ontology for semantic information retrieval in supply chain management, traffic information retrieval, and intelligent transportation systems fields [19–21].

Leite and Ricarte presented a framework to encode a geographic knowledge base composed of multiple-related ontologies whose relationships were expressed as fuzzy relations. This knowledge organization was used in a fuzzy method to expand the user initial query [22].

An ontology-based spatial query expansion method was proposed by Fu et al., which considered a geographical ontology to expand geographic terms. Various factors are taken into account to support intelligent expansion of a spatial query, including types of spatial terms as encoded in the geographical ontology, types of nonspatial terms as encoded in the domain ontology, as well as the semantics of the spatial relationships and their context of use [23]. Bratsas et al. used a fuzzy query expansion and a fuzzy thesaurus

to solve the Medical Computational Problem (MCP). In the experiments, the system was capable of retrieving the same MCP for distinct descriptions of the same problem. The system uses a unique fuzzy thesaurus for query expansion [24].

Ogawa et al. used the keyword connection matrix model which the knowledge about the relevant keywords or terms is encoded as a single fuzzy relation that expresses the degree of similarity between the terms. The information retrieval process uses the fuzzy keyword connection matrix to find similarities between the query terms as a way to improve query results [25].

Pereira et al. presented the fuzzy relational ontological model in information search systems that considers knowledge base as a fuzzy ontology with concepts representing the categories and the keywords of a domain. When the user enters a query, composed of concepts, the system performs its expansion and may add new concepts based on the ontology knowledge. After expansion, the similarity between the query and the documents is calculated by fuzzy operations [26].

The fuzzy rough set-based web query expansion method was represented by Cock and Cornelis, which used the tight upper approximation in fuzzy rough set theory to find terms to be added to the query. The knowledge base is a thesaurus that consists of a term-term relation. In this approach, a term Y will only be added to a query represented by the fuzzy set A if all the terms that are related to Y are also related to at least one keyword of the query. The proposed technique results were promising when the query had ambiguous terms [27].

Calegari and Sanchez proposed a fuzzy ontology-approach to improve semantic information retrieval and introduced an information retrieval algorithm that allows to derive a unique path among the entities involved in the query in order to obtain maxima semantic associations in the knowledge domain [28]. Bahri et al. implemented fuzzy ontology for query answering on databases. They proposed a language to define fuzzy ontology schema and to query fuzzy ontology databases and an inferential engine to infer fuzzy concept instances and their membership degrees [29].

By reviewing the studies in fuzzy ontology implementation, it can be observed that the proposed approaches do not have adequate ability in uncertainty representation of concepts and have not considered natural semantic relations between ontology concepts. These weaknesses may cause some problems in fuzzy semantic information retrieval. Since concepts belong with a specific membership degree to a certain domain and similarly relations belong with a certain membership degree to concepts, this paper proposes a fuzzy ontology with two degrees of uncertainty, both in concept expression and relation expression.

3. Fuzzy Membership Function and Linguistic Variables

The special structure of fuzzy numbers makes calculations very time-consuming and sophisticated. Generally for facilitating calculations and practical usage, particular fuzzy
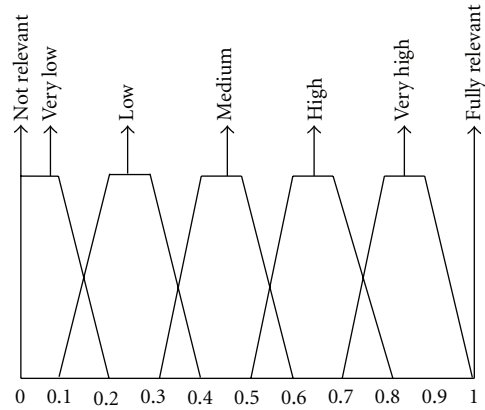

Figure 1: The membership function of linguistic variable.

numbers are used. In this paper, the experts' opinions are described by linguistic variables which have been expressed in trapezoidal fuzzy numbers. In order to determine the relevancy of ontology elements to related specific domain (concept's membership degree to main domain and relation's membership degree to concepts), seven linguistic variables have been used as: "not relevant", "very low relevant", "low relevant", "medium relevant", "high relevant", "very high relevant", and "fully relevant". Figure 1 presents these linguistic variables and their corresponding trapezoidal fuzzy numbers.

4. Fuzzification the "Modification Process Ontology"

Software maintenance happens in a relatively disorganized way and naturally leads to the deterioration of software systems' structure. Lacking a complete knowledge of all the implementation details, apply modifications that will result in a loss of structure, which in turn makes the systems more difficult to understand fully and, therefore, to maintain [30]. To break this vicious circle, we aim at developing a knowledge management approach for software maintenance domain. This approach will be modeled as fuzzy ontology.

The initial modification process ontology, which has been proposed by Dias, is presented in Figure 2. This ontology organizes concepts from the modification request (and its causes) to the maintenance activities in SME domain [24]. In this study, the existing uncertainty in concepts and relations of the modification process ontology is modeled in two phases. Firstly, an uncertainty degree for describing the concepts of ontology is obtained using linguistic variables. Secondly, the relations of existing ontology have been fuzzified, and finally, a new ontology with a unique membership degree for each relation is proposed by implementing fuzzy composition. Assume that U and V are two collections of objects. An arbitrary fuzzy set R, defined in the Cartesian product $V \times W$, will be called a fuzzy relation in the space $U \times V$. R is thus a function defined in the space $U \times V$, which takes values from the interval $[0, 1]$. Let R_1 be a fuzzy relation in $U \times V$ and R_2 a fuzzy relation in $V \times W$. For all
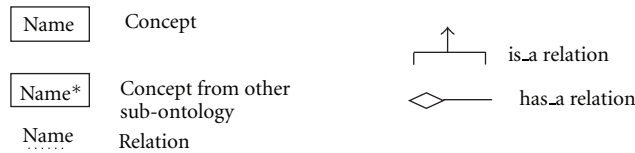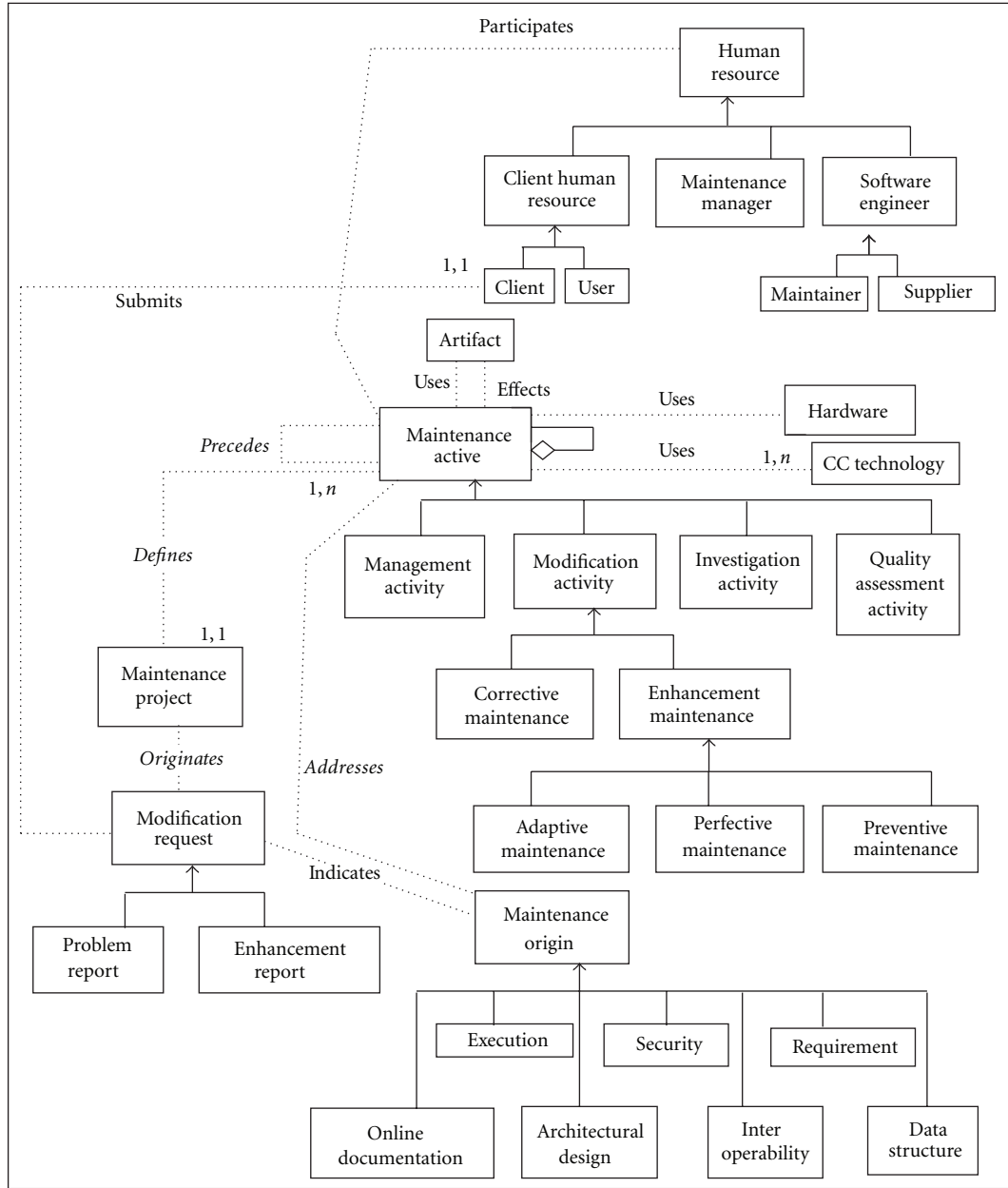
FIGURE 2: The modification process ontology.

(u, w) in $U \times W$, max-min, max-product, and max-average compositions are defined as follows [31]:

$$\text{Max-Min}(R_1, R_2)(u, w)$$
$$= \text{Max}(\text{Min}(R_1(u, v), R_2(v, w))) \text{ over all } v \text{ in the set } V, \tag{1}$$

$$\text{Max-Product}(R_1, R_2)(u, w)$$
$$= \text{Max}(R_1(u, v) \times R_2(v, w)) \text{ over all } v \text{ in the set } V, \tag{2}$$

$$\text{Max-Average}(R_1, R_2)(u, w)$$
$$= \left(\frac{1}{2}\right) \text{Max}\{(R_1(u, v) + R_2(v, w))\} \text{ over all } v \text{ in the set } V. \tag{3}$$

TABLE 1: Concept's fuzzification results of the modification process ontology.

Row	Concepts	Average of expert's opinions	Defuzzification value
1	Human resource	(0.80,0.87,0.93,1)	0.9
2	Client human resource	(0.7,0.8,0.9,1)	0.85
3	Maintenance manager	(0.7,0.8,0.9,1)	0.85
4	Software engineer	(0.9,0.93,0.97,1)	0.95
5	Client	(0.7,0.8,0.9,1)	0.85
6	User	(0.63,0.73,0.83,0.93)	0.78
7	Maintainer	(0.9,0.93,0.97,1)	0.95
8	Supplier	(0.63,0.73,0.83,0.93)	0.78
9	Artifact	(0.43,0.53,0.63,0.73)	0.58
10	Maintenance activity	(0.63,0.73,0.83,0.93)	0.78
11	Hardware	(0.63,0.73,0.83,0.93)	0.78
12	CC technology	(0.37,0.47,0.57,0.87)	0.52
13	Management activity	(0.80,0.87,0.93,1)	0.9
14	Modification activity	(0.80,0.87,0.95,1)	0.92
15	Investigation activity	(0.80,0.87,0.93,1)	0.9
16	Quality assurance activity	(0.80,0.87,0.93,1)	0.9
17	Maintenance project	(0.63,0.73,0.83,0.93)	0.78
18	Corrective maintenance	(0.80,0.87,0.93,1)	0.9
19	Enhancement maintenance	(0.80,0.87,0.93,1)	0.9
20	Modification request	(0.80,0.87,0.93,1)	0.9
21	Adaptive maintenance	(0.7,0.8,0.93,1)	0.87
22	Perfective maintenance	(0.7,0.8,0.9,1)	0.85
23	Preventive maintenance	(0.7,0.8,0.9,1)	0.85
24	Problem report	(0.7,0.8,0.9,1)	0.85
25	Enhancement report	(0.7,0.8,0.9,1)	0.85
26	Maintenance origin	(0.80,0.87,0.93,1)	0.9
27	Execution	(0.7,0.8,0.9,1)	0.85
28	Security	(0.7,0.8,0.9,1)	0.85
29	Requirement	(0.7,0.8,0.9,1)	0.85
30	Online documentation	(0.7,0.8,0.9,1)	0.85
31	Architectural design	(0.7,0.8,0.9,1)	0.85
32	Interoperability	(0.7,0.8,0.9,1)	0.85
33	Data structure	(0.7,0.8,0.9,1)	0.85

4.1. Concepts' Fuzzification of the "Modification Process Ontology".
Although fuzzification of ontology's relations has been focused in almost all the revised papers, concept's fuzzification has been considered rarely [2, 4, 9, 11, 18–21]. So, this section considers fuzzification of ontology concepts particularly. The supportive idea is that existing concepts in ontology are not related completely to the considered SME domain, so each concept has a specific membership degree to the specific domain. For fuzzification of the concepts' ontology, a special questionnaire was designed and related experts were asked to determine the relevancy degree of the modification process ontology's concepts to the domain with following linguistic values: "not relevant", "very low relevant", "low relevant", "medium relevant", "high relevant", "very high relevant", and "fully relevant". These values were fuzzified by the membership function shown in Figure 1. In order to reach a unique membership degree for each concept of the ontology, trapezoidal fuzzy variable defuzzifier has

been used. The defuzzification value of a trapezoidal fuzzy number $\widetilde{A} = [a_1, a_2, a_3, a_4]$, $a_1 \leq a_2 \leq a_3 \leq a_4$, is defined as [32]

$$D = \frac{(a_1 + a_2 + a_3 + a_4)}{4}. \tag{4}$$

The results have been presented in Table 1.

4.2. Relations' Fuzzification of the "Modification Process Ontology".
Resembling the mentioned process in Section 4.1, another questionnaire was designed for fuzzification of ontology relations in Figure 2. The experts were asked to determine the relevancy degree of each relation to the existed concepts with these values: "not relevant", "very low relevant", "low relevant", "medium relevant", "high relevant", "very high relevant", and "fully relevant". These linguistic variables were fuzzified with the membership function which has been shown in Figure 1. Table 2 shows the results of

TABLE 2: Relation's fuzzification results of the modification process ontology.

Row	Ontology relations	Average of expert's opinions	Defuzzification value
1	participates	(0.7,0.8,0.9,1)	0.85
2	is a	(1,1,1,1)	1.00
3	is a	(1,1,1,1)	1.00
4	is a	(1,1,1,1)	1.00
5	is a	(1,1,1,1)	1.00
6	is a	(0.9,0.93,0.97,1)	0.95
7	is a	(0.9,0.93,0.97,1)	0.95
8	is a	(0.7,0.8,0.9,1)	0.85
9	submits	(0.9,0.93,0.97,1)	0.95
10	uses	(0.9,0.93,0.97,1)	0.95
11	affects	(0.9,0.93,0.97,1)	0.95
12	precedes	(0.9,0.93,0.97,1)	0.95
13	has a	(1,1,1,1)	1.00
14	uses	(0.7,0.8,0.9,1)	0.85
15	uses	(0.7,0.8,0.9,1)	0.85
16	defines	(0.9,0.93,0.97,1)	0.95
17	addresses	(0.7,0.8,0.9,1)	0.85
18	is a	(0.9,0.93,0.97,1)	0.95
19	is a	(0.9,0.93,0.97,1)	0.95
20	is a	(0.9,0.93,0.97,1)	0.95
21	is a	(0.9,0.93,0.97,1)	0.95
22	originates	(0.7,0.8,0.9,1)	0.85
23	is a	(0.9,0.93,0.97,1)	0.95
24	is a	(0.9,0.93,0.97,1)	0.95
25	is a	(0.7,0.8,0.9,1)	0.85
26	is a	(0.7,0.8,0.9,1)	0.85
27	is a	(0.7,0.8,0.9,1)	0.85
28	is a	(0.7,0.8,0.9,1)	0.85
29	is a	(0.7,0.8,0.9,1)	0.85
30	indicates	(0.63,0.73,0.83,0.93)	0.78
31	is a	(0.9,0.93,0.97,1)	0.95
32	is a	(0.9,0.93,0.97,1)	0.95
33	is a	(0.9,0.93,0.97,1)	0.95
34	is a	(0.9,0.93,0.97,1)	0.95
35	is a	(0.9,0.93,0.97,1)	0.95
36	is a	(0.9,0.93,0.97,1)	0.95
37	is a	(0.9,0.93,0.97,1)	0.95

modeling the ontology relations with trapezoidal numbers. In order to reach a unique membership degree for each relation of the ontology, the trapezoidal fuzzy variable defuzzifier (4) has been used. Column 2 of Table 2 shows the ontology relations which have been presented in Figure 3.

4.3. Proposing the New Combined Model Based on Fuzzy Set Theory. The final step of this phase is the combination of concepts and relations' membership degrees and proposing a unique membership degree for ontology relations in the SME domain. Because the relevancy degree of the ontology relations to the concepts has been fuzzified and the concepts of the SME domain are fuzzified as well, the membership

degree of relations in the modification process ontology can be calculated by using fuzzy compositions. Implementing the proposed fuzzy ontology shows better results for max-product composition (2) in comparison to other fuzzy compositions, therefore this fuzzy composition type is used. The final fuzzy ontology has been shown in Figure 3.

5. The Query Expansion Algorithm

In this section, topics including the proposed query expansion algorithm, the algorithm in a pseudocode form and computational complexity of the crisp, and fuzzy ontology query expansion algorithms are discussed.
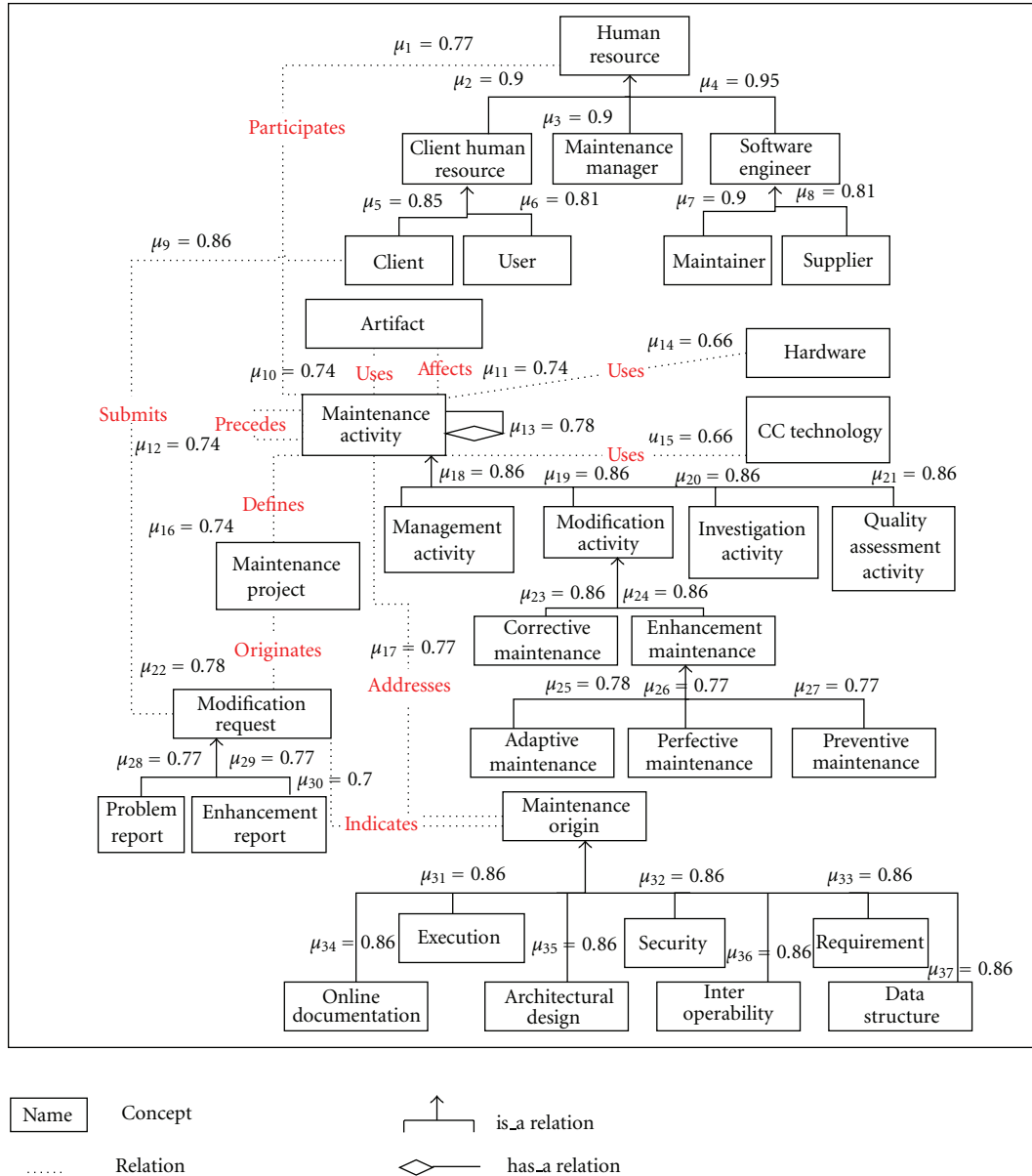
FIGURE 3: The final fuzzified modification process ontology.

5.1. The Proposed Query Expansion Algorithm Based on the Ontology. In this algorithm, it is assumed that entered user query terms are existed in the modification process ontology. If the query terms were not in the related ontology of a specific domain, the most common approach is adding synonym terms based on a general dictionary (commonly English comprehensive ontology WordNet). General dictionaries do not consider any specific domain, hence a satisfied accuracy will not be obtained [33]. Therefore, in the proposed query expansion algorithm, those situations are considered that the entered query terms exist in the modification process ontology. Different approaches are used for query expansion based on the existence or the lack of relations between query terms in the ontology. These approaches are explained in the following.

(a) Query Expansion Based on Crisp Ontology. With this supposition that query terms exist in the modification process ontology and these terms have or do not have a semantic relation with each other, two situations are possible.

(i) Query terms exist in the modification process ontology and have semantic relation. The associated relations of terms in an ontology are extracted (it is possible that one term is father or child of another term). For each father term, all its father terms (generalized terms) are extracted, and for each child term, all its child terms (specialized terms) are extracted as well and then these extracted terms are entered in the expanded query, consequently. For example, consider query = ("modification activity", "enhancement maintenance"). These terms have "modification activity" > "enhancement

```
//situation (a.i)
for (){
get(query);
if (query words exist in modification process ontology && have relation together){
    if (the query word is parent of other words)
      add(other parent of this word to the query terms);
    else if (the query word is child of other words)
      add(other child of this word to the query terms);
}//end if
}//end for
* * * * * * * * * * * * * * * * * * * * * * * * *
//situation (a.ii)
for (){
get(query);
    if (query words exist in modification process ontology && have no relation together){
    add(all parents and children of this words to query);
}//end for
* * * * * * * * * * * * * * * * * * * * * * * * *
//situation (b.i)
for (){
get(query);
  if (query words exist in modification process ontology && have relation together){
    if (the query word is parent of other words && their membership degree >=threshold limit)
      add(other parent of this word to the query terms);
    else if (the query word is child of other words && their membership degree >=threshold limit)
      add(other child of this word to the query terms);
}//end if
}//end for
* * * * * * * * * * * * * * * * * * * * * * * * *
//situation (b.ii)
for (){
get(query);
  if (query words exist in modification process ontology && have no relation together && their
  membership degree >=threshold limit){
    add(all parents and children of this words to query);
}//end for
* * * * * * * * * * * * * * * * * * * * * * * * *
```

PSEUDOCODE 1

maintenance" relation in the ontology; thus, instead of the term "modification activity", generalized terms such as "maintenance activity" will be entered in the expanded query. Similarly, instead of the term "enhancement maintenance", specialized terms such as "adaptive maintenance", "perfective maintenance", and "preventive maintenance" are added in the expanded query.

(ii) Query terms exist in the modification process ontology and do not have semantic relations. In this situation, all related father and child terms will be entered in the expanded query. As a case in point, consider query = ("corrective maintenance", "maintenance project"). Because there is no relation between these query terms in the ontology, all of their father and child terms are added in the expanded query; so, instead of the term "corrective maintenance", the terms "modification activity", and instead of the term "maintenance project", the terms "maintenance activity" and "modification request" will be added in expanded query.

(b) Query Expansion Based on Fuzzy Ontology. This status is very similar to status (a) with the difference that terms added in the query have a membership degree equal or greater than a certain threshold limit. As is discussed in the next section, the maximum average precision belongs to threshold limit 0.78, so this limit is considered. For example, consider query = ("modification activity", "enhancement maintenance"). These terms have "modification activity" > "enhancement maintenance" relation in the ontology and obey the situation (i); thus, instead of term the "modification activity", generalized terms with a membership degree equal or greater than 0.78 such as "maintenance activity" will be entered in the expanded query. Similarly, instead of the term "enhancement maintenance", specialized terms with membership degree equal or greater than 0.78 such as "adaptive maintenance" are added in the expanded query.

As a more illustration, consider query = ("corrective maintenance", "maintenance project"). Because there is no relation between these query terms in the ontology, all of

their father and child terms are added in the expanded query based on the situation (ii); so, instead of term "corrective maintenance", the term "modification activity" and instead of term "maintenance project" the term "modification request" which have a membership degree equal or greater than 0.78 will be added in expanded query.

5.2. Pseudocode Form of the Algorithm. The pseudocode form of the algorithm based on crisp and fuzzy ontology is as shown in Pseudocode 1.

5.3. Computational Time of the Crisp and Fuzzy Ontology Query Expansion Algorithm. Computational time of the crisp and fuzzy ontology query expansion algorithm in worth case of both situation (a) and (b) of our pseudocode will be $O(n)$, because in the worth case the user enters all of terms of the crisp and fuzzy ontology in the query and in normal case is $O(\text{constant})$.

6. Performance Evaluation of Information Retrieval System

The two most frequent and basic measures for information retrieval effectiveness are precision and recall [34, 35]. So, we used these measures for the ontology performance evaluation. Precision and recall are defined in terms of a set of retrieved documents (e.g., the list of documents produced by a web search engine for a query) and a set of relevant documents (e.g., the list of all documents on the internet that are relevant for a certain topic),

$$
\begin{aligned}
\text{Precision} &= \frac{\#(\text{relevant items retrieved})}{\#(\text{retrieved items})}, \\
&= P(\text{relevant} \mid \text{retrieved}) \\
\text{Recall} &= \frac{\#(\text{relevant items retrieved})}{\#(\text{relevant items})} \\
&= P(\text{retrieved} \mid \text{relevant}).
\end{aligned}
\tag{5}
$$

The average precision plots at each standard recall level across all queries and evaluates overall system performance on a document/query corpus [36, 37].

For performance evaluation of information retrieval system in local searches, the Google Desktop search engine has been implemented, and a lot of related documents to the Software Maintenance Engineering domain have been entered in. For appropriate threshold limit determining in described situation (b), various thresholds from 0 to 0.9 have been considered. Furthermore, both unique and multiple query terms were entered in the search engine. Figure 4 shows the average precision of fuzzy ontology for different threshold limits. As it is obvious in this figure, the maximum average precision belongs to threshold limit 0.78.

As a result, the threshold limit 0.78 has been used for the situation (b) in the proposed query expansion algorithm. In Figure 5, the average precision for different levels of the recall values has been depicted for the three
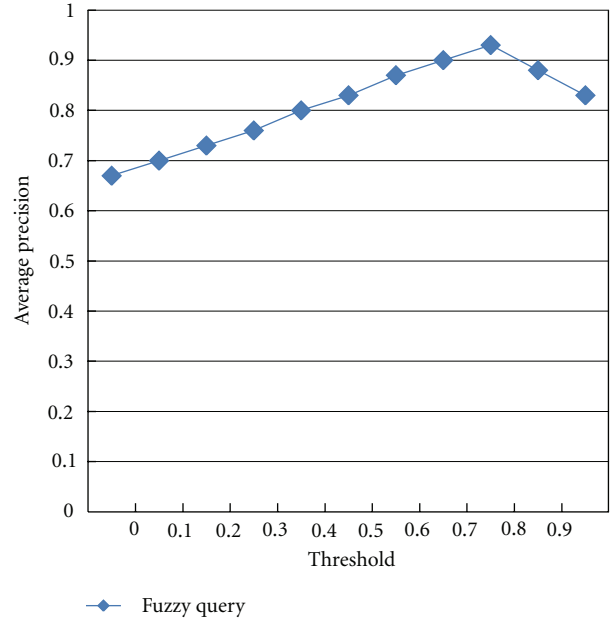

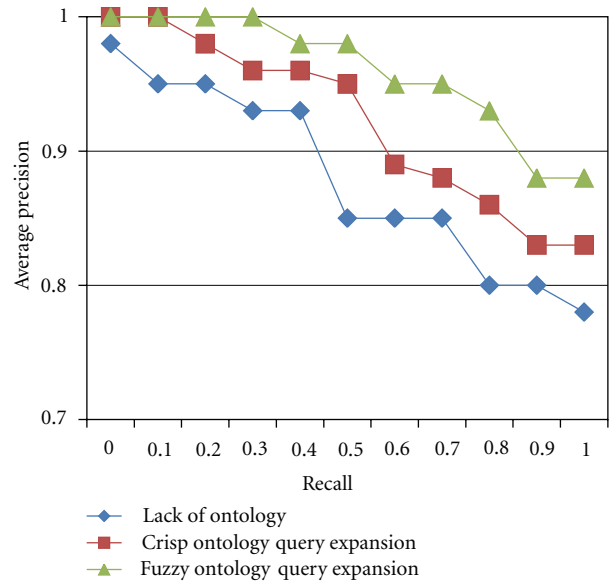FIGURE 4: The average precision plot for different fuzzy query's threshold limits.

FIGURE 5: The average precision plot for three situations: lack of ontology, crisp ontology query expansion and fuzzy ontology query expansion.

situations: the lack of ontology situation, the crisp ontology query expansion situation, and the fuzzy ontology query expansion situation. These situations are selected to evaluate the fuzzy ontology performance rather than the lack of ontology situation and the crisp ontology query expansion situation. As shown in Figure 5, the fuzzy query expansion situation has the best average precision in retrieving relevant documents.

7. Conclusion

In this paper we proposed a new approach for effective retrieving of information by implementing fuzzy ontology generation technique with two uncertainty degrees. Taking fuzzy ontology can tackle the uncertainty of relations in comparison to taking crisp ontology and make it possible to find uncertain information in a specific domain. For performance evaluation of information retrieval system, we considered three different situations (two situation with ontology existence (crisp or fuzzy) and one without any ontology). The empirical results of using proposed query expansion algorithm showed that crisp ontology's average precision increased 5% in comparison to take no ontology for expansion job and fuzzy ontology's average precision increased 3% in comparison to use crisp ontology query expansion situation. Furthermore, fuzzy ontology's average precision increased 8% in comparison to take no ontology for expansion. These results showed that the quality of information retrieval system using fuzzy ontology query expansion method improves remarkably.

In the future, we will implement fuzzy theory and neural network methods to build fuzzy ontology from unstructured data automatically. Also, we will compare the results with the proposed approach in this paper.

Acknowledgment

References

[1] M. Swain, J. A. Andersen, N. Swain, and R. Korrapati, "Study of information retrieval using fuzzy queries," in *IEEE Southeastcon 2005: Excellence in Engineering, Science and Technology*, pp. 527–535, April 2005.

[2] M. A. A. Leite and I. L. M. Ricarte, "Fuzzy information retrieval model based on multiple related ontologies," in *Proceedings of The 20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI '08)*, pp. 309–316, November 2008.

[3] J. Li, J. Y. Song, and H. Zhong, "Ontology-based query division and reformulation for heterogeneous information integration," *Journal of Software*, vol. 18, no. 10, pp. 2495–2506, 2007.

[4] N. Guarino and P. Giaretta, *Ontologies and Knowledge Bases:Towards a Terminological Clarification. Toward very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, IOS Press, 1995.

[5] N. Shadbolt, W. Hall, and T. Berners-Lee, "The semantic web revisited," *IEEE Intelligent Systems*, vol. 21, no. 3, Article ID 1637364, pp. 96–101, 2006.

[6] B. C. Chien, C. H. Hu, and M. Y. Ju, "Intelligent information retrieval applying automatic constructed fuzzy ontology," in *Proceedings of The 6th International Conference on Machine Learning and Cybernetics (ICMLC '07)*, pp. 2239–2244, August 2007.

[7] Q. T. Tho, S. C. Hui, A. C. M. Fong, and T. H. Cao, "Automatic Fuzzy ontology generation for semantic Web,"

[8] R. Bandarian, "Measuring Commercial Potential of Technology with fuzzy Logic," *STP Journal*, vol. 1, no. 1, pp. 15–32, 2008.

[9] B. Zhang, H. Li, Y. Du, and Y. Wang, "Query expansion based on topics," in *Proceedings of The 5th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD '08)*, pp. 610–614, October 2008.

[10] H. Hagras, "Type-2 FLCs: a new generation of fuzzy controllers," *IEEE Computational Intelligence Magazine*, vol. 2, no. 1, pp. 30–43, 2007.

[11] C. S. Lee, Z. W. Jian, and L. K. Huang, "A fuzzy ontology and its application to news summarization," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 35, no. 5, pp. 859–880, 2005.

[12] C. S. Lee, M. H. Wang, C. Y. Hsub, and H. Hagras, "Novel type-2 fuzzy ontology and its application to diet assessment," in *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 2009.

[13] C. S. Lee, M. H. Wang, and H. Hagras, "A type-2 fuzzy ontology and its application to personal diabetic-diet recommendation," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 2, Article ID 5409608, pp. 374–395, 2010.

[14] T. T. Quan, S. C. Hui, and A. C. M. Fong, "Automatic fuzzy ontology generation for semantic help-desk support," *IEEE Transactions on Industrial Informatics*, vol. 2, no. 3, Article ID 1668074, pp. 155–163, 2006.

[15] A. Bahri, R. Bouaziz, and F. Gargouri, "Fuzzy ontology implementation and query answering on databases," in *Proceedings of The Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS '09)*, June 2009.

[16] J. Z. Pan, G. Stamou, G. Stoilos, S. Taylor, and E. Thomas, "Scalable querying services over fuzzy ontologies," in *Proceedings of The 17th International Conference on World Wide Web (WWW '08)*, pp. 575–584, April 2008.

[17] D. Parry, "A fuzzy ontology for medical document retrieval," in *Proceedings of The Australasian Worhhop on Data Mining and Web Intelligence*, 2004.

[18] J. Zhai, Y. Liang, Y. Yu, and J. Jiang, "Semantic information retrieval based on fuzzy ontology for electronic commerce," *Journal of Software*, vol. 3, no. 9, 2008.

[19] J. Zhai, Q. Wang, and M. Lv, "Application of fuzzy ontology framework to information retrieval for SCM," in *Proceedings of The International Symposium on Information (ISIP '08)*, pp. 173–177, May 2008.

[20] J. Zhai, Y. Cao, and Y. Chen, "Semantic information retrieval based on fuzzy ontology for intelligent transportation systems," in *Proceedings of The IEEE International Conference on Systems, Man and Cybernetics (SMC '08)*, pp. 2321–2326, October 2008.

[21] J. Zhai, Y. Yu, Y. Liang, and J. Jiang, "Traffic information retrieval based on fuzzy ontology and RDF on the Semantic Web," in *Proceedings of The 2nd International Symposium on Intelligent Information Technology Application (IITA '08)*, pp. 779–784, December 2008.

[22] M. A.A. Leite and I. L.M. Ricarte, "Document retrieval using fuzzy related geographic ontologies," in *Proceedings of the International Conference on Information and Knowledge Management*, pp. 47–54, 2008.

[23] G. Fu, C. B. Jones, and A. I. Abdelmoty, "Ontology-based spatial query expansion in information retrieval," *Lecture Notes in Computer Science*, vol. 3761, pp. 1466–1482, 2005.

IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 6, pp. 842–856, 2006.

[24] C. Bratsas, V. Koutkias, E. Kaimakamis, P. Bamidis, and N. Maglaveras, "Ontology-based vector space model and fuzzy query expansion to retrieve knowledge on medical computational problem solutions," in *Proceedings of The 29th IEEE Annual International Conference on Engineering in Medicine and Biology Society*, vol. 2007, pp. 3794–3797, 2007.

[25] Y. Ogawa, T. Morita, and K. Kobayashi, "A fuzzy document retrieval system using the keyword connection matrix and a learning method," *Fuzzy Sets and Systems*, vol. 39, no. 2, pp. 163–179, 1991.

[26] R. Pereira, I. Ricarte, and F. Gomide, "Chapter 20 Fuzzy relational ontological model in information search systems," *Capturing Intelligence*, vol. 1, no. C, pp. 395–412, 2006.

[27] M. D. Cock and C. Cornelis, "Fuzzy rough set based web query expansion," in *Proceedings of Rough Sets and Soft Computing in Intelligent Agent and Web Technology, International Workshop at WI-IAT*, pp. 9–16, 2005.

[28] S. Calegari and E. Sanchez, "A fuzzy ontology-approach to improve semantic information retrieval," *InternationalJournal of Computer Science and Applications IV(II)*, pp. 125–144, 2007.

[29] A. Bahri, R. Bouaziz, and F. Gargouri, "Fuzzy ontology implementation and query answering on databases," in *Proceedings of The Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS '09)*, Cincinnati, Ohio, USA, June 2009.

[30] M. G. B. Dias, N. Anquetil, and K. M. De Oliveira, "Organizing the knowledge used in software maintenance," *Journal of Universal Computer Science*, vol. 9, no. 7, pp. 641–658, 2003.

[31] A. A. Molai and E. Khorram, "An algorithm for solving fuzzy relation equations with max-T composition operator," *Information Sciences*, vol. 178, no. 5, pp. 1293–1308, 2008.

[32] C. H. Cheng and Y. Lin, "Evaluating the best main battle tank using fuzzy decision theory with linguistic criteria evaluation," *European Journal of Operational Research*, vol. 142, no. 1, pp. 174–186, 2002.

[33] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to wordnet: an on-line lexical database," *International Journal of Lexicography*, vol. 3, no. 4, pp. 235–244, 1990.

[34] B. Teimourpour, M. M. Sepehri, and L. Pezeahk, "A new method for intelligent categorization of scientific texts (case of Iran's nanotechnology papers)," *STP Journal*, vol. 2, no. 2, pp. 5–6, 2009.

[35] A. M. Farough and L. S. Loghman, "Evaluation of procedures and methods of university researches commercialization; the case of Tabriz University," *STP Journal*, vol. 3, no. 4, pp. 15–29, 2011.

[36] H. Tbatabaeian and M. Entezari, "Institution mapping of power industry as an example," *STP Journal*, vol. 1, no. 1, pp. 53–64, 2008.

[37] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press Cambridge, England, UK, 2008.