

Fuzzy-Ontology-Enrichment-based Framework for Semantic Search

Hajer Baazaoui-Zghal and Henda Ben Ghezala

Riadi-GDL Laboratory, ENSI, Manouba University, Tunis, Tunisia

Keywords: Fuzzy Ontology, Web Search, Query Reformulation.

Abstract: The dominance of information retrieval on the Web makes integrating and designing ontologies for the on-line Information Retrieval Systems (IRS) an attractive research area. In addition to domain ontology, some attempts have been recently made to integrate fuzzy set theory with ontology, to provide a solution to vague and uncertain information. This paper presents a framework for semantic search based on ontology enrichment and fuzziness (FuzzOntoEnrichIR). FuzzOntoEnrichIR main components are: (1) a fuzzy information retrieval component, (2) an incremental ontology enrichment component and (3) an ontology repository component. The framework aims on the one hand to capitalize and formulate extraction-ontology rules based on a meta-ontology. On the other hand, it aims to integrate the domain ontology enrichment and the fuzzy ontology building in the IR process. The framework has been implemented and experimented to demonstrate the effectiveness and validity of the proposal.

1 INTRODUCTION

Ontologies are defined as an explicit formal specification of a shared conceptualization. They can be classified as lightweight ontologies gathering concepts and relations' hierarchies which can be enriched by classical properties called axiom schemata (algebraic properties and signatures of relations, abstraction of concepts, etc.) and heavyweight ontologies which add properties to the semantics of the conceptual primitives and are only expressible in axiom domain form. The axioms schemata describe the classical properties of concepts and relations (subsumption, disjunction of concepts, algebraic properties and cardinalities of the relations, etc.). The domain axioms characterize domain properties expressible only in an axiom form. They specify the formal semantics constraining the conceptual primitive interpretation.

Currently, ontologies are playing a fundamental role in knowledge management and semantic Web. Building ontology manually is a long and tedious task. Thus, many approaches have been proposed during the last decade to make this task easier. Information Retrieval (IR) deals with models and systems aiming to facilitate accessibility to sets of documents and provide to users the corresponding ones to their needs, by using queries. Generally, Information Retrieval System (IRS) integrates techniques allowing selection of relevant information. The first research on ontologies for the IRS dates back to the late 90s

(McGuinness, 1998), and aims to remedy limits of the traditional IRS based on the keywords. This research topic presents one of the main actual axes of the semantic Web.

In addition to domain ontology, the integration of the fuzzy logic shows that it presents an interesting way to solve uncertain information problems. In fact, fuzzy logic is used in IR to solve the ambiguity issues by defining flexible queries or fuzzy indexes (e.g., (Baazaoui-Zghal et al., 2008)). A fuzzy ontology can be considered as an extension of domain ontology by embedding a set of membership degrees in each concept of the domain ontology and adding fuzzy relationships among these fuzzy concepts (Zhou et al., 2006).

In this paper, we present a framework for semantic search based on fuzzy ontologies. It includes an ontology repository (meta-ontology generating a domain ontology and ontology of domain services), incremental approach of domain ontology learning and fuzzy ontology enrichment method. The framework has been implemented to demonstrate the proposal effectiveness and to evaluate it.

The remaining of this paper is organized as follows. Section 2 presents related works to information retrieval and fuzzy ontologies. Section 3 describes our proposal. Section 4 presents and discusses some experimental results of our framework. Finally, section 5 concludes and discusses directions for future research.

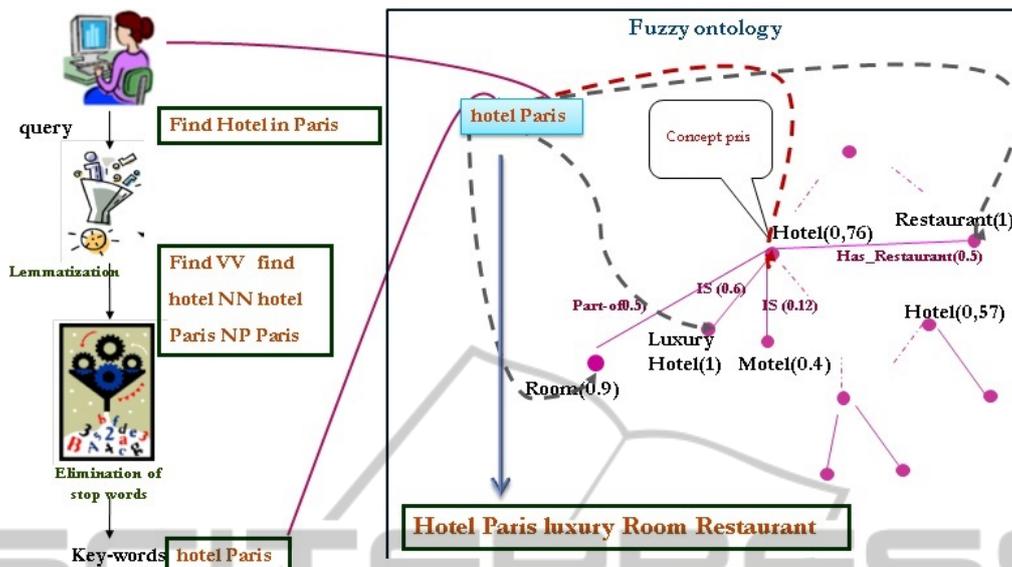


Figure 1: An example of a fuzzy ontology.

2 RELATED WORKS AND MOTIVATIONS

Several studies were presented showing how the fuzzy logic could be integrated to IRS, to solve uncertain information problems (Zhou et al., 2006). We precise here, that in a fuzzy ontology, each index term or object is related to every term (or object) in the ontology, with a degree of membership assigned to the relationship and based on fuzzy logic.

The fuzzy membership value μ is used for the relationship between terms or objects, where $0 < \mu < 1$, and μ corresponds to a fuzzy membership relation such as "strongly", "partially", "somewhat", "slightly"

etc, where for each term: $\sum_{i=1}^{i=n} \mu_i = 1$;

n is the number of relations that a particular object has, $n = (N - 1)$, with N representing the total number of objects in the ontology (Lee et al., 2005).

The insertion of the fuzzy logic and the ontology in the process of information retrieval has improved the quality and the precision of the returned results. Thus, integration of the fuzzy ontology into the IR process is an interesting area of research and can lead to more relevant results than in the case where ontology and fuzzy logic are used separately (Chien et al., 2010; Bordogna et al., 2009; Calegari and Ciucci, 2006). Several existing IRSs (Zhou et al., 2006; Chien et al., 2010; Calegari and Ciucci, 2006) generally use semi-automatic or automatic methods, which allow the fuzzification only of the "IS-A" rela-

tion. In addition, from the state of the art (Chien et al., 2010; Colleoni et al., 2009), it is noticed that there is a lack of information retrieval system integrating fuzzy ontology allowing a document classification and assisting users in their searches.

First, in (Sayed et al., 2007), document classification is not based on fuzzy ontologies.

In fact, classification based only on domain ontology could not take into account dynamic aspect of fuzzy ontology, mainly when the aim is to improve query reformulation and information retrieval results. Both in (Parry, 2006; Akinribido et al., 2011), only "IS-A" relations are taken into account. Nevertheless, all relations are important mainly in case of query reformulation.

From the conducted survey made on methods for fuzzy ontology construction, we have noticed that automatic methods can take as input a database (Lee et al., ; Quan et al., 2006), a documentary corpus (Widyantoro and Yen, 2001) or an existing ontology (fuzzification) (Parry, 2006),(Sayed et al., 2007), (Chien et al., 2010), (Calegari and Ciucci, 2006).

Figure 1 illustrates an example of a fuzzy ontology. The number related to relations represents the membership value of the relationship between the concept "Hotel" and other concepts (room, Suite). The related value to a concept describes the importance of this concept into the ontology. These different relations and concepts will have different membership values depending on the context of the query, and particularly the user's view of the world.

In this paper, our main objective is to develop a

framework allowing ontology's building for the semantic Web. The proposed framework includes an ontology repository (meta-ontology generating a domain ontology and ontology of domain services, and fuzzy ontology), incremental approach of domain ontology learning, fuzzy ontology building method and ontology based retrieval process.

So, the aims of this paper can be summarized in:

- The integration of the domain ontology enrichment and the fuzzy ontology building in the IR process.
- The capitalization and formulation of extraction ontology rules based on a meta-ontology aiming to explicitly specify knowledge about the concepts, relationships, instances and axioms extraction, the learned patterns and frames, and the semantic distance

3 FUZZY-ONTOLOGY-ENRICHMENT-BASED FRAMEWORK FOR SEMANTIC SEARCH

The general structure of the framework (FuzzyOntoEnrichIR) is given by Figure 2. FuzzOntoEnrichIR framework is based on a fuzzy ontology building, an incremental ontology learning and an ontology repository.

The main FuzzOntoEnrichIR components are: a fuzzy ontology building component, an incremental ontology learning component and an ontological repository component.

The first component is composed of:

- Two methods of information retrieval based on a domain ontology and an individual fuzzy ontology.
- An automatic method of fuzzy ontology building: allowing the fuzzification of all existing relations in the initial domain ontology initial, and assuring the updating of membership values at the end of every information retrieval, which is made dynamically by the user
- A classification of documents by service using a domain ontology

This first component aims to automate the collection of the relevant documents which will be used as entry of the second component

The second component is based on a meta-ontology (which is a high-level ontology of abstraction (Baazaoui-Zghal et al., 2007a)) and an incremental ontology learning which may require enrichment

phase. It allows incremental construction of ontologies from the Web documents. Thus, this component proposes a composite ontological architecture of three interdependent ontologies: a generic ontology of web sites structures, a domain ontology and a service ontology.

These offer a representation of the domain and services behind Web content, which could be exploited by the semantic search engines. This latter is instantiated with the contextual information of concepts and relations of the ontology extracted incrementally from texts. The semi-automatic construction of the domain ontology is the main objective of this component.

The third component is composed of a meta-ontology generating a domain ontology and ontology of domain services, and fuzzy ontology.

The details related to each component will be given in the next subsections.

3.1 Fuzzy Ontology Building Component

FuzzyOntoEnrichIR integrates an automatic fuzzy ontology building method, with an automatic fuzzification of all the existing relations in the domain ontology, not restricted to "Is-a" relations. Indeed, in conventional ontologies, particular objects may occur in multiple locations. So, a simple expansion that does not understand the intended location of the query term may lead to many irrelevant results being returned. A fuzzy ontology membership value could therefore be used to identify the most likely location in the ontology of a particular term. Each user would have own values for the membership assigned to terms in the ontology, reflecting their likely information need and worldview.

Then, the use of an individual fuzzy ontology approach allows the convenient representation of the relationships in a domain according to a particular view, without sacrificing commonality with other views; the ontology framework is common, just the membership values are different. An individual fuzzy ontology using an automatic fuzzification based is built.

Initialization of Membership Values

We propose to build an individual fuzzy ontology using an automatic fuzzification based on Jiang-Conrath similarity measure (Jiang and Conrath, 1997). To calculate *IC* (Information Content), we use the formula presented by (Seco et al., 2004) which is based on the structure of the ontology hierarchy. In fact, this frequency has the advantage of bringing the occurrence frequency of the concept itself and the

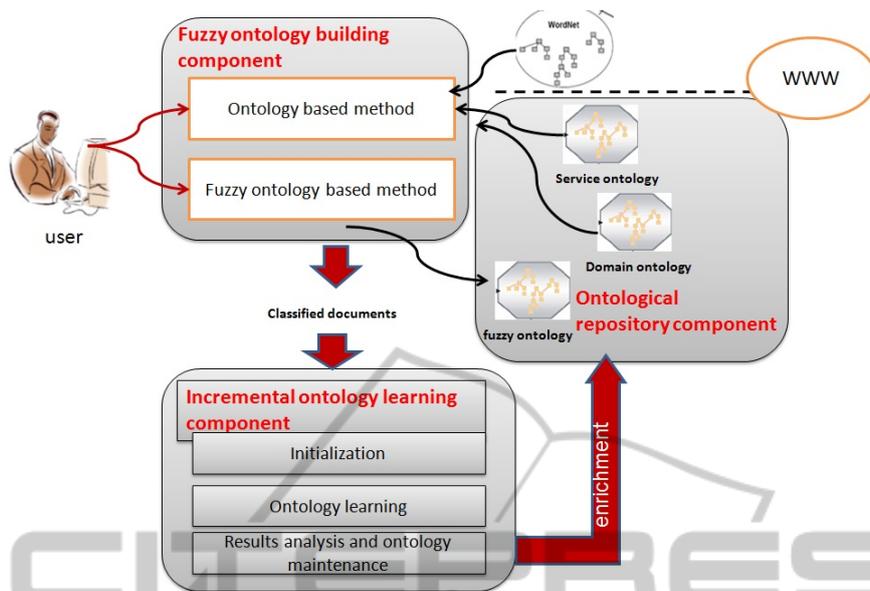


Figure 2: FuzzyOntoEnrichIR framework architecture.

concepts it subsumes, which allows supporting all relations' types.

Updating the Membership Value of Concepts and Relations

We suppose that a defined fuzzy ontology is not available in any context. Thus, it is necessary to define an update process of fuzzy values, taking into account the user's needs. The membership value should consider the previous values, the retrieved documents and the query. In the literature, there are researchers that have presented similar ways of updating membership value (Calegari and Ciucci, 2006; Parry, 2006). Inspired by these methods, we have integrated in our case two updating membership values respectively for concepts and relations.

$$\mu_{new} = \mu_{old} + \frac{\mu - \mu_{old}}{Q + 1} \quad (1)$$

where μ_{old} is the current membership value, Q is the number of update performed to this value and μ_{new} is the new value. μ is a value that evaluates the new change added to the relation or the concept. μ must take into account the query and the returned documents content that have been selected by the user.

The fuzzy ontology is used for query reformulation and for documents and query indexing. A fuzzy ontology is an individual an ontology owned by each user.

To show the purpose of the given formulas, we take as example the query sent by the user containing the concept "Hotel". The framework computes

the measures of this concept and all its related concepts (like: Rate, Restaurant, Motel...) using the returned documents selected by the user. Finally, the membership value of relations using the formula 5 are also updated. In this example the membership value of the relation "has-restaurant" between "Hotel" and "restaurant" will be updated (cf. Figure 3).

3.2 Ontological Repository Component

A dedicated architecture is proposed based on two interdependent ontologies to build a knowledge-base of a particular domain, constituted by a set of Web documents, and associated services. Thus, two ontologies are distinguished: domain ontology and ontology of domain services, which are in interaction. These two ontologies are built in accordance with the meta-ontology. Domain ontology is a set of concepts, relations and axioms that specify shared knowledge concerning a target domain. Ontology of domain services specifies for each service, its provider, its interested users, possible process of its unrolling, main activities and tasks that constitute this service (Baazaoui-Zghal et al., 2007b).

This ontology contains axioms specifying the relations between domain services and precise main domain concepts which identify each service. These ontologies are semantically linked and relationships between them are defined. The meta-ontology is a specification of meta-models of domain ontology and ontology of domain services. Besides, knowledge concerning the semi-automatic construction of domain

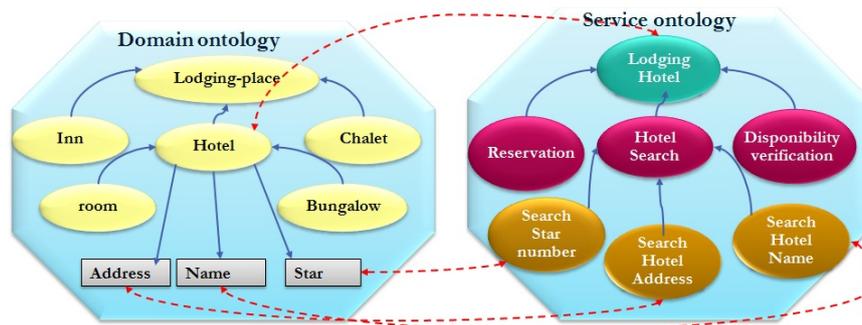


Figure 3: Relation between domain ontology and service ontology.

ontology is also specified by this meta-ontology.

The proposed architecture is composed of three ontologies, namely a generic ontology of web sites structures, domain ontology and service ontology.

The generic ontology or meta-ontology contains knowledge representation related to each ontology, which required knowledge for ontology construction, knowledge representation specified by the meta-model related to these ontologies. It is mostly based on generic concepts: "meta-concept", "meta-relation", and "meta-axiom". The class of "meta-concept" is divided into subclasses which represent respectively the domain meta-concept, the meta-concept of domain services and the meta-concept of element. Besides, the class "meta-relation" and the class "meta-axiom" are designed in the same way. The meta-ontology is, consequently, made up of three homogeneous knowledge fields. The first field is a conceptualization of knowledge related to learning concepts, relations and axioms related to a target domain.

Besides, for each instance of the class *Domain-Concept* and the class *Domain-Relation*, the technique leading to its discovery is specified.

The service ontology specifies the common services that can be solicited by web users and can be attached to several ontologies defined on subparts of the domain (*cf.* Figure 3 showing the relation between domain ontology and service ontology)

3.3 Incremental Ontology Learning Component

The incremental ontology learning component is based on a process of ontology learning from Web content according to LEO.By.LEMO (LEarning Ontology BY Learning Meta-Ontology) approach (Baazaoui-Zghal et al., 2007b). Our approach is based on learning rules of ontology extraction from texts in order to enrich ontologies in three main

phases:

- Initialization phase
- Incremental phase of learning ontology
- Result analysis phase.

The initialization phase is dedicated to data source cleaning. The input of this phase is constituted by a minimal ontology, the meta-ontology, the terminological resource "Wordnet" (Miller", 1995) and a set of Web sites delivered by a search engine and classified by domain services. A minimal ontology is designed and built to be enriched in the second phase. It is called "minimal domain ontology" as the number of concepts and relations are reduced. Consequently, data source preparation consists of: searching Web documents related to the domain corresponding to a query based on concepts describing a target service (these concepts are obtained from the projection of the corresponding service specified in the ontology of domain services), selecting Web sites provided by a search engine tool (the number of chosen sites is limited because analyzing an important number of Web pages requires very important execution time), classifying Web pages according to domain services.

Finally, cleaning Web pages by eliminating markup elements and images, text segmentation and tagging in order to obtain a tagged textual corpus. One hypothesis is that we deal with Web documents written in a target language. The meta-ontology adjustment is thus done according to linguistic knowledge related to the target language. The second phase is a learning iterative process. Each one of the iterations is made up of two main steps. The first one is the meta-ontology instantiation and the second one enables us to apply the meta-ontology axioms related to the learning of ontology. An iteration is processed in two steps. In the first step, techniques are applied to the corpus. In this context, we have adapted the construction of a word space (Baazaoui-Zghal et al., 2008) by applying the N-Gram analysis instead of a 4-

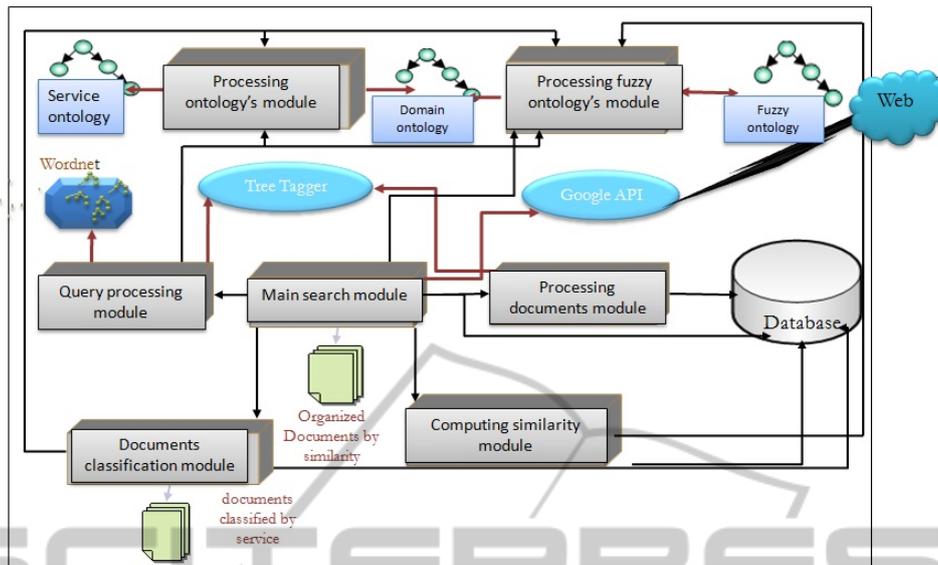


Figure 4: Modular architecture framework.

gram analysis. We have also proposed a disambiguation algorithm (Baazaoui-Zghal et al., 2007b). It aims to determine the right sense of a lexical unit. This algorithm is based on the study of term co-occurrence in the text and the selection of the adequate sense. Besides, we propose to use many similarity measures to build the similarity matrix which describes the contextual similarity between concepts.

The last phase is useful to verify the coherence of enriched ontology by analyzing learning results. We admit that the maintenance of the meta-ontology allows the readjustment of the rules according to the results obtained in order to improve the ontology construction during a further execution of the second phase of the process. Moreover the correction of meta-ontology generates a more valid ontology scheme and richer.

4 EXPERIMENTATION AND RESULTS ANALYSIS

The implementation and experimentations of the proposed framework, have been done in order to evaluate the proposed architecture. Figure 4 gives an idea about the developed modules and the application structure. FuzzEnrichIR is composed of the construction and updating module which allows manipulation of the fuzzy ontology. The processing module regroups classes assuring the different treatments done on the request and on the documents (as indexing and downloading). The class module regroups the

different useful classes to the document classification by service. A pre-processing module of data sources which pre-process textual corpus, its POS (Part-of-Speech) tagging and importation of terminological and conceptual resources (minimal ontology, Ontology of domain services and terminological resource "Wordnet"). An editing module of the meta-ontology which allows concepts and ontology axioms update by integrating the Plug-in of Protege-OWL tool. A module for domain ontology generation. An alimentation module of the meta-ontology which consists of conceptual elements in the meta-ontology from text and implements the incremental process of ontology domain construction proposed by the "LEO-By-LEMO" approach. Finally, a module of domain ontology learning which is the result of the association and the development of learning techniques set (concepts and relations).

The framework was implemented in Java, providing an online service and using the Jena Api to handle ontologies and Google Api to search through the Web.

Several experiments were conducted to investigate the performance of our proposal and to evaluate:

- The impact of ontology enrichment on information retrieval relevance
- The impact of fuzzy ontology enrichment on information retrieval relevance with and without update

The adopted protocol is centered on users, and the used data for the experimentation and the evaluation was composed of a domain ontology and users' requests. Fifteen queries in the tourism's domain and

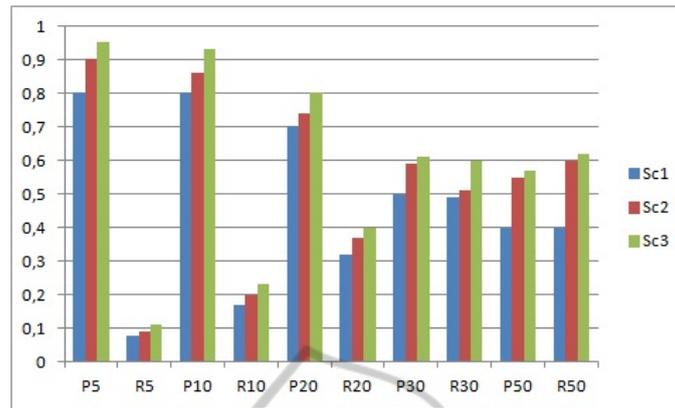


Figure 5: Scenarios' results.

ten users were considered. Users evaluate the results obtained by using the domain ontology, the individual fuzzy ontology and the updated individual fuzzy ontology.

Three scenarios were designed to evaluate the proposed framework:

- Scenario based on domain ontology (*Sc1*)
- Scenario based on individual fuzzy ontology, without update (*Sc2*)
- Scenario based individual fuzzy ontology, with 4 updates (*Sc3*)

Figure 5 shows the results in terms of precision (P) and recall (R), for Top 5, 10, 20, 30 and 50 retrieved documents. To evaluate the recall values, we consider the same queries used for the precision. We analysed the first 50 relevant URL's returned by every scenario. The obtained precision and recall related with FuzzOntoEnrichIR for scenario *Sc3* are clearly higher than the ones obtained by *Sc1* and *Sc2*. Indeed, results related to the comparison of exact precision obtained with FuzzOntoEnrichIR based on domain ontology only, FuzzOntoEnrichIR based on fuzzy ontology with and without update, present a precision of 0.95 which is superior to the two other scenarios. These results show that the use of fuzzy ontology supporting update process increases the precision more than the use of simple domain ontology.

To complete these results we computed the improvements given by Table 1, which show that individual fuzzy ontology updated four times is reported to achieve 18,75% precision and 37,50% recall. Indeed, the results show that the use of fuzzy ontology increased the precision more than the use of simple domain ontology.

In the experimentations, the initial used ontology and fuzzy ontology are composed of 8 concepts, enriched with 20 concepts after the first iteration, with

Table 1: The improvement of the average recall and the average precision of the FuzzEnrichOntoIR framework.

Precision/ Recall	Precision improvement TS3 vs. TS1 (in %)	Recall improvement TS3 vs. TS1 (in %)
P/R 05	18,75	37,50
P/R 10	16,25	35,29
P/R 20	14,28	25,00
P/R 30	12,00	22,44
P/R 50	07,50	17,50

20 concepts, enriched with 27 concepts after the second iteration, enriched with 40 concepts after the third iteration, and enriched with 100 concepts after the fourth iteration. From the obtained results after the experimentations, we note that the incremental enrichment of the domain ontology improves the relevance.

However, relevance becomes stable after the third iteration, when the size of the ontology is enough great to cover a unique complex query. Indeed, the 100 concepts don't cover the field of the same request, but they serve to other composed requests in the same domain. For this reason, the variance of the relevance of the first iteration of the enrichment has a remarkable impact on the relevance of improvement.

5 CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed a fuzzy-ontology-enrichment framework based on fuzzy ontology, namely FuzzOntoEnrichIR. Since ontologies have proven their capacity to improve IR, fuzzy ontology-based IR is becoming an increasing research area. FuzzOntoEnrichIR's framework takes place in four

main phases:

- Initialization of membership values,
- Updating the membership value of concepts and relations,
- Updating the membership value of the existing concepts in the user's query
- Updating the membership value of relations related to the existing concepts in the user's query.

Fuzzy ontologies building method is integrated to IR process, and returned results are classified taking into account fuzzified relations.

So, in this work, our first contribution concerns the fuzzy ontology's building process. Our method considers automatic fuzzification of a domain ontology taking into account both taxonomic and non taxonomic relations, however, all relations are important mainly in case of query reformulation.

Second contribution concerns the integration of our fuzzy ontology method into the IR process. Indeed, query reformulation is based on the weights associated to all the relations existing in the fuzzy ontology, and this fuzzy ontology is used to classify documents by services.

Finally, the obtained results establish the great interest and FuzzOntoEnrichIR's contribution to improve the performance of the retrieval task. Experiments and evaluations have been carried out, which highlight that overall achieved improvement are obtained thanks to the integration of fuzzy ontologies into IR process, integration of update and classification. These components contribute to significantly increase the relevance of search results, by enhancing documents ranking as shown by the obtained results.

As an evolution of this work, integration of modular ontologies in order to facilitate the updates is in progress. Otherwise, the ontology will be extended to different domains so that architecture will support a multi-domain use of the ontology. A multi-domain retrieval based on modular and fuzzy ontologies will be possible.

REFERENCES

- Akinribido, C. T., Afolabi, B. S., Akhigbe, B. I., and Udo, I. J. (2011). A fuzzy-ontology based information retrieval system for relevant feedback. In *International Journal of Computer Science Issues*.
- Baazaoui-Zghal, H., Aufaure, M.-A., and Mustapha, N. B. (2007a). Extraction of ontologies from web pages: Conceptual modelling and tourism application. *Journal of Internet Technology (JIT), Special Issue on Ontology Technology and Its Applications*, 8:410–421.
- Baazaoui-Zghal, H., Aufaure, M.-A., and Mustapha, N. B. (2007b). A model-driven approach of ontological components for on-line semantic web information retrieval. *Journal of Web Engineering*, 6(4):309–336.
- Baazaoui-Zghal, H., Aufaure, M.-A., and Soussi, R. (2008). Towards an on-line semantic information retrieval system based on fuzzy ontologies. *JDIM*, 6(5):375–385.
- Bordogna, G., Pagani, M., Pasi, G., and Psaila, G. (2009). Managing uncertainty in location-based queries. *Fuzzy Sets and Systems*, 160(15):2241–2252.
- Calegari, S. and Ciucci, D. (2006). Towards a fuzzy ontology definition and a fuzzy extension of an ontology editor. In *ICEIS (Selected Papers)*, pages 147–158.
- Chien, B.-C., Hu, C.-H., and Ju, M.-Y. (2010). Ontology-based information retrieval using fuzzy concept documentation. *Cybernetics and Systems*, 41(1):4–16.
- Colleoni, F., Calegari, S., Ciucci, D., and Dominoni, M. (2009). Ocean project a prototype of aiwbases based on fuzzy ontology. In *ISDA*, pages 944–949.
- Jiang, J. J. and Conrath, D. W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. *CoRR*, cmp-lg/9709008.
- Lee, C.-S., Jian, Z.-W., and Huang, L.-K. (2005). A fuzzy ontology and its application to news summarization. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 35(5):859–880.
- Lee, C.-W., Shih, C.-W., Day, M.-Y., Tsai, T.-H., Jiang, T.-J., Wu, C.-W., Sung, C.-L., Chen, Y.-R., Wu, S.-H., Hsu, and Wen-Lian. Asqa: Academia sinica question answering system for ntcir-5 clqa.
- McGuinness, D. L. (1998). Ontological issues for knowledge-enhanced search. In *Proceedings of Formal Ontology in Information Systems*.
- Miller, G. A. (1995). "wordnet: A lexical database for english". *Commun. ACM*, 38(11):39–41.
- Parry, D. (2006). Chapter 2 fuzzy ontologies for information retrieval on the {WWW}. In Sanchez, E., editor, *Fuzzy Logic and the Semantic Web*, volume 1 of *Capturing Intelligence*, pages 21 – 48. Elsevier.
- Quan, T. T., Hui, S. C., Fong, A. C. M., and Cao, T. H. (2006). Automatic fuzzy ontology generation for semantic web. *IEEE Trans. Knowl. Data Eng.*, 18(6):842–856.
- Sayed, A. E., Hacid, H., and Zighed, D. A. (2007). Using semantic distance in a content-based heterogeneous information retrieval system. In *MCD*, pages 224–237.
- Seco, N., Veale, T., and Hayes, J. (2004). An intrinsic information content metric for semantic similarity in wordnet. In *ECAI*, pages 1089–1090.
- Widyantoro, D. and Yen, J. (2001). A fuzzy ontology-based abstract search engine and its user studies. In *Fuzzy Systems, 2001. The 10th IEEE International Conference on*, volume 3, pages 1291–1294.
- Zhou, L., Zhang, L., Chen, J., Xie, Q., Ding, Q., and Sun, Z. X. (2006). The application of fuzzy ontology in design management. In *IC-AI*, pages 278–282.