Applied Mechanics and Materials Vols. 519-520 (2014) pp 853-856
Online available since 2014/Feb/06 at www.scientific.net

doi:10.4028/www.scientific.net/AMM.519-520.853

An Enhanced Fuzzy Information Retrieval Model based on Linguistics

Zeinab E. Al-Arab[a], Ahmed M. Gadallah[b] and Hesham M. Hefny[c]

[a]eng.zeinabezz@gmail.com, [b]ahmgad10@yahoo.com, [c] hehefny@ieee.org

Keywords: Fuzzy Ontology, Fuzzy Ontology Information retrieval, Information Retrieval, Ranking documents, Linguistic based query answering system.

Abstract. The paper proposes a linguistic based fuzzy ontology information retrieval model. The model deals with linguistic based queries in multi domains. Such linguistics are user defined, reflecting his subjective view. The model also proposes a ranking algorithm that ranks the set of relevant documents according to some criteria such as their relevance degree, confidence degree, and updating degree.

Introduction

An information retrieval system (IR) consists of a document collection, a user query, a retrieval engine, and a ranking module. It stores and annotates documents such that when users express their information needs in a query, the ranking module will show a set of ranked relevant documents. This set of documents is retrieved by the retrieval engine associating a score to each one. The higher the score is, the greater the document relevance [1]. So, the challenge in IR is to find a number of the most relevant documents according to user's query.

Researchers deal with this challenge using two different approaches. These approaches are keyword based approach and concept based approach. In the keyword based approach, documents are returned when they are annotated by terms specified in the searching query. However, this approach neglects many related documents that are not annotated with the query terms [1]. In the concept based approach, documents are returned according to their relevance to the searching query. This approach is a domain specific approach. It can be classified into ontology based approach and fuzzy ontology based approach. Many works are done using these two approaches which are known as Information Retrieval model, IR model.

The paper proposes a linguistic based Fuzzy Ontology Information Retrieval model. The rest of the paper is organized as follow; section 2 presents fuzzy ontology based Information Retrieval. Some related works are showed in section 3. Section 4 presents the proposed Linguistic based Fuzzy Ontology Information Retrieval model. The paper is concluded in section5.

Fuzzy Ontology Based Information Retrieval

Fuzzy ontology represents uncertain information commonly found in many application domains in a human understandable, machine readable format [2]. It is used as a standard knowledge representation for the semantic web [3].

Fuzzy Ontology based Information Retrieval System, FOIR, is an IR system that semantically retrieves a set of relevant documents with respect to a certain query in a specific domain. This domain is represented using a fuzzy ontology. This fuzzy ontology is used to expand the user query and to annotate the document collection with a set of weighted keywords [4, 5, 1].

Related work

Leite model [6] semantically retrieves a set of query's relevant documents in multi-domains. It uses fuzzy ontology to represent these domains. It deals with crisp queries connecting multidomains.

Fuzzy Relational Ontology Model, FROM, [4] semantically retrieves a set of relevant documents with respect to a user query in a specific domain. It represents this domain using fuzzy ontology. It

considers fuzzy ontology as a set of concepts, terms, and relations between concepts and terms. FROM deals with crisp queries.

Fernández model [7] proposed an ontology based information retrieval model. This model deals with crisp queries in open environment.

Unfortunately, none of these models can deal with lingiuistic based queries. They only handle crisp ones. The proposed fuzzy ontology IR can deal with both linguistic based and crisp user queries.

The proposed Linguistic based Fuzzy Ontology Information Retrieval model

The proposed model is a linguistic based fuzzy semantic document retrieval model that uses a fuzzy ontology. It semantically retrieves relevant documents in multi-domains according to a user's linguistic based query. The proposed model's main features are:

- Build a subjective fuzzy ontology model that describes a specific domain in a more generalized and flexible manner with a linguistic based querying system.
- Allow users to express their needs about a certain information using linguistic terms, e.g., select all papers that are *very related* to *data mining*. This will give users flexibility in representing their needs.
- Allow users to define their subjective profile about linguistic terms.
- Retrieve relevant documents semantically in multi-domains according to each user subjective view.
- Rank the resulted relevant documents according to some criteria.

1.1.The proposed Information Retrieval Structure

The proposed information retrieval model's main components are a set of annotated documents, users' profiles, users' queries, retrieval engine, and ranking module. It depends mainly on fuzzy ontology methodology and a stemmer NLP tool.

Figure1 shows the structure of the proposed model. Firstly, each user should create a profile to define all his linguistic terms. Now, the user can build his query. This query is a set of keywords each is associated with its importance degree. This importance degree is expressed in linguistic terms. For example, select all papers that are very relevant to data mining, here the user searches for papers that are *very related* (linguistic term) to the keyword *data mining* (keyword).This query is then passed on some operations, which are:

- Interpreting each linguistic term according to the user's subjective view,
- Expanding each keyword with its related keywords using the predefined fuzzy ontology.

Then, this expanded list enters the retrieval phase that semantically retrieves a set of matched documents each associated with a matching degree. This set is then ranked according to some criteria using the proposed ranking algorithm. Finally, the ranked relevant set of documents is displayed to the user.

1.2.The Proposed model Phases

The proposed model phases are as follows:

1.2.1. User profile creation

User should create an account before building his/her query. Using this account, he/she can define any linguistic term according to his subjective view. Figure 2 shows the scheme of storing users' linguistic terms definitions. When a certain user creates an account, this account is stored in the *Users* table. Then user can define any linguistic term, e.g., "related" is a linguistic term, using the *userLinguisticTermFunction* table. This table specifies which membership function is used to define a certain linguistic term according to the user subjective view. This membership function is also defined according to the user's subjective view and stored in a table correspond to its name, e.g., *triangularUserLinguisticTerms* table for triangular membership function, *piUserLinguisticTerms*

table for pi membership function. Hedges, e.g., very, more or less are hedges, can also be defined according to user's subjective view and stored in *userHedgeDefinition* table. In *userHedgeDefinition* table user specifies the hedge name and its power. Also, users have

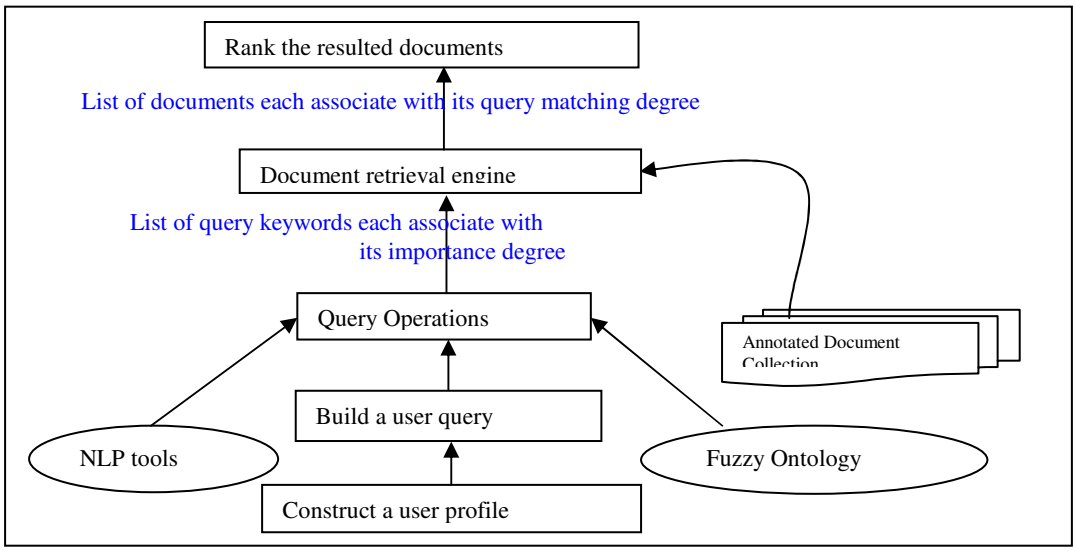

Figure 1: The proposed model phases

the ability to specify which method is used to interpret a conjunctive or disjunctive query, by determining the conjunctive and disjunctive methods and storing them in *userConjunctiveDisjunctiveMethod* table.

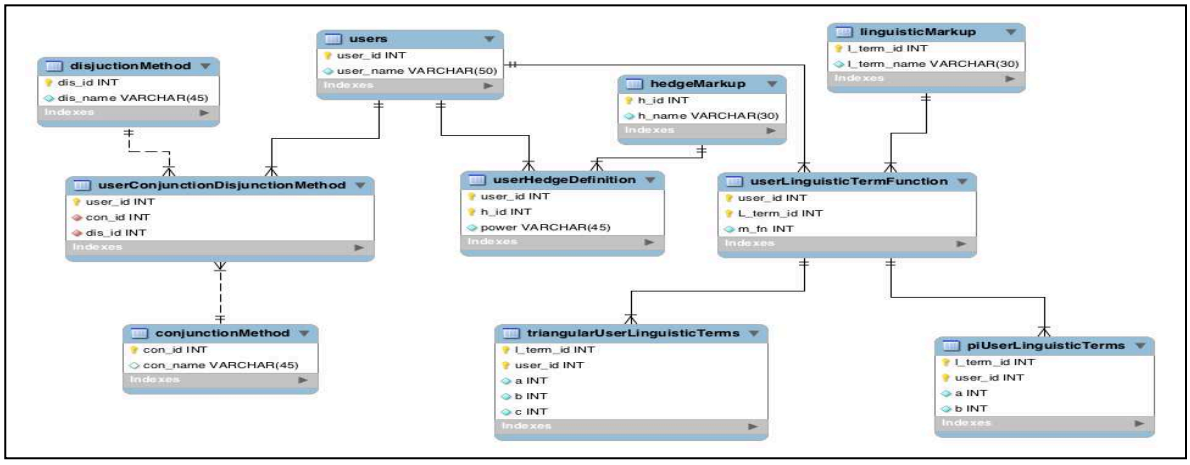

Figure 2: Storing a user linguistic terms definition

1.2.2. Constructing a linguistic based query

Now, the user can build his query. This query can be crisp, fuzzy, or linguistic based-query. For example, the query statement, "select all data mining papers" represents a crisp query. On the other hand, the query statement, "select all data mining papers with membership degree 0.6" and "select all data mining papers with membership degree around 0.6" are examples of the fuzzy query. A linguistic based query may be like:

select all papers very related to data mining

where "*data mining"* is the keyword that the user searches for. "*very related"* is a user linguistic term that reflects his needs for the keyword "data mining". This linguistic term is previously defined by the user according to his subjective view and stored in his account.

1.2.3. Applying the Query Operations

After user submits his query, some operations are performed on it. First the query is parsed, such that each searched keyword is extracted with its importance degree that is expressed using linguistic terms and hedges. All linguistic terms and hedges are then interpreted according to the user's subjective view. All keywords are then expanded using the domain fuzzy ontology. The importance

degree of any expanded word is the product of its relation with the original keyword, that is extracted from the fuzzy ontology, and the importance degree of the original keyword, defined by the user in his query.

1.2.4.Retrieving a set of relevant documents

It semantically retrieves a set of relevant documents with respect to a certain user query through calculating document matching degree . A document matching degree is calculated as the max product composition between the list of weighted keywords that annotate this document and the list of query's weighted expanded keywords.

The result of this is a list of semantically relevant documents each associated with its matching degree.

1.2.5. Ranking the resulted documents

It ranks the resulted semantically relevant documents from the retrieval phase based on some criteria:

- The document's matching degree with user needs. The higher the matching degree is, the more document relevance with respect to user's needs.
- The document's confidence degree. This degree is extracted from the document's authors, the confidence degree of the journal, or conference that the document is published in. This factor reflects to what extent does the knowledge in this document is trusted. The higher the journal impact degree is, the more confidence that the knowledge in this document is correct,
- The document's updating degree. This degree is extracted from the document publishing date. This factor reflects to what extent does the knowledge in this document is new and updated, not out of date.

The resulted ranked list of relevant documents is then displayed to the user in the same order.

Conclusion and Future work

This work presents an improvement in the fuzzy semantic information retrieval through:

- Building a Linguistic based query system. This gives users more flexibility while building their queries.
- Allow users to define all their linguistic terms according to their subjective view. This helps in retrieving documents according to their linguistic terms definitions not to our definitions.
- The resulted set of documents are ranked according to some criteria which are their relevance degree with respect to user's query, confidence degree and updating degree.

The future direction to work in this area would be build a document annotation algorithm.

References

[1] M. A. A. Leite and I. L. M. Ricarte," A Framework for Information Retrieval Based on Fuzzy Relations and Multiple Ontologies," Springer, pp. 292-301, 2008.

[2] J. Zhai, M. Li, and J. Li, "Semantic Information Retrieval Based on RDF and Fuzzy Ontology for University Scientific Research Management," Affective Computing and Intelligent Interaction, AISC 137, pp. 661–668, 2012.

[3] Q. T. Tho, S. C. Hui, A. C. M. Fong, T. H. Cao," Automatic Fuzzy Ontology Generation for Semantic Web," IEEE transaction on knowledge and data engineering, vol. 18, No.6, June 2006.

[4] R. Pereira, I. Ricarte, F. Gomide, " Information Retrieval with FROM: The Fuzzy Relational Ontological Model," International Journal Of Intellegent Systems, vol. 24, 340-356, 2009.

[5] A. Nawaz and A. Khanum," Ranked Neuro Fuzzy Inference System (RNFIS) for Information Retrieval," Springer, ISNN 2011, Part I, LNCS 6675, pp. 578–586, 2011.

[6] M. A. A. Leite, I. L. M. Ricarte, "Relating ontologies with a fuzzy information model," KnowlInfSyst, pp. 619-651, 2013.

[7] M. Fernández, I. Cantador , V. López, D. Vallet, P. Castells, E. Motta ," Semantically enhanced Information Retrieval: An ontology-based approach," Web Semantics: Science, Services and Agents on the World Wide Web 9 ,pp. 432-452, 2011.

Computer and Information Technology
10.4028/www.scientific.net/AMM.519-520

An Enhanced Fuzzy Information Retrieval Model Based on Linguistics
10.4028/www.scientific.net/AMM.519-520.853