Handling uncertainty in semantic information retrieval process

Chkiwa Mounira[1], Jedidi Anis[1] and Faiez Gargouri[1]

[1]Multimedia, InfoRmation systems and Advanced Computing Laboratory
Sfax University, Tunisia

m.chkiwa@gmail.com, jedidianis@gmail.com, faiez.gargouri@isimsf.rnu.tn

Abstract. This position paper proposes a collaboration method between Semantic Web and Fuzzy Logic aiming to handle uncertainty in the information retrieval process in order to cover more relevant items in result of search process. The collaboration method employs OWL ontology in query enhancement, RDF in annotation process and fuzzy rules in ranking enhancement.

1 Introduction

In the information retrieval process, there are returned documents which are relevant to the query but they focus in addition of query main interest on others additional topics. To deal with this imprecision we propose to valorize in the ranking process relevant documents which deal mainly with query themes. Another source of imprecision in the search process is the user queries; we propose to enhance it in order to come near the intention of the user. This paper is organized as follows: in the next section we present our proposition to enhance the query background expression then we explain how Semantic Web and Fuzzy Logic collaborate to enhance ranking process. In Section 3, we present some related works and Section 4 concludes the paper.

2 Handling uncertainty by semantic/fuzzy collaboration

2.1 The semantic/fuzzy query enhancement

A main cause of uncertainty in the information retrieval process comes from the user's queries. In order to return more relevant results, the information retrieval system has to indentify the user's intention behind the query. To do it, we propose to enhance user queries by adding semantically related terms. In this purpose, we use the Web Ontology Language OWL and then we employ some fuzzy rules in order to weight up the query terms importance. In Figure 1, we present our semantic/fuzzy query enhancement.

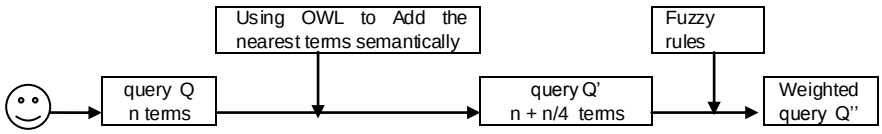

Fig. 1. Semantic/fuzzy query enhancement

The Semantic query enhancement passes through the enrichment of the query by new terms syntactically different but semantically near; the new added terms are not picked to derive the query meaning but to find terms expressing more the user intention. Several works as [1- 3] are proposed to express the semantic similarity between ontology concepts. After eliminating empty terms from the query, we can reuse the algorithm presented in [3] to find the semantically nearest term to each query term using OWL ontology. The number of added terms must not be constant; it can derive the query meaning if it is large or useless if it is few. So we decide that the number of added terms be proportional to the query length. Hence, we propose to add only n/4 terms having the highest similarity to query terms. Also, we propose that the information retrieval system is interactive and allows users to highlight certain query terms in order to reflect their importance. Finally, to weight the query terms, we apply some fuzzy rules; those rules define the priority of weighting:

- If a query term is added from the ontology then it will has low weight priority.
- If a query term is not bold, then it will has a medium weight priority.
- If a query term is bold then it will has a high weight priority.

2.2 The semantic/fuzzy ranking enhancement

The semantic/fuzzy ranking enhancement aims to manage uncertainty about the output of classic querying process and to valorize documents focusing specially in the same user query interests. It aims principally to limit the number of relevant documents dealing with several topics. The semantic/fuzzy ranking enhancement passes through two fundamental concepts: the "meta-document" which allows annotating semantically the collection of documents and the "themes clouds" which enhance the ranking process based on Fuzzy Logic. The meta-document is introduced in [4] and it is able to annotate semantically multimedia objects as well as web documents. A meta-document uses RDF metadata to annotate web resources in a way that ensures its reusability. The querying process matches the user query with the meta-documents in order to identify the score relevance of the resources to the query. We define the "theme cloud" as groups of weighted terms concerning a given theme. Simply, we collect potential terms representing a given theme to construct a theme cloud. The terms' weights express the ability of each term to represent the theme. After running a usual querying process matching the query and the meta-documents, we get the relevance score for each annotated resource or document. At this point, the theme clouds are used to enhance ranking results in the benefit of relevant documents focusing

mainly on query interests. The Figure 2 gives a simple presentation of the structure of the semantic/fuzzy ranking enhancement:

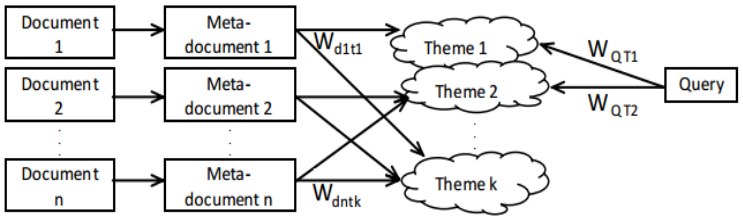

Fig. 2. The semantic/fuzzy ranking enhancement

To run the ranking enhancement, first, we establish the meta-document/theme weighted links $W_{DT.}$ W_{DT} expresses the potential themes mentioned by the meta-document. To assign a weight W_{DT} to a meta-document/theme link, we simply sum the weights of theme terms existing in the meta-document. Then we establish query/themes weighted links which express the ability of each theme to represent the query. To assign a weight to a query/theme link, we use the classic similarity measure between two weighted terms vectors:

$$W_{QT} = sim(Q,Ti) = \frac{\sum_{j=1}^{t} W_{qj} * W_{t_{ij}}}{\sqrt{\sum_{j=1}^{t} (W_{qj})^2 * \sum_{j=1}^{t} (W_{t_{ij}})^2}} \quad (1)$$

The next step of ranking enhancement is to calculate for each document his theme similarity with the query in order to increase or decrease its relevance score in terms of the value of the theme similarity. The theme similarity TS is calculated as follow:

$$TS\ (Q,D) = \sum_{i=1}^{k} \left| W_{QT_i} - W_{DT_i} \right| \quad (2)$$

The main goal of the ranking enhancement is to increase relevance of documents focusing on the same query themes and to decrease relevance of document dealing with different themes vis-à-vis the query. The TS (Q, D) value is optimal when its value is minimal; this means that the query and the document are focusing on the same themes with approached values. Contrariwise, if the TS is high, this means that the document deals with other themes in addition to the query themes. Finally, the increase or the decrease Rate R affected to a document Relevance Score RS is based on the following fuzzy rules:

- If RS is high or medium and TS is low then R is high
- If RS is low and TS is low then R is medium
- If RS is low or medium and TS is high then R is negative

3 Related work

Several approaches considering both uncertainty and the Semantic Web have been proposed in the information retrieval issue. [5, 6] propose to fuzzify in different ways RDF triples, likewise [7, 8] propose to fuzzify OWL ontology statements. A common point in those works is the use of formal ways to express the assignment of a truth degree to RDF triples or OWL axioms. In our proposition, numerical membership values identification is done in background using mathematical deduction without the need of formal expressions (e.g. weight priority of a query term). Some other works are near our proposition: [9] shows that it is useful to express a fuzzy proximity values between terms of a query. By using a fuzzy set of rules [10] shows the usability of a ranking system based on fuzzy inference. In the query enhancement issue, many works are proposed [11-12]; our method is characterized by its simplicity and flexibility.

4 Conclusion

In this paper we studied two interoperable axes in the information retrieval process: the Semantic Web and the Fuzzy Logic. We propose to enhance query background expression and also to enhance ranking process using fuzzy rules. Given that Numerical inputs of fuzzy rules are deduced from the meta-documents characteristics, it remains to identify in the short run, the numerical limits to fuzzy sets on which we will apply the fuzzy rules set. Equally, we plan to extend the current proposition and to investigate the concept of user profile in order to cover more relevant result document.

5 References

1. B. Y. Liang, J. Tang, J. and Z. Li. Semantic Similarity Based Ontology Cache. Heidelberg: Springer-Verlag, 2006, pp.250-262.
2. Z. Yang. Semantic similarity measure matching between ontology concepts based on heuristic rules. Journal of computer applications, vol.27, no.12, pp.2919-2921, 2007.
3. Xiao Min; Zhong Luo; Xiong Qianxing. Semantic Similarity between Concepts Based on OWL Ontologies. 2nd International Workshop on Knowledge Discovery and Data Mining, 2009. WKDD 2009. vol., no., pp.749,752, 23-25 Jan. 2009.
4. Jedidi A. « modélisation générique de documents multimédia par des métadonnées : mécanismes d'annotation et d'interrogation » Thesis of « Université TOULOUSE III Paul Sabatier », France. July 2005.
5. Simou, N., Stoilos, G., Tzouvaras, V., Stamou, G., and Kollias ,S. Storing and Querying Fuzzy Knowledge in the Semantic Web. In Proceedings of the Fourth International Workshop on Uncertainty Reasoning for the Semantic Web, Karlsruhe, Germany,, 2008
6. Mazzieri, M., and Dragoni, A. F.. A fuzzy semantics for Semantic Web languages. In ISWC-URSW, pages 12-22. 2005

7. Calegari, S. and Ciucci, D. Fuzzy Ontology, Fuzzy Description Logics and Fuzzy-OWL. WILF '07 Proceedings of the 7th international workshop on Fuzzy Logic and Applications: Applications of Fuzzy Sets Theory Pages 118-126. 2007.
8. Bobillo F., Straccia U. Fuzzy Ontology Representation using OWL 2. International Journal of Approximate Reasoning 52(7):1073-1094, 2011.
9. Beigbeder, M. and Mercier, A. Application de la logique floue à un modèle de recherche d'information basé sur la proximité. In Actes 12th rencontres francophones sur la Logique Floue et ses Applications. Nantes, France, pp. 231-237. 2004.
10. Rubens, N. The application of Fuzzy Logic to the construction of the ranking function of information retrieval systems. Computer Modeling and New Technologies, Vol.10, No.1, 20-27. 2006.
11. Neda A., Latifur K. and Bhavani T. Optmized ontology-driven query expansion using map-reduce framework to facilitate federated queries. Comput. Syst. Sci. Eng. 27(2) (2012)
12. Min S, Il-Yeol S, Xiaohua H, Robert B. Allen. Integration of association rules and ontologies for semantic query expansion. Data Knowl. Eng. 63(1): 63-75 (2007).