Enhancing semantic search using case-based modular ontology

Nesrine Ben Mustapha, Hajer Baazaoui, Marie-Aude Aufaure, Henda Ben Ghezala

Enhancing Semantic Search using Case-Based Modular Ontology

Nesrine Ben Mustapha
Laboratoire RIADI ENSI
Campus Universitaire de la Manouba 2010

nesrine.benmustapha
@riadi.rnu.tn

Hajer Baazaoui Zghal
Laboratoire RIADI ENSI
Campus Universitaire de la Manouba 2010

hajer.baazaouizghal
@riadi.rnu.tn

Marie-Aude Aufaure
Ecole Centrale Paris
Laboratoire MAS Chaire
SAP Business Objects
Grande Voie des Vignes
92 295 Chatenay-
Malabrymarie-
aude.aufaure@ecp.fr

Henda ben ghezala
Laboratoire RIADI ENSI
Campus Universitaire de la Manouba 2010

Henda.benghezala@riadi.rn
u.tn

ABSTRACT

In this paper, we present a semantic search approach based on Case-based modular Ontology. Our work aims to improve ontology-based information retrieval by the integration of the traditional information retrieval, the use of ontology and the case based reasoning (CBR). In fact, our recommender approach uses the CBR with semantic Web language markup -by ontology- for case representation and indexing. Ontology-based similarity is used to retrieve similar cases and to provide end users with alternative recommendations. The main contribution of this work is the use of a CBR mechanism and an ontological representation for two purposes: Resource Retrieval from Web and ontology enrichment from cases.

Categories and Subject Descriptors

I.2.4 [**Knowledge Representation Formalisms and Methods**]- *Representation languages*

General Terms

Design, Theory.

Keywords

Semantic Web, ontology, modular ontology, semantic search, information retrieval, case-based reasoning.

1. INTRODUCTION

Over the past few years and with the continuous and rapid growth of Web information volume, information access and knowledge management has become challenging. Thus, adding a semantic dimension to the Web, with ontology deployment, contributes to solve many problems in many domains (Information retrieval, knowledge sharing, communication between Web agents, etc.). Besides, the problem of query formulation is becoming a challenging point. A good query formulation must include the necessary features to retrieve the relevant information, which is not an evident task and especially on the first attempt [2]. Thus, the use of an iterative a process of trial and error is necessary to improve query formulation. Therefore, using a Case-Based reasoning (CBR) model in an IR process is an ambitious area. In fact, CBR is a problem-solving method [1] based on the concept of case which is the description of a problem and its solution. The main idea under CBR consists in storing experiences as cases and problem-solving processes as instances of cases. When a new problem is encountered, the system uses the relevant past stored cases to interpret or to solve it [2]. Since intelligent retrieval is

one of the main application of Case-base reasoning paradigm (CBR) [1,2,6], semantic formalization in CBR systems has also become an increased research area [4]. In CBR systems, semantics are the main source of reasoning, similarity calculation and case adaptation. We describe in this paper our approach which is based on the composition of an enhancing semantic search approach based on case-based modular ontology. We propose to combine a CBR mechanism with multi-domain modular ontologies which are classified by topic in order to ameliorate query formulation, dynamic semantic indexing and eventually, the results precision.

2. ENHANCING SEMANTIC SEARCH APPROACH USING CASE-BASED MODULAR ONTOLOGY
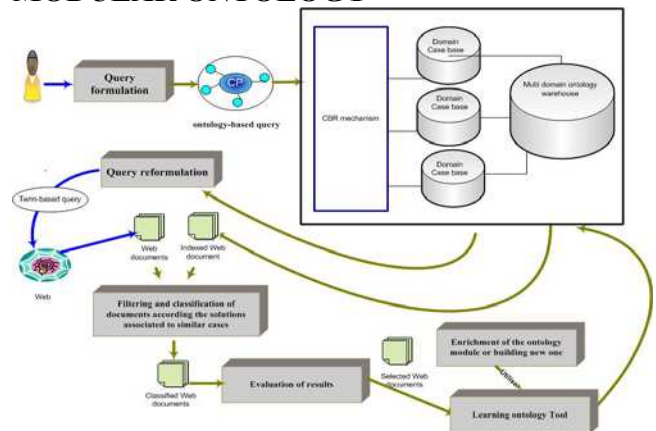

Figure 1. Approach combining modular ontologies with CBR.

The main components of the proposed approach are (figure 1): An iterative IR process, a multi-layer ontology-warehouse for indexing Web resources and domain cases, a CBR mechanism using in the IR process. The idea behind this approach is (1) to use the previously discovered ontology-aided semantic metadata representation in OWL, and (2) to ask the user about resource characteristics and (3) to respond to queries with ranked cases. At each step, a ranked set of recommended queries related to previous similar cases is provided. After the case selection, new documents are imported and classified by using the relevance feedback [5]. This process narrows down results and incrementally eliminates cases which are proven to be irrelevant and enrich the cases base with new ones. Modular ontologies used

for the domain case representation and indexing are also incrementally enriched using text mining techniques. These ontologies are designed according to a previous work [3] where a multi-layer ontological warehouse was designed to annotate discovered resources.

3. CASE STUDY AND EVALUATION

In this section, tow use cases related to tow users that submit respectively two queries *Q1* and *Q2* that appear as two independent requests, but according to domain knowledge, the results of one is the answer to the other. The query 〈Q1〉 is **What is the BMI in medicine?**. Q2 is 〈**how to measure the nutritional disorder?** 〉 We suppose that the first user submits the query 〈Q1〉and no related ontology module is found in the ontology warehouse and it is not possible to find similar cases as the Query keywords are not clear. The user selects the relevant documents to be saved and a new case 〈C1〉is inserted in the case base (Table 1). The explication of this common result is described as following; the ontology module indexing the case C1 is strongly related to the ontology module indexing the case C2 (figure 2) because BMI measures the percentage of body fat, it may be a useful tool to estimate a healthy body weight and a high BMI is linked to nutritional disorder. The relation between the two ontology modules was discovered in the step of ontology enrichment of the case C1. So, BMI is one of the measurements of nutritional problem. There are others answers such as 〈serum albumin measurement〉 (C2). Then, we apply the documents filtering by using the proposed approach. We remark that majority of removed Web documents include the term BMI, but they have not the same mining of 〈body mass index〉 In fact, there are other senses of BMI such as (〈BMI as Commercial noun of the British Midland Airways〉and BMI as Web site that collects license fees on behalf of songwriters and composers). The system supporting the proposed approach is developed using the service 〈GOOGLE_API〉 and 〈JENA〉 library.
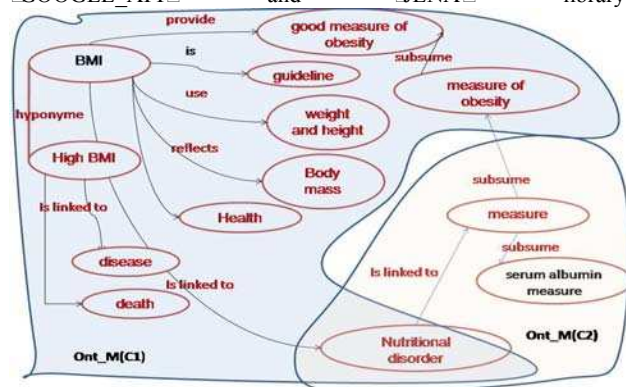

Figure 2. Relation between two ontology module indexing respectively two similar cases.

To evaluate the proposed approach, we define three scenarios to be compared. The first scenario presents search results from Google. The second scenario represents the results delivered by the proposed approach and where there are 〈nearly similar〉cases in the case bases. Otherwise, no 〈strongly similar cases〉 are founded. The research is based on the vector model to import similar results and new cases are added. The third scenario is to

search for information based on the selected 〈strongly similar Cases〉 the reformulation of the request and utilization of vector model for classification. The results are shown in Figure 3.
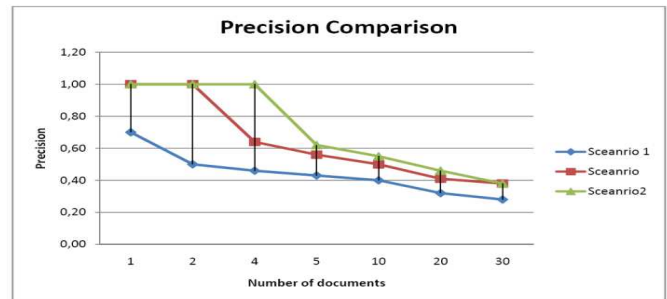

Figure 3. Precision comparison

4. CONCLUSION and FUTURE WORK

In this paper, we discussed an enhanced semantic search based on case-based modular ontology, by which the traditional information retrieval, ontology and CBR can be integrated. Our recommender approach uses CBR, with semantic Web language markup (ontology) for case indexing. Ontology-based similarity is used to retrieve similar cases and provide the users with alternative recommendations. Case base are also used in ontology enrichment. Our observations can prove that combining Ontology technology and CBR has a positive impact on search results and the more cases are stored the better the system performs. Our future Work will deal with further evaluation in the case of search on large digital libraries taking in account the use of social networks.

5. REFERENCES

[1] Aamodt, A., and Plaza, E. Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications 7*, 1 (1994), 39〈59.

[2] Aktas, M., Pierce, M., Fox, G., and Leake, D. A web based conversational case-based recommender system for ontology aided metadata discovery. In *Proceedings of the Fifth IEEE/ACM International Workshop on Grid Computing (GRID 2004)* (2004), IEEE Computer Society Press.

[3] Ben Mustapha, N., Baazaoui Zghal, H., Aufaure, MA. and Ben Ghezala, H. Combining Semantic Search and Ontology Learning for Incremental Web Ontology Engineering, International Workshop on Web Information Systems Modeling (WISM 2009) (Held in conjunction with CAiSE 2009), Amsterdam, juin, 2009.

[4] Gao, J., and Deng, G. The research of applying domain ontology to case-based reasoning system. In *Proceedings of International Conference on Services Systems and Services Management* (2005), pp. pp.1113 〈1117.

[5] Rocchio, J. J. Relevance feedback in information retrieval. In *The Smart retrieval system - experiments in automatic document processing*, G. Salton, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1971, pp. 313〈323.

[6] Vehvilainen A., Olli Alm, E. H. Combining case-based reasoning and semantic indexing in a question-answer service. In *1st Asian Semantic Web Conference* (2006).