

Ontology learning from Web: survey and framework based on semantic search

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

▶ **To cite this version:**

Nesrine Ben Mustapha, Hajer Baazaoui, Marie-Aude Aufaure, Henda Ben Ghezala. Ontology learning from Web: survey and framework based on semantic search. Second International Conference on Web and Information Technologies(ICWIT'09), Jun 2009, Tunisia. pp.217-223. hal-00831683

HAL Id: hal-00831683

https://centralesupelec.hal.science/hal-00831683

Submitted on 7 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ontology learning from Web

Survey and framework based on semantic search

Nesrine BEN MUSTAPHA*, Hajer Baazaoui Zghal *,** Marie-Aude Aufaure** Henda Ben Ghezala***

** Laboratoire RIADI ENSI,*
Ecole Nationale des Sciences de l'Informatique
Campus Universitaire de La Manouba 2010
{nesrine.benmustapha, hajer.baazaouizghal, henda.benghezala }@riadi.rnu.tn

*** Ecole Centrale Paris*
Laboratoire MAS Chaire
SAP Business Objects Grande Voie des Vignes
92295 Chatenay-Malabry
marie- aude.aufaure@inria.fr

RÉSUMÉ. *L'ajout d'une dimension sémantique sur le Web, par le déploiement des ontologies, contribue à résoudre de nombreux problèmes. Les ontologies permet l'amélioration de l'exploitation des ressources Web par l'ajout des connaissances consensuelles. L'utilisation des ontologies de domaine pour la recherche d'information (RI) a été explorée par plusieurs travaux afin de mieux répondre aux requêtes des utilisateurs. Toutefois, les systèmes d'RI sont généralement basées sur peu de l'ontologie de domaine qui ne peut pas être prorogé. Les ontologies dans les systèmes d'RI nécessitent, alors, une mise à jour régulière, en particulier l'ajout de nouveaux concepts et de. Ce papier propose une étude de plusieurs approches de l'apprentissage de l'ontologie à partir du Web afin de proposer la collaboration du processus de recherche sémantique et le processus d'apprentissage d'ontologies peuvent collaborer pour construire les ontologies dans un moteur de recherche sémantique multi-domaine en utilisant le raisonnement à base de cas.*

ABSTRACT. *Adding a semantic dimension to the Web, by the deployment of ontologies, contributes to solve many problems. Ontologies improve the exploitation of Web resources by adding a consensual field of knowledge. The need for using domain ontology for information retrieval (IR) has been explored by some approaches to better answer users' queries. However, IR systems are generally based on few number of domain ontology that cannot be extended. Then, ontology in IR system requires a regular updating, especially the addition of new concepts and relationships. This paper presents a survey of several approaches of ontology learning from Web. Collaboration between the processes of semantic search and ontology learning is proposed for multi-domain search engine using case-based reasoning.*

MOTS-CLÉS: *Web sémantique, apprentissage d'ontologies, recherché sémantique, raisonnement à base de cas.*

KEYWORDS: *Semantic Web, Ontology learning, semantic search, case-based reasoning.*

1. Introduction

Adding a semantic dimension to the Web (Berners-Lee and *al.2001)*, by the deployment of ontologies, contributes to solve many problems: knowledge sharing, semantic access to Web resources and information retrieval. In fact, ontologies improve the exploitation of Web resources by adding a consensual knowledge. The need for using domain ontology for information retrieval (IR) has been explored by some approaches to better answer of users queries. However, ontologies in IR system cannot be extended. During this last decade, several approaches of ontology learning have appeared and proposed a partial automatization of knowledge acquisition from structural, semi structural or unstructured data sources (data base, knowledge base, texts, etc.). In this paper, knowing that a unique data source cannot cover all concepts of a target domain of knowledge and that Web is a rich textual source, we have chosen to consider the Web as learning corpus from which domain ontologies are extracted. These ontologies will be used in semantic search systems. The main objective of this work is to make the semantic search engine more flexible and autonomous to construct their domain ontologies from relevant documents in an incremental manner. Then, we choose to combine ontology learning from text and semantic search technology to propose a domain-independent approach to automate ontology learning from Web documents.

This paper is organized as follows. Section 2 presents the related work to ontology learning approaches from Web. Limits and open issues are presented later. In section 3, a semantic search framework is described with an illustration scenario. Finally, we conclude and give some perspectives for this research work.

2. Ontology building from Web

Ontology learning (OL) is defined as an approach of ontology building from knowledge sources using a set of machine learning techniques and knowledge acquisition methods. OL from texts is a specific case of OL from Web and has been widely used in the community of engineering knowledge since texts are semantically richer than the other data source type. These approaches are generally based on the use of textual corpora. This one should be a representative of the domain for what we are trying to build ontology. By applying a set of text mining techniques, granular ontology is enriched with concepts and relationships discovered from textual data. In such approach, human intervention is required to validate the relevance of learned concepts and relationships. In the last decade, with the enormous growth of Web information, Web has become as important source of information for knowledge acquisition: due to its huge size and heterogeneity. This has been the cause of mainly many categories of OL approaches: ontology learning from textual content of the Web, ontology learning from online Web ontologies, from web dictionary and from Web heterogeneous sources.

2.1. *Ontology learning approaches from Web documents*

OL from Web documents require the same techniques used before for ontology extraction from texts. Several approaches are based on eliminating tags from documents to obtain plain texts on which traditional text mining texts could be applied. We propose to classify these approaches to domain-dependant OL and incremental OL.

2.1.1. *Domain- Dependent approach for Ontology learning from textual documents*

OL approaches from Web content consists generally in enriching a small ontology called "minimal" or "granular" with new concepts and new relationships using text mining techniques. Learning ontologies from texts has been widely used in the community of knowledge engineering. This is in particular the work of: (Aguirre and *al.*, 2000) (Alfonseca and *al.*, 2002) (Aussenac-Gilles and Jacques, 2006) (Aussebac-Gilles and *al.*, 2003) (Bachimont, 2000) (Faatz and Steinmetz, 2002) (Hahn and Marko, 2001) (Hearst, 1998) (hwang, 1999) (khan and Luo, 2002) (Kietz and *al.*, 2000) (Lonsdale and *al.*, 2002) (Moldovan and Girju, 2000) (Nobécourt, 2000) (Roux and *al.*, 2000). However, no sufficiently detailed methodology has been presented to assist the learning process ontology. Indeed, the literature is limited to the presentation of guidelines more or less general. Thus, for each approach, it is important to know the aims and scope of the learning process, its main stages, the sources of knowledge used in learning, the main techniques applied in the process, re-usability of ontologies existing and the study of its feasibility. These approaches to ontology learning from text are generally based on the use of a corpus of texts. Using a set of techniques, we try to project in the ontology knowledge contained in texts by extracting concepts and relations. We distinguish mainly five categories of text mining techniques: linguistic techniques (Aussebac-Gilles and *al.*, 2003) and lexico-syntactic patterns (Hearst, 1998), clustering techniques and / or classification techniques (Aguirre and *al.*, 2000) (Alfonseca and *al.*, 2002) (Lin and hovy, 2000) (Faure and *al.*, 1998), statistical techniques (Resnik, 1995) (Sekiuchi and *al.*, 1998) (Sugiura and *al.*, 2003) Verladi and *al.*, 2002), association rule (Maedche and Staab, 2000), and hybrid ones. Besides of ontology learning from texts, ontology learning from Web documents appears to be a second category in domain-dependant OL. The most known approaches exploit the textual Web content to enrich concepts using Wordnet. Several approaches described in (Aguirre and *al.*, 2000) and (Faatz and Steinmetz, 2002) enrich ontologies from Web documents. Another approach is proposed in order to reduce the terminological and conceptual ambiguity among members of a virtual community. This approach proposes the discovery of concepts and relations from the Web sites and lead to the development of the system OntoLearn (Missikoff and *al.,* 2002). In these approaches, domain knowledge a priori is required. For this reason, they are called dependent to the domain of the ontology and the collection of Web documents related to this domain need user intervention.

2.1.2. *Incremental approach for Ontology learning from Web documents*

On the other hand, other approaches are dedicated to the ontology building from Web, which is based on the generation of taxonomies without the use of knowledge or a priori or processing techniques of natural language and use of large corpus or thesaurus. The same approach were improved in (Sanchez and Moreno, 2004) to an incremental approach of ontology learning from Web. In (Sanchez and Moreno, 2007), a study of several types of available Web search engine and how they can be used to assist the learning process (searching web resources and compute IR measures). The learning process proposed by this approach is based on four steps. The first one is a Taxonomic learning step where the user starts to specify keyword used as a seed for the learning process from Web using a web search engine, the output of this step is one-level taxonomy, a set of verbs appearing in the same context as extracted concepts. Secondly, no-taxonomic learning is carried out. Verb list and keywords are used as bootstrap for construction domain related patterns and to construct query to search engine. The third step is the recursive learning task where the two previous learning stages are recursively executed for each discovered concept. Finally, post-processing step consists in refining and evaluating the obtained ontology. This approach is domain independent and incremental. In the same context, our previous work was done (Baazaoui-Zghal and *al.,* 2007). We have proposed an incremental approach of ontology learning from Web. We combined many text mining techniques and use an ontology-based IR System to classify the web documents.

2.1.3. *Web Structure mining-based approach for ontology learning from Web*

In (Shirenand and Tat-Seng, 2007), the underlying assumption is that the noun phrases appearing in the headings of a document as well as the document□s hierarchical structure can be used to discover the concepts and taxonomic relations from documents. A system that supports this approach is implemented and applied on a set of Arabic agricultural extension documents. It takes as input a root concept, analyzes all input documents□heading structure, extracts concepts from headings and builds a taxonomical ontology (Stojanovic and *al.*, 2002).

In this section, several approaches of ontology learning from web were detailed. Many semantic Web documents appeared on the Web and new semantic search engines are developed to search them, several approaches are interested to ontology construction by aggregating on-line ontologies. This will be the subject of the next section.

2.2. *Ontology learning approaches from Web ontologies*

The idea about online ontology building from Web is not a new one. Harnessing RDF files on the Web might be the first step towards achieving true reuse. In (Deitel and *al.*, 2001) an approach for learning ontology from RDF annotations of Web resources was proposed. (Stojanovic and al., 2002) presents an approach for an

automated migration of data-intensive web sites into the Semantic Web. They extract light ontologies from resources such as XML Schema or relational database schemata and try to build light ontologies from conceptual database schemas using a mapping process that can form the conceptual metadata annotations that are automatically created from the database instances. In (Manzano-macho and *al.*, 2008), a framework for integrating multiple ontologies from structured web pages into a common ontology is proposed. According to (Missikof and al., 2002), the out¬put ontology will follow users configuration such as their preferred structure and filtering threshold. It facilitates deep annotation and interoperation in structured web pages from heterogeneous systems. Several approaches were proposed to use ontology search engines or ontology meta-search engines to build ontologies by aggregating many searched domain ontologies. There are an increasing number of online libraries for searching and downloading ontologies. Examples of such libraries include Ontolingua, Protégé, and DAML. Few search engines have recently appeared that allow keyword-based search for online ontologies, such as Swoogle and OntoSearch. In (Allani, 2006), a new approach consists in searching online ontologies for representations of certain concepts, ranks the retrieved ontologies according to some criteria, then extract the relevant parts of the top ranked ontologies, and merge those parts to acquire the richest domain representation as possible.

We don t deny that such approaches could lead easily to have many domain ontologies but some problems still remain. In fact, we are still worry about many issues: the reliability of existent Web ontology, the availability of ontologies to be reused in terms of numbers and domain variety, the quality of output ontology depends on the quality of input ontologies, and the use of Ontology searching, ontology ranking, ontology mapping, ontology merging, and ontology segmentation methods make this approach more complex.

2.3. *Ontology building from Web dictionary*

Wikipedia mining is a new research area which is recently addressed. In (Nakayama and *al.*, 2007), Web thesaurus construction method based on Wikipedia mining is proposed. By analyzing 1.7 million concepts on Wikipedia, a very large scale association thesaurus which has more than 78 million associations was constructed. To avoid NLP problems, link structure mining is applied to Web-based dictionaries.

2.4. *Other hybrid approaches*

In (maedche and Staab, 2000), a method for learning ontologies combining heterogeneous sources of information and various processing techniques associated with each of them to improve the detection of potential useful knowledge. First, it extracts the core vocabulary to the domain using a parsing process. The underlying idea of the method is that the combination of all these additional sources of evidence improves the accuracy of the OL process. Thus, the extracted terms are analyzed at

five different levels at this moment: chunk, statistical, syntactical, vis²ual and semantically. The experimental results obtained processing a set of HTML documents belonging to two domains, Universities and Economics, have shown the potential benefit of its use to learn or enrich ontologies following an unsupervised learning approach.

2.5. *Limits and opens issues*

The state of the art presented in the previous section allowed us to release the limits of most of approaches which are based on text mining techniques. We notice the absence and the difficulty of evaluation of the approaches and the tools of ontological engineering: Indeed, each approach is developed by applying techniques allowing the enrichment of ontology with new concepts and new relations from texts. These techniques are then implemented in a tool. In this work, we did not find a comparative study of the used techniques to deduce the best. This is explained by the fact that there must be the experimentation of these approaches for the same corpus relating to the same field and written in a given language. So the Web could be a common corpus for testing such techniques and offer to ontology engineers to adjust extraction rules of ontology for each domain. Until now, it difficult to propose a domain-independent approach for learning of networked ontologies. Besides, modularity is not respected in these approaches. Then, a motivation to use semantic search for ontology learning is explained in the following. Our study on ontology learning process and semantic search process enabled us to conclude that collaboration between the two processes could be useful to have both "incremental ontology building" and "performed search". We illustrate the different relations of collaboration or resemblance that could exist between the two processes in figure 1. In fact, recent approaches tend toward building a graph-based query for query formulation. Thereby, in the case of absence of the appropriate domain ontology for user within the semantic search system, this first submitted query would be assimilated to a "seed ontology" regarding ontology learning process. Moreover, the step of semantic disambiguation problem is one of the problems handled in ontology refinement. Lexical resources as linguistic ontology or thesaurus are used to fulfill this task. Moreover, finding relevant document for a query represents the same problem for ontology enrichment. Besides, query reformulation with enriched ontology could ameliorate the search by providing users with additional information to constraint his query. The enriched ontology can contain more relevant concepts, relations, instances, or axioms. So, a further collaboration between these two steps will be profitable. Finally, ontology validation in the combined process would be the result of tow type of collaboration: collaboration between searchers having same search goals and an indirect collaboration between searchers and ontology engineer. The idea behind the collaboration between these two processes consists in enabling each contextual semantic search engine to be more flexible and autonomous by discovering others domain ontologies from Web documents. For instance, ontology-based query is the result of the mapping of the ontological concepts with a query written in natural language. In the case of absence of target domain ontology, a

possible ontology could be extracted from text-based query. So, we assimilate a possible query to an initial minimal ontology. This one could be enriched from the selected Web documents tagged as relevant ones by users. Other domain ontologies could be discovered and existent ones will be enriched with the use of terms in query formulation and relevant Web documents selected by target users.

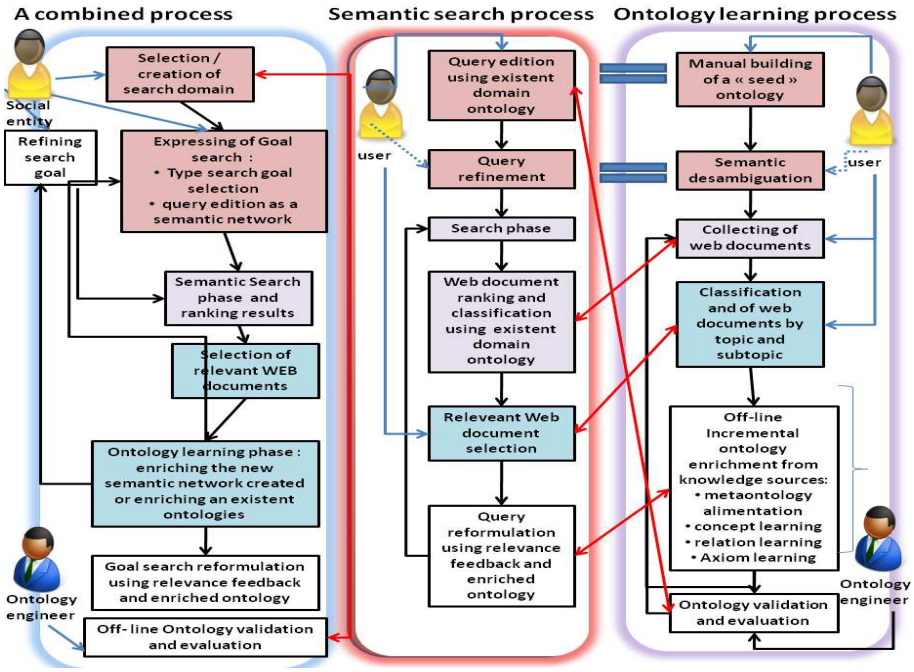


Figure 1. *Combining semantic search process and ontology learning process*

2.6. *Motivations based on our previous work*

On one hand, any process of ontology learning from text depends on the relevance of the textual corpus besides of applied machine learning techniques. On the other hand, the main purpose of semantic search is providing users with the most relevant Web documents according to their query and with the use of specific domain ontology. Starting from this fact, we can affirm that semantic search can be a useful way to perform ontology learning from Web content. In this context, an approach presented in (Baazaoui-zghal and al., 2007) was proposed to use ontology-based search engine (Aufaure and al., 2007) to collect textual sentences from which new concepts and new relations are discovered. In (Baazaoui-zghal and al., 2007), we have proposed a distinguishable and incremental process based on three phases: an initialization phase, an incremental phase of domain ontology learning and finally, a phase of analysis of the results. Indeed, the initialization is dealing with

the preparation and the pretreatment of the data sources which are made of a minimal ontology, a metaontology, the linguistic ontology "Wordnet" and a set of Web documents relating to the target domain. The second phase is characterized by its incremental and iterative aspect. Each iteration is made of two successive steps.

The first one is the alimentation of a metaontology (BenMustapha and al., 2008) and the second consists in applying the axioms related to ontology element learning. The first step consists in applying the techniques specified by the Metaontology to instantiate metaconcepts and metarelations. These techniques are applied according to process described in (Baazaoui-zghal and al., 2007). The second step consists in discovering new concepts, new relations, and new axioms related to a domain. Our approach leads to the implementation of the OntoCosemWeb prototype (Ben Mustapha and al., 2007) and we have used it to build tourism ontology. We have also developed an online information retrieval based on this ontology to collect and classify the results selected by users. These results are used it as the input of "OntoCosemWeb prototype". For this reason, our motivation lies in to integrate an ontology learning task into the semantic search process and to define how the two processes could collaborate to build more domain ontologies from selected documents and, by the way, ameliorate the semantic search.

3. Towards semantic search Framework for incremental ontology learning from Web

According to (Esmaili and *al*., 2006), the problem in contextual semantic search systems resides on building a new domain ontology which has not been defined before. Standing from the fact that the Web is an enormous information source and a dynamic, we have the idea to integrate ontology learning process in the search process. To fulfill this motivation, many objectives are fixed such as: modularity and reuse of learned ontologies, scalability and evolution of ontology building, facility of learning axioms on ontology modules by linking the search request to search results, and Personalization of the built ontology. In fact, to have networked ontologies in a multi contextual search engine is a key requirement to cover user needs. However, when many domain ontologies are used by a semantic search system, taking consideration of modularity aspect make easy the management task of these ones. In many cases, a search query can be translated to an ontology module (a sub-part of ontology). These modules could be reused by other users to express a similar query or to enrich it with new concepts, instances, or relations. So, any search system will become multi-contextual and more adaptable to user's queries. The searcher will participate also in ontology building by selecting the more relevant documents. These ones will be the input of ontology learning process to enrich the initial submitted query.

3.1. *The ontology Warehouse*

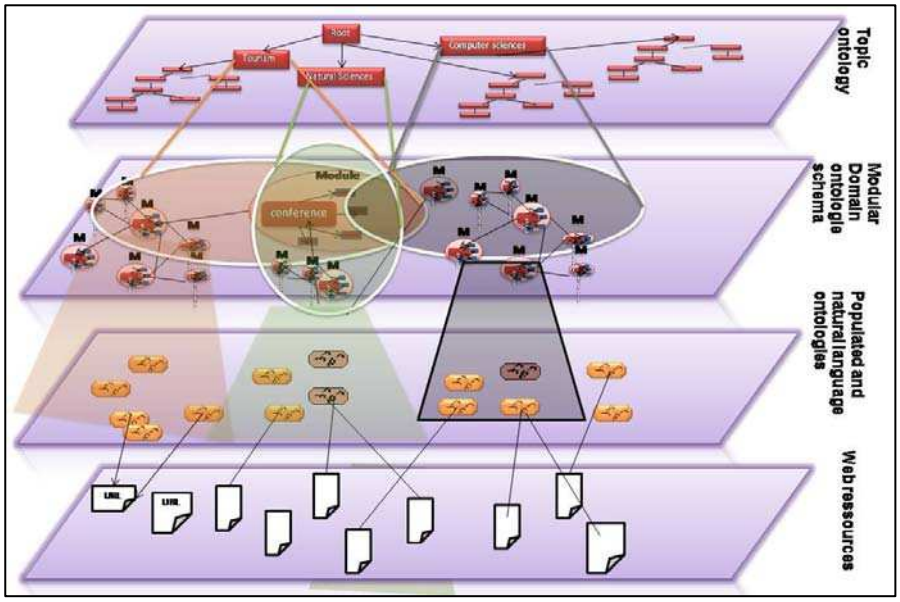


Figure 2. *Multilayer ontology warehouse***.**

Ontologies Warehouse is made up of four levels of ontologies (figure 2). The first layer represents the topic ontology. It is an ontological classification of topics, domain and contexts, regardless of the used language. Each topic T can be the subject one or more domains D, it depends on the position of the topic in the hierarchy. The second layer represents a set of networked domain ontology schema. Each Domain ontology Od is a networked modules M. a Module M is seen as a dimension in the domain ontology which consists of a main concept C with its common properties (relations with others concept i). Proprieties of a concept C1 are defining as the more frequent relations that characterize C1 and that are used in query interfaces and relevant Web document. So, a Module M1 could be in many ontologies and in relation with other modules. For example, the module having as main concept □conference□ could be in many domain ontology (computer science, physics, mathematical, etc.), as we can find conferences related to many domains. A concept C is the following tuple (id, {(ti, language, context)} i=1..n, state, credibility Degree) where:

□ Id: is a concept identifier associated to a sens regardless of the terminological labels and the language referencing it.

 {(ti, language, context)}i : is a set of triple (t, language, contexte) where t is nominal phrases referencing the concept in a targed language and used in specific context which can be the topic that represent the concept role in a specific domain.

 State: is the state of the discovered concept. A discovered concept from text could be new candidate, validated, rejected, average candidate.

 Credibility degree: is a degree of the correctness of the concepts according to his module, we are working on our future work to determine this degree with the observation of the usability of a concept in semantic search. Then, to each user, a personalized view of domain ontology is associated which represents the most used ontology fragment in their search activity besides of their used terminology.

3.2. *Query Graph pattern vs ontology module pattern*

According to the type of search goal selected by user, a set of graph patterns were designed. For example, if the user goal is a navigational search then the patterns presented in figure will be instantiated and the searched node will be marked by X?

3.3. *Case bases*

On one hand, the ontology learning tool needs Web documents classified by search request. On the other hand, indexing document with ontological elements is important for semantic search system. For this reason, we have designed a cases base indexing document with instantiated graph patterns which is an ontology module. A case is a couple of a problem and solution. The structure of problem is not a fixed one, it represents the following tuplet (T, SG, C, G) when T: the topic, SG: type of search goal, C: the main concept which is subject of the search, G: instantiated graph). The formalism of representation will be treated in future Work. A case is represented by a problem and its solution. A problem is equivalent to the search request. The solution is the set of URLs found by the user. The base case is displayed to the user. Otherwise, if similar cases exist, a new case is added to the base case. A query is sent to a search engine. When the user selects the relevant documents, a new event is added to the base case and a process of text mining is applied to selected documents to enrich the ontology module on the associated request.

3.3. *Learning process based on CBR*

The combined process is represented by this algorithm. The user selects an existing topic from the Topic ontology, if it is a new one, he can create the topic and place it in the appropriate position in the ontology. We will use the type of search goal selected by user to better understand the search purpose.

To each goal search, a set of graph patterns are affected. These graphs will be instanced by users according to their request with the ontological elements. This step is important to construct an initial core ontology module that will be enriched

by relevant Web documents. We insist on the fact that each search goal will be translated into ontology module characterized by a target concept that we called "main concept" and others concepts that characterize this one and restrict the search. After some iteration of using semantic search system, ontology learning process doesn't enrich immediately the underlying ontology with all discovered concepts. When there is a doubt about adding these concepts, some searchers submit their graph-based query by enriching an antecedent similar query with one of these concepts. This act could be a hidden way of validation of some discovered concepts. The indirect collaboration consists on the maintenance of domain ontology and ontology views classified by search goals by ontology engineers. Starting from this convergence, we can imagine that the combined process will be as described in figure 1.

4. Illustrating example

We suppose that a user wants to know the URL of the workshop WISM 2009. The type of search goal is a navigational search. The user selects from the topic ontology, by searching the term workshop, he will find that there are no topics related to system modeling. So, a new insertion of a new topic in the topic ontology is done. We suppose also that we have modular domain ontology related to computer science in the ontology Warehouse. But the concept workshop doesn't exist. So the formulation of this first request will be a new core ontology module to be enriched in computer science ontology. The main concept of this request is "workshop" (figure 3).

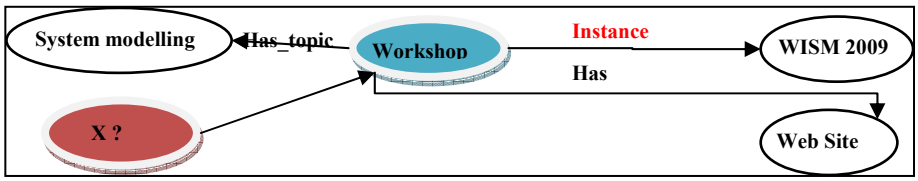


Figure 4. Initial goal search formulation.

The disambiguation step will be held by user with the senses delivered by Wordnet or using online linguistic resource as Wikipedia. This step is important to collect the synonyms and hyponyms in order to instantiate the metaontology (Ben mustapha and *al.*, 2008) with this contextual information. Since, the case base is empty, a query is submitted to a search engine and the user selects the Web document corresponding to the WISM workshop. This document will be the input of ontology learning phase of the process described in. Then, we have chosen to illustrate the application using one of the learning techniques specified by our meta-ontology which is syntactic patterns and verb based pattern. The results of these techniques are illustrated by table 1 when new concepts and relations are discovered.

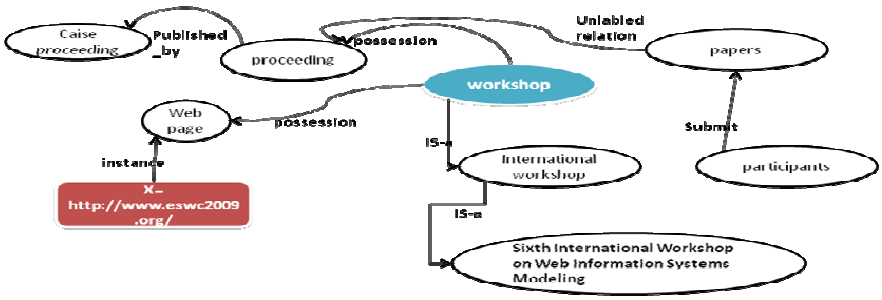


Figure 5. Enriched ontology module.

5. Conclusion

In the present paper, we focused on the possible combination of semantic search approaches and ontology learning methods to facilitate the integration of personalized and evolutionary ontology building in semantic search systems. We have proposed a framework with an illustration scenario. The originality of our proposal consists in appling ontology technology with information retrieval based on case base reasoning and combining ontology learning with semantic search based on case base reasoning. The main contribution of this work is to facilitate the Web semantic engineering using semantic search and ontology learning from Web document and to link the request of users to ontology modules constructed by using their selection of relevant documents.

REFERENCES

Agirre E., Ansa O., Hovy E. et Martinez D. « Enriching very large ontologies using the WWW »., actes de l'atelier *sur la construction d'ontologies de la conférence européenne de l'intelligence artificielle* (ECAI-00), 2000.

Alfonseca E., Manandhar S. « An unsupervised method for general named entity recognition and automated concept discovery ». Actes de la 1er *conférence internationale sur General WordNet*, India, 2002.

Allani H., Position paper: ontology construction from online ontologies, *International World Wide Web Conference* , 2006.

Aufaure M, Soussi R. et Baazaoui Zghal, H. Ben Ghezala H., : SIRO: On- Line Semantic Information Retrieval using Ontologies, *The Second International Conference on Digital Information Management* (ICDIM'07), october 28-31 lyon, France, 2007.

Aussenac-Gilles, Jacques M-P. Designing and Evaluating Patterns for Ontology Enrichment from Texts. *EKAW 2006*, 2006, pp. 158-165..

Aussenac-Gilles N., Biébow B. et Szulman S. D'une méthode à un guide pratique de modélisation des connaissances à partir de textes, *Actes des cinquièmes journées Terminologie et Intelligence Artificielle (TIA)*, Strasbourg. Ed. F. Rousselot, ENSAIS, Avril 2003.

Berners-Lee, T, Hendler, J, Lassila, O.: The Semantic Web, Scientific American ; 2001.

Baazaoui-Zghal H., M.-A. Aufaure, N. Ben Mustapha (2007) "A Model-Driven approach of ontological components for on-line semantic Web information retrieval", *Journal on Web Engineering, Special Issue on Engineering the Semantic Web*, Rinton Press, vol. 6, n°4, pp 309-336.

Bachimont B. Engagement sémantique et engagement ontologiqueconception et réalisation d'ontologies en Ingénierie des connaissances. *Ingénierie des connaissances: évolutions récentes et nouveaux défis*, chapitre 19, 2000.

Ben mustapha, R. Soussi, H. Baazaoui zghal, MA. Aufaure (2008) «A Metaontology for Domain Ontology Enriching in an Information Retrieval System», *2 ème journée francophone sur les ontologies*, pp.63-72, Lyon, 1-2 decembre 2008.

Ben Mustapha, N., Baazaoui Zghal, H. et Aufaure, MA.. A Prototype for knowledge extraction from semantic Web based on ontological components construction, *Internet technology, and Web information System and Technologies* (WEBIST07), 2007.

Deitel A., Faron C., Dieng R. : Learning ontologies from RDF annotations. In *Proceedings of the IJCAI Workshop in Ontology Learning*, Seattle, 2001

Esmaili K. S. and Abolhassani H.. A Categorization Scheme for Semantic Web Search Engines. In 4th ACS/IEEE *International Conference on Computer Systems and Applications (AICCSA-06)*, Sharjah, UAE, 2006.

Faatz A. et Steinmetz R. Ontology enrichment with texts from the WWW. *Semantic Web Mining 2nd Workshop at ECML/PKDD-2002*, 20th August 2002, Helsinki, Finland, 2002.

Faure D. et Nedellec, C. A corpus-based conceptual clustering method for verb frames and ontology acquisition, *LREC workshop on adapting lexical and corpus resources to sublanguages and applications*, Granada, Spain, 1998.

Hahn U. et Markó K. Joint knowledge capture for grammars and ontologies. *Proceedings of the First International Conference on Knowledge Capture K-CAP* 2001, Victoria, BC, Canada, 2001.

Hearst M.A. Automated Discovery of WordNet Relations. "Wordnet An Electronic Lexical Database". *MIT Press*, Cambridge, MA, 132-152, 1998.

Hwang CH. Incompletely and imprecisely speaking: using dynamic ontologies for representing and retrieving information. *Proceedings of the 6th International Workshop on Knowledge Representation meets Databases* (KRDB'99) 14–20, 1999.

Khan L. et Luo F. Ontology Construction for. Information Selection. *Proceedings of 14th IEEE International*, 2002.

Kietz JU., Maedche A. et Volz R. A Method for Semi-Automatic Ontology Acquisition from a Corporate Intranet. In: Aussenac-Gilles N, Biébow B, Szulman S (eds) EKAW'00 *Workshop on Ontologies and Texts. Juan-Les-Pins,* France. CEUR Workshop Proceedings 51:4.1–4.14. Amsterdam, The Netherlands, 2000.

Lin C-Y et Hovy E.H The Automated Acquisition of Topic Signatures for Text Summarization. *Proceedings of the COLING Conference*. Strasbourg, France. Aout, 2000

Lonsdale D., Ding Y., Embley D.W. et Melby A. Peppering Knowledge Sources with SALT; Boosting Conceptual Content for Ontology Generation. *Proceedings of the AAAI*

Workshop on Semantic Web Meets Language Resources, Edmonton, Alberta, Canada, July 2002.

Maedche A. et Staab S. Mining ontologies from text. Proceedings of the *12th International Conference on Knowledge Engineering and Knowledge Management*, Springer Lecture Notes in Artificial Intelligence, 2000.

Manzano-Macho D., Gómez-Pérez A.and Borrajo D., Unsupervised and Domain Independent Ontology Learning: Combining Heterogeneous Sources of Evidence, Proceedings of *the Sixth International Language Resources and Evaluation (LREC'08*), 2008, 28-30.

Missikoff, M., Navigli, R., & Velardi, P. (2002). Integrated approach to web ontology learning and engineering, *IEEE Computer* Vol 35(11) pp.60-63.

Moldovan D. I. et Girju R. Domain-Specic Knowledge Acquisition and Classification using WordNet. Proceeding *of FLAIRS2000 Conference*, Orlando, 2000.

Nakayama K., Hara T., Nishio S., "A thesaurus construction method from large scale web dictionaries", Proc. of IEEE *International Conference on Advanced Information Networking and Applications* (AINA2007), pp. 932-939, 2007

Navigli R., Velardi P.: Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites Computational Linguistics, *MIT press*, juin 2004.

Nobécourt J. A method to build formal ontologies from text. In: EKAW-2000 *Workshop on ontologies and text*, Juan-Les-Pins, France, 2000.

Resnik P. Using information content to evaluate similarity in taxonomy. Proceedings of the *14th joint conference in Artificial Intelligence*, 1995

Rose D. and Levinson D., "Understanding User Goals in Web Search," In: *13th International Conference on World Wide Web*, pp.13-19. USA, ACM Press, 2004.

Roux C., Proux D., Rechermann F. et Julliard L. An ontology enrichment method for a pragmatic information extraction system gathering data on genetic interactions. Proceedings of the *ECAI2000 Workshop on Ontology Learning* (OL2000), Berlin, Germany, août 2000.

Sánchez D. et Moreno A. Creating ontologies from Web documents. *Proceedings of the Setè Congrés Català d'Intel.ligència Artificial*, IOS Press 113, 2004.

Sekiuchi R., Aoki C., Kurematsu M. et Yamaguchi T., DODDLE: A Domain Ontology Rapid Development Environment, PRICAI98, 1998.

Shiren Y. Tat-Seng C., Automatically Integrating Heterogeneous Ontologies from Structured Web Pages, , Int'l *Journal on Semantic Web & Information Systems, 3(2)*, 96-111, April-June 2007.

Stojanovic N., Stojanovic L., Volz R.: A Reverse Engineering Approach for Migrating Data-intensive Web Sites to the Semantic Web. In: *17th World Computer Congress*, *Kluwer Academic Publishers*, 2002, pp. 141-154..

Sugiura N., Masaki K., Naoki F., Noriaki I. et Takahira Y. A Domain Ontology Engineering Tool with General Ontologies and Text Corpus. Proceedings *of the 2nd Workshop on Evaluation of Ontology based Tools*, 2003.

Velardi P., Fabriani P. et Missikoff M. Using text processing techniques to automatically enrich a domain ontology, Proceedings of the *ACM Conference on Formal Ontologies and Information Systems*, pp 270-284, 2002.

Xu F., Kurz D., Piskorski J. et Schmeier S. A Domain Adaptive Approach to Automatic Acquisition of Domain Relevant Terms and their Relations with Bootstrapping. Proceedings of LREC, la troisième conference sur language resources and evaluation, Las Palmas, Canary island, espagne, mai 2002.