Understanding and Supporting Ontology Evolution by Observing the WWW Conference

Nicolas Guelfi[1], Cédric Pruski[1,2], Chantal Reynaud[2]

[1] Laboratory for Advanced Software Systems – University of Luxembourg campus Kirchberg 6, rue Coudenhove-Kalergi, L-1359 Luxembourg
[2] LRI – University of Paris-Sud, CNRS & INRIA-Futurs, 4 rue Jacques Monod, Parc Club Orsay Université, 91893 Orsay Cedex (France)
{nicolas.guelfi, cedric.pruski}@uni.lu, chantal.reynaud@lri.fr

Abstract. Ontologies which represent domain knowledge in information systems are efficient to enhance information retrieval. However, domain knowledge is evolving over time and thus it should be also expressible at ontology level. Unfortunately, we consider that ontology evolution is barely study and its basic principles have not been yet precisely defined according to our notion of evolution. In this paper, we have followed a bottom-up approach consisting in a rigorous analysis of the evolution of a particular domain over a significant period of time (namely the WWW series of conference over a decade) to highlight concrete domain knowledge evolutions. We then have generalized and we present a precise set of evolution features that should be offered by ontology metamodels. We also evaluate the modelling capabilities of OWL to represent these features and finally, we show the contribution of ontology evolution support to improve Web information retrieval.

Keywords: Ontology Evolution, Domain Analysis, OWL, Web Information Retrieval

1 Introduction

Although being firstly introduced in philosophy, ontologies have recently appeared in the field of computer science as the cornerstone of the Semantic Web paradigm [3]. The later aims at giving a sense to the Web what will allow computers to "understand" Web data. If this goal is achieved, computers will be able to unload users of many tedious tasks like searching relevant documents or services. The Semantic Web implements ontology that models a part of the real human world, mainly to annotate Web data or to facilitate information retrieval. Nevertheless, since ontologies represent the knowledge of a particular domain, they have to smoothly follow the evolution of that domain otherwise their use will lead to unwanted effects. Therefore the ontology evolution problem [11], [15] has recently been deep studied since it becomes rapidly of utmost importance.

In our previous work we have defined the O^4 approach [6], [8] that aims at improving the results of a Web search in terms of relevance. This is done mainly through the use of ontology-based query expansion rules. In order to optimize the search, we need to select the adequate terms from the ontology to enrich the query. Actually, if the ontology does not reflect the knowledge of the domain associated to the submitted query, the search results will not be those awaited by users. We are thus facing the problem of ontology evolution.

In this paper we propose a set of modelling features for ontology evolution. These features have been defined after the rigorous study of the evolution of a particular domain (in our case, the domain defined by the WWW series of conference topics) over a ten years period of time. In consequence, we will first have to present in detail the construction of a corpus of documents that is representative of the domain we have studied. This requires the definition of relevant criterion and tools that will facilitate the analysis of the domain. The results of this analysis will lead directly to the definition of the various kind of evolution that can appear [7] which in turn will allow the proposition of modelling features that aims at designing evolving ontologies. The proposed primitives will first allow us to understand the evolution of ontologies and will aid to predict future versions of ontologies. They can be used to describe a structural evolution on one hand and a progressive evolution on the other hand. Since this work has been carried out in a context covering Web information retrieval, we will highlight the contribution of such ontologies through an example implementing ontology-based query expansion techniques [8] to improve the relevance of documents when searching the Web.

The remainder of the paper is structured as follows. Section 2 presents the characteristics of the domain we have studied in order to define the new modelling features devoted to ontology evolution. In Section 3 we detail the proposed modelling primitives as well as their properties. Section 4 illustrates an application of our work through a basic example dealing with Web information retrieval. In Section 5 we discuss related work in the ontology evolution field. Finally the paper wraps up with concluding remarks and our future work.

2 Domain of Study Definition and Ontologies Construction

The first step towards the proposition of modelling features devoted to ontology evolution concerns the construction of a significant corpus of documents that will allow us to highlight the various kinds of domain evolution. In this section we present the characteristics of such a corpus and the ontologies we have built from that pool of documents.

2.1 Domain Selection

Since we want to derive modelling features for ontology evolution from the analysis of the evolution of a particular domain, the selection of such domain is of utmost importance. Many domains, like bioinformatics through the Gene Ontology [16], are already modelled using ontologies. Unfortunately, these ontologies are either young or built using only domain-specific relations. Therefore we considered that the study of their evolution will not be relevant enough and we decided to construct ontologies from a set of descriptions of an evolutionary domain. To this end, we have chosen the domain covered by the World Wide Web series of conference which is reflected in the calls for papers and in the accepted papers. Thus, papers accepted for publication at these events together with the calls for papers, which are online and can be retrieved via a web search engine, form our “Micro Web” (i.e. the corpus of documents we will analyze). In order to see the evolution of the domain of the Web over a significant period of time, we decided to harvest the accepted papers of the last 10 WWW events. The so-called *Micro Web* consists of good case study since the chosen conference is world famous and known to be one of the most representative events in the domain of the Web. Therefore, its successive calls for papers reflect the various fields in vogue in the corresponding domain. Moreover, the quality, quantity and homogeneity of the submitted papers as well as the high level of selectivity (less than 20%) set by reviewers reinforce this idea. As a result, we have a corpus made of 622 documents all stored in a relational database which will facilitate their future analysis.

2.2 Methodology for Ontology Construction

We have designed, from the calls for papers and the accepted papers of the different conferences, the ontology of the domain of the corresponding event for each year using the Protégé[1] ontology editor. It means that we have built 10 ontologies, one concerning each event of the WWW series. These various ontologies represent in fact the same domain that is evolving over time. The ontologies are constructed following a rigorous process inspired from the ARCHONTE methodology [2]. The three steps of this methodology consist in a semantic normalization of the terms introduced in the ontology, followed by a formalization of the meaning of the knowledge primitives obtained and an operationalization using knowledge representation languages. The so built ontologies will allow us to identify the different evolutions of the domain to try, in the next phase, to explain the changes. The construction of the different ontologies has been done manually following the process described hereafter. We first model the knowledge of the domain and then we formalized it using the Web Ontology Language (OWL). As modeling is the main purpose here, we use the most expressive flavor of OWL (i.e. OWL Full).

First of all, we stated that every topic of a call for papers denotes a concept in the corresponding ontology. This means that our ontologies are small and made of about 30 classes. For instance, the topic *multimedia* in 1998 provides the concept *multimedia* in 1998’s ontology. Furthermore, topic like *social and cultural* gives rise to two concepts (i.e. a concept *social* and another one *cultural*) in the ontology. Indeed, we decided to split expressions according to the conjunction “and” which regularly appears. Nevertheless, the conjunction of words using this proposition indicates that words involved in that particular topic are linked. This is the first step towards the construction of the set of relations that bind concepts of the ontology.

[1] http://protege.stanford.edu/

Then, to determine these relations, we base on the content of the accepted papers with a particular attention devoted to the abstract of the papers. We first localize an occurrence of the concepts we tried to bind in a document of our corpus, and then we tried to identify manually from the text the relation of the domain. In order to validate our choice, we reiterated the operation on several other papers. This basic but rigorous process provides us the ontologies (one is depicted in figure 1) of the domain that will be the material of our study devoted to domain evolution.

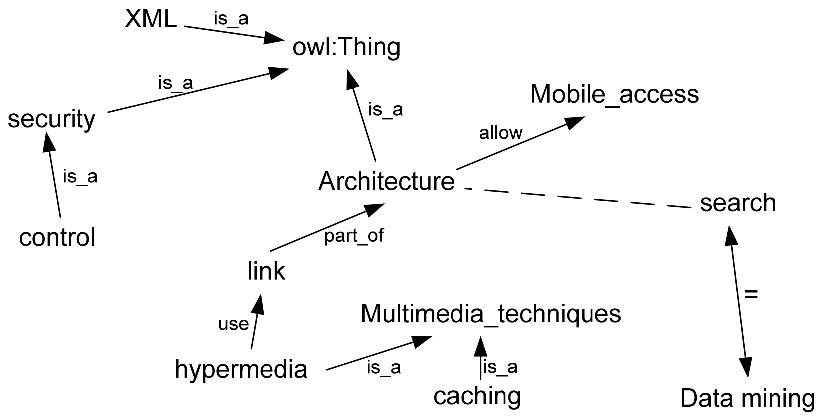

Fig. 1. Part of the Ontology representing WWW 2000's call for papers

As partially illustrated on the ontology in figure 1, we used very elementary relations like subsumption, equivalence or meronymy (i.e. part-of) to design our ontologies. Nevertheless, we also introduce some particular relations resulting from the analysis of the content of the papers like *use* or *allow*. All the constructed ontologies can be downloaded from our Web site[2].

3 Domain Analysis: Towards Ontology Evolution

The analysis of the evolution of the domain represented by WWW conferences' topics made it possible to define various kinds of evolution that affect the domain. In this section, we will detail these evolutions and as a result, we will give the corresponding modelling primitives for the design of evolving ontologies.

3.1 Domain Evolution

In this first subsection, we discuss the various kind of evolution that stand out during the analysis phase of our micro web. The analysis we made is at two levels. The first one, called general observation, defines the macroscopic evolution of the domain over a long period of time (10 years in our case). In contrary, the second level, called local observation, highlights the microscopic variation of the domain. This second kind of observations is made on a very short period (i.e. 1 or 2 years). These observations

[2] http://se2c.uni.lu/tiki/tiki-index.php?page=TargetTool

made it possible to emphasis different kinds of evolution among which one finds concept persistence, emergence of new concepts, concepts removal, generalization and specialization of concepts. However, our observations have also permitted to define other important features like the importance of concepts in the domain but also the resistance to modification and the variation of a distance between concepts over time as well as the speed of evolution. All these characteristics will be detailed in the remainder of this section. However, in order to explain the various highlighted kinds of evolution, we needed to distinguish between the ontology built from the calls for papers which represent the conferences chairman point of view and the content of the papers which represent the authors' interpretation of the calls. Unfortunately, we did not have access to the reviews. These would have permitted to understand if the authors' interpretations of the calls were consistent with the chairs point of view of the domain.

Concept Persistence

This first kind of evolution affects some particular concepts of the domain. Actually, we observe that special concepts like *security* or *search* are present in the ontology over the whole period of observation. It means that since its appearance in the domain, the concept remains in the domain. We called this constraint on evolution **concept persistence**. Our personal knowledge of the domain lets us claim that these two concepts denote key notions of the domain (recall that we study the domain covered by World Wide Web series of conference). Thus, we can say that concepts that are part of the ontology over a predefined long period of time constitute the core of the ontology because the semantics of the concept is still covered by the semantics of the domain. This is particularity important for approach exploiting ontologies like techniques for indexing data or data retrieval. In fact, these concepts are the most relevant and in consequence should be favoured in their usage. For instance, the *search* concept is present in the domain for the whole period of study, whereas other concepts which seem to be less important like *social* remains in the domain only for one year. We will illustrate this particular point in Section 4 hereafter.

Concept Emergence

The second observation in the evolution of a domain concerns the addition of knowledge at a particular moment. This **emergence of concept** was particularly true for the *Semantic Web* in 2002. Since this paradigm was defined in 2001 by Tim Berners-Lee [3], and its associated semantics was close enough from the semantics covered by the domain defined by the topics of the WWW series of conference, it rapidly appeared as a concept of the domain and as a result, one year later in the topics of the WWW conference. This is why it takes place in our ontology representing the domain covered by WWW 2002 topics. Our survey has shown that 79 concepts have emerged in the domain of the WWW series of conference between 1997 and 2007. Moreover, there are about 11 concepts in average that emerge each year in an ontology that contains about 30 concepts. Recall that we have one ontology per conference.

Concept Removal

A concept can be removed from a domain for several reasons. The first one is related to its semantics. Actually, by virtue of knowledge evolution, the semantics of that concept could not be covered by the semantics of the domain (described by the ontology) anymore and therefore should be removed from that domain. Moreover, a concept can be either not precise enough (i.e. the concept is too abstract) or too precise (i.e. specific concept). This would also require some domain refinement which in turns will lead to the removal of concepts for the benefit of more abstract or specific concepts. Another reason concerns the properties of the concept. For instance, if the concept is no more "popular" or "profitable" (if we are in a business domain) for the domain it can also be removed. We can speak about obsolete concept. In our case study this kind of evolution arose several times. For instance, concepts like *social* and *cultural* appear in the 1998 WWW conference topics but are removed in the 1999 conference topics and does not appear anymore. Moreover, our study revealed that in the WWW 1998 conference, no papers containing these two words were submitted which proves a kind of irrelevance. This is probably the reason why both concepts have been removed from the domain from that moment.

Concept Abstraction

Our observations revealed that a concept or several concepts can be substituted along time axis by a more general concept. We call this phenomenon **concept abstraction**. This can be done when the semantics of a concept is completely covered by the semantics of a concept that is directly link to it. However, we observe that this phenomenon usually turns up when a concept is becoming less relevant for the domain. For instance, in our ontologies concepts like *browser* and *tool* are generalized into the more general concept *application*. This substitution give less importance to the two concepts that have been generalized which in turns give more freedom for the future authors in their interpretation of the call for papers. Actually, since this evolution in the call for papers took place, there have been more submitted papers dealing with a wider range of applications than papers discussing only the use of Web browsers. Our study has permitted to emphasis this idea. Concretely, there are 351 occurrences of the word *browser*, 173 occurrences of the word *tool* and 351 occurrences of the term *application* in the documents. Moreover, 30% of these words are cited in the same papers and in most of the cases, the words *browser* and *tool* can be replaced by the term *application* without a loss of semantics (i.e. the sentences where this phenomenon appears have the same meaning after terms substitution). Therefore the concepts of the ontology representing these notions (i.e. *browser* and *tool*) have been substituted by a more general one (i.e. *application*) which gave place to a wider variety of papers on Web applications. We have identified 5 concepts that have been abstracted. However, the time needed for a concept to become more abstract varies. Actually, some abstractions are very fast, only one year for the abstraction of concepts like *browser* and *tool*, other highlighted abstractions can be longer.

Concept Specialization

In the contrary, our empirical study has shown that a concept or a group of concepts can evolve in a more **specific concept**. Contrary to concept abstraction presented above, this phenomenon is possible only if the more specific concept on one hand shares a part of the semantics of its super concept and on the other hand, offers some specifics axioms that make it possible to represent the domain (or the subpart of it) at the right level of abstraction. In this particular case (i.e. concept specialization), the main objective is to bring more precision in the description of the domain by introducing new concepts. In our domain of study, to know the Web, this has been the case for the concepts *language*, *programming languages*, *markup language* and *metadata system* in 1998. Indeed, they have been transformed in a more specific concept: *XML* the year after. This modification that brought more precision in the call for papers has have an important impact on the submitted papers since 23 papers dealing with XML have been accepted in 1999. However, this rapidity in the change (only one year) can be explained by the analysis of the content of the papers submitted in 1998. We first observe that the *XML* word appears mostly in the paper of the track corresponding to the concepts that have changed (*language, programming language* …). Concretely, there are 162 occurrences of the term *XML* in papers related to programming languages, metadata systems and markup languages for a total of 205 occurrences of *XML* in all the accepted papers. Furthermore, the study of the abstract of these papers has highlighted that the concept *XML* was directly linked to the concepts *metadata, languages and markup languages* through a subsumption relation and terms *metadata, markup language*, and *programming languages* refer in most of the case to *XML* in the content of the papers. This phenomenon combined with the relevance of the XML language at this period of time has probably egg WWW 1999 chairman to adapt the call for papers. This observation underlines the relations between the interpretation of the domain (i.e. the content of the papers) and the evolution of the domain itself (i.e. the call for papers). 7 concepts have been specialized over the period of study and this evolution gave place to 16 new concepts of the domain. Moreover, as it is the case for abstraction, the operation required more or less time depending of the nature and the importance of the concept in the studied domain.

Semantic Weight

Another important kind of evolution that has been highlighted by our study deals with the notion of importance of the concepts in the domain. We call this phenomenon **concept emphasis.** This property put the stress on the punctual tendency of the evolution. In fact, at some time, concepts are more relevant for the domain than other ones. Depending on the domain of interest, this “relevance” can be popularity, profitability, technological improvements, etc. In our study, this is the case for concepts like *search*, *hypermedia* or *Semantic Web* in 2002 but also *ontologies* recently in 2006. This turns up at two different levels. First, it appears in the tracks of the conference. In fact, since there have been so many accepted papers dealing with these notions, two tracks were organized which underlines the importance or the **semantic weight** assigned to these topics. Second, 83% of the accepted papers of the other tracks contain at least one occurrence of the involved word which is also an indication concerning the important aspect of the concept *ontology* in the domain. We believe that this notion is really important and we will give an illustration in Section 4.

Semantic Distance

A more meticulous observation of the evolution of the domain of the Web through the calls for paper of WWW's events has permitted to emphasize the notion of **semantic distance** between the concepts of the ontology. However, the distance we highlighted is different from those proposed by Hirst-St-Onge [9], Jiang-Conrath [10] or Resnik [14]. Actually, these metrics measure the distance between concepts that are linked by at least a path composed by more than one arc in the graph of an ontology and the objective is to estimate the closeness given their localization in the graph and the number of arcs that separate them in this ontology. Nevertheless, we found, through our empirical study, that the distance between concepts directly linked by the same arc in the graph of the ontology varies. Actually, some concepts seem to be "closer" (from the semantic point of view) than other ones although they are linked by the same relation in the ontology. This turns up in the use of the words denoting these concepts in the documents of the corpus. For instance concepts *browser* and *application* appear more frequently in the papers than concepts *tool* and *application* in 1999 and both couple of concepts are bounded by the same relation (in this case the relation of subsumption). Nevertheless, adequate metrics (different from those cited in this subsection) are needed to catch this notion of semantic distance. For the time being, we decide to consider words frequency. It means that we measure how many times two concepts of a relation are cited together in the same kind of documents (i.e. documents published the same year) and in the same context. Moreover, this distance plays a key part to explain for instance the removal of concepts from the ontology. In fact, concepts which are not relevant anymore for the domain, are getting further and further form other concepts of the domain (i.e. the semantic distance is increasing). Therefore, when a predefined threshold is reached, concepts are removed from the domain. In the contrary, when concepts are very close, they can be replaced by a more abstracted or specific concept if another appropriate threshold is reached.

Resistance

This other kind of evolution, called **resistance to change**, is a bit different from the other characteristics presented so far. Actually, it has the particularity to be opposed to evolution. This appears in our study in the ontology of 1998 and 1999. It reflects also in the documents of the corpus. Indeed, there are 49 occurrences of the words security in the papers accepted in 1999 which is very few. Furthermore, one paper contains 26 occurrences of that word. This reveals that the notion of security was not of utmost importance in 1998. Thus, following the natural aspect of the evolution process, this concept should have been removed from the ontology representing WWW 1999's call for papers which is not the case as the concept security remained in the call for papers in 1999. This proves that the chairman of WWW 1999 has considered this notion as important for the field. The resistance to changes is also present in other field mainly knowledge management [4], [12]. Nevertheless, the resistance seems to vary according the concepts involved. Each concept resists differently to evolution. The "coefficient" of resistance to change affected to each concept is different. This introduces a notion of degree of freedom in the evolution of the ontology. In fact, using this property, one can partially control the evolution of the ontology. Thus, this newly introduced metrics should be determined rigorously by domain experts. Furthermore, this phenomenon turns up under various forms in every day's life. For instance, for approximately 80% of the population, whales are seen as fish and for only 20% of the people whales are mammals. In consequence, if we follow the natural evolution process, in a significant period of time, all the people should classified whales under fishes. Nevertheless, among the 20% of the people are biologists (i.e. domain experts) that will permanently reject this evolution. This proves the existence of such resistance to change and should be taken into account in the ontology representing the domain mainly using adapted coefficient as shown in section 3.2.

Speed of Change

The evolution of the domain takes place at different speeds. Some changes are rapid; others are very slow and required several years. For instance, the specialization concepts *metadata systems, programming languages, and markup languages* into *XML* (as presented before) has taken only one year in the contrary, concepts *browser* and *tools* have been abstracted into *application* in 2 years. We believe that the speed of change is function of the coefficient of resistance to changes presented before. In fact, if the coefficient is high, it means that the ontology should not change (or change very little) which ensures a kind of stability in the ontology. However, if the same coefficient is low, it will allow more flexibility in the evolution of the ontology. The speed of change depends also on external factors like technology improvements. This was the case for the concept *Semantic Web;* only one year after its definition it became a key concept of the domain.

3.2 Modeling Features for Ontology Evolution

The various kinds of evolution we have highlighted through our empirical study, have led us to the proposition of modeling primitives for ontology evolution. In this section we describe these various modeling elements.

The proposed features can be classified into two different sets. Actually, we have primitives that act on concepts (i.e. vertices of the ontology graph) and also primitives that apply on relations (i.e. edges of that graph). The first set is made up of primitives that put the stress on concepts emergence, concepts persistence and concepts importance. So, first when a new concept emerge in a domain, it is important to know the exact date at which the concept has appeared in the ontology. Second, concerning persistence, two things are needed. On one hand, the **emergence date** and on the other hand a **duration** determined manually by domain experts. The latter correspond to a constraint of time the concept has to satisfy in order to be considered as persistent. The last modeling feature that applies on concepts is related to concept importance. We have decided to model this property using a coefficient called **importance**. For the time being, this coefficient is computed based on the occurrences and the repartition of the given concept in our corpus of documents. In consequence, the more frequent its associated term is cited and the better the repartition of this term in the corpus is, the higher the coefficient of importance will be. We decided to limit the coefficient between 0 and 1 (1 representing a very important concept).

The second set is formed by modeling primitives affecting relations between concepts. These are related to the semantic distance and the resistance to changes. Both notions are represented using **coefficients**. The semantic distance between two concepts measures the evolution of the joint use of these two concepts in the corpus of documents. This coefficient can vary from 0 to infinite but a maximum distance is set by domain experts and if the distance reaches this particular value, the relation between the two concepts is removed (i.e. an edge of the ontology graph is removed). Moreover, if one concept becomes isolated in the ontology (i.e. it is no more linked to any other concepts) it can be removed from the ontology. Concerning the coefficient of resistance to changes, it must be defined by domain experts. This coefficient takes its values between 0 and 1 where 1 denotes a very strong coefficient which prevents every relation that is affected to evolve.

As OWL is the *de facto* standard for designing ontologies, we decided to study how to represent the modeling elements presented in this paper using the OWL language. Due to its powerful expressivity, OWL offers enough characteristics to take all the presented features into account. Table 1 hereafter presents the various modeling features, their associated datatype and the ontology notions they are applied to. Observe that the types we use are the same than those contained in XML schema definition.

Table 1. Modeling Features Summary

Element Name	Datatype	Affect
Emergence date	xsd:dateTime	concept
Persistence	xsd:duration	concept
Importance	xsd:float	concept
Semantic Distance	xsd:nonNegativeInteger	relation
Resistance	xsd:float	relation

However, OWL metamodel[3] [11 p.83] should be enriched in order to integrate the above modeling features as basic OWL primitives. Concerning all features that apply on concepts, three attributes should be added to the class *Class* of the OWL metamodel. One attribute for representing the emergence date of a concept in the ontology, a second one to express the persistence duration of a concept and finally a third one for the importance of a concept. Moreover, these attributes must have the same type than their associated elements (see table 1).

The two modeling elements related to relations, can be integrated to the OWL metamodel by adding two attributes to the class *Property*. A first non negative integer concerning the semantic distance and a second float for expressing the notion of resistance to changes are needed. Nevertheless, the expressivity of OWL makes it possible to easily express properties concerning concepts of an ontology without

[3] The OWL metamodel we refer to is the one described using UML by Klein in his PhD Thesis.

modifying the OWL metamodel mainly using OWL Datatype properties but for elements related to OWL properties it would be more problematic.

Another way to proceed would consist in using annotation properties or datatype properties via datatypes defined in accordance with XML Schema datatypes to express our concepts at ontology level. Nevertheless, this would require the expressivity of OWL Full.

4 An Application to Web Information Retrieval

In this section we describe a real contribution of adaptive ontologies in the context of information retrieval. Our formerly mentioned O^4 approach [6], [8] implements an ontology-based query expansion technique in order to improve the results, in terms of relevance, when searching the Web. Actually, the query expansion phase is made according rigorous expansion rules defined by taking into account terms of the query, the form of the initial query and the relations that link the concepts of the query in the ontology. The ontological relations implemented in this approach are on one hand the well-known equivalence and subsumption relations which are already implemented in OWL and on the other hand part-of and opposition relations which have all been formalized in first-order logic and added to the Web Ontology Language as primitives [8]. A first basic rule consists, given a basic query constituted by only one keyword, in adding all the equivalent concepts of this keyword in the ontology. Nevertheless, the ontologies we implemented so far were not able to evolve over time and thus do not reflect the knowledge evolution of the domain they model. In consequence, the choice of the right terms to put in the query was not fine enough. Due to the properties of the evolution features we have presented in this paper, and mainly the semantic distance and the semantic weight assigned to concepts of the ontology, we will be able to refine even more this choice by selecting concepts which weights are the highest since they are considered as the most relevant concepts of the domain. The results of such a search will be more relevant because the more relevant concepts of the domain will be added to the query.

Assume to illustrate this argument that an initial query "Web" will be submitted to a Web search engine. If, for instance, the ontology we use to perform query expansion contains two equivalent concepts for Web that are "WWW" and "Internet" with a semantic distance from "Web" of 1 and 10 respectively. The system will select the term "WWW" to put in the query since it is closer to the initial term "Web" than "Internet" is close to "Web". So, the expanded query "Web WWW" will be submitted. Such expansion is judicious if we compare the different search results associated to both queries "Web WWW" and "Web Internet". Actually, pages returned when the query "Web Internet" is entered are older and probably out of date than pages returned corresponding to the other query. That shows that the integration of domain evolution at ontology level will improve Web information retrieval at least by giving right up to date information. Another basic example consists in filtering the returned pages using the emergence date and the persistence duration of concepts that constitute the query. Assume that the query "modem Internet" is submitted to a Web search engine. The system would be able to return pages dealing with modems that were published from the emergence date of *modem* in the domain of the *Internet* and for the persistence duration of the *modem* concept.

This is all the more true for approaches implementing ontologies for tagging or indexing information. Since the vocabulary for indexing or tagging is extracted from ontologies, it has to be selected rigorously. Moreover, tags are usually chosen by taking their popularity or any other properties that is domain dependent into account. However, this kind of information was not provided by static ontologies. Nevertheless, we have proposed an approach that has the advantage to integrate such properties directly at ontology level. Therefore, if the concepts presented in this paper will be integrated directly in ontologies, they will have a huge impact on approaches dealing with tagging or information indexing.

5 Related Work

In the field of ontology evolution, relevant work has been carried out but two main different approaches stand out. The first one, inspired by the work done in the database field, considers ontology versioning. This problem has mainly been tackled by Michel Klein [11]. He compared ontology evolution with database schema evolution. The framework he proposed contains a set of operators, on the form of an ontology, useful for modifying another evolving ontology. Klein also proposes a change specification language based on the ontology of change operations. Moreover, Avery and Yearwood, through their extension of OWL called dOWL [1], have proposed a set of primitives to improve ontology versioning by facilitating the design of dynamic ontologies.

The second approach for ontology evolution deals with consistency during the evolution process. To this end, Ljiljana Stojanovic proposed a general methodology for managing ontology evolution [15]. The process can be divided in 6 different phases occurring in a cyclic loop. It enables handling the required ontology changes; ensures the consistency of the underlying ontology and all dependent artifacts; supports the user to manage changes more easily; and offers advice to the user for continual ontology reengineering. Recently, Peter Plessers [13] described another framework for managing consistent changes in ontology. This is done through the definition of a *Change Definition Language* and the notion of *version log*. The former is a temporal logic based language that allows ontology engineers to formally define changes whereas the latter stores for each concept ever defined in an ontology the different versions it passes through during its life cycle.

Besides, another interesting work has been carried out by Giorgios Flouris [5]. It consists in applying approaches related to *belief change* to the ontology evolution problem. The set of modeling features we propose introduces a new dimension in ontology mainly by the introduction of the *Semantic Distance* between concepts of the ontology. Nevertheless, our approach is different from the two approaches existing in the literature which are reviewed in this section to know ontology versioning and ontology evolution management. In our approach, we represent the knowledge related to domain evolution in an ontology and show how this knowledge can be exploited in Information Retrieval. Moreover, these new properties will allow first to understand the evolution and will make it possible to anticipate future evolution. Nevertheless, the dynamic ontologies we obtain can support ontology versioning and moreover, since we formalized our ontologies in OWL, techniques for change management can be applied too.

6 Conclusion

In this paper we have presented a domain analysis over a significant period of time leading to a set of ontologies corresponding to the same views of a same domain over different periods. We analyzed this set of ontologies in order to define new modelling elements dealing with ontology evolution. Moreover, we also illustrate the potential contribution of our proposition through an example dealing with information retrieval. We believe that the evolution features we have defined consist in an important step towards automatic ontology evolution. This will be possible if we find a way to analyze the corpus of documents automatically. Nevertheless, our approach needs to be strengthened mainly through the proposition of good metrics that will be able to characterize as faithfully as possible the status of knowledge in a corpus of documents from an evolution point of view. Therefore, our future work will concern on one hand the definition of such metrics and on the other hand, the proposition of a formal set of operators able to, given a corpus of documents, update automatically the appropriate elements of the ontology we have introduced in this paper.

References

1. Avery, J., Yearwood, J.: dOWL: A Dynamic Ontology Language. In: Proceedings of the IADIS International Conference WWW/Internet 2003, Algarve, Portugal, IADIS (2003) 985-988
2. Bachimont, B., Isaac, A., Troncy, R.: Semantic Commitment for Designing Ontologies: A Proposal. In: 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW'02). Volume LNAI 2473., Sigüenza, Spain (2002) 114-1213.
3. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American **284**(5) (2001) 34-43
4. Biscaccianti, A., Renard, P.: The Cooperative Contextual Change Model: a Systemic Approach to Implement Change while Preserving Stability. Cahiers du CEREN **4** (2003) 1-16
5. Flouris, G.: On Belief Change and Ontology Evolution. PhD thesis, University of Crete, Heraklion (2006)
6. Guelfi, N., Pruski, C.: On the use of Ontologies for an Optimal Representation and Exploration of the Web. Journal of Digital Information Management (JDIM) **4**(3) (2006)
7. Guelfi, N., Pruski, C., Reynaud, C.: Towards the Adaptive Web using Metadata Evolution. In Calero, C., Moraga, M.Á., Piattini, M., eds.: Handbook of research on Web information systems quality. Idea Group Publishing (2007)
8. Guelfi, N., Pruski, C., Reynaud, C.: Les ontologies pour la recherche ciblée d'information sur le web: une utilisation et extension d'owl pour l'expansion de requêtes. In: Proceedings of the Ingenierie des Connaissances 2007 (IC07) french conference, Grenoble (July 2007)

9. Hirst, G., St-Onge, D.: Lexical Chains as Representation of Context for the Detection and Correction Malapropisms. In Fellbaum, C., ed.: WordNet: An electronic lexical database and some of its applications, Cambrige, MA, The MIT Press (1998) 305-332
10. Jiang, J., Conrath, D.: Semantic Similarity based on Corpus Statistics and Lexical Taxonomy. In: Proceedings on International Conference on Research in Computational Linguistics, Tapei, Taiwan: Academia Sinica (1997) 19-33
11. Klein, M.: Change Management for Distributed Ontologies. PhD thesis, Vrije Universiteit Amsterdam (2004)
12. Maurer, R.: Beyond the Wall of Resistance: Unconventional strategies that build support for change. Bard Press (1996)
13. Plessers, P.: An Approach to Web-based Ontology Evolution. PhD thesis, Vrije Universiteit Brussel (2006)
14. Resnik, P.: Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In: IJCAI. (1995) 448-453
15. Stojanovic, L.: Methods and Tools for Ontology Evolution. PhD thesis, University of Karlsruhe, Universität Karlsruhe (TH), Institut AIFB, D-76128 Karlsruhe (2004)
16. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene ontology: tool for the unification of biology. The Gene Ontology consortium. Nat Genet **25**(1) (May 2000) 25-29