Indexing Corporate Memories through Ontologies

Djida Bahloul, Youssef Amghar, Pierre Maret
INSA de Lyon – LISI, 7 avenue Jean Capelle.
69621, Villeurbanne – France.
dbahloul@lisi.insa-lyon.fr, {amghar,maret}@insa-lyon.fr

Abstract

In the context of Knowledge Management, we carry out a Corporate Memories (CM) project for the Company CIRTIL[1]. Our purpose is to focus on the modelling of the application domain. It is built as a domain ontology with a structure supporting a semantic model based on ontological relationships. In this paper we, present our S^3 model which permits to model knowledge and to index documents. We also show how semantic indexing of technical documents can be improved by mean of the domain ontology. We show finally the interest of our model with the implementation of a prototype.

Keywords

Semantic indexing, ontology, corporate memory, knowledge modelling.

1. INTRODUCTION

Corporate Memories (CM) require abilities to manage disparate information and heterogeneous sources in order to make knowledge accessible to the adequate users of the enterprise. CM must also consider the integration and the storage of knowledge contained into electronic documents or contained into knowledge bases. According to us, to permit the success of using and maintaining the CM, it is important to consider employees as the hard core of the system. Indeed, employees can participate to capture knowledge and to structure it into the CM. The objective of our research works aims to offer to users a methodological assistance and tools enabling the knowledge management. To achieve this objective, our approach takes advantage of both: the contribution of ontologies as proposed by the Artificial Intelligence (AI) community and the documentary indexing such as defined in the domain of the library science. The role of ontology is the representation and the modelling of knowledge. "*An ontology is an explicit specification of a conceptualisation"* [Gruber, 1993]. The role of documentary indexing is the matching between the represented knowledge and users' queries. In this paper, we propose an indexing model which is more efficient than a simple taxonomy of concepts. Therefore, we build a domain ontology which has a significant capacity of expression thanks to the possibility of introducing semantic links, structural links and subsumption links between concepts. We show in this paper how to facilitate the indexing of technical documents through the ontology. We outline our approach for the construction of an ontology based on ontological relationships. This is followed in section 2 by a description of our model called S^3 allowing both representation and indexing of knowledge. In the section 3, we apply this model to build an ontology and to index the CM of a company.

2. RELATED WORKS AND POSITIONING

Ontologies can constitute a component of a CM: they can be explored by the end-user to discover the organization processes and business objects of the enterprise (e.g. enterprise ontology), or to study methodically a specific technical domain (e.g. domain ontology), etc. Ontologies are then used as a coherent support to describe and to share knowledge. "*Ontologies constitute the glue that binds knowledge subprocesses together. Ontologies open the way to move from a document-oriented view of Knowledge Management to a content-oriented view, where knowledge items are interlinked, combined, and used.»* [Staab and al., 2001]. In fact, ontologies are used more and more in KBS (Knowledge Management System) development. For example, projects such as SHOE [Heflin and Hendler, 2000] and Ontobroker [Benjamins and Fensel, 1998] use ontologies to improve the searching abilities on the World Wide Web. Both systems provide logical reasoning based on ontological definitions. In [Gandon, 2001]; [Gandon and al, 2002], CoMMA project offers a solution to implement a CM based on ontologies and agent technology. It promotes a wide vision of the document retrieval issue that could be applied to several cases. The memory is composed of heterogeneous evolving documents, structured using semantic annotations expressed with concepts and relationships provided by a shared ontology. In others approaches, ontologies are exploited to organize the knowledge and to support computational design. For instance, the approach for ontology-based knowledge management [Staab and al., 2001] includes a tool suite and methodology for developing ontology-based Knowledge Management systems. Among these tools, OntoAnnotate tool allows users to create objects and describe them with their attributes and relationships.

This outline of the state of art shows that modelling based-ontology is interesting in the frame of CM. Nevertheless, the difference between our approach and related works consists in using ontologies to carry out the documentary indexing of formalized knowledge. We understand by documentary indexing, *"the operation which consists to describe and to characterize contents of documents by using representative concepts."* In our approach, this operation is carried out by the ontology (instead of a documentary language). We argue that ontologies can guarantee a sufficient indexing because, they introduce a host of structural and conceptual

[1]CIRTIL : Centre d'Informatique Régional du Traitement de l'Information Lyonnais which supports this work.

relationships including super class/subclass/instance relationships, property values time relationships, and others depending on the used language representation [Saadani and Bertrand-Gastaldy, 2000]. In addition, an ontology can also infer that one concept is a special case of another because the logical definition of each concept can be compared. If the concept C2 satisfies the requirements being a specialization of concept C1, then C2 can automatically be classified below C1. We consider thus, the ontology adequate to allow indexing knowledge.

3. TOWARDS AN INDEXING APPROACH BASED ON ONTOLOGY

3.1 Complexity of the domain of knowledge

Our approach is proposed in the frame of the design of a CM for the company CIRTIL. This company aims to save and capitalize its knowledge and its know-how concerning the production activities related to technical incidents on database servers and applications. We aim at providing users with support tools for indexing knowledge especially contained into technical documents. Knowledge of this company has many senses according to the context (administrator of database *vs.* administrator of administrative department). As well, synonymous terms are used in different situations (application *vs.* software), and words have a large variety of different links and interrelationships (report is a kind of documents and chapter is a part of a document). This requires a model based on a formal domain ontology which considers all those characteristics. The interest of such model relies in the role of these relationships in the semantic representation. We studied some ontological relationships composing ontologies and we concluded that these relationships define and enrich the semantic between the concepts. Indeed, the type of relationship can change the semantic as showed in the following example. Thus, two concepts C1 and C2 linked by two distinct relationships R1and R2, give two different semantic:

R1(C1,C2) gives S1

R2(C1,C2) gives S2

Let us consider two concepts *"Incident"* and *"Employee"* and two relationships *"reported-by"* and "treated-by" (Figure 1). Semantic can be variable according to the relationship between each of these two concepts. The meaning of the first representation is different from that of the second representation. Thus, *Reported-by (Incident, Employee)* means that a given incident has been notified by an employee during his task; and *Treated_by (Incident, Employee)* meaning that an employee has resolved the incident.

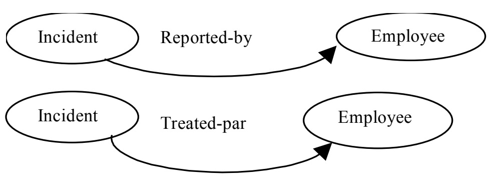

Figure 1 - Relationships for Semantic

This complexity requires an indexing model that is most possible representative of the domain.

3.2 An Indexing Model on Ontological Relationships: The S3 model

To elaborate the S3 model, we first built an ontology called ***OntoCIRTIL*** that intends to model technical incidents, employees who treat these incidents, products concerned by these incidents and the entities required for the characterisation of these incidents, products and employees. The particularity of *OntoCIRTIL* is to model and to represent knowledge according to various facets and viewpoints by taking advantage of conceptual relationships. The S^3 model proposed in this paper is based on three views giving three spaces for organizing semantic of knowledge fragments represented by concept of domain ontology. These three spaces are described below:

1. Semantic space: This space gives a view allowing discovering concepts through semantic links. This space is domain dependent because relationships used here are defined and interpreted by users. Our environment offers primitives to define such relationships such as: create, delete, and rename …a semantic link. The graphs path algorithms make possible to navigate in this space in order to discover concepts: to displace from a concept to another according to the semantic link defined by the expert. This space embeds *semantic links.*

2. Structural space: This space makes proposes to structure knowledge considering the structural dimension of the knowledge fragments. These fragments are represented by concepts related according to aggregation links (derived from the structuring mechanism of the Object-Oriented Approach). Besides the graphs path algorithms, it is interesting to define a zooming function. Its goal is to discover various levels of the selected concepts. This zooming function allows exploring deeply the various parts composed one concept. This space allows representing *structural links*.

3. Subsumption space: This space makes it possible to organize the definitions of concepts by using subsumption links. This space increases semantic space by using a new dimension which allows improving concepts in term of related definitions. The graphs path algorithms will make it possible to find a concept about a large definition. In fact, this space allows organization of knowledge in manner to permit employees to retrieve concepts (and related documents) from definition, which concerns their ancestors in the subsumption graphs. This space includes *subsumption links.*

3.3 Description of link types of S3

In our model each concept is linked with one (or several) other (s) concepts by at least a type of links: *Semantic link*, *Structual link*, *Subsumption link.* Each type of relationships interprets a particular semantic between two concepts. To model these relationships, our approach is based on manual linguistic analysis. We applied a same method adopted to define the concepts, to choose the representative links, i.e. we proposed several lists, which were modified then validated by employees according to the ontological commitment.

1. Semantic links*:* The main characteristic of this type of relationship is such as the nominations of links related to the usual language (natural language) of the community for which ontology is available. Note that, representation of semantic links, it means the nomination of relations, is not standardized. It varies according to each designer. It can be represented in: verbs or prepositions [Sherratt and Schlabatt, 1990]; verbs or nouns [Heeren and Collis, 1993]; verbs or logical connectors [Malone and Dekkers, 1984]. This facilitates their use in especially for retrieving knowledge in the CM because, these relationships are considered as keywords. Semantic relationships express clearly "evident" and no ambiguous knowledge. Examples of semantic links used in our ontology are: *treated-by, written, organized*, *causes, necessitates, implicates.* Note that those relationships allow the bi-directional and opposite (inverse) semantic expression.

2. Aggregation links: Structural links constitute a kind of relationships, which expresses a strong property between the whole and the parts, as well as subordination between the existence of the parts and the whole. The parts associated to a non fixed multiplicity can be created after the composite itself, but once created, they "lives" and "dies" with him (i.e. they share its duration life). Such parts can also be explicitly withdrawn before the death of the composite. The composition can be recursive. This type of relationship represents links of dependence between concepts. The choice of the nomination of the links (*"Part-of"*, *"component-of"*) depends on the meaning that we wish to represent. For example, to express that the *Functional incident* and the *Technical incident* belong to the *Incident Family,* the link *component-of* is better adapted than *part-of.*

3. Subsumption links:

In the simplest case, ontology describes a hierarchy of concepts related by subsumption relationships. This relationship is largely used and it is considered as the foundation of ontology, because all ontologies are presented in taxonomies [Guarino, 1997]. In such taxonomy more general concepts *subsume* more specific concepts. This allows information about concepts to be associated with their most general concept, and it allows information to filter down to more specific concepts in the taxonomy via inheritance. To name the relationship, we can use the *"is-a"* link or *"kind-of"* according to the adequate semantic of the context. Formally, an "*is-a"* link is an implicit subset/superset relationship between the two concepts. A simple example of this relationship is the *Incident is-a Event.* Note that, this relationship characterizes the hierarchical relationships between the key words into thesaurus.

In addition of these three types of conceptual relationships, our indexing model proposes a mechanism, which allows deducting new knowledge in order to enrich a semantic. In some case, knowledge is also defined with rules*: a Chief Project Manager is a person who manages project*. These rules permit to express implicit useful knowledge of information retrieval. We write some constraints that permit to control semantic relationships between concepts in order to support clarification reasoning. Inference link functionality is supported by inference engine, which uses concepts of ontology to derive factual knowledge that is only provided implicitly. The inference engine is used to derive information that is implicitly present in resources of knowledge without requiring that all information is complete materialized by relationships. Unfortunately, the inference link which is very important in the context of information retrieval, is not present in the thesaurus.

4. EXPERIMENTATION

This part focuses on the Knowledge Base and indexing tool "*KnowIndex*". We explain how we partially implement structural space of the model S^3; then we present the mechanism of indexing process.

4.1 The Knowledge Base

Knowledge about technical dysfunction (failure server, unavailable application, defective networks, etc.) and documents of resolution of these incidents represent a category of crucial knowledge for the company. Employees and decision-makers need to capitalize this knowledge, within sight of access and re-use this knowledge in order to improve quality of services for customers. In order to contribute to build an environment for CM we developed a tool for hitches management called ***MaTIP*** for *"**Ma**nagement of the **T**echnical **I**ncident **P**roject"*. The main objective is to capitalize data, information and knowledge allowing, as well as, the identification, the management and the anticipation of dysfunctions and technical anomalies that occur during the exploitation time of the software applications by customers. This *Knowledge Base* supports the ontology *OntoCIRTIL* (concepts and relationships identified during the conceptualization process). Validated concepts and relationships are modelled by the class UML diagram. In class diagram of UML, the concepts are represented by class and relationships by arcs. The model formalized by UML is translated during this stage of the development of the Ontology into RDF. Compared to existing knowledge representation standards, e.g., KIF [Genesereth and Fikes, 1992], KRSL [Allen and Lehrer, 1992], or DAML+OIL [Broekstra and al., 2001], RDF can be considered as a not very expressive formalism. Nevertheless, it offers an adequate expressiveness for our application needs. RDF uses a simple data model expressed in XML syntax to represent properties of resources and their relationships. The main element of RDF is the notion of *rdfsResource*. This one is at the top of the *class* hierarchy and is *subclass* of itself. A *rdfsResource* has a *rdfstype* relationship with a *rdfsClass,* and a *rdfslabel* relationship with a *rdfsLitteral* and may also have some literals as *comments.* One of the resources in our model is "***Relation". It*** is a super class for ***Structural link***, ***Semantic link*** and ***Subsumption link.*** Also, each type of these links has a value which is its object. For example, ***linked*** is the value of the resource "***Relation".***

Figure 2 presents an example of the representation of the relationships and its properties.

```
<!-- Representation of relationships -->

    <rdfs:Class rdf:ID="relation" />
    <rdfs:Class rdf:ID="structural_relationship"
    />
        <rdfs:subClasseOf rdf:resource="#
    relation" />
    </rdfs:Class>
    <rdfs:Class rdf:ID="linked" />
    <rdfs:Class rdf:ID="component of">
        <rdfs:subClasseOf rdf:resource="#
        linked" />
        </rdfs:Class>

<!--  Properties -->
    <rdf:Property rdf:ID="link_ name">
    <rdfs:domain rdf:resource="# relation" />
    </rdf:Property>
```

Figure 2: Example of a representation using RDF

3.2 Indexing tool

We designed a prototype based on Java and XML technologies to implement the indexing engine *"KnowIndex"* that supports the semantic S^3 model. One of the objectives of our contribution is to integrate the indexing operation into the daily activities of the employees' company. In fact, *"KnowIndex"* is an application that employees can use easily to index or to retrieve formalized knowledge. The indexing technique can be achieved through the ontology considered as an indexing resource. The indexing mechanism is divided into three steps; we address each of them in the following paragraphs.

1. Selection of knowledge: The selection of knowledge is achieved in the usual environment of employees' work, for instance during the use of *Microsoft Word application (it could be Lotus Notes or any* traditional employees' office tool. This selection concerns as well as, all content of a document or only a part of a document (paragraph, sentence,…). Once the text is selected, *"KnowIndex"* generates automatically the selected knowledge into an XML file.

2. Selection of representative's concept: This step consists in choosing relevant concepts in the ontology in order to associate them to the selected knowledge. Notice that the employee himself achieves this delicate phase. To help employees, we reserved the left side of the *"KnowIndex"* interface to display the ontology. Thus, users can select either node or leaf by navigating in the tree of the ontology. We conserve the right side of the interface to the *"Editor"*. Its role is to visualize for each concept all knowledge identifying these concepts, particularly: the definition of concept, properties and relationships. When the user displays a concept, the concept definition, super-concepts, sub-concepts, relationships and instances of the concepts are displayed in the frame on the right**.** These characteristics and information about concepts and relationships guide the user to choose a concept compared with another concept.

3. Indexing and Classification: This step is based on the conceptual model. Indexing task is initiated by the user, thanks to the function *"Add"* associated to each concept. By clicking *"Add"* on one of top-level concepts or the sub-concept, *"KnowIndex"* creates automatically correspondences between this concept and selected text. In addition, to link synonyms of a concept with the indexed document, the system supports an automatic mechanism which permits the linguistic analysis.

3.3 Knowledge Retrieval

To show the double role of the ontology *OntoCIRTIL*, i.e. the restitution of knowledge and its indexing, we propose a single user interface. Ontology can be seen as a semantic interface for accessing CM. For that, the ontology appears on the left of the user interface in a tree view representation. To simplify the indexing task to the end-users, we associate to each concept two operations *"Add"* and "*Show"* (see Figure 3). These operations facilitate the correspondence between a document (and part of document) and the appropriate concepts. As previously mentioned, thanks to the *"Add"* operation, knowledge can be captured and classified into the CM. The *"Show"* operation is used to search for knowledge. Selected text and their links address are classified and stored into the knowledge base. The choice of concepts is manual because it is achieved directly by the end-user. This user, can be able to determine the representatives concepts while consulting a document. Furthermore, the *"KnowIndex"* module allows the linguistic treatment, makes correspondence between selected text and concepts, and classifies indexed documents and their concepts in the knowledge base. Thus, the integration of contents or news documents into the Knowledge Base can be relevant, because it is ensured by the users for themselves.

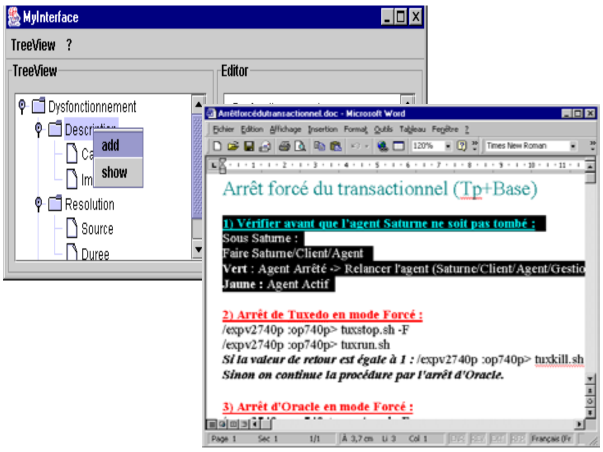

Figure 3 - *"KnowIndex"* : the indexing operation

We observe on the right side of this figure a technical document in Word format. The user locates by traversing the document a paragraph describing a given dysfunction. To record this fragment of Knowledge, the user selects in a traditional way by using the mouse. Then, he browses the ontology in the application "KnowIndex " in order to find, the representative concept. Once the relevant concept is added, the system automatically connects the concept to the paragraph.

5. CONCLUSION

The advantage of the ontology according to knowledge engineering researches is that the concept must be unambiguous and unique. In addition, it is necessary and important to take advantage of the richness of the ontological relationships. This what guided our approach which leads to the development of a semantic model. In this paper we have. We presented the S^3 model and its components. We explained that the ontological relationships allow a strong semantic. In this context, we proposed three type links, which we analyzed and modeled. The experimentation applied to a project of CM permits first, to expose real needs and then to test and validate our approach. Thanks to the indexing tool *"KnowIndex"* dedicated to assist actors, we developed a process which, allows:

- the selection of knowledge;
- the selection of representatives' concepts;
- and the Indexing and Classification.

The advantage of our contribution is to develop an indexing model which exploits the ontological relationships in order to allow a level of abstraction and a semantic expressiveness. Nevertheless, the major problem of building our model is the semantic formalisation of the cognitive knowledge. This is accentuated by the fact that the realization of the model is time-consuming. The application of the model to a small corpus showed that the approach based-ontology can proven time-consuming in particular when the ontology must be built. Also, only the structural space was formalized and implemented. Thus, our proposal will become more interesting while we formalize and implement the semantic space and the subsumption space and obviously by integrating the inference mechanism. We also plan -in the future works- to evaluate the relevance of indexing process based on the ontology compared with that based on the thesaurus.

6. REFERENCES

[Broekstra and al., 2001] D. Broekstra and al. Enabling Knowledge Representation on the Web by extending the RDF Schema. *10th International WWW Conference,* Hong Kong 2001.

[Gandon, 2001] Gandon. F. Engineering an Ontology for a Multi-Agents Corporate Memory System In *proceedings of ISMICK 2001 Eighth International Symposium on the Management of Industrial and Corporate Knowledge,* 22-24 October 2001, Université de Technologie de Compiègne, France, pp 209-228.

[Gandon and al, 2002] Gandon.F, Dieng-Kuntz.R, Corby.O, Giboin.A. Semantic Web and Multi-Agents Approach to Corporate Memory Management. *17th IFIP World Computer Congress IIP Track - Intelligent Information Processing*, Eds Musen M., Neumann B., Studer R., p. 103-115, August 25-30, 2002, Montréal, Canada.

[Genesereth and Fikes, 1992] M. R. Genesereth and R. E. Fikes (eds.). 1992. Knowledge Interchange Format*, Version 3.0 Reference Manual*. Computer Science Department, Stanford University, Technical report Logic-92-1.

[Gruber and Poli, 1993] Nicola Guarino and Roberto Poli, Toward Principles for the Design of Ontologies Used for Knowledge Sharing, *In Formal Ontology in Conceptual Analysis and Knowledge Representation,* Kluwer Academic Publishers, in press. Substantial revision of paper presented at the International Workshop on Formal Ontology, March, 1993, Padova, Italy.

[Guarino, 1997] Guarino, N. 1997b. Understanding, Building, and Using Ontologies: A Commentary to "Using Explicit Ontologies in KBS Development", by van Heijst, Schreiber, and Wielinga. *International Journal of Human and Computer Studies (in press)*(46): (in press).

[Heflin and Hendler, 2000] Heflin, J. and Hendler, J. Semantic interoperability on the web. In *Extreme Markup Languages 2000*.

[Heeren and Collis, 1993] Heeren, E. and Collis, B. Design considerations for telecommunications-supported cooperative learning environments : concept mapping as a telecooperation suppport tool. *Journal of educational multimedia and hypermedia*, 4(2), 1993, 107-127.

[Malone and Dekkers, 1984] Malone, J. and Dekkers, J. The concept map as an aid to instruction in science and mathematics. *School science and mathematics*, 84(3), 1984, 220-232.

[Saadani and Bertrand-Gastaldy, 2000] Saadani.L, Bertrand-Gastaldy.S. Conceptual Maps and Thesauri: A Comparison of Two Models of Representation from Different Disciplinary Traditions CAIS 2000. Canadian Association for Information Science *Proceedings of the 28th Annual Conference*.

[Staab and al, 2001] Staab S., Schnurr H.-P., Studer R., Sure Y.(2001): Knowledge Processes and Ontologies. *IEEE Intelligent Systems*. 16(1), Special Issue on Knowledge Management, January/February 2001.

[Sherratt and Schlabatt, 1990] Sherratt, C. S. and Schlabach, M. L. The application of concept mapping in reference and information services. *RQ*, 30, 1990, 60-69.