IJCAI'2005 Workshop on Knowledge Management and Organizational Memories

Rose Dieng-Kuntz (INRIA, France)

Nada Matta (UTT, France)

Co-chairs

Knowledge Management and Organizational Memories

Knowledge Management (KM) is one of the key progress factors in organizations. It aims at capturing explicit and tacit knowledge of an organization in order to facilitate the access, sharing, and reuse of that knowledge as well as creation of new knowledge and organizational learning. KM must be guided by a strategic vision to fulfill its primary organizational objectives: improving knowledge sharing and cooperative work inside the organization; disseminating best practices; improving relationships with the external world; preserving past knowledge of the organization for reuse; improving the quality of projects and innovations; anticipating the evolution of the external environment; and preparing for unexpected events and managing urgency and crisis situations. One approach for KM consists of building a corporate memory or organizational memory (OM). Several techniques can be considered, according to the type of organization, its needs and its culture: knowledge-based approaches, document-based approaches, workflow-based approaches, CBR-based approaches, CSCW and cooperative approaches, ontology-based approaches, corporate Semantic Webs, Web-based approaches, agent-based approaches, distributed OMs, etc.
Several scenarios of KM can be tackled through OMs: project memory, skills management, communities of practice, strategic or technological watch, etc. The workshop aims at gathering researchers from multiple disciplines, industrial participants and students in order to discuss models, methodologies, techniques and application scenarios useful for building, using, managing, evaluating and evolving corporate memories.

Organizing Committee

Jean Paul Barthès, University of technology of Compiegne (France)
Rose Dieng-Kuntz, INRIA Sophia Antipolis (France) (co-chair)
Knut Hinkelmann, University of Applied Sciences Solothurn (Switzerland)
Ann Macintosh, International Teledemocracy Centre (United Kingdom)
Nada Matta, Université de Technologie de Troyes (France) (co-chair)
Ulrich Reimer, Bauer & Partner (Switzerland)
Carla Simone, Universita' di Milano Bicocca(Italy)

Program committee

Andreas Abecker, FZI, Karlsruhe University, Germany
Jessica Chen-Burger, University of Edinburgh, UK
John Domingue, Open University, (UK)
Jean-Louis Ermine, INT (Paris, France)
Fabien Gandon, INRIA-Sophia-Antipolis (France)
Michel Grundstein, Mgconseil (France)
Robert Jasper (USA)
Gilles Kassel, LARIA, Amiens (France)
Myriam Lewkowicz, Tech-CICO, UTT (Troyes, France)
Frank Maurer, University of Calgary (Canada)
David G. Schwartz, Bar-Ilan University (Israel)
Camille Rosenthal-Sabroux, LAMSADE, Paris-Dauphine, (France)
Carla Simone, University of Torino (Italy)
Steffen Staab, University of Koblenz, (Germany)
Manuel Zacklad, Tech-CICO, UTT (Troyes, France)

Acknowledgements

We thank very much the members of the program committee and of the organization committee for their careful rereading of the 14 papers submitted to the workshop. We also thank very much Patricia Maleyran (INRIA Sophia Antipolis) for her logistic support for the management of the submitted papers.

Table of Contents

A generic model of corporate memory: application to the industrial systems

Lotfi ADMANE

Institut National d'Informatique

BP 68 M, Oued Smar, 16270, Algiers, Algeria

l_admane@ini.dz

Abstract:

This paper presents an industrial system model and a model of corporate memory supporting these models of systems. The corporate model sees the latter through two types of knowledge. Skill Knowledge, which constitutes the main capital knowledge of the company and refers to its basic skill. Theme Knowledge represents a specialized knowledge or a knowledge relating to a given field.

The objective of the corporate memory model, named ReCaRo, is to capitalize knowledge by allowing its systematic re-use. ReCaRo builds corporate memories which have a multi memory architecture. This architecture means that every memory will be made up of five communicating memories. We will answer to two questions: 1) how can we model an industrial system? and 2) how can we implement the re-use principle in such systems.

Key words: Corporate memories, meta-model, re-use, ReCaRo

1. INTRODUCTION

This work was motivated by the observation of the strong tendency of the today's company to be specialized. To be effective, the company standardizes its processes and its resources. it, often, handles the same entities for different actions. These entities can be physical objects, rules, processes, etc. Progressively and using these entities, the company constitutes, in the long run, a capital knowledge. However, this capital knowledge is often scattered on the experts of the company and in documents. It is very volatile. The objective is to collect it, organize it and preserve it for re-use purposes. This preservation is done, very often, through the concept of corporate memory (Brooking, 2000; Pomian, 1996; Vanheijst, 1996). This capital is then re-used in different situations in order to reduce the costs and the times of development. It is the concept of the company learning (knowledge creating company) (Nonaka, 2002) and the working knowledge (Davenport, 2000).

In this paper, we propose a corporate model and a generic corporate memory model supporting it. The corporate memory model is named ReCaRo which is the acronym of REsource, CAse and ROle which are the main concepts of the memory. We emphasize, particularly, n the re-use problem of the knowledge contained in the corporate memory.

To build the corporate memory, we propose an approach in two stages. In the first stage, the objective is to propose a modelling of the company (in this research, we are interested in the industrial companies which we will call industrial systems) which supports the development of corporate memories allowing the easy and systematic re-use of capital knowledge. In a second stage, the objective is to propose a generic corporate memory deduced from the industrial system.

This paper is divided into four main parts. The first one presents the concept of knowledge and its reuse. The second one presents the model of the industrial system, the third presents the proposed generic model of the corporate memory and, finally, the last one presents the application of the ReCaRo model for the conception of a corporate memory in the design of industrial systems for liquid hydrocarbons transportation.

2. CORPORATE KNOWLEDGE AND REUSE

In this section, we give some theoretical elements regarding the knowledge management in the company as well as the concept of knowledge re-use. We will not try to define the corporate knowledge, but we will position ourselves directly within a framework of corporate knowledge management while trying to determine its main characteristics. Among these characteristics, we have the usual distinction between data, information, comprehension and knowledge (Boersma, 1996). The aspects of puting in context the knowledge were particularly developed in sociology (Vink, 1999) and in cognitive psychology (Poitou, 1996). Pomian (1996) introduced the distinction between " knowledge ", " information " and " data ". Hatchuel and Weil (1995) introduced the distinction between know-how and the knowledge. Danveport and Prusac (2000) evokes the link between knowledge and context. Finally Alquier (2003) and Ermine (1996) introduce the concept of knowledge system. The definition of the concept of knowledge in the organization remains very prone to discussion. For us, we remain in the optics of (Davenport, 2000) which is interested in the problem of the re-use of knowledge.

There is a multitude of classifications of corporate knowledge. In it, we find the typology of KADS (Breuker, 1995), which classifies knowledge according to the specialization. The typology of Brooking (2000) and Pomian (1996) classify knowledge according to the type. The typology of Grundstein (2000) classifies knowledge according to the mode of use. The typology of Colins (1992) classifies knowledge according to the degree of exactitude and distribution in the company.

An important point in this paper is the re-use of this knowledge. In the most general case, " To re-use " means, to use again existing elements. In the case of knowledge management, the term re-uses means, to use one or several existing components resources in order to create new components with a minimum search time and few adaptations. Thy have to be lower than those necessary to the construction of new components offering the same functionalities (Demourioux, 1998). Most works on the re-use principe introduce the concept of reusable component (Bushmann, 1998). These works introduce the reusable component as being an object of the organization described through a set of characteristics, often descriptive.

For simplification needs, these reusable components are often gathered in classes. Demourioux (1998) proposed a typology of reusable components for the design of information systems. Projects DECIDE (Alquier 1997) and PRIMA (Alquier, 2000) propose a classification of the reusable components for cost management in the design of new products.

3. THE INDUSTRIAL SYSTEM MODEL

In this section, we will answer the following question: how can one model an industrial system from a point of view of a corporate memory?

The model of the company that we propose sees the company through the entities it re-uses. Any industrial system, therefore, will be modelled as a system constituted of, or handling, two types of components:

The skill components: They represent every physical or logical object which constitutes the basic skills of the company.

The theme components: They represent all that is necessary to operate the system, such as the consumed resources, the inputs and the outputs of the system etc.

The two components' types are in interaction and are connected according to a logic suitable for the system. These components constitute the main objects on which a capital knowledge is developed. The corporate memory model that we propose aims to capitalize this knowledge with a re-use goal.

4. THE CORPORATE MEMORY MODEL

The corporate memory model that we propose as a support to the model of the industrial system handles two types of knowledge and has a multi-memory architecture. In this model, every memory is built around five main ones that we present below:

Reusable resources' memory (R. R. Memory): In the first stage, it is necessary to take an inventory of the capital corporate knowledge and the capital skill and theme knowledge. This is done through an inventory of the industrial system's components. Each component will be listed in the form of what we will call the reusable resource and will constitute the memory of the reusable resources. Each reusable resource answers a well defined model. To conceive the memory of the reusable resources comes down finally to take an inventory of the various types of components and modelling each one of these types.

Roles' memory: In a second stage, it is necessary to build the memory of the roles. A role describes an element of the reusable resource context. The objective of the roles is to ensure the connection of a resource's use to its context of use. Very often a role comes down to taking an inventory of the set of roles that describes and comments the capital knowledge, and proposing a model for each type.

Cases' memory: A case of use represents the description of the use or the re-use of a reusable resource in a given context. It is defined by a reusable resource, to which it was decided to add a set of roles. Each role is carrying a single semantic which relates to the described part of the context. In a third stage, it is necessary to take an inventory of all the experiences, around these reusable resources. Each experience is represented by a case of use. To constitute the memory of the cases, it is necessary to take an inventory of all the types of cases. Each type of case will be represented by a model.

Networks of cases' memory: A network of cases represents the description of the coordinated use of several cases for the realization of a common and single goal. It is an assembly of several cases using roles. In this situation, the roles keep exactly their function which is that of connectors carrying a semantic. This memory is used to describe experiences which are too complex to be described by cases.

Contexts' memory: The context is a description of the situation in which the case was carried out. The concept of context is complementary to that of a role. It is used to describe situations which are either too rare or too complex to be described by roles.

According to this, every corporate memory will be made up of those five memories. These memories are connected according to the logic described by figure 1. The cases' memory consists of the connection of the reusable resources, the memory of roles and the memory of contexts. The memory of the cases network consists of the connection of the cases' memory, the roles' memory and the contexts' memory. The global design of the corporate memory is thus summarized as follows: make, for each of the five concepts, an inventory of its various forms and propose for each form an adequate model.

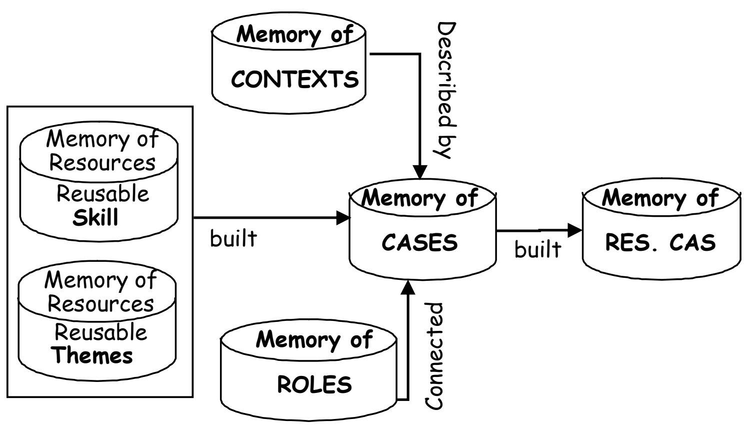


Figure 1 : General architecture of the corporate memory

5. APPLICATION OF THE MODEL TO THE SH_TRC PROJECT

The goal of this section is to validate the proposed model on a real case of corporate memories' modelling. We will present the SH_TRC project as well as the built corporate memory.

5.1. PRESENTATION OF THE SH_TRC PROJECT

This project aims to set up a corporate memory of all the capital experience and knowledge acquired during the design, the renovation, the maintenance or the extension of the transportation of liquid hydrocarbons. Each one of these actions is done through a study which gives rise to a specification, describing for example the requirements and the technical features of the future industrial system. This specification constitutes for us, in addition to the experts, one of the main sources of knowledge.

The goal of this project is to help the designers of future industrial systems to carry out their tasks in best times by assisting them in the design phase. This assistance will be done by proposing to the designers a set of components, resulting from the memory of components, and approaching the most their needs.

5.2. THE INDUSTRIAL MODEL OF THE SH_TRC PROJECT

As described in section 3, the industrial system of the SH_TRC project is seen through two types of components:

Skill components: we have listed two types of skill components. The component of type product which represents any physical element entering in the composition of an industrial facility (for example: pump, circuit breaker, pipe etc.) and the component of type process which represents all the dynamics of the industrial system.

Theme components: Among the themes studied in the project SH_TRC, the topic retained is the conception of the specification draft. The development of the latter offers to the users an assistance in the specification of future installations. In this theme, we have found two types of components: the component of type portion of text and the component of type graphic element. The component of type portion of text gathers all that was written around a skill component and the component of type graphic element gathers all the diagrams and graphs associated to a component.

5.3. ARCHITECTURE OF THE PROPOSED MEMORY IN THE SH_TRC PROJECT

In accordance with the generic architecture proposed in section 4, the memory of the SH_TRC project will be made up of five memories that we will present below:

5.3.1. Memory of Reusable Resources

This memory gathers the four types of reusable resources: product type, portion of test type, graphic element type and process type.

In this article, we are, particularly, interested in the first three reusable resources' types. The reusable resource of process type is described in (Admane, 2004).

Memory of the reusable resources of type *product*: It is the main resource among the ones of the skill type. Almost all the corporate memory requests are on it. The product is regarded as the element of the finest granularity. It can be used in the composition of an industrial facility, or in the composition of another product. The model of figure 2 describes the product isolated from any use. The structural properties describe the resource. The properties of environment describe the interaction of the resource with its environment. The properties of re-use describe the possible re-use forms of the resource.

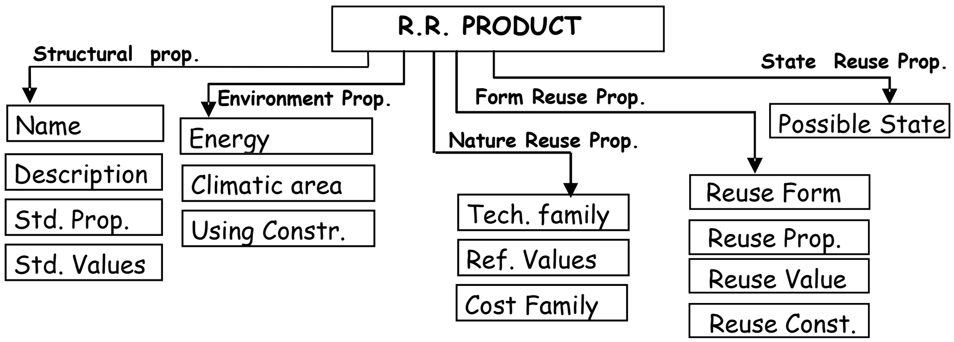


Figure 2 : Model of reusable resource type Product

Each instance of a component's model will become an element of the memory of the reusable resources of type product.

The example of figure 3 is an element of this memory. We represent this component as a record. This example describes a programmable pump. The characteristics of the re-use inform that this pump can be used as an amplifier of flow. It can also be used in manual or automatic mode.

Reusable Resource Product: pompe P217	
Name:	**ZPHMP2002-1**
Description:	**Programmable Hydraulic pump**
Standard Properties:	**Double pumping, …**
Standard Values	**Max Power: 100 bars, diameter of entry: 300 mm, …**
Energy:	**Diesel**
Climatic zone:	**Arid, tropical**
Constraints of reuse:	**Product not corrosive**
Technical Family:	**Pumping, amplification flow …**
Cost Famiy:	**High**
Reference value:	
Form of reuse:	**Amplifier of flow**
Properties of reuse:	**Flow parameters, modification section …**
Values of reuse:	
Constraints of reuse:	
Possible State:	**Automatic, manuel …**

Figure 3 : Example of a reusable resource of type PRODUCT

Memory of reusable resources of type *portion of text* : It is a resource of the theme resources' memory. This resource represents any portion of text that seems interesting. It is described by figure 4. The re-use properties of type nature describe the form or the length of the text and those of the re-use type of usage gives the type of the text (descriptive, modifying, etc.)

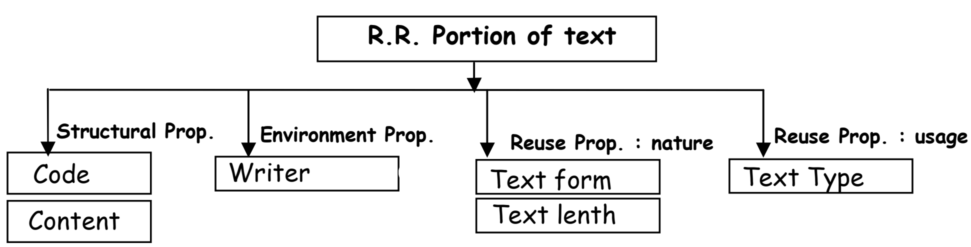


Figure 4 : Model of reusable resource of type PORTION OF TEXT

The example of figure 5 describes a reusable resource of type potion of text. It is an assembly notice of a reusable resource. It is a short technical text which describes a process. The text itself is in the

Content property.

Reusable resource PORTION of TEXT : note of assembly	
Name :	**Notice 111**
Description :	**Note of assembly of the programmable hydraulic pump**
Content :	**<Text >**
Writer :	**Name of the writer**
Text Form :	**Technical Text**
Lenth of text :	**Short**
Type of text :	**Processus**

Figure 5 : Example of a reusable resource of type PORTION OF TEXT

Me3mory of reusable resources of type *graphic element* : The reusable resource of type graphic element is, also, a resource of the theme resources memory. This resource represents any graphic element extracted from the specification which seems interesting. It is described in figure 6. The properties of re-use of the type nature describe the form or the dimensions of the graph and those of type of usage give the type of the graph (descriptive, modifying, assembly, etc;)

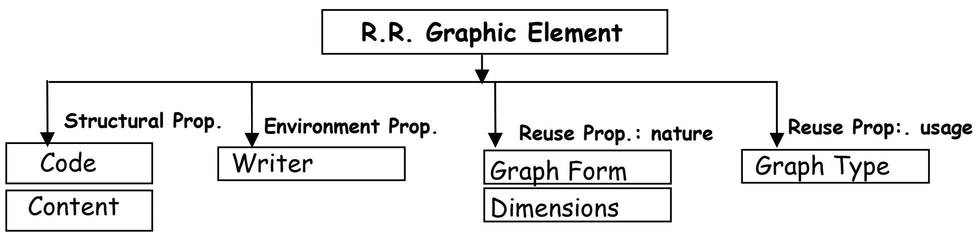


Figure 6 : Model of reusable resource of type GRAPHIC

A reusable resource of graphic type can be modelled exactly like a reusable resource of textual type.

5.3.2. Memory of Roles

The roles serve to describe all or a part of a reusable resource within a particular use. All the semantics carried by the role relates to the evoked reusable resource. The model of roles is described by figure 7.

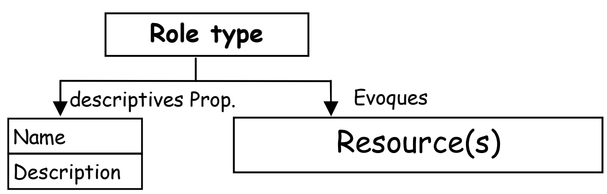


Figure 7. Role's model

In the SH-TRC project, we have proposed three models of roles.

Model of Internal role: These roles implement only one reusable resource: the current resource. The global model of this role is described in figure 7. The internal role is described by a set of properties which are: The name which is used as an identifier and a short description of the role. As an example of internal roles, we can inventory all the roles which aim are to describe the called upon reusable resource. Among these roles, we find the Mechanical role which describes the reusable resource as being a mechanical component, the Manuel_mode role which means that the component is used in a manual mode, etc...

Model of connection's role: The roles of the type connection are roles which are used to describe all the interaction that the called upon reusable resource can have with another reusable resource. This latter is called dependent reusable resource. The general model of this role is identical to the internal role except that it implements two reusable resources. These roles can be for example:

be_connected_to which informs that the current resource was connected to another, adapted_on which means that the component can be adapted on another component or commented_on which means that the component iscommented on by another resource (of type portion of text or person that we can define).

Model of mediation's role: The roles of mediation type are used to describe the way with which the called upon reusable resource comes between two dependent reusable resources. This resource is used as a mediator between the two dependent reusable resources. The role of mediation, thus, calls upon three reusable resources. We give, as an example, the following roles: To assemble which means that the reusable resource is used to assemble two other reusable resources; to annotate which means that the reusable resource (of textual type) is used to put a comment on a portion of text associated to a resource etc.

The table of figure 8 presents some elements of the roles' memory .

Types	**Roles**
Internal Roles	Physical Obj, obj Informational, machine, module, hydraulic, mechanics, electric, location, make modification, section reduction, descriptif text, operating mode Position-fonction, Modify- characteristic, suppress -componant, (joint), etc.
Connection's Roles	Adapt on, connected to, connectable with, non compatible with, used in, described by, describe, schematized by, schematize, Add-component, Obligatory composition, Optionel Composition, Specialization, Generalization, Reuse, Derivation, Equivalence, Obligatory need, Optional need, Induction, precede, following, before, after, etc.
Mediation Roles	To assemble, to adapt, annotate, informs, etc.

Figure 8. Extraction from the roles' memory

5.3.3. Memory of Models of contexts

We thought of contexts' models which make it possible to describe situations in the form of texts. Practically, we built a single model of context. This model is illustrated in figure 9.

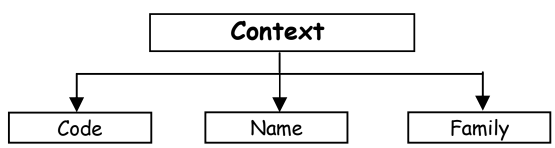


Figure 9 : Model of context

It represents the context as being a portion of text described by the context code, context name, and a family of context. We define for example the contexts:

Desert : knowledge is valid for desert regions

Renovation : knowledge is valid in a situation of renovation

Reduction of section: which means that the described knowledge is valid for problems of conducts section reduction.

5.3.4. Memory of Models of cases

The goal of these models is to represent the cases of use of reusable resources. Each model describes a type of a well defined case. A model of case is composed of the triplet: reusable resource, role and context. It is described by figure 10.

The properties Name and Description describe the case itself. The property Resource represents the resource implemented in this particular case. It results from the memory of the reusable resources to be added as a case in the memory of the cases. The property Role is used to document the way in which the reusable resource was used in this case. The role is invoked by its name. The property context describes the context in which the case was built. And the property action explains the action defined by the role on the reusable resource

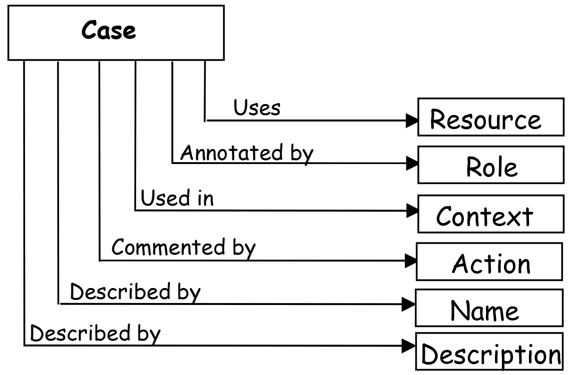


Figure 10 : Model of cases

The example of figure 11 describes a case of use of a programmable pump like a regulator of pressure.

Case: **amplification of pressure**	
Description:	<Description of the case>
Reuse Res.:	**pump P217**
Role:	Position-fonction, Modify- caracteristic, Add-component (joint)
Context:	Increase power
Actions:	Position the pump in automatic mode, increase its section, remove the joint of origin and replace it by a hermetic one.

Figure 11 : Example of a reusable resource of type CASE

5.3.5. Memory of Scenarios of cases

There exists, in reality, situations which are meaningful only after the description of several cases of use of several reusable resources. For these situations, we introduce the concept of scenarios of cases. The latter are defined like the description of the interaction of two or several cases for the realization of a common objective. The general model of the scenario of cases is defined in figure 12.

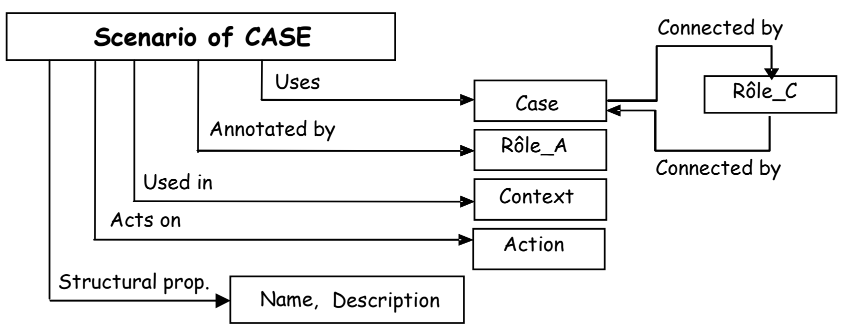


Figure 12: : model of a scenario of cases

The cases are connected by sequence or logic roles (and, or, oux etc).

For example, the connection of two conducts cannot be described by a case. It is the combination of four cases of use. The built scenario can be schematized by figure 13.

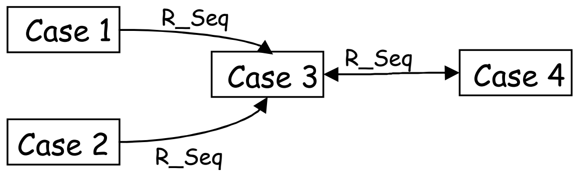


Figure 13 : Example of a scenario of cases

In this example, Cas1 means the assembly of a core of reduction (to reduce the diameter of a conduct), Cas2 means assembly of a core of increase (to increase the diameter of a control), and Cas3 means assembly of a ring of connection (connection of two conducts) and Cas4 means regulating assembly of pressure (regulation of the pressure)

6. CONCLUSION

In this article, we presented a meta-model and a methodological way for the design of corporate memories based on the re-use principle. Two ideas were developed.

The first relates to the architecture of the corporate memory. We chose an architecture multi memories which means that every memory developed according to the ReCaRo model will be made up of five communicating memories. We have the memory of the reusable resources, the memory of the roles, the memory of the cases and the networks of cases and the memory of the contexts.

The second idea relates to the implementation on industrial cases. The main problem that we had to solve was that of the definition of the concept of reusable resource in the field of the hydrocarbon transport. For that, we proposed a modelling of the industrial system through two classes of components: skill knowledge and theme knowledge.

The memory of corporate knowledge offers to the technicians all the help with the industrial systems design. The connection of the corporate memory to the documentary theme memory offers to them the assistance with specification when designing new installations.

We chose to implement the corporate memory as a data base. The set of models and reusable resources was implemented as a set of data bases. Admane & Al. (2002), Admane & Al. (2002a) and Admane & Al. (2003) give all the details for this modelling.

To capitalize, in the long run, this work, the idea is to develop mechanisms to collect produced knowledge. In our case, this knowledge is:

Generic models of reusable resources: they are collected throughout the dissemination of the suggested method. These models are standardized, and given to the designers of corporate memories as reusable generic models;

Generic models of reusable roles: they are collected in the same way as the reusable models of resources. The models of roles are standardized and classified;

Listing of the most usual roles: one could collect the roles themselves because they can be reusable in their state. Their capitalization becomes taking an inventory of those roles, organizing them and proposing them to the users.

7. REFERENCES

Admane, L. Gherbi, Hacene, A. (2003) “Réutilisation des connaissances dans une mémoire de projets : application à l’aide à la conception de cahiers de charges”. *Internal report : SI2003/007*, INI, Algérie.

Admane, L., Benferhat, R. & Hendel, S. (2002a), “Conception et réalisation d'une mémoire de projet pour l'aide à l'élaboration des cahiers des charges”. *Internal report SI2002/23,INI,* Algérie.

Admane, L., Benferhat, R. Boukraa, D., Ghomari, A., Bouchama, S., Terr S., Alquier, A.M. & Tignol,

M. H. (2002), "Un modèle de projet et une mémoire de projet pour la conception de cahier de charges**"**, *Proceedings of 7th MCSEAI,* Algérie. (pp103-110)

Admane, L., Benferhat, R., Ghomari, A. & Terr, S. (2004), "Modélisation de projets de mémoire d'entreprises**"**. *Proceedings du premier colloque international sur l'optimisation et les systèmes d'information***,** Algérie. (pp67-75).

Alquier, A. M. (2000), *Project Risk Management*, *www.esi2.us.es/prima*.

Alquier, A. M.(1997), *DECIDE : DECIon Support for Optimal Bidding in a Competitive Business Environment*, projet n°22298. Commission des communautés européennes (CCE): Programme ESPRIT.

Alquier, A.M. (2003), *Systèmes d'information et systèmes de connaissance*, Internal Report, University of Toulouse1, France.

Boersma, J. & Stegwee, R. (1996), "Exploring the issues of Knowledge Management*", Proceedings of the Information Resources Management Association International Conference*, (pp 127-135).

Breuker, J. Van de Velde, W. (1995), *CommonKADS Library for expertise modelling, Reusable problem solving components, Frontiers in Artificial Intelligence and Applications***,** IOS Press, Amsterdam.

Brooking, A. (2000), *Corporate memory: strategies for knowledge management,* Thompson Business Press, London.

Bushmann, F., Meunier, R., Ronhert, H., Sommerland, P. & Stal, M. (1998), *Pattern oriented software architecture: a system of pattern*, John Willey and Sons Ed, London.

Collins, H. M. (1992), *Experts artificiels, machines intelligentes et savoir social*, Ed. Seuil, Paris.

Davenport, T. & Prusac, L. (2000), *Working knowledge, how organizations manage what they know,* Harvard Business School Press, Boston.

Demourioux, M. (1998), *Une méthodologie d'aide à la conception des systèmes d'information fondée sur la réutilisation*, Thèse université Paris Dauphine, France.

Ermine, J.L. (1996), *Les systèmes de connaissances,* Ed organisation, France.

Grundstein, M. (2000), *Le Management des Connaissances dans l'Entreprise Problématique, Axe de progrès et Orientations*, Research Report #050207, http://www.mgconseil.fr.

Hatchuel, A. & Weil, B. (1995), *L'expert et le système,* Ed. Economica, France .

Nonaka, I. & Takeuchi, H. (2002), *The Knowledge-creating Company: How Japanese Companies Create the Dynamics of Innovation*, Ed. Hardcover.

Poitou, J.P. (1997), "La gestion collective des connaissances et la mémoire individuelle**"**, in : *J. M. Fouet, Connaissance et savoir faire en entreprise*Ed. Hermes, (pp 238-274), France.

Pomian, J. (1996), *Mémoire d'entreprise : techniques et outils pour la transmission du savoir*, Ed. Sapientia, Paris.

VanHeijst, G. VanDerSpek, R. & Kruisinga, E. (1996), "Organizing corporate memories", *Proceedings of workshop on Knowledge Acquisition for Knowledge-Based Systems* Alberta, Canada, (pp. 87-96).

Vink, D. (1999), *Ingénieur au quotidien : Ethnographie de l'activité de conception et d'innovation*: Presses Universitaires de Grenoble, France.

Annotations for managing knowledge in the Electronic Health Record

Sandra Bringay[1], Catherine Barry[1], Jean Charlet[2]
[1] Laboratoire LaRIA, Université d'Amiens,
5 rue du Moulin Neuf, 8000 AMIENS (France)
{sandra.bringay,catherine.barry}@u-picardie.fr
[2] Laboratoire STIM/DSI/AP-HP, Paris,
Hôpital Broussais, 96, rue Didot, 75014 Paris
Jean.Charlet@spim.jussieu.fr

Abstract

Practitioners still do not have at their disposal tools of health record management, allowing them to reproduce all the practices they carry out with the paper record. By positioning us in the paradigm of a documentary approach, we adopt an original vision on the human computer interfaces of the electronic health record: we consider the electronic documents with their annotations. The aim of this publication is to show the interest of the annotations to work on the documents of the health record and on the functionalities which result from these annotations (edition of documents, filtering, message, etc.). This paper is an extension of the paper presented to the IC conference[1].

1 Introduction

1.1 DocPatient project

The evolution of the information systems is one of the stakes of the health system. The mutualisation of the patients' medical data became one of the main objectives of the medical authorities. They want to simplify the exchanges of knowledge in a hospital unit, between several hospitals and others medical organisations (connection between the hospital and the city with the care networks). Indeed, the more the number of health professionals around the patient increases, the more the flow of data, information and knowledge must be fast and coherent.

Traditionally, practitioners use a collection of paper documents, the Health Record, to convey medical knowledge. From now on, this record shows its limits and, in particular, concerning the traceability and the filing. Many research teams had worked on the design of an Electronic Health Record (EHR) since the Eighties. In the central issue of this work, we find the problems of the knowledge distribution to all the actors of the organisation. Indeed, many categories of practitioners must have access to the medical knowledge but their objectives, their missions are different. So, it is difficult to propose them adapted knowledge, in the adequate format, at the right time.

Since 2002, we were part of the DocPatient project. We try to computerise the hospital health record according to this documentary approach. This project of the University of Amiens is financed by the Picardy region. It gathers a multi-field team composed of sciences for the engineer (data-processing) and social sciences (law, management, psychology). We work in collaboration with a pilot site[2] and an industrial partner[3]. We develop documentary functionalities making easier the manipulations of the electronic documents.

After a theoretical research on the concept of document applied to the medical documents [Bringay et al., 2004-b], we went out into the field (in the paediatric unit) for a multidisciplinary study: observations of the practices, analyses of the needs, study of the various supports of knowledge, specifications, realisation and evaluation of models. Thank to this study, we understood how the practitioners currently transcribe knowledge on perpetual supports (paper or electronic documents, personal notes) and re-use it in contexts sometimes very distant in time and space from their creation. As specified by Hardstone et al. [2004], the creation of an electronic record, medical or not, is complex. We must take into account the daily actions and social interactions on which practitioners base their multidisciplinary work. With this study, we get some directives about the construction of the EHR and we justified our choice of a documentary approach.

In addition, we find annotations in the paper documents of the health record. These annotations support a part of the daily actions and interactions of the practitioners. We have chosen to maintain information in these annotations which is relevant to the record. Therefore, we went out into the field a second time to analyse this particular practice. We

[1] French Knowledge Engineering Conference http://www-sop.inria.fr/acacia/afia2005/IC.html

[2] The unit of paediatric intensive care and neonatal medicine of Amiens hospital

[3] The company UNI-MEDICINE http://www.uni-medecine.com/

realised a theoretical study of the concept of annotation and we wrote a definition of an annotation [Bringay et al., 2004a]. We studied the existing software of annotations. We got some directives about a software component of annotations, which can be link to an EHR. We affirm now that a functionality of annotations is one of the solutions to solve some problems of handling of the electronic medical documents.

1.2 Annotations in the literature and annotations in the EHR

In the community of people aiming at building research tools, the term "annotation" generally designates the metadata (e.g. Dublin Core[4]) and the descriptors of contents (e.g. XML tags) used to qualify and describe a resource in order to retrieve it in a set of resources and to re-use it (presentation of the same content with several layouts, creation of tables of contents or index). Either human beings, helped or not by the machines, or machines create these annotations. The machines treat them to answer the users' requests. For example, with the traditional Web, human beings look for knowledge thanks to a search engine (full text research) and exploit themselves the lists of links through HTML pages. New techniques of health information retrieval are also based on these annotations [Hersh, 2003]. For example, within the framework of the CISMeF[5] project, the hospital of Rouen (France) built a catalogue of the principal French medical sites and documents. This catalogue organises the resources thanks to annotations, according to several points of view and, in particular, according to a classification of topics. This classification includes the principal biological and medical specialities recognised in France. The thesaurus Medline[6] inspired this classification [Douyere et al., 2004; Soualmia and Darmoni, 2005]. They are *formal annotations* the machines can understand.

In addition, new collective practices linked to the practices of annotation arise from the Web technical evolutions. We find, for example, the forums which are virtual places on a network where several users can converse more or less freely. We find community portals built by groups of Net surfers around a topic [Sack et al., 2004]. We find also Wiki, these Web sites freely modifiable by the visitors who create a common work [Aronsson, 2002; Chawner and Lewis, 2004]. In well defined specialities, some software of collaborative design exist [Darses et al., 2001], as in automotive engineering or aeronautics. This type of software allows the users to share knowledge through documents. In all these new spaces of exchanges, we find annotations. Human beings create these annotations and use them. They generally exploit them as a base for the creation of a perpetual object as a document, for the construction of common references, for acting, etc. They are *informal annotations* which are comprehensible by human beings.

[4] http://dublincore.org/

[5] http://www.chu-rouen.fr/cismef/

[6] http://medlineplus.gov/

Finally, we oppose these two categories of annotations more by the use made with them than by their formality. Formal annotations are created by humans or machines to be used mainly by machines and informal annotations are created by humans to be used mainly by humans. But if we want to exploit these informal annotations, we must add to them formal annotations to carry out calculations.

In the EHR, we will find these two types of annotations. Indeed, the quantity of medical documents in a record can be very important. An EHR must be equipped with a system which helps practitioners to arrange documents in the record and to look for knowledge. In this aim, we use formal annotations interpreted and treated by machines. In addition, as it is a collective work, practitioners add to the electronic documents informal annotations to exchange knowledge. Thanks to our multi-field study of the medical practices carried out with the paper health record [Bringay et al., 2004, a-b], we have noted that the health professionals already use this type of annotations during paper documents writing and reading.

In this paper, we will focus our attention mainly on the informal annotations and the processing which can be carried out on them. Our definition of an informal annotation elaborated thanks to [Chahuneau et al., 1992; Churchill et al., 2000; Denoue et al., 2003; Golovchinsky, 2003; Koivunen and Swick., 2001; Zacklad, 2004; Wolfe, 2000] is:

> *An annotation is a particular note linked to a target. The target can be a collection of documents, a document, a segment of document (a paragraph, a group of words, an image, and a part of image ...), and another annotation. Each annotation has a content, materialised by an inscription. It is the trace of the mental representation elaborated by the annotator about the target. The content of the annotation can be interpreted by another reader. The anchor links the annotation to the target (a line, a surrounded sentence ...).*

Our objective is to show the interest of such functionality in the particular context of the EHR. In section 2, we show why the study of the medical documents justifies the documentary approach to present the EHR. In section 3, we analyse the limits of the electronic medical documents and we explain how formal and informal annotations can solve some problems encountered by the practitioners when they use them. In section 4, we expose the processing authorised on the informal annotations.

2 A documentary approach for the EHR

2.1 The paper health record

2.1.1 Documentarisation of the medical documents

In a hospital unit, there are many situations of *transaction*[7]. Various *creators* carry out these transactions for various

[7] We use in this section the terms of [Zacklad 2004], who based him on the transactional theories to describe the medical

beneficiaries (patients, fellows, physicians, nurses, pharmacist …). They are linked by *social relations* (they work for the patient). They have a *common project*: the patient's recovery. However, each actor has his own project. The physician who writes a prescription does not have the same concerns as the pharmacist who manages all the drugs of the hospital. Fortunately, during the transactions, they will use a *representational common ground* (on the patient, medicine) as well as their own *competences* in order to disambiguate the communication. These transactions are realised in *heterogeneous spatiotemporal frameworks* (offices, conference rooms, corridors, by phone) with various *environmental conditions* (the actors have a table, little time).

Most of these transactions are oral. Yet, traditionally, practitioners developed a significant culture of writing words. In order to keep the most traces of the exchanges, they transcribe or record them on a perpetual medium, the paper or electronic documents of the health record. So, practitioners can handle (complete, annotate, read) this knowledge. It is reactivated in various contexts and will be the support of new transactions. Considering the number and the complexity of the situations of transaction, the hospital units organise a real process of documentarisation" [Zacklad, 2004]. The heads of department designed the architecture of the record and the organisation of the documents. So, the management of the documents is easier: in the paper health record, we know where to retrieve the surgical report. Likewise, their physical manipulations are easier: thanks to the predefined outline of the patient discharge summary, a writer knows where to look for the paragraph he wants to fill in and a reader knows where to retrieve the paragraph with the required knowledge. Besides, during their studies, practitioners learn how to write and read these documents. "The medical record is a tool (…) it does not 'represent' the work, but it feeds into it, it structures and transforms it in complex ways: it structures that communication between healthcare personnel, shapes medical decision making, and frames relations between personnel and patients." [Berg, 1998]

2.1.2 Creators and beneficiaries of the medical documents

Several creators can produce a document and several beneficiaries can read it. For example, the "interns' card", built collectively by the interns and daily update, is written for all the practitioners tacking care of the patient. Of course, the interns who write it are the first beneficiaries because the fact of summarising their knowledge in the card helps them to memorise the patient's state. However, most of the documents are written for precise beneficiaries. For example, a physician writes the patient discharge summary for all the specialists who take part into the patient recovery when he leaves the hospital. However, these documents can be the support of non foreseeable exchanges, when unexpected beneficiaries take advantage of the knowledge stored in them. For example, a physician places an imagery report in the record. A teacher-physician incites his students to read it. The intentionality of the document moves. A document written with a particular aim can satisfy a need of communication not envisaged during its design. The reader distorts the initial intentionality of the document. He *recontextualises* the knowledge present in the document according to his reading objectives. Berg and Goorman [1999] affirm that this way of using and re-using the record is linked to the contextual nature of the medical knowledge. "Medical information is entangled with the context of production: medical data are tied to the purpose of their generation and they are part of an evolving array of medical data which continually reshapes their meaning (…) Doctors are aware of the constantly evolving nature of the data they produce and they generate their data accordingly".

2.1.3 Two types of medical documents

In order to documentarise the record, the head of department (designer) have organised the documents in the record and the knowledge in the document. The designer structures the document and gives indications in the headings on the knowledge to be captured. Their internal articulation[8] is explicit. This decomposition into hierarchical fragments, organised in a particular way, highlights the semiotic productions and gives meaning to their organisation. The writers analyse the designers' indications to fill in the documents. In the health record, there are two types of documents:

- The *forms written in real time*, in the room of the patient, contain primarily raw data related to the cares. The designer structures the forms finely and gives precise information in the headings of the fields on the knowledge to be captured. The writer interprets these indications to fill in the fields. These two authors have a joint project: to keep traces of the knowledge used during stereotyped transactions. They have also their own project. The designer wants to codify knowledge in the forms to re-exploit it easily. The writer wants, as fast as possible, to keep the most traces of the medical event in which he took part. In these structured forms, we can identify easily small semiotics productions and the precise links relating them. The writers' works (their captures are limited to some words or sentences segments) and the readers' works (they learned how to retrieve knowledge in such forms) are easier. An example of form is the "administrative document of entry".
- The *documents of synthesis written after the medical acts* contain interpretations of the practitioners. There are also two authors. The outline of the document,

documents as documents for action. These terms are in italic in this publication.

[8] We described in [Bringay *et al.,* 2004a] the internal articulation of the document (the logic design [Bachimont, 2001]), i.e. the decomposition into fragments (in little semiotics productions) and their layout which is itself significant.

created by the designer, structures the document into paragraphs (and not into fields as for the forms). The designer gives indications on the contents of the paragraphs. The writer remains free of the knowledge he captures. Generally, he uses the natural language. According to Bachimont [2001], only this linguistic layout allows the expression of the various levels of information (the factual, the potential and the intentional), necessary to the knowledge contextualisation. The two authors have a joint project: to consign the most knowledge resulting from an analysis. They have also their own project. The designer, by imposing an outline, wants to organise the writing to make it exploitable. This outline cannot be as precise as the forms outline because it is impossible to predefine the knowledge resulting from a reflection. The writer wants to keep traces of his analysis which will help him during the cares. These documents are semi-structured. So, important semiotics productions and the links relating them can be identifying easily. An example of document of synthesis is the "patient discharge summary", summarising the hospitalisation of the patient.

Of course, there are also documents written without a predefined model, such as the diagram improvised by the surgeon to explain his operation to the patient. However, these documents are rare. Finally, we can oppose these two categories of documents by the type of knowledge captured (predefined knowledge vs unforeseeable knowledge), by the type of writing (a rigid writing vs a free writing) and by the level of structure of the document (structured documents vs semi-structured documents). Bachimont [2001] affirms that the predefined textual types fix the rules of reading and writing. These rules allow readings in distant contexts in time and space from the creation.

2.2 The EHR and the documentary approach

Traditionally, we distinguish two types of applications: document-based applications and data-based applications.

- The aim of the *document-based applications* is to manage documents. These ones permit to place and retrieve a document in a collection (surf from a document to another and move in a document). In such an application, the health record is considered as a collection of electronic documents. That seems to be adapted for the care practices. Documents are structured, semi-structured or non-structured and this structure is used for processing on documents.
- The aim of the *data-based applications* (built with a database) is to manage knowledge. These ones allow to capture data, to place and retrieve them in a database and to make calculations on them. In such an application, the health record is considered as a set of data. That seems to be adapted for the management and research practices. Documents are structured (sometimes semi-structured) but this structure is not used for processing on documents.

In those two types of applications, we capture and consult knowledge through documents. So, we choose to present the health record as a hypermedia (a collection of electronic documents). Moreover, it is not relevant to oppose these two approaches. These ones merge into themselves. According to the exploitation of the data wished, we must build structured or semi-structured documents. The more the system has a fine logic data structure and the more it can realise processing on data. The more the system has a fine internal documents articulation and the more it can realise processing on documents to make easier their manipulations (creation of index, synopses). Whereas a weak structure of the documents makes easier a daily use (the writer fills in the documents as he want), it is impossible without a strong structure to integrate tools with a strong added value for management and research. The solution comes from a compromise resulting from the study of the documents use. The documents use determines the type of electronic documents presented to the user:

- For the forms written in real time, we will use electronic structured and fixed documents (no freedom of capture for the writer) with helps for speeding the capture. They are similar to the forms in data-based applications.
- For the document of synthesis, we will use electronic semi-structured and non-fixed documents (freedom of capture for the writer) because the conservation of the writing context is the essential criterion. They are similar to the forms in document-based applications.

Computerising the health record according to a documentary approach consists of transforming this record into a hypermedia, by taking into account the uses of the practitioners with the documents. On this point, some authors like [Arnott Smith, 2004; Berg and Goorman, 1999; Charlet, 2003; Lovis et al., 2003] agree with us.

3 Needs of annotations

In spite of the *documentarisation* effort of the medical authorities to simplify the writing and the reading of the medical documents, those documents are not sufficient to allow the practitioners to really work on all the knowledge they create. The change of medium from paper to electronic medium stresses these difficulties. We show in this section that the contribution of the traditional formal annotations is not sufficient. Practitioners still suffer from difficulties when they work with the EHR. Informal annotations can solve a part of these problems.

3.1 Contributions and limits of the formal annotations

By using an EHR, the practitioners have a lot of problems. [Nygren et al., 1992, a-b] describe the reading problems specific to the health record. Most of them are linked to the loss of the spatiality. On a screen, we can visualise at the same time only a little number of documents whereas on paper medium, we can spread out the record on a table to build a global vision of its contents. The readers also suffer from confusion in the hypermedia. The memory is overloaded by too many requests: buttons, links. So, we must improve the functionalities of hypertextual navigation. The formal annotations can be used for indexing the documents and retrieve them.

In an EHR, most of the formal annotation can be automatically filled in. We can collect metadata through the application as the author of the document, the date and the place of production, etc. In addition, as most of the documents are predefined, we can mark their contents thanks to descriptors which allow to identify the logic outline of the document as the title, the quotations, etc.

These formal annotations allow the improvement of the hypertextual navigation functionalities. For example, in the Hospitexte project[9], the authors automatically generated documents of navigation (index, tables of contents, synopses) thanks to calculations carried out on the formal annotations marking the documents contents of the health record. With these documents of navigation, we can reach the knowledge in the documents by trying to reduce the cognitive costs related to the knowledge search. They offer new means of finding his way in the hypermedia, to build reading roads.

However, we cannot generate easily formal annotations on the semantic contents of a document without human's intervention, in spite of the improvement in natural language processing (NLP) [Laforest and Flory, 2000]. In particular organisations, people are used to designate one or more actors of the group to enrich the documents with annotations relating to an interpretation of their contents. For example, in traditional document management, librarians appoint one of them to index the documents thanks to the addition of formal annotations. This indexation profits to all the users of the documentary unit. In the case of the health record, we cannot imagine to ask to the practitioners to put a lot into the addition of such information in the documents, even to facilitate later reading road. They have no time for that.

3.2 Contributions of the informal annotations

The informal annotations allow to solve other problems in the EHR.

Some difficulties encountered by practitioners are related to the type of the documents - too rigid for writers' writing - and to the unpredictability of medical knowledge - where practitioners can consign the knowledge not envisaged by the designers and which emerge during the writing and the reading? On paper, practitioners currently use informal annotations to complete their capture in the forms and to keep traces of their readings. Indeed, even if the designer leaves textual fields for non foreseeable knowledge, the writers will prefer to annotate the rigid forms for the writing. With a graphical way (an arrow, an underlined part), they connect the comment and the part of the document having caused the comment. We do not find annotations written by writers in the documents of synthesis, less rigid, because they have sufficient freedom to write in the paragraphs. *A posteriori*, whatever the type of the document is, readers leave traces of their comprehension in annotations. The annotations thus allow contextualise the knowledge not envisaged by the designer of the documents, produced during the writing and the reading. The reader enters in the constitutive process of the document. With the annotation practice, he can reappropriate the document, rewrite it according to the desired use. Consequently, he becomes the "author of his reading" [Bachimont, 2004].

With annotations, practitioners also link documents, guide the reading from a document to another document. For example, to justify his argumentation in an imagery report, an expert adds the comment "cf. thorax radiograph n°2".

The study of the paper health record showed that the practitioners need helps for the creation of synthesis. Indeed, numerous documents result from the combination of annotations target (part of documents). Let us take the example of a physician who wants to write the part "Hospitalisation causes" and "Disease History" of the entry record. He reads the maternity record and looks for knowledge relative to the family antecedents of the new born. He annotates the important points and gathers them in the concerned paragraphs. He rewrites the result to build something understandable. We must propose the same functionality on the electronic medium: the combination of annotations or parts of documents in a new editable document. In such a scenario, the reader becomes a "reader-writer" [Stiegler, 2000] because he carries out two tasks: the reading of the documents used for the writing of the new document.

In addition, some authors as Hardstone et al. [2004] have shown the importance of the informal communications, most of the time oral, within the framework of the medical collaborative activities. Thanks to these informal communications, practitioners exchange knowledge often partial, speculative, provisional, incremental, etc. They do not want to consign this knowledge in the documents of the record which have a statute too much "public" (all the practitioners working with the patient consult the documents of the health record) and too much "formal" (the practitioners consign knowledge according to writing rules' fixed by the documents designers). Therefore, annotations are a relevant support, less "official". Practitioners know that annotations will be read as incomplete and subject to revision, contrary to the predefined documents of the health record. The passage to the electronic medium accentuates this way of thinking of the practitioners. Indeed, they often consider that all things captured in a computer must be finalised because they can be distributed to a wide audience. We could observe a particular behaviour of the practitioners, writers of documents, in our pilot site. A data base equips this service. Practitio-

[9] This project was carried out by the DIAM (collaboration of V. Brunie, B. Bachimont and J. Charlet) (Charlet et al. 1998) http://www.biomath.jussieu.fr/Hospitexte/.

ners often delay the data-processing capture in this data base until the moment they validate collectively information. But in the first step, they consign knowledge in their personal notes or in the annotations added to the documents of the paper record.

From these scenarios, we can conclude that a person annotates because:

- *She cannot*, without annotations, add her semiotic production to the document because most of the data captures do not allow the writer to enter the desired data. It is the case of the medical forms too rigid for allowing the writer to add knowledge not envisaged by the designer. Therefore, annotations are an escape clause if we have no current method for extending forms.
- *She does not want* to add her semiotic production to the document because this one is written with an intention of communication different from the initial intention of the annotated document. In so doing, the annotator is adding meta-information, i.e. information about the document rather than information that belongs in the document itself. It is the case when a reader annotates to keep traces of his reading, when a person wants to build a new document using her readings or when several practitioners collaborate by taking as support the documents they annotate. Consequently, annotations are also an escape clause if we have no means to code the comments about the documents.

These examples show how the practitioners use annotations to act: either to enrich the annotated document or to be the transitory support of knowledge used to create new knowledge (recorded or not in a document). Therefore, annotating is already an action.

We will show in section 4 that they are also sources of solutions for the problems of hypertextual reading. Indeed, in the Hospitexte project, yet evoked in section 3.1, the protagonists stopped with the processing of the documents and so with the processing of the formal annotations to build documents of navigation (index, table of contents). They just consider the prospects related to the use of the informal annotations left during the writing and the reading of the EHR. In DocPatient project, we go further. We propose calculations on these informal annotations to improve the functionalities of hypertextual navigation and the handling (creation, reading) of the electronic documents.

4 Exploitation of the annotations

Now, we will focus our attention on the processing we can carry out on the informal annotations thanks to their properties. This properties can be seen as formal annotations on the informal annotations. The aim of such a specification is to provide the user with a software component, an annotation system, to supplement an EHR (navigation, creation of documents, collaboration).

In this aim, we study the meaning of the annotations, for what they are used, what they make possible to express, their properties and the authorised processing we can carry out on them according to these properties (potentially valid combinations, actions, etc.). In order to build this conceptual user requirement and the specification of a software component, we studied the behaviours of the practitioners during the multidisciplinary studies undertaken within the framework of the DocPatient project [Bringay et al., 2004a-b]. We also studied the existing applications of annotations[10].

We have reused a first model realised by our industrial partner to present the documentary approach in the DocPatient project (model of the record used in our pilot site presented as a hypermedia). We have completed this model with an implementation of the processing carried out with the informal annotations. We have realised preliminary tests with the practitioners of our pilot site.

In this section, we indicate the processing made with annotations. We enumerate the properties necessary to them. To finish, we give the first practitioners' feedbacks.

4.1 Three types of processing

The figure 1 presents the processing we apply to the informal annotations.

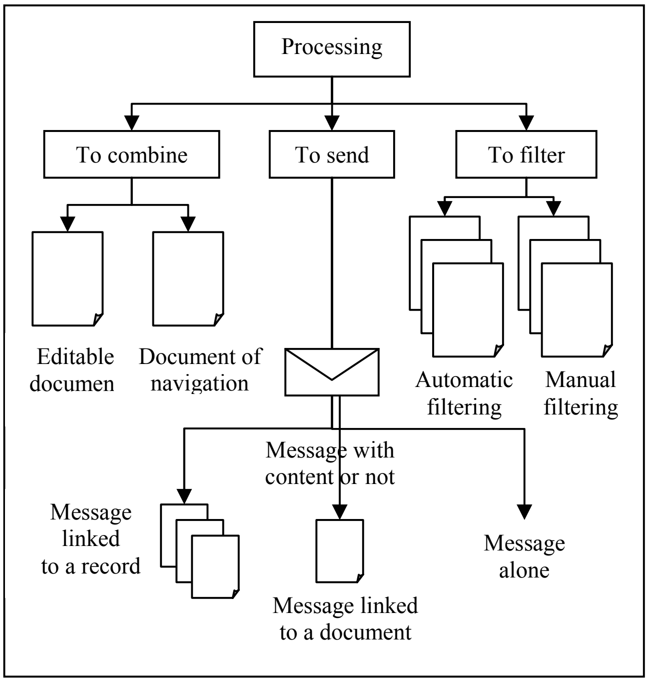


Fig. 1 - Tree types of processing on the informal annotations

[10] iMarkup (http://www.imarkup.com), Xlibris (http://www.fxpal.com/xlibris), Anchored conversation (FXPAL laboratory de Palo Alto), Annotea (W3C project), TheBrain (http://www.mines.inpl-nancy.fr/~tisseran/tsie/02-03/etudes/thebrainssite/)

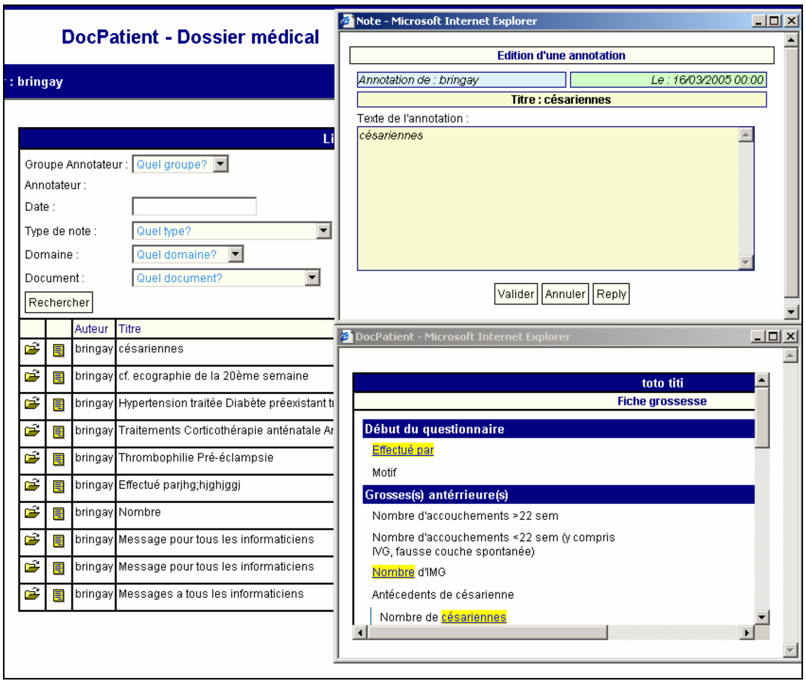


Fig 2 - Example of document of navigation

4.1.1 To combine annotations

We can combine annotations to form documents of navigation (just for reading) and documents editable (modifiable).

4.1.1.1 Document of navigation

A document of navigation is a new document added to the record. Such a document allows to retrieve the initial documents of the record (summary, index). A document of navigation contrived from annotations corresponds to a list of items leading to annotations, selected by the user according to one or more criteria. From these annotations, a reader has access to the annotated documents. So, we offer him new reading roads. The items correspond to the titles of the annotations. These titles can be completed or not by information as the annotator's name or the date resulting from the selection criteria of the annotations. We will detail this selection latter in this section. This information will help the reader to choose the reading roads.

The list can be *flat*. For example, during a hospitalisation, a patient suffers from a cardiovascular problem. A practitioner generates a new document allowing him to visualise all the annotations produced this day and dealing with the cardiovascular system. Thus, he can rebuild partly the history of this event thanks to the reading roads provided by this new document.

In figure 2, we see how we can present this flat list. In the top of the document, a form allows the user to choose the criteria of selection. He can choose to build the list with annotations:

- written by a special group of annotators (field "groupe annotateur" which lists the different professional categories: fellow, intern, …),
- written by a specific annotator (field "annotateur" which lists the persons working with the record),
- written a particular day (field "date"),
- written in a particular form (field "type de note" which lists the types of annotations: comment, link, message, …),
- related to a particular topic (field "domaine" which lists the most important specialities in a paediatric unit: cardiovascular, neurological, …)
- related to a unique document (field "document" which lists the documents of the record).

Behind this form, we find the list got in function of these criteria. The user has opened an annotation (window opened in the upper right corner of the image) and the target of this annotation (window opened in the bottom right corner of the image).

The list can be *hierarchical*. For example, a practitioner wants to remember the information exchanges he had with a colleague through annotations. These annotations form threads. We can organise them into a hierarchy with the link "Reply to" (as in a forum).

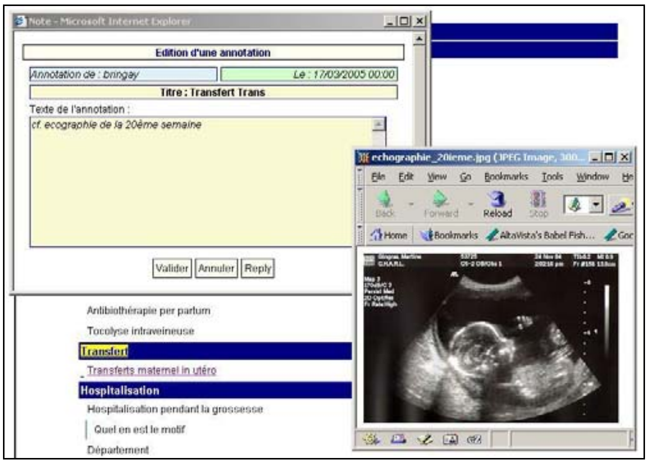


Fig. 3 - Annotation link

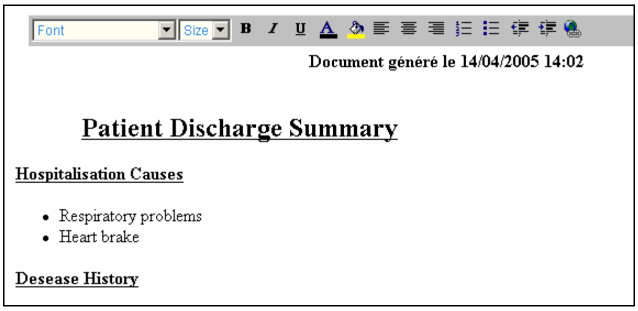


Fig. 4 - Document of synthesis

The list can be represented by a *graph*. In figure 3, we have an annotation link which guides the reading from a document to an ultrasound. Thanks to this type of annotations (and/or with predefined links between documents), we can make a graph. From each node, the user reaches the documents. Such a graph can be dynamically built when a user selects a resource (a document or an annotation) and chooses to visualise all the links leading or going away from this resource. For example, a practitioner uses a new protocol. He looks for all the documents, in a set of health records, which have led his colleagues to use this protocol. So, he can know the way his colleagues have used it, in which situations, etc.

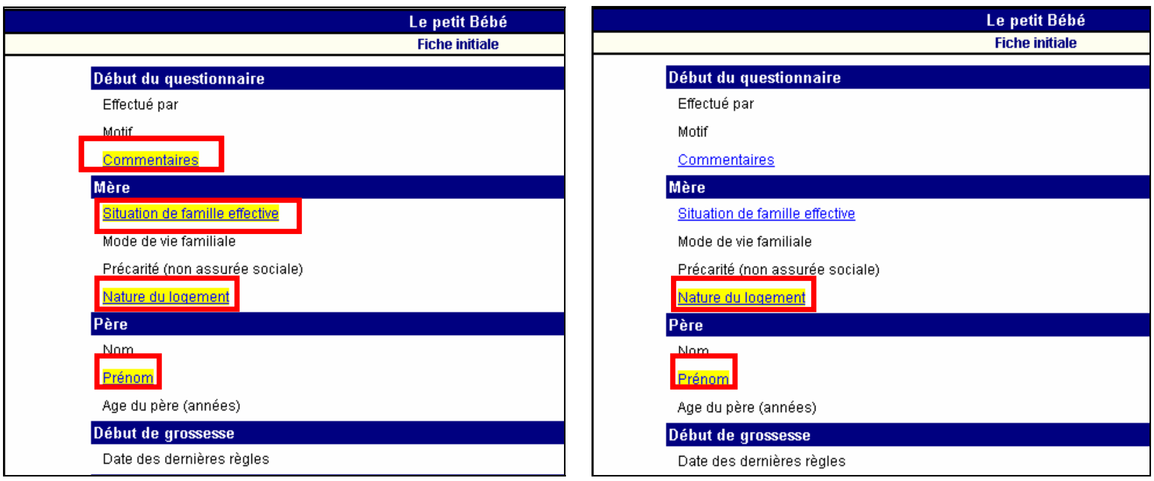
4.1.1.2 Editable document

To build a new document editable, we place the contents of the annotations selected by the user the ones after the others. Then, he can work the generated document, add knowledge and a page setting. For example, in figure 4, a practitioner writes the patient discharge summary. He reads the record, selects and comments some parts with annotations intended for this document. Gathering all these annotations in a new document gives him a base to write this report.

4.1.2 To filter annotations

During the creation of an annotation, the annotator can specify the recipients of his annotation by imposing access rights: himself, a group of users, all the readers. During the consultation of the record, there is an automatic filtering of these annotations. Consequently a reader visualises only the annotations he has the right to see.

In addition, during his reading of the record, the reader himself can choose to visualise only a collection of annotations selected according to criteria (annotator's name, date). For example, in figure 5-a, a practitioner reads a document with four annotations. He is the author of two of them. By filtering this document according to the criterion "My annotation", he gets the document of the figure 5-b, with only two annotations highlighted



a. Document without filtering b. Document with filtering

Fig 5 - Manual filtering.

4.1.3 To send a message

A user can send an annotation to one or more recipients. The message can have a content (a comment) or not. There are several kinds of messages:

- a message can be produced in connection with an element of a record. For example, a practitioner reads an analysis and detects an anomaly. He comments the document and decides to indicate it to all the practitioners concerned by the record. When a recipient receives this message, he must be able to retrieve the annotated document.
- a message can be produced in connection with a patient and thus in connection with a record taken as a whole. For example, a practitioner encounters difficulties to establish a diagnosis. He decides to ask his opinion to a colleague. He sends him a message linking the annotation to the record of the patient. When the recipient receives this message, he must be able to retrieve the record of the patient.
- a message can be produced in response to another message. When the recipient receives this message, he must be able to retrieve the previous message, as well as the source, if it exists, at the origin of the first message.

A software component of annotation supplementing an EHR application must be able to manage all these messages more those which do not have a relationship with a particular patient.

In figure 6, we can see an example of an interface of messages management. On the left part, in the top, the user can retrieve all the messages he received. He can open them in the bottom of this frame. On the left side, he can visualise the document linked to the annotation he consults.

4.1.4 Impacts of the processing on the contents of the record.

The software component of annotation act on the annotations, by filtering them and by sending them as messages. This component also can modify the contents of the record by adding new documents. These latter results from calculations on the annotations combined or not with human intervention (to choose the criteria of selection of the annotations).

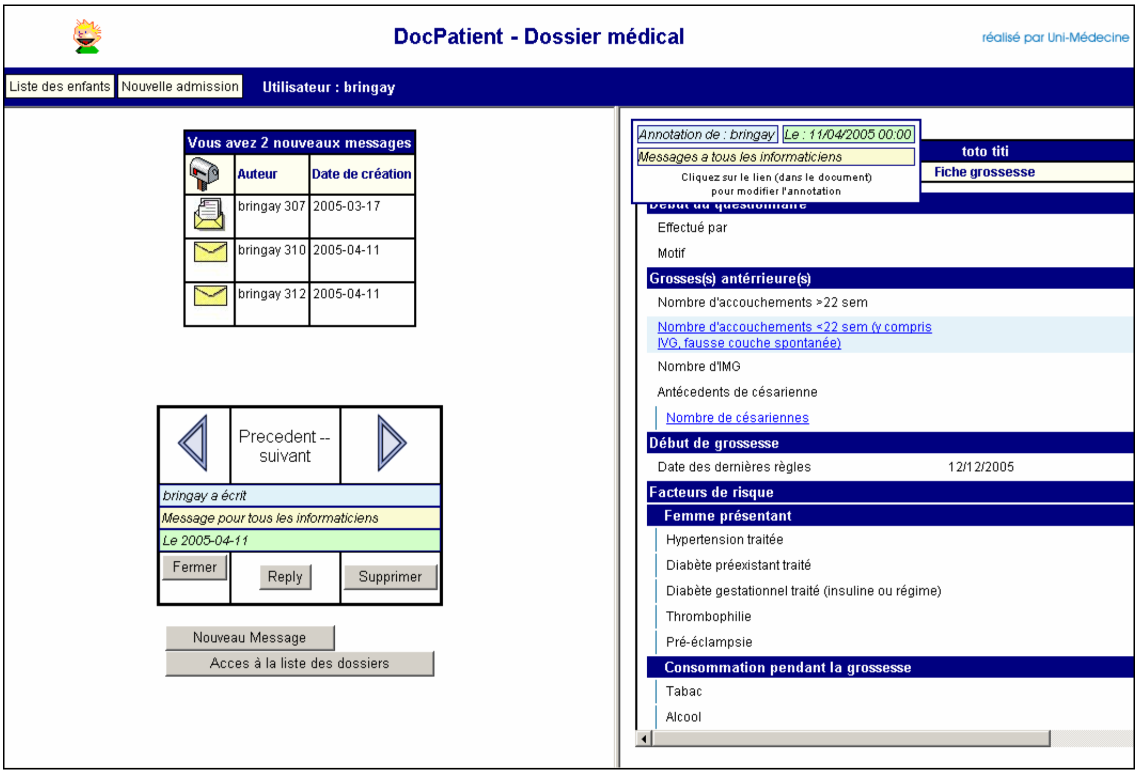


Fig 6 - Interface of messages management

Properties /Processing	Combine	Filter	Send
Organisational properties	×	×	×
Category	× To create a hierarchical list (respectively a graph), we treat only the annotations of the category message (respectively link)		× We send only the annotations of the category message
Sphere		× To carry out automatic filtering we use the access right	× We send a message only to the recipients located in the sphere
Topic	×	×	

Fig. 7 – Properties of the annotations and processing

4.2 Annotation properties

We now enumerate the properties necessary to the processing previously described. We elaborated a first list from this list of the processing and the traditional list of the Dublin core metadata. This list includes properties related to the event at the origin of the annotation, properties related to the action of annotation, properties related to the audience of the annotation and properties related to the semantic contents of the annotation. The table in figure 7 summarises these properties and their importance compared to the processing of the section 4.1.

Properties related to the event at the origin of the annotation creation (where, who, when, etc.) are: the annotator, the date (of creation or modification), the document and the target. The target is useful to assign an identifier to the annotation. This identifier results from an interpretation of the target. According to the annotated object, the identifier corresponds to text (if the target is textual), to the name of the image (if the target is an image), etc. As specified by Lewkowicz et al. [2005], this knowledge corresponds to the organisational dimension of the annotation which allows to determine the place and the role of the annotation in the organisation.

The properties related to the action of annotation determine the category of the annotation: a comment, a link between two documents (a document of the record or an external document), an annotation created in order to write a synthesis, a response to an annotation, a message for a precise recipient.

The properties related to the annotation audience correspond to the annotation recipients, i.e. to its sphere: the annotator himself (private sphere), all readers of the document (public sphere), one or a group of precise recipients (sphere of the group) [Zacklad et al., 2003].

The properties related to the semantic contents of the annotation allow to know the domain referred by the annotation. The domain corresponds to a set of knowledge forming a common reference for a group of persons (as an informal ontology). The user can choose a topic relative to this domain to qualify the annotation. On this subject, Lewkowicz et al. [2005] uses the expression of the specific dimension of the field. Assign a topic to an annotation consists of typifying it with specific knowledge of the users' field (e.g. some key words for a medical speciality)[11].

4.3 First feedbacks of the practitioners and future works

Our model was presented to the practitioners for use tests. During individual interviews, we showed them how to use the interfaces of the model and we collect their first remarks. They particularly appreciate the possibility of creating synthesis (combination of annotations in editable documents) thanks to the annotations because this activity is very important for them. Thanks to these discussions, we validated the processing and we specified the annotation properties.

Of course, other possibilities remain to be explored before getting a stabilised specification (processing and properties). Our model must be tested with real record, in real situations of care. That is why we need a more robust tool. Unfortunately, the economic context obliged our industrial partner to reorient his work on the building of a viewer. This viewer will allow the users to visualise the electronic documents of the EHR settled in a warehouse. It will include a functionality of annotations. However, as this tool will not be used for the production of the medical documents, we will retrieve only the annotations produced during the reading of the records. Our industrial partner plans to test this tool in real conditions during the summer 2005.

Consequently, this work is a first stage to create an annotation tool adapted to the health record. The last validation will be, in any case, related to the way practitioners use annotations and the possibilities offered by the calculations. We still have a lot of questions about this practice. We must study the impact of such functionality on the record. Is there a risk of impoverishment of the record if the practitioners prefer to annotate rather than write in the documents? How would the others actors (researchers and medical managers) use these annotations? How can users be motivated to make

[11] To make easier the choice of keywords, we can use thesaurus of specialty or tools of natural language processing (based on an ontology) to parse the annotation and its target in order to propose adapted keywords.

annotations with an electronic document as intuitively as with paper document?

In this context, we must pay attention to the reactions and to the way of answering them: an older project DOME [Séroussi et al., 1996] showed us that we must be reactive to the requests we cannot anticipate. We must answer them by reflections and finally by software developments perfectly adapted to the care activities. The aim of this project was to offer linguistics services to the practitioners and it appeared that they wanted to have an hypertextual electronic health record.

4 Conclusion

The health record is the privileged partner of the medical practice. Its computerisation has many consequences on the medical actors and their organisation. As documents are the most adapted support to handle medical knowledge during the cares, we affirm that a documentary approach is adapted to build interfaces in adequacy with the various uses of the health professionals.

In addition, we propose the integration of a particular documentary functionality: informal annotations. Indeed, from now on, the tools of annotation are common. On the one hand, most of the word processing software and the collaborative software include functionalities of annotations. On the web, we find virtual places allowing annotating (Wiki, forum). On the other hand, the practitioners already annotate the paper documents of the health record. Consequently, such functionality in the EHR seems natural. The originality of our work comes from the way we exploit these annotations to make easier the manipulations of the electronic documents. The preliminary tests realised by the practitioners on the tool developed by our industrial partner, validate our assumption that an annotation tool is useful for their care mission. About the theoretical prospects, this work underlined a need, which consists of defining the meaning of the calculations carried out on the annotations added to the documents of the health record. This list of processing must be completed. In addition, we foresee to generalise our definitions with contexts broader than the health record because experts use annotations in many others fields as genetics, architecture...

Finally, the change of medium, which upsets the practitioners' works, influences the way they read and write. Like Tom Thumb [Stiegler, 2000], they can leave traces of their manipulations of the documents to build their own vision of the record. So, annotations are real objects of action. Thanks to them, readers become increasingly active in the process of documents design. The writing joined the reading in the form of a new activity the "writing-reading" [Soubrié, 2001].

Acknowledgments

We thank Physician G. Krim, in charge of the Unit of paediatric intensive care and neonatal medicine (CHU of Amiens), for his invaluable help. We also thank H. El Moutcharay for her help during the construction of the model.

References

[Arnott Smith, 2004] Catherine Arnott Smith. The Document in the Electronic Health Record: A Model Made of Paper. *Proceeding of DOCAM "Document Academy"*, San Francisco (USA), 2004. http://thedocumentacademy.hum.uit.no/events/docam/04/program.html

[Lars Aronsson, 2002] Lars Aronsson. Operation of a Large Scale General Purpose Wiki Website. Elpub 2002: Technology Interactions. *Proceedings of the 6th International ICCC/IFIP Conference on Electronic Publishing*, November 6-8 2002, Karlovy Vary (Czech Republic), edited by J.A. Carvalho, A. Hübler, and A. Baptista, Berlin: Verlag, 27-37, 2002.

[Bachimont, 2001] Bruno Bachimont. Dossier et lecture hypertextuelle : problématique et discussion : Exemple autour du dossier patient. *Les cahiers du numérique, special issue of the revue "Information médicale numérique" (Electronic medical information)*, Vol. 2/2, P. Le Beux et D. Boullier, 105-123, 2001.

[Bachimont, 2004] Bruno Bachimont. Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle. *Habilitation report*, 2004. http://www.utc.fr/~bachimon/Livresettheses_attachments/HabilitationBB.pdf

[Berg, 1998] Marc Berg. Medical work and the computer-based patient record: a sociological perspective. *Methods Inf Med 1998*;37:294-301, 1998.

[Berg and Goorman, 1999] Marc Berg, Els Goorman. The contextual nature of medical information. *International Journal of Medical Informatics* 56, 51-60, 1999.

[Bringay et al., 2004a] Bringay Sandra, Catherine Barry and Jean Charlet. The Health Record: Kernel of a Medical Memory. *Proceeding of the Workshop "Knowledge management and organisational memories"*, Workshop of the conference ECAI European Conference on Artificial Intelligence, 23-27 August 2004, Valencia (Spain), 2004.

[Bringay et al., 2004b] Bringay Sandra, Catherine Barry and Jean Charlet. Annotations: A new type of document in the Electronic Health Record. *Proceeding of DOCAM "Document Academy"*, San Francisco (USA), 2004. http://thedocumentacademy.hum.uit.no/events/docam/04/program.html

[Chahuneau et al., 1992] François Chahuneau, Christophe Lécluse, Bernard Stiegler, Jacques Virbel. Prototyping the Ultimate Tool for Scholarly Qualitative Research on Texts. *Proceedings of the 8th Annual Conference New Oxford English Dictionary,* October 18-20 1992, Waterloo (Canada), 1992.

[Charlet et al., 1998] Jean Charlet, Bruno Bachimont, Vincent Brunie, Sawsan El Kassar, Pierre Zweigenbaum and

Jean-Francois Boisvieux. Hospitexte: towards a document-based hypertextual electronic medical record. *JAMIA Journal of American Medical Informatics Association*, 5(suppl), 1998. http://www.biomath.jussieu.fr/Hospitexte/

[Charlet, 2003] Jean Charlet. L'ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales. *Habilitation report*, 2003. http://www.biomath.jussieu.fr/~jc/CharletHDR2002.pdf

[Chawner and Lewis, 2004] Brenda Chawner and Paul H. Lewis. WikiWikiWebs: New Ways of Interacting in a Web Environment. *Handout prepared for LITA National Forum*, October 8-10 2004, Sheraton West Port Lakeside Chalet, St. Louis, Missouri (USA), 2004.

[Churchill et al., 2000] Elizabeth Churchill, Jonathan Trevor, Sara Bly, Les Nelson, Davor Cubranic. Anchored Conversations. Chatting in the Context of a Document. *Proceedings of CHI 2000 Conference on Humans Factors in Computing Systems*, April 1-6 2000, The Hague (Netherlands), 454-461, 2000.

[Darses et al., 2001] Françoise Darses, Françoise Détienne, Pierre Falzon and Willemien Visser. COMET: A method for Analysing Collective Design Processes. *Research Report*, INRIA, no 4258, September 2001. http://www.inria.fr/rrrt/rr-4258.html

[Denoue et al., 2003] Laurent Denoue, Patrick Chiu, Tohru Fuse. Shared Freeform Input for Note Taking across Devices. *Proceedings of Humans Factors in Computing Systems*, April 5-10 2003, Fort Lauderdale (Florida), 794-795, 2003.

[Douyere et al., 2004] Magali Douyere, Lina Soualmia, Aurelie Neveol, Alexandrina Rogozan, Badisse Dahamna, Jean-Philippe Leroy, Benoit Thirion, Stéfan Darmoni. Enhancing the MeSH thesaurus to retrieve French online health resources in aquality-controlled gateway. *Health Info Libr J. Dec;*21(4), 253-61, 2004.

[Golovchinsky, 2003] Gene Golovchinsky. Emphasis on the relevant: Free-form Electronic Ink as a mechanism for relevance feedback. *Proceedings of RIAO2000 Recherche d'Information Assistée par Ordinateur* (Computer-Assisted Information Retrieval), April 12-14 2000, Paris (France), 180-1995.

[Hardstone et al., 2004] Gillian Hardstone., Mark Hartswood, Rob Procter, Roger Slack, Alex Voss and Gwyneth Rees. Supporting Informality: Team Working and Integrated Care Records. *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW) 2004*, November 6-10 2004, Chicago (USA), 142-151, 2004.

[Hersh, 2003] William Hersh. Information Retrieval: A Health and Biomedical Perspective. *2nd Edition Hardcover*, 528 pages ISBN 0387955224 Springer-Verlag January 2003.

[Koivunen and Swick, 2001] Marja-Riitta Koivunen, Ralph Swick. Metadata Based Annotation Infrastructure offers Flexibility and Extensibility for Collaborative Applications and Beyond. Presentation for the WorkShop K-CAP First International Conference on Knowledge Capture, October 21-23 2001, Victoria B.C (Canada), 2001. http://semannot2001.aifb.uni-karlsruhe.de/papers/1_annotea.pdf

[Laforest and Flory, 2000] Frederique Laforest and André Flory. The Electronic Medical Record: Using Documents for Information Capture. *In MIE2000 Medical Informatics Europe*, Hanover (Germany). Also available in Stud Health Technol Inform 2000; IOS Press, Vol 77:617-21, 2000.

[Lewkowicz et al., 2004] Myriam Lewkowicz, Gaëlle Lortal, Amalia Todirascu, Manuel Zacklad and Mohamed-Foued Sriti. A Web-based annotation system for improving cooperation in a care network. Matera, M., Comai, S., *Engineering Advanced Web Applications*, Rinton Press 2004. http://www.ii.uam.es/~rcarro/AHCW04/Lewkowicz.pdf

[Lovis et al., 2003] Christian Lovis, Alexander Lamb, Robert Baud, Anne-Marie Rassinoux, Fabry P. and Antoine Geissbühler. Clinical Documents: Attribute-Values Entity Representation, Context, Page Layout and Communication. *Proceeding of AMIA, America Medical Informatics Association*, Washington (USA), 254-258, 2003.

[Nygren and Henriksson, 1992a] Else Nygren and Peter Henriksson. Reading the medical record I: analysis of the physicians' ways of reading the medical record. *Computer Methods and Programs in Biomedicine*, 39 pages, 1-12, 1992.

[Nygren et al., 1992b] Else Nygren, *Mats* Johnson and Peter Henriksson. Reading the medical record II: design of a human computer interface for basic reading of the computerized medical record. *Computer Methods and Programs in Biomedicine,* 39 pages, 13-15, 1992.

[Pédauque, 2003] Roger Pédauque. Form, Sign and Medium, As Reformulated for Electronic Documents. *Working paper*, 2003. http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/05/94/sic_00000594_01/sic_00000594.pdf

[Sack et al., 2004] Warren Sack, Françoise Détienne, Jean-Marie Burkhardt, Flore Barcellini, Nicolas Ducheneaut and Dilan Mahendran. A Methodological Framework for Socio-Cognitive Analyses of Collaborative Design of Open Source Software. *Proceedings of the workshop "Distributed Collective Practice: Building new Directions for Infrastructural Studies"*, Workshop of the conference Computer-Supported Cooperative Work (CSCW) 2004, November 6-10 2004, Chicago (USA), 2004. http://tech-web-n2.utt.fr/cscw04/Detienne.pdf

[Seroussi et al., 1996] Brigitte Seroussi, Rober Baud, Marc Moens, Andrei Mikheev, Peter Spyns, Werner Ceusters and Pierre Zweigenbaum. *Final report Dome*, Deliverable MLAP-Dome 8, DIAM-SIM/AP-HP, 1996.

[Soubrié, 2001] Soubrié, T. Apprendre à lire grâce à l'hypertext. *PHD report*, 2001. http://marg.lng2.free.fr/documents/the0010_soubrie_t/the0010.pdf

[Soualmia et al., 2005] Lina Soualmia and Stefan Darmoni. Combining Different Standards and Different Approaches for Health Information Retrieval in a Quality-Controlled Gateway. *International Journal of Medical Informatics*; pp 74(2-4), 141-50, 2005.

[Stiegler, 2000] Bernard Stiegler. Annotation, navigation, édition électronique : vers une géographie des connaissances. *Ec/arts, n°2*, 2000.

[Wolfe, 2000] Joana Wolfe. Effects of Annotations on Student Readers and Writers. *Proceedings of the International Conference on Digital Libraries*, June 2-7 2000, San Antonio (USA), 19-26. http://www.cwrl.utexas.edu/~wolfe/wolfe_dl00_final.pdf

[Zacklad et al., 2003] Manuel Zacklad, Myriam Lewkowicz, Jean-François Boujut, Françoise Darses and Françoise Détienne. Forme et gestion des annotations numériques collectives en ingénierie collaborative. *Proceedings of IC Ingénierie des Connaissances* (Knowledge Engineering), Laval (France), 207-225, 2003.

[Zacklad, 2004] Manuel Zacklad. Documents for Action (DofA): infrastructures for Distributed Collective Practices. *Proceedings of the workshop "Distributed Collbective Practice: Building new Directions for Infrastructural Studies"*, Workshop of the conference Computer-Supported Cooperative Work (CSCW) 2004, November 6-10 2004, Chicago (USA), 2004.

Experimentation of a socially constructed "Topic Map" by the OSS community

Jean-Pierre Cahier[1], L'Hédi Zaher[1], Xavier Pétard[1,2], Jean-Philippe Leboeuf[2,1], Claude Guittard[3]

[1]Tech-CICO, ISTIT - Université de Technologie de Troyes
12, rue Marie Curie - BP 2060, 10010 Troyes cedex - France
[2]LIMSI - BP 133, 91403 Orsay cedex
[3]BETA, ULP Strasbourg

{jean-pierre.cahier,hedi.zaher,xavier.petard}@utt.fr
jean-philippe.leboeuf@limsi.fr
guittard@cournot.u-strasbourg.fr

Abstract

We describe, in this article, a "topic map" system applied in the Open Source Software (OSS) community. Our approach is deliberately open and based on the HyperTopic model created by Tech-CICO lab. Our collective experimentation aims at the construction of a shared information platform that would be visible and useable for the OSS community. Thanks to this platform, OSS community members can describe and find software applications, by browsing multi-point of view "topic maps". Everyone may declare the characteristics of a software project following an index structure made of several tree diagrams. Thus, the community will build, in an ideal case, a dynamic and collective meaning

We present this project as an example of a "Socio-Semantic Web" (S2W). We also detail the HyperTopic model, on which is based our application, and the AGORAE platform which brings this software application into play. The HyperTopic model is inspired by the simple semantic models of "topic maps", but it enriches this approach by two new dimensions. First, it aims at improving the representation of shared meaning artifacts (in this case: software applications), of the social actors and of their activities. Second, this HyperTopic model is a support for debating the meaning of these representations.

1 Introduction

To find their way in "*territories*" of complex activities with huge, specific and moving glossaries and shared meanings, the human actors need *maps*. It is important to help them use, organize and modify again and again the "*topic map*" linking topics describing their domain, their activity and their position as actors. Besides, there often exist several points of view; the topic map has to take this plurality into account.

Nowadays many researches are focusing on a "Semantic Web" to provide better services. In this context of improving the Web standards, they should provide "topic maps" to improve the deposit and the finding of information. According to us, the more efficient and useful semantic Web, understood as a shared meaning artifact using Web standards, is the one created within and by communities.

Thus, our vision of the "Semantic Web" is more precisely a "Socio-Semantic Web" (S2W), because social and human aspects are central in our approach. In this article we would like to give a concrete example of this concept. Indeed, we will present the permanent distributed co-construction of a "topic map" by an open, large and already existing community: the Open Source Software (OSS) community. For now, the experience is only beginning and limited to the French speaking OSS community[1]. This project aims at giving an example of the "Socio Semantic Web" and to experiment concrete co-construction of "topic maps" with many distant actors. The construction step by step of this map will enable to acquire the knowledge of the different members of the OSS community. Thus, the topic map could be a support for the OSS knowledge and particularly the know-how. This process is promoted by dedicated multiple forums where people may share their experiences and discuss each topic of the map.

Article outline

Part 2 of this article brings elements to introduce Socio Semantic Web (S2W) and Knowledge-Based MarketPlaces (KBM) which is an example of S2W application family. We justify our approach in the fields of Computer Supported Cooperative Work (CSCW) and of Knowledge Engineering. Cooperative work in a KBM depends strongly on many

[1] http://www.yeposs.org (Yellow Pages for Open Source Software)

problems of semi-formal ontology sharing. We set quickly, at a very general level, the approach of "semi-formal ontology" that we use to facilitate the writing and maintaining of the S2W application directly by the cooperating experts themselves.

In Part 3 we will present the methodological, conceptual and practical tools that will be used in this study. These tools are mainly the KBM based on the HyperTopic model and the software platform Agorae.

In part 4 of this article we describe the Topic Map of the OSS industry and the software portal used to build and maintain this map by the community.

We conclude by giving an overview of our current development and some perspectives for further research.

Background and motivations

2.1 A stake for a cooperative Knowledge Management

In order to keep a shared vision of their world, groups build and maintain continuously a lot of "maps" and landmarks, of various means. The particular context of some activities (collective design, choice between competing products, adaptation to quickly changing contexts, sharing of document repositories…) justify some debates and divergences on terms used in the organization, whose solution is not imperatively to unify these contentious questions and to align vocabulary, according to a centralized and neat directory.

Particularly, in a system for sharing a topic map, such as the one we shall propose latter on, two crucial management objectives will be the learning by the community of a set of "Points of view", and the naming and categorizing of these Points of view by Topics. Related to domain or business entities, topics are terms characterizing important heuristic attributes, not in a perspective of universal or academic knowledge, but for activities which are crucial for actors at a given time.

Managing collectively Topic maps with several Points of view represents a particular Knowledge Management stake. The index structures on which the proposed system is based must answer to questions such as: how do a set of Points of view, including the topics and their relations, emerge from the interactions of the members? How do they evolve within actors' activities and discussions ("forums", etc.)? The actors are both *using* an existing shared meaning and *co-building* it. We follow Vygotsky's Theory of Activity [Vygotsky, 1997, Engeström *et al.,* 1999], stating that a loop does exist between language and activity. A shared meaning is built within the collective activity, by community members who are both users and "co-designers" of this cultural "socio semantic web". Our conviction is that, especially in communities of all sizes, this human "natural" practice will take advantage of unexplored potentialities of the Web, as a support for document and content management, for communication, for Computer Supported Cooperative Work (CSCW) and *in fine* for the participatory design of shared meaning artifacts, in the perspective of a "socio semantic Web" that we must now introduce.

2.2 Socio-Semantic Web (S2W)

Several critic arguments are presented to underline a lot of bottlenecks and weaknesses of the mainstream approach of the Semantic Web field, as summarized by Tim Berners-Lee and the W3C [Berneers-Lee, 2001]. For instance, today's semantic Web main perspective deals with meaning in a very restricted sense, and solutions offered are too static [Veltman, 2004]. Inside the Semantic Web field, our "socio semantic Web" (S2W) proposition appears on the contrary as a promising field of research, tools and applications [Cahier and Zacklad, 2004]. S2W does not imply a high level of "automation of the meaning" with formal ontologies built by ontologists and processed by software agents using automated inferences.

On the contrary S2W focuses on situations where an emerging shared meaning indeed needs support of Information Technologies, and Knowledge Engineering, but with human beings highly required to stay in the process, interacting during the whole lifecycle of applications, for both cognitive and cooperative reasons. Note that this S2W vision is not contradictory to Semantic Web classical automated techniques. As we will see it in the detail of our model and tool (§3), S2W does use the low-level open standards (XML, RDF) of the semantic Web "Cake" proposed by W3C. The two approaches could be complementary in a lot of applications.

From a Computer Supported Cooperative Work (CSCW) point of view, S2W deals with a very large spectrum of collective activities, especially in the context of the Communities of Action [Zacklad, 2003], characterized by coordination mechanisms based on Symbolic Communicational Transactions. In this context, "socio semantic" preoccupation emphasizes the symbolic level as an important coordination component. Communities of Action theory contrasts with the theories of situated action [Suchman, 1987], of distributed cognition [Hutchins, 1995] and with the "Social Web" approach, which emphasize more tacit knowledge or more direct "awareness" mechanisms. It differs also from the approach of Coordination Mechanism, based on protocols and artifacts implementing models or workflows schemes for the articulation of the cooperative work [Schmidt and Simone, 1996, Simone, 2000].

S2W aims at supporting Communities needing to collectively elicit, in a continuous manner, a crucial part of the knowledge, especially of the "locally-situated" semantic structure underlying both business objects and collective work of the community. For the business objects it can arise through artifacts such as thesauri, maps, yellow pages, catalog directories, structures of indexes, etc. Some of these cases of semantic resources can be considered as "semi-formal ontologies" to manage with the help of the Hypertopic and KBM models, as we shall see it below.

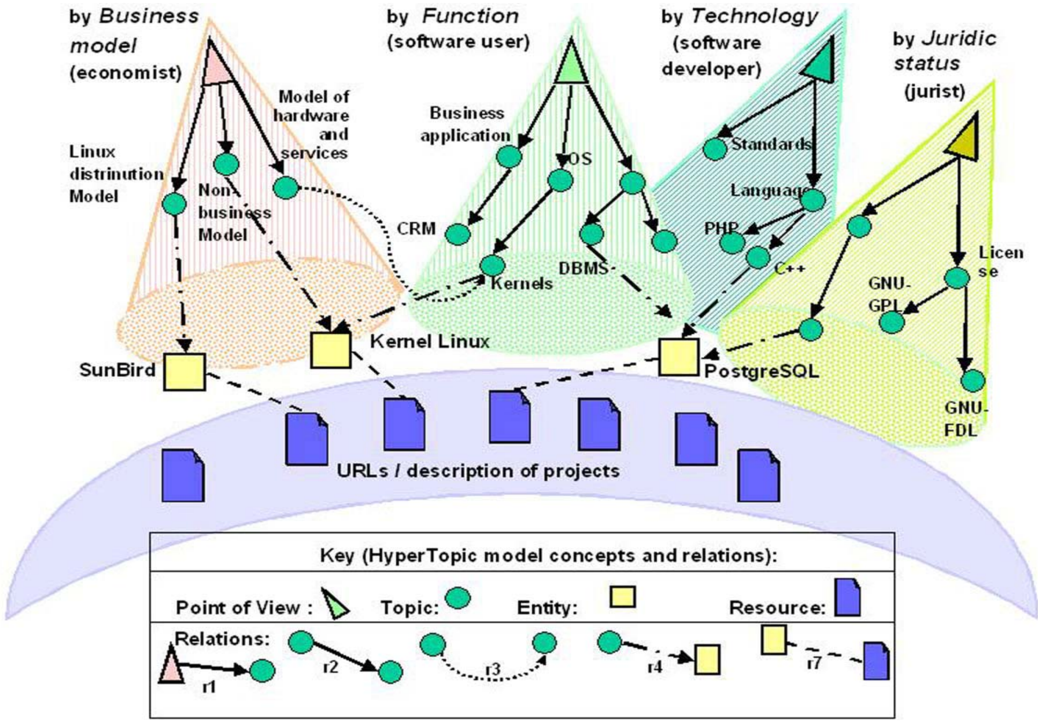


Figure 1 - Multiple points of view in a socio semantic web semi-formal ontology, according to HyperTopic model (exemple of Yeposs application)

At the level of a Community (or of an inter-Community, e.g. business extranets associating Clients, Sellers and Subcontractors, etc.), the "located" shared meaning is collectively and continuously "auto-constructed", tacitly or explicitly, by and for the actors in their activity. In such a process, "users" are not only consumers of externally-designed semantic resources, but they are users and creators in a constructive manner of a local and living "ontology", with inherent semiotic characteristics [Zacklad, 2005], pragmatically managed at the Community level. As a consequence, in the cases where there is a strong need to make explicit a part of the underlying shared meaning, it is a better solution - in many cases it is the only one - for this shared meaning to be managed by the concerned people from multiple points of view.

[Bénel *et al.,* 2001] argue that in digital libraries there is no meaning in a universal ontological consensus between all readers. On the contrary, the conceptual structures which describe document contents in a digital library must allow the clarification of multiple points of view, tolerate conflicts between humans and help them overcome these conflicts by communication.

2.3 A way for co-building large-scale semi-formal ontologies

The HyperTopic model and the Agoræ tool and method, that we want to propose (see §3) for co-building topic maps based on multiple viewpoints, could solve in certain cases the difficulties often noted for ontology learning and maintaining [Uschold and Jasper, 1999] [Maedche and Staab, 2003]. From a cooperative Knowledge Management point of view, front-end tools have also been proposed to build, align and merge formal ontologies [Tennison and Shadbolt, 1998] [Euzenat, 1996]. But with regard to acquisition and learning stakes, a particular bottleneck often subsists for building and maintaining these formal computable domain ontologies, especially in large domains with frequent changes : project memories, electronic marketplaces, skills yellow pages [Cahier et al., 2004], web content management systems, etc.

In such cases S2W is better-suited, because socio semantic approach aims at constructing in a continuous manner cooperative shared meaning artifacts, expressed according to a precisely defined model. Such topic maps can be understood as "semi-formal ontologies", referring to the [Uschold and Gruninger, 1996] classification. As noted also by [Kassel and Perpette, 1999], semi-formal approaches articulating terms, notions and objects could be more suitable than formal solutions to build cooperatively the meaning. A S2W semi-formal ontology cannot (generally) be used to compute automatic inferences, but it constitutes a semantic network which is structured at an epistemological level [Brachman, 1979] and which has to be understood pragmatically as "semiotic" [Zacklad, 2005], depending on the human interpretation context through "Points of view".

The HyperTopic model that we propose to ground S2W applications and that we shall see in detail further (§3), is a knowledge representation model that takes place at an epistemological level. HyperTopic gives to a shapeless non-formal semantic network a structured topic map form tuned to the HyperTopic standard concepts and rules (like the *key* components within a roadmap, cf. Figure-1). But these topics and their relations within the map need a high intervention of the human actors to fully complete the meaning in context. According to [Ribes and Bowker, 2004; Bowker and Star, 1999], who have studied communities with actors such as experts or scientists, it is necessary "*to be aware of processes of the constructive ambiguation of concepts - what Leigh Star has referred to the creation of boundary* objects *which can sit between multiple communities and share just enough meaning for the purpose at hand while being understood quite differently*".

The S2W could be a new and original way to tackle the difficult issue of creation and updating of ontology when it does not aim at automatic logic inference computing. This new approach could be a good way to fill the ontology learning gap. The actors should be able to share their knowledge and to take part in at this semantic creation. To be efficient in this process, it is important that the actors work only on their topic of competences and not anonymously. So the phenomenon of reputation and recognition by peers will be the engine of their self commitment into the project.

2.4 "KBMs" as examples of S2W applications

In the electronic commerce field, e-Marketplaces can be studied as places for cooperative work between suppliers and buyers, involving knowledge and creation of new knowledge. In a preceding paper [Cahier and Zacklad, 2002], we started a work from a theoretical point of view, in order to build a model of cooperation that we have entitled "Knowledge-Based Marketplace" (KBM). In that perspective, e-Marketplaces catalogues and Web content management systems proceed from a twofold problem of modeling information and knowledge from multiple points of view and from multiple experts.

In our present focus, a KBM can be seen as a particular type of socio semantic Web application, in which the semantic framework proposed (Points of view about Entities organizing a Topic Map) appears strongly "structuring" on a few generic given Roles (in S2W systems, roles, objectives and representation models can be very various). A KBM include three main roles, that we have called "KBM-roles": to *consult* the topic map and the information (the "client" role), to *contribute* (to describe a domain entity and to index it according to the map, *i.e.* the "seller" role), and to *structure* the topic map (complete and change topics names and places as a "semantic editor").

3 Methodology, models and tools

3.1 General methodology

A general guideline to build S2W applications is to give users basic affordances to understand, analyse and model the threefold activity which is necessary for them to build a shared meaning in their group: *domain objects*, *actors* and *activities* have to be concurrently taken into account. As showed in Figure-2, issues are numerous, about these three dimensions.

Many unforeseeable actions have to be undertaken for the modeling of the domain objects, as well as for the organizing of designers groups and roles. Context and goals of the actors can change, implying a pragmatic methodology. In particular, a community must face a continuous growth of the information influencing members activity. According to Peirce's definition of information [Peirce, 1868], information growth implies both *independently* in width growth and in depth growth. In our case, adding new domain Entities represent "in width" (extension) information growth, while adding new Points of view, Topics and map links represent "in depth" (intension, comprehension) information growth. The knowledge representation framework that we propose can be very helpful at the methodological level, to take into account the relation between multiple actors and these two facets of information growth.

For these reasons it would be convenient to use a knowledge representation framework furnishing appropriate concepts to construct initial knowledge map according to various strategies (incremental, brainstorming, top-down, in width / in depth alternately) especially in the bootstrap phase. In the KBM applications, semi-formal ontology learning and maintaining is particularly facilitated by the ability to build the map (new topics, etc.) simultaneously by examining new domain objects or new required tasks (bottom-up approach, ontology-learning "guided by instances"). In fact all types of ontology building methods could be used, for instance those inventoried by [Uschold and Jasper, 1999]. But for semi-formal ontologies, major opportunities and methodological changes could come in addition from the cooperative perspective. Different members of the group, at different steps of the design of the topic map, can use different methods, for example preferring to work at a more theoretical level rather than to use an inductive or abductive method.

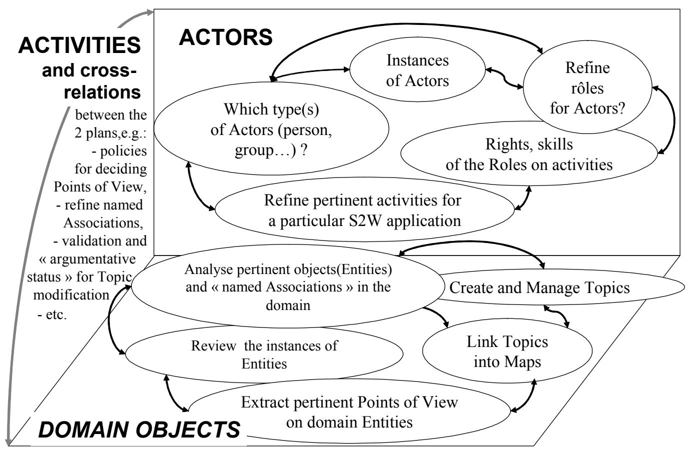


Figure 2 - S2W applications requires a threefold analysis

In such a collective design problem solving situation, an efficient way is to lean on the artifact itself, which in our case is the Topic map. Because of our *a priori* principle of participatory design within the Community, a lot of actions have to be carried out by the actors themselves and together. Participatory design axiom implies in our case for modelers the ability to make visible the representation of their actions, as actors of the world and as modelers. S2W methods and tools must provide the keys to make progressively explicit within the group the landmarks and the "map" not only of domain, but also of roles and actions needed to co-build the shared meaning. Fig.8 at the end of the paper illustrates a possible way for users to use HyperTopic for that.

As "end-users", community expert / actors are not ontologists. *A fortiori* they are not specialists of computerized domain ontology building. To model knowledge and act according to the three folder of Fig.2, they have to be strongly helped by the method implicitly carried by the proposed tool.

Here is the reason why the HyperTopic-based knowledge representation framework we propose below is so important, because it organizes simple key concepts useful for community members:

- to identify, name and articulate sub-components of this threefold activity,
- to re-use know-how and best practices of other KBMs building, such as KBM roles and frequent actions,
- to arrange and register at each step the co-design process results, so that all members can consult the last state of the deliverable, annotate and debate, contribute and structure the "topic map" artefact in progress .

3.2 The Knowledge representation framework

We will now present the "HyperTopic" model which is the formal framework clarifying these two overlapping facets of the whole of all the activities:

- the expression of the explicit shared meaning - by topic maps – clarifying objects in a field, for various roles of the actors;
- and the co-construction of this explicit meaning by the actors, including the shared vision of the activities and roles.

HyperTopic is the generic model suggested to face this problem, keeping in mind reusability. HyperTopic includes different types of concepts and relations. They are the key elements the end-user needs to understand the Topic Map, (for instance to interpret a schema such as Figure-1). The HyperTopic model is used as a Knowledge Representation language and as a core for building a set of semi-formal elements around it, such as a topic map built by users of the system.

Followed methodology is strongly structured by the components of HyperTopic. Moreover, it supports the roles organization in the architecture of the Knowledge-Based Marketplace (KBM). HyperTopic is also used as a support for more specialized models, like KBM, to adapt S2W concepts to particular kinds of activity which could be topic map-based, like collective drafting, annotation, negotiation, content management on the Web, etc. In the present case, the knowledge-based marketplace model (KBM) proposed in §2.4 brings at the same time a whole preset of roles and a cooperation model. At the higher levels of Figure-3, we contrast between the topic map and the data-processing level itself, which involves the data and the digital information resources memorized or referred by the system.

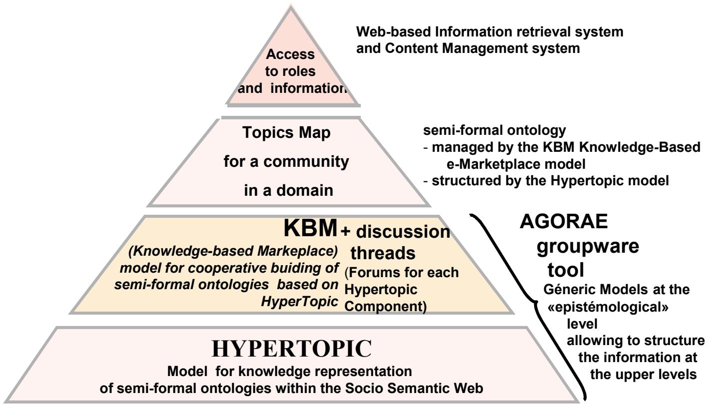


Figure 3 - Architecture of a socio semantic Web application based on HyperTopic and KBM models

Afterwards, we will describe HyperTopic concepts (§3.3) before giving some elements about the Agoræ platform, which was used for the realization of the OSS application topic map (§3.4).

3.3 HyperTopic Model

The figures 4a and 4b summarize the HyperTopic components, respectively in its first version (V1) implemented in Agoræ and in the future version (V2) which extends the knowledge map to the actors and activities.

The general objective is, as detailed in the Topic Map Organization ISO standard [TM, 1999] and in its XML implementation [XTM, 2001], to provide the elements to describe a map of topics, to which Web resources are *in fine* attached. These resources are linked to the objects from the world.

Topics are not concepts but simple or complex linguistic expressions expressing "subjects". According to the TM ISO standard [TM, 1999], "*in some sense, a topic link reifies a subject*" and "*in the most generic sense, a 'subject' is any thing whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever.*"

We also use this definition for the Topic in HyperTopic in the context of the socio semantic Web (cf. §2.2 and 2.3). In HyperTopic, the basic set of elements used to structure a map is improved compared to the TM ISO standard, in particular to facilitate their handling. For that, in addition to the *topics*, *associations* and *resources* which take again standardized concepts of the topic maps, HyperTopic defines the concepts of *entity* and *point of view*.

The *entity*, and not the documentary resource, is connected to the topics. We introduce this concept because in a lot of applications the information retrieval is applied initially to "objects" having a generic structure. Entities, like objects, include some descriptors allowing their "primary" characterization. Standard attributes and one or more occurrences of material resources carrying target information are associated to these descriptors. For instance, in the OSS application, *Entities* are software projects. Their associated *Resources* may be description cards or URLs of projects homepages, possibly with a link to download the software.

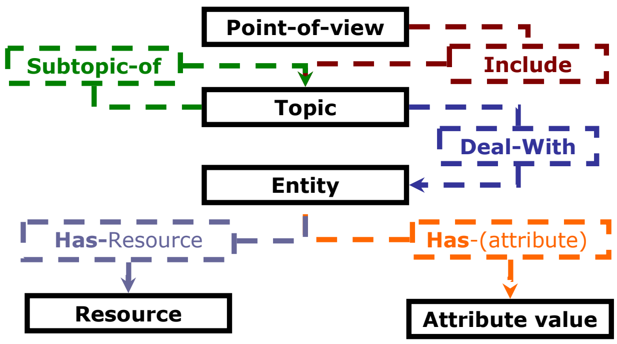


Figure 4a - HyperTopic Basis (associations are dashed)

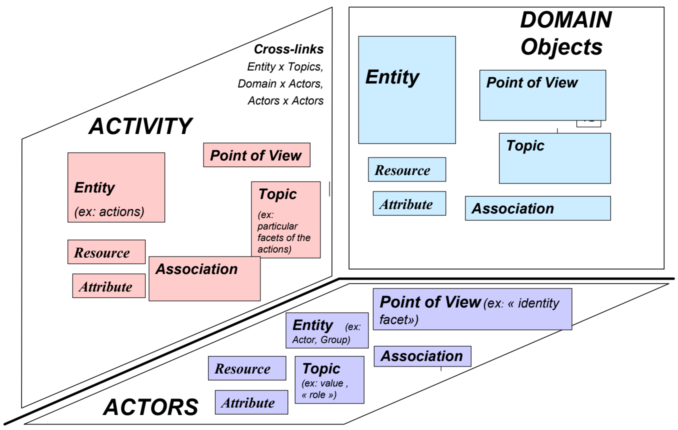


Figure 4b - The HyperTopic Model (v2), emphasizing the 3 dimensions of the socio semantic Web

The *point of view* is a descriptor to contextualize entities corresponding to a vision of certain actors. It corresponds to a set of characteristics of the entity, gathered and treated on several hierarchical levels, according to a vision meaningful for an actor or a group of actors (*e.g.* a point of view corresponding to a "business" or a "community").

This definition of the Point of view distinguishes HyperTopic from the concept of "facet". In models such as FacetMap [XFML, 2002] or Topic Map [TM, 1999], facets allow assigning property-value pairs to information resources. In Hypertopic, this feature corresponds to the "attribute" concept.

It is subtler to distinguish between Hypertopic *Point of view* and Topic Map *Scope*:

- Technically, at a syntaxic level, the HyperTopic *Point of view* is similar with the *Scope*. In the XTM specification, a *Scope* uses a *Theme* (which is a Topic), to help contextualize another Topic. But in Hypertopic, **i)** the Point of View is always a Scope *at the higher level*, and **ii)** a restriction is brought : the same Theme is used to contextualize an important set of Topics, so that inside a Point of View, it is not allowed to introduce low-level different *scopes*.

- Thus, at a semantic level, Points of View permit to resume homogeneous Knowledge Management collective choices in the organization. They organize in a simple manner the Topics, to produce a classification schema (a set of Points of view) easy to share in the Community.

For these reasons, in our approach, Points of View, which are semantic, are not contradictory with Scopes. At a technical level, Hypertopic Points of View can be represented using Topic Maps, RDF or other semantic web standards.

In HyperTopic, a point of view is a way of looking at a particular type of entity. In the OSS application studied below, there is only one entity "software project", to which five points of view (cf. §4.2) are related.

Let us insist on the fact that topics are not only "facets" or simple attributes of software, but often important "heuristic" properties in the experts' points of view. For instance, as showed in Figure 1, a particular software could be in conformity with a standard, a programming language or a label (interoperability in industrial sphere), follow a business model (economic level), be placed under patent mode (legal stake), etc. As a consequence, topics are linguistic expressions with often a high heuristic content: in practice Topics can require up to ten or more words to express their subjects.

Points of view and Topics as "heuristic attributes" condense a real expertise and can create controversies during the co-design of the map. In the OSS application for example, they highlight several dimensions of software evaluation in the knowledge map: the selection criteria between competitor tools, the complex structure of the Total Cost of Ownership (TCO) for software, the technical feasibility of a components assembly, etc. A stake is to develop, at the same time, the community shared culture (but not always unanimously) and the framework of terms, standards and rules, resulting from clarification of often implicit knowledge, shared at the beginning in the collective.

The relation used between Topics in a same Point of view is the generic association "sub-type-of". Moreover, HyperTopic allows transverse associations from Topic to Topic. They are named relations, for instance in OSS application, only one of these transverse relations is used: "see-also". To keep the Fig-4a easy to read, this relation is not represented. But one can see examples on Fig.1 ("r3 relation") or Fig 8a.

In its second version, HyperTopic allows specification and search of what are the business objects, and how they change, in terms of external definition by heuristic attributes (Topics associated with an Entity instance). In its actors / activities part, the model can express *who* modifies objects of the entities collection, when, how, with which certainty for actors and with which degree of validation for the organization / community, etc. Actors must not have the same rights or competences to contribute at the various stages, that is why it is important to also have a malleability margin in the definition of the roles of a given actor.

Thus, we can describe the way in which several actors contribute to sequences or actions complementary. For instance let us consider the creation of a new entity instance. In the OSS application map, a sequence would be initialized by a developer wanting to describe new software to associate it to several topics, according to various points of view (Figure 5a), and to create a new topic or to move an existing one if the knowledge map is incomplete (Figure 5b). Topic moving or removing can imply seeking and finding actors having created it and having attached entities to it, and alerting them or starting a discussion.

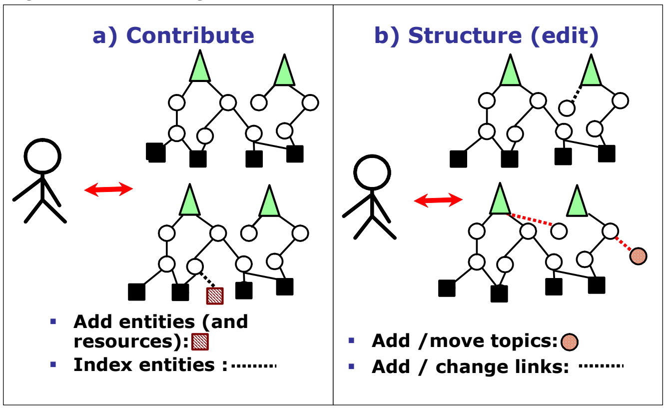


Figure 5 - Contributor (a) and semantic editor (b) actions with HyperTopic

Moreover, not always the same actor will be implied in all events of this scenario. For instance, in the contribution activity, a developer can be qualified to index the software in the “features” and “software engineering” points of view, but less in “legal” and “business model” points of view, therefore he can wish to discuss with another actor to relay or to ask him about topics and choices.

Actors and activities’ dimensions in HyperTopic V2 alleviate management of these socio semantic collaborative activities by using HyperTopic knowledge maps. Without these dimensions, the maintenance of the topic map under operational conditions is more difficult, particularly with great number of actors and entities.

HyperTopic (V2) model is designed to improve the management of these stages of cooperation between actors, as indicated in Fig.3 and 4b. It is conceived to allow easy definition of roles according to a fine granularity, by computer sciences non-specialists using knowledge map (for instance, to manage authorizations).

3.4 The Agoræ tool

In conformity with the first version of HyperTopic (V1), but soon adapted to V2, the toolbox platform Agoræ currently used for OSS application is developed by the Tech-CICO team with free software components, according to an open architecture and a modular source code, with portability and generics concerns.

Although it is a research prototype, especially intended to develop and validate socio semantic Web concepts, Agoræ V1 supports today the main KBM features: semantic structures creation and management by points of view and topics tree, entities and associated resources creation and management, topic-located threads for discussion, etc.

Agoræ implements also mechanisms for cooperative contribution to several features (structuring tree, resources inclusion…) and actions log.

Agoræ propose interface elements adapted for designer role, in initial stage of HyperTopic based application design. Designers have access to all features of the other roles, in particular editors for *ex-nihilo* creation of points of view and topics tree structure. They also have importation feature allowing the merging of tree structure elements from various sources (of thesaurus, ontologies) like other Agoræ applications, XML dump or other formats (Excel, Mindmanager, etc.).

A knowledge map created with Agoræ can be exported automatically in XML according to HyperTopic XML schema. A topic map, a point of view or a subset of tree structure from the map can be represented and exported with dedicated feature to XML representation in special format[2] we propose calling XHT (XML HyperTopic). That takes a part of our effort to propose HyperTopic as a standard for the socio semantic Web. Such standardization, supplementing XML and possibly RDF syntax level), can be very useful, including for the very practical needs recovering and easily merging topic maps in a shared format. This facilitates exchanges and accelerates manual handling. Bridges and conversations are also considered from XHT towards XTM and XFML.

4 The context and the “open source software” application

4.1 Goals, field work and motives of actors

This part aims at clarifying the goals and the specific method applied, using the general elements outlined previously (§3), and applied while building the topic map for open source software. As well as developing an operational Web application, the final goal is indeed to build a singular socio-technical system, this work of construction mainly being done by the actors. The makers of a socio-technical groupware (like this topic map), even if they can start the process, and take part in it, they can neither put all the knowledge in the system, nor cooperate in place of the ac-

[2] For more details, see www.sociosemanticweb.org.

tors. That is the reason why we want to precise quickly some goals and choices we have made in the context of the community studied (dealing with management and knowledge management).

The “open source software” topic map targets all developers and users of this kind of software: in this system, everybody can propose, classify and easily find online complete software or components; an actor can describe a software he has built, or also a software he knows of (because he uses it or he tested it), according to a very complete and varied set of themes. A few weeks after its beginning, the topic map contains already several hundreds themes, and a hundred open source software are linked to it.

At this first level, the aim is to propose a simple tool easily usable in order to search and describe open source software: a kind of marketplace, getting back to the original inspiration of the KBM model (cf. §2.4). The resulting topic map must also condense the technical expertise shared by the community, while staying open, commercially independent and aiming quality.

In comparison with these ideal objectives, building a semantic structure for a map is a real difficulty, especially at the beginning for the priming team. To reduce this difficulty, the team has competences in knowledge engineering and open source software (people expert in this field must have had complete but different experiences). The team involved in the initial conception of the topic map is composed of about ten people (including the authors of the present article) member of three laboratories (BETA-ULP in Strasbourg, Tech-CICO in Troyes, LIMSI in Orsay).

This double competence condition seems important to be successful in the points of view priming phase, and it is worth noting that the majority of open source software experts are rarely competent in knowledge engineering.

The motivations for building this system must be strong during the priming phase, but also in the widening one, in

Figure 6 – www.yeposs.org homepage for the French-speaking Open Source Software Community

order to conduct to a real participation of many actors as map editors. How to motivate these actors, beyond a simple curiosity for an innovating concept of knowledge map co-construction? The act of giving a structure to a topic map, even just a little structure, requires standing back from the problem, which is difficult, unusual, and indeed inhibitive for many actors, when the actors we are searching for are overwhelmed by their work. Building a shared structure also requires being part of a group exchange (even if the tool aims at making this pleasant, less difficult and not time consuming). We must forget to ask many real world actors to build the topic map only for that purpose without compensation for all the work needed: thinking, classifying, discussing, and negotiating.

However, these compensations exist. At the beginning, the membership or the desire to be part of a community is important and even more essential. In this context, an actor can regard as gratifying to be an “author” of knowledge elements, if he (or she) knows that the trace of his (or her) creative contribution will survive in the system one way or another. There is a common goal and a “win-win” strategy: if the actor puts knowledge in the system, he will derive knowledge from the other members.

These first aspects of motivation are real but are not always enough to create a strong motivation. According to us, they are reinforced by the notion of entity in the HyperTopic model. This is a concrete issue which conducts not to build (only) “the knowledge for knowledge’s sake”. Describing an entity (a specific software, not any software) linked to a work related issue is also a strong way to motivate somebody (in order to introduce this software, to promote it, to diffuse it, to sell associated services, to bring the author to attention as a critic, etc.). To place a theme on the topic map or to discuss a displacement of theme involves an important cognitive and social effort. However the actor will consent to this effort more easily, because of the relation he creates between the necessary effort and the expected return for a good description of the product, therefore with its work related goal, which is equally important.

This “principle of reality” naturally conducts the actor not only to consult and take part in the topic map construction, but also to structure the map and to exchange points of view with other editors, through the proposed system: discussions, negotiations (convergences, micro-conflicts dealing with diverging points of view, lexicon, theme meaning, entity description, etc.). Thus the actor takes part in a second level of community of “semantic co-construction” which must equip itself with all ways of communication and arbitration. This second community, which we have analyzed more precisely as a community of action [Zacklad, 2003] [Cahier and Zacklad., 2004], sets itself as an internal goal (goal of service) to expand and update the topic map, taking into account the evolution of indexed entities (here, open source software), knowledge, environment, and external forces. The open source software community is not staying inactive[3] to elaborate change tracking systems, in a context where the number of open source projects is becoming astronomic, and we ought to place ourselves as a source of experimental ideas, potentially complementing these initiatives.

In our case, the open source software particularity conducts ourselves to formulate the hypothesis of a sensible overlapping of these two community levels: in the world of open source software, users are often in the same time developers and end-users, reinforcing the fact that the same people are inclined to consult, contribute and edit the topic map, and are competent for all these. At least at the first time, we thought it would be useful to distinguish these two levels of communities. This is clearly an advantage in order to develop the system more easily (there is no need to integrate a complex system of right and authorization management). The responsibility and tradition of trust in the open source community were also arguments to select this field for an application. When someone wants to contribute, he automatically receives a password corresponding to his roles (to limit risk of wrong manipulations, backups are made regularly).

After the priming phase, if the system arouses the adhesion of actors, the tradition of exchange and creativity characterizing the open source community should permit to further elaborate the application inside the community itself (for example to incorporate a right management system, or to bring further the groupware functionalities) .

4.2 Which points of view to look at OSS?

As we have seen, the topic map aims at representing a meaning both familiar in the community and efficient for classifying software. Then, it behaves like the index of a yellow pages directory for OSS, easing the access through the Web to descriptive resources about each of them. Users will try to compare or rate tools depending on business goals (for instance, integrate some software in an application, make a long-term choice, choose a mature product, evaluate future enhancements, etc.

To take all these dimensions of activity, the topic map is organized at the end of the priming step, along the following points of view:

- *themes/features*: software development, system tools, multimedia, games…
- *software engineering*: methods and tools for development, integration, deployment;
- *business models* (model of hardware and services, defensive model, Linux distribution model, dedicated software model, non-business model);
- *legal aspects*: legal point of view, licenses used, third parties rights, patents…
- *actors/stakeholders*: organizational point of view, dealing with software communities, companies, institutions establishment, or research project.

These points of view, which can't be reduced one to another, show that OSS form a complex, fast-pace evolving domain. They are neither far from being given at start, nor from being consensual, even in the small group who initiated the process. Choosing them is then a real knowledge creation about the field.

These points of view have several functions. They are used as bridges, to translate between "supply languages" and "demand languages", between more or less specialized languages, or between business dialects. In this context, stakeholders may belong at the same time to several communities: firms, geeks or computer scientists' community, OSS community, and other epistemic communities. Then, the business lexicons can be ambiguous, unstable and not consistent, and experts have to explicitly link them to the contexts in which terms are used.

The points of view (and the complete paths from each viewpoint to a topic, through every child topic) can be used to reduce meaning conflicts and to expand expressiveness. Each viewpoint matches specific languages of stakeholder roles, for instance one can distinguish between people providing software (developer, software vendor) and people wanting to get software, to use them, integrate them, make forks, etc.). Business terms are important for key activities of stakeholders, such as promotion, comparison, choice… They thus define the properties and identity of OSS tools.

The background of the semantic map is, at first, made of the knowledge of the OSS community, *i.e.* a common ground, but neither consensual, nor explicit or as consistent as a scientific model of the domain should be.

4.3 Interaction features, depending on roles

a) Home page, reading: users have a general view upon all viewpoints, and can browse among several hundred topics.

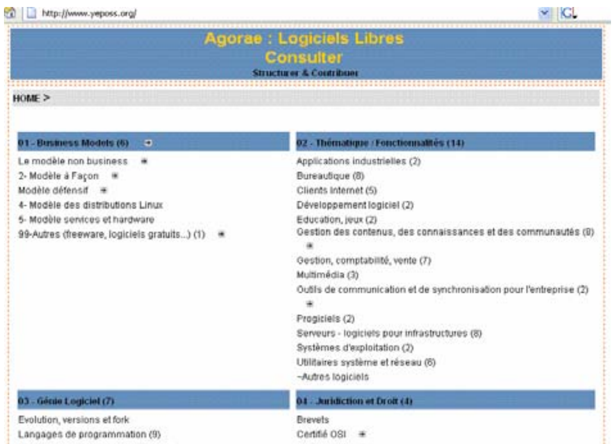


Figure 7a – OSS application, home page reading

[3] Some projects, modern XML extensions of RPM, or tools like Gentoo's *emerge*, are currently under development, by Edd Dumbill (XML Europe chairman) among others, who wants to integrate a project description inside each open source code.

b) For each topic, users can see corresponding products, and related topics.

Figure 7b– OSS application: an entity linked to the "CMS - Groupware" topic

c) Creation of an entity instance: the user describes the software by filling a text box, and by linking it to any topic, for instance the "Conflictual ontology" topic, under the "features" viewpoint. Contrary to reading pages, this is restricted to authenticated members.

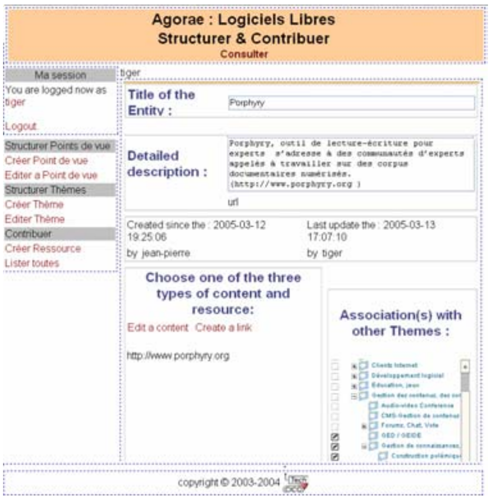


Figure 7c – OSS application, contributing

d) Creation or modification of a topic: users having the "reviewer" role can create a topic, modify its name and comments (definition, remark), and its location in the tree.

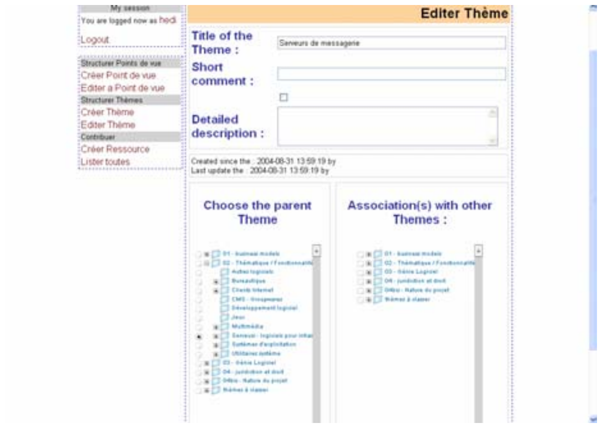


Figure 7d - OSS application, reviewing

5 Future work

At the current stage of the project, we have not definitively fulfilled our goal to "map" the OSS industry. It will be longer than expected to transfer this R&D project toward real stakeholders of the OSS Community, for instance well-known Associations. However, we think that we should reach this goal quite easily thanks to the multiple points of view. Thus, we have already tested the possibility for the actors to cooperate through the Internet in order to create a Topic Map. Besides, they also can communicate trough a forum organized according to topics and points of view from the Agoræ application. But conversations on the forum are still rare.

In the near future, we want to guaranty conditions for a safe scaling. The aim is to observe the system in wider open settings, where classification and meaning conflicts will sharpen. We would like more than twenty members to review topics regularly, and many others contributing or editing the map from time to time

Other extensions of the Agoræ generic tools are planned or under development, to enable the socio-semantic Web. For instance, to better enable the co-construction "at run-time" of a shared meaning, and better understand the conditions that may level difficulties, we want to add measuring tools to the system, in order to trace and analyze communications and discussion threads by topics on the micro-forums.

According to the evolution towards an enhanced version HyperTopic V2 mentioned above, we explore alternatives for a semantic specification of roles and rights. (cf. Figures 8a and b). As for now, the user/role association is done by a standard access matrix. We plan to let users edit these relations just the way they do for topics in HyperTopic which will allow us to use the Agoræ toolbox for administration purposes. In particular, for specific cases such as the OSS application, it could be useful to enhance the KBM roles, by defining more precise one and permit initial roles organization and modification by the members of the community themselves. For instance, to set multi-lingual maps, some translator-reviewer should be allowed, and only them, to translate topics.

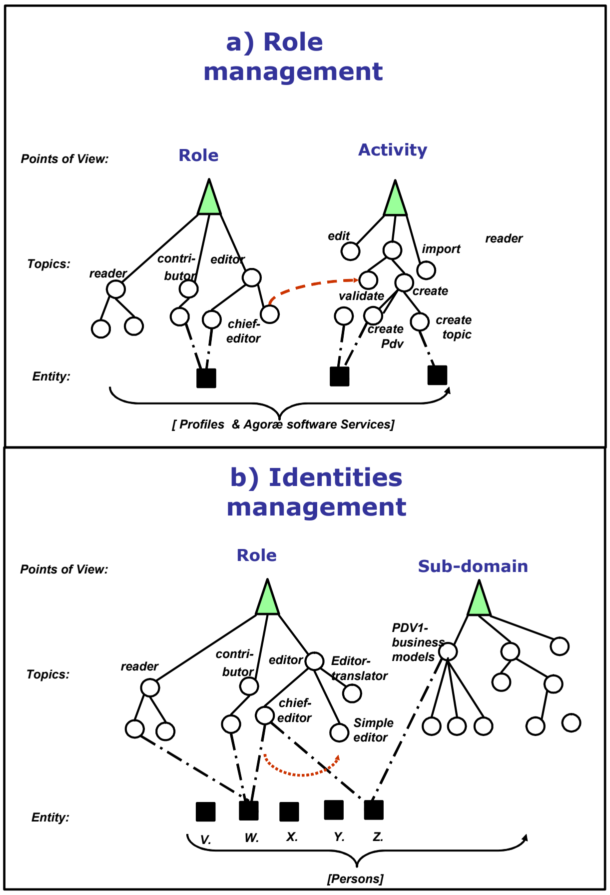


Figure 8 –Flexible edition of KBM roles (a) , of identities and rights (b), based on HyperTopic.

6 Conclusion

This OSS application is a first step that allows exploring many interesting subjects. This study has given us a framework of objectives for further research. Besides, we have the opportunity to understand and describe the groupware tools necessary for a real and concrete appropriation by the actors.

This step of our experiment encourages our hypothesis that "islands" of lively shared meaning on the Web could be created and maintained directly within and by communities. Standards such as HyperTopic could permit to link or merge such islands on a more flexible manner. But developing engineering tools and methods for co-building by actors in large scale social groups remains a very difficult challenge. Towards socio semantic Web, obstacles are numerous, needing strong R&D effort jointly on the cooperative (CSCW), Knowledge Engineering and Management facets of the solution.

On the CSCW side, the OSS application we have presented has in this first step permitted to a few distant co-designers from different universities (less than ten people) to discuss and work together in an asynchronous manner, to use the tool and to effectively "bootstrap" with HyperTopic a first draft of the OSS topic map. In a second step, it will be more arduous and long to accompany a lot of professional actors to transform this successful "demo" into a social application, giving the ability to the French-speaking Open Source community to appropriate and to complete the concepts and the tool, and possibly to use it at a larger scale.

This attempt could be an opportunity to ameliorate the tool, especially to support actors, roles and actions representation by the designers themselves. Above all, it will permit for the first time to better evaluate the socio semantic Web concept – and thus to validate the underlying HyperTopic model we propose as a standard – in the real-size context of a mature, complex and contentious community. In this perspective, management and CSCW issues will be crucial – especially the enhancement and the use of annotation and discussion threads, and other collaborative functions and complementary helps to propose to users. Will these services really facilitate members' expression, debate, mutual confidence and involvement? Can S2W standards favor a new participative dynamic to manage shared meaning artefacts? Only real-scale experiments will give answers. Future functions we develop in Agoræ, like logging of users and trace analysis of activity, could be useful to better understand the socio semantic activity in the co-builders group, and propose best communication, discussion and building services in the tool.

On the Knowledge Management side, as we begin to observeit in our small group, the continuous creation of a shared meaning in the S2W appears possible, as a crucial process allying discussion, clarification of the knowledge at a fine-grained level, creation of new knowledge, creation and management of the points of view to take in account all types of actors and needs. It could be interesting to verify in the future of the OSS application if such a S2W application can be also a source of innovation and of organisational learning. For example, in the OSS Community, we make the hypothesis that the Yellow Pages web site could progressively become a source of shared knowledge for members and visitors, experts and novices, in knowledge dimensions of width and depth.

Reference

[Bénel *et al.,* 2001] Aurélien Bénel, Elöd Egyed-Zsigmond, Yannick Prié, Sylvie Calabretto, Alain Mille.Truth in the Digital Library: From Ontological to Hermeneutical Systems. In *Proceedings of the fifth European Conference on Research and Advanced Technology for Digital Libraries*, Darmstadt, September 4-9, 2001. Lecture Notes in Computer Science #2163. Berlin : Springer-Verlag. p.366-377.

[Berneers-Lee, 2001] Tim Berneers-Lee. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, May 2001.

[Bowker and Star., 1999] *Sorting Things Out, Classification and its consequences,* Geoffrey C. Bowker, Susan L. Star. Cambridge, MA, The MIT Press, 1999.

[Brachman, 1979] Ronald J. Brachman. On the Epistemological status of semantic networks,. In N.V. Findler, *Associative Networks*, pp. 3-50, Academic Press, New York 1979.

[Cahier and Zacklad, 2002] Jean-Pierre Cahier and Manuel Zacklad. Towards a Knowledge-Based Marketplace model (KBM) for cooperation between agents. In *Proceedings COOP'2002*, St Raphael, 4 - 7 june 2002, IOS Press.

[Cahier *et al.,* 2004] Jean-Pierre Cahier, Manuel Zacklad, and Anne Monceaux. Airbus Engineering Agoræ : application du Web socio sémantique à la définition d'un annuaire métier en ingénierie,. In *Actes de la Conférence Ingénierie des Connaissances IC'2004*, Lyon Mai 2004.

[Cahier and Zacklad, 2004] Jean-Pierre Cahier and Manuel Zacklad. Socio-Semantic Web applications: towards a methodology based on the Theory of the Communities of Action. In *COOP'04 Workshop on Knowledge Interaction and Knowledge Management*.

[Engeström *et al.,* 1999] Yrjo Engeström, Miettinen Reigjo, and Punamaki Raija-Leena. *Perspectives on Activity Theory*, Cambridge University Press, New York, 1999.

[Euzenat, 1996] Jérôme Euzenat, Corporate memory through cooperative creation of knowledge bases and hyperdocuments, Proceedings of the KAW'96 Workshop on Corporate Memory and Entreprise Modelling

[Kassel and Perpette, 1999] Gilles Kassel and Sébastien Perpette. Co-operative ontology construction needs to carefully articulate terms, notions and objects,. In *Proceedings of the International Workshop on Ontological Engineering on the Global Information Infrastructure*, Dagstuhl Castle (Germany), 1999, pages 57-70.

[Hutchins, 1995] Edwin Hutchins. *Cognition in the wild,*. Cambridge, Ma, MIT Press.

[Maedche and Staab, 2003] Alexander Maedche and Steffen Staab. Ontology Learning,. In S Steffen Staab and Rudi Studer (eds.) *Handbook on Ontologies in Information Systems*. Springer 2003.

[Peirce, 1868] Charles S. Peirce. Upon Logical Comprehension and Extension, (1869) (published in " Proposed memoirs on Minute Logic", Carnegie Inst.,1902). Peirce, *Collected Papers*, vol.2, Elements of Logic, CP 2 .391-2.434, Harvard University Press

[Ribes and Bowker, 2004] David Ribes and Geoffrey C. Bowker. Ontologies and the Machinery of Difference: Towards a Sociology of Knowledge Representation. Journal of the Association of Information Systems (JAIS), Special Edition on Ontologies, 2004.

[Schmidt and Simone, 1996] Kjield Schmidt and Carla Simone. Coordination Mechanisms: Towards a conceptual foundation for CSCW system design. *Computer Supported Cooperative Work* (CSCW) An International Journal, vol.5, n°2-3, 1996

[Simone, 2000] Carla Simone. "Unifying or reconciling when constructing Organisational Memory? Some Open Issue". xtd version of ECAI 2000 Workshop on KM/OM

[Suchman, 1987] Lucy Suchman. *Plans and Situated Actions, the problem of Human Machine Interaction*. Cambridge University Press, 1987

[Tennison and Shadbolt, 1998] Jenifer Tennison, & Nigel R.Shadbolt, APECKS: a Tool to Support Living Ontologies, In B.R. Gaines & M. Musen (eds.) Proceedings of the 11th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW'98) , Banff, Canada, April 18-23, 1998

[TM, 1999] *ISO/IEC 13250:2000 Topic Maps: Information Technology -- Document Description and Markup Languages*, Michel Biezunski, Martin Bryan, Steven R. Newcomb, ed.3 Dec 1999. http://www.y12.doe.gov/sgml/sc34/document/0129.pdf

[Uschold and Gruninger, 1996] Michael Uschold and Michael Gruninger. Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, 11(2), 1996, pp 93-136

[Uschold and Jasper, 1999] Michael Uschold and Rob Jasper. A Framework for Understanding and Classifying Ontology Applications. *Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods* (KRR5) Stockholm, Sweden, August 2, 1999, (V.R. Benjamins, B. Chandrasekaran, A. Gomez-Perez, N. Guarino, M.Uschold, eds.)

[Veltman, 2004] Kim H.Veltman. Towards a Semantic web for Culture. *Journal of Digital Information*, Volume 4 Issue 4, Article No. 255, 2004-03-15

[Vygotsky 1978] Lev S. Vygotsky. *Mind in society*. Cambridge, MA: Harvard University Press

[XFML 02] eXchangeable Faceted Metadata Language, XFML Core Specification, 8 oct.2002 http://xfml.org,

[XTM 01] XML Topic Maps (XTM) 1.0, TopicMaps.Org Specification, 2001. http://www.topicmaps.org/xtm/1.0]

[Zacklad, 2003] Manuel Zacklad. Communities of Action: a Cognitive and Social Approach to the Design of CSCW Systems. in *Proceedings of GROUP'2003*, pp. 190-197, Sanibel Island, Florida, USA.

[Zacklad, 2005] Manuel Zacklad. Vers le Web Socio Sémantique : introduction aux ontologies sémiotiques. IC'2005 Ingénierie des Connaissances, Nice, june 2005

An Operational Learning System for the French Textile - Clothing Institute

Oswaldo Castillo, Nada Matta
University of Technology of Troyes
Laboratory Sciences and Information technologies Institute of Troyes
Equip Cooperation Technology for the innovation and the organizational change
12 rue Marie Curie, BP 2060, 10010 Troyes cedex, France
{oswaldo.castillo,nada.matta}@utt.fr

Abstract

The management of knowledge and know-how occupy a more and more important place in organizations. The construction of corporate memories in a purpose of conservation and sharing of knowledge became a rather common practice. However, we forget too often that the efficiency of these activities is strictly connected to the capacities of appropriation and learning of the organization actors. It is through this learning that new skills can be acquired. In this paper, we propose general guidelines facilitating the process of creation and appropriation of professions memories built by means of methods stemming from the knowledge engineering and from the techniques of the educational engineering.

1 Introduction

In the framework of definition of a training platform for workers in the knitting industry, the French Textile - Clothing Institute (IFTH) is leading a project of knowledge capitalization on the "knitting of 3D pullovers". The object is to train the textile companies of a department of France in this type of knitting in order to improve the manufacturing process. Our role in this project is to define and conceive an operational learning system on the knitting of 3D pullovers. The system we defined is based on knowledge engineering and educational techniques.

2 Knowledge management

With the purpose to take highest advantage of their intellectual capital, at present, the IFTH and most of the others companies develop and apply strategies of knowledge. Thereby it becomes more and more easy to realize a management, an innovation and a creation appropriate of their knowledge.
Dieng-Kuntz [Dieng-Kuntz *et al.,* 2001] presents some definition on the knowledge management. "The knowledge capitalization in an organization has for objectives to favor the growth, the transmission and the preservation of the knowledge in this organization" [Steels, 1993]. "It can carry both the theoretical knowledge and the know-how of the company. It requires the management of the resources of knowledge of the company to facilitate their access and their re-use" [O'Leary, 1998]. "It consists in capturing and in representing the knowledge of the company to facilitate their access, their sharing and their re-use. This very complex problem can be approached by several points of view: socio-organizational, economic, financial, technical, human, and legal" [Barthès, 1996].

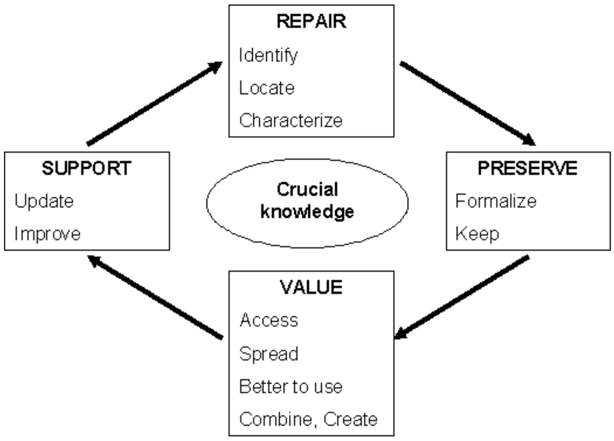


Figure 1. Knowledge management cycle of life

We adopted the cycle of life of the knowledge management proposed by Grundstein (figure 1), where, according to him, "for any operation of knowledge capitalization, it is important to identify the strategic knowledge to be capitalized" [Grundstein and Barthès, 1996].

3 Corporate memory and profession memory

In organizations, the formalized knowledge is more and more represented in the form of corporate memory. We can define this memory as the "explicit, persistent and disembodied representation of knowledge and information in an organization" [Van Heijst *et al.*, 1996]. A corporate memory should supply "the good knowledge or information to the good person at the right time and to the good level" [Dieng-Kuntz *et al.*, 2001]. One of the types of the corporate memory [Tourtier, 1995] is the profession memory, which clari-

fies the repository, the documents, the tools and the methods employed on a given profession. Our study is based essentially on the appropriation of the knowledge from a profession memory.
For the knowledge capitalization in the IFTH on the knitting of 3D pullover, we are going to focus on the design and the use of knowledge based profession memory. This is, based on the collection and the explicit modeling of knowledge of a number of experts and specialists of the company, or knowledge that come from a document [Dieng-Kuntz *et al.*, 2001].

4 Appropriation of a profession memory

One of the main motivations for the construction of a memory in a company is the improvement of employee's learning. This learning can be individual, group or organizational level [Dieng-Kuntz *et al.*, 2001]. In the case of our work in the IFTH, learning will be approached in an individual way for their employees. By looking at our preliminary experiences under construction of profession memory and more specifically the learning from such memories, we noticed that the learning from a profession memory is not easy. These memories are generally presented under several points of view (classifications, constraints, processes, problem solving strategies, etc.). The links between these views are put in background because the knowledge formalization shows the nature of the knowledge [Castillo *et al.*, 2004]. Learning and following the learning progress in such a memory can be easy for an information works or a knowledge engineer but it is complex for an employee who is specialist on his profession and who wants to learn a know-how formalized by an expert in his domain.
To facilitate the learning from a profession memory, we adapted techniques from the educational engineering by modifying the way of building the profession memory, and especially, by showing this memory, to the employee in a different way.

4.1 Educational engineering

According to Paquette [Paquette, 2002], educational engineering or training engineering contains all principles, procedures and tasks that allow:

- defining the contents of a training by means of a structural identification of the knowledge and the skills recounted,
- realizing an educational scenario of the training activities and to define the context of use and the structure of the materials of learning,
- defining infrastructures, resources and services necessary for the distribution of the lessons and for the preservation of their quality.

We use the general architecture proposed for the educational engineering, constituted by four interdependent modules:

- Expert module: represents the knowledge of the taught subject;
- Student model: represents the students needs, his level and his difficulties, which represents, the state of his knowledge, as well as the history of his interactions with the system;
- Educational module: contains the educational system, its purpose is to plan and to pass the learning;
- Interface module: concerns the management of support and the system's communication modes with the student.

In other terms, the educational engineering allows us to develop a learning system [Rolland, 2000; Paquette, 2002]. Our particular interest is the construction of a system based on a profession memory. By taking into account of the practical knowledge (that is, problem solving) of the training contents, our system becomes an operational learning system.

5 COLS: for the Construction of an Operational Learning System

In this article, we are only going to present our proposition of the operational learning system's expert module for the knitting of 3D pullover to the IFTH.

5.1 Proposition of the expert module

Our expert module is represented as a guide named "activity process". It is also used to focus on detailed explanation of every step of the process.

Representation of the activity process
For the definition of the expert module, the knowledge engineer has to find and emphasize a guide to assure the progress in learning, with the expert's help. This guide has to show the objectives, the deep knowledge and the links between the knowledge. Based on MASK [Ermine, 2002], the proposed guide (figure 2) in the operational learning system on the knitting of 3D pullover in the IFTH, is an arrangement which allow to better understand the succession of stages to be realized within the activity.

From this process, the knowledge engineer, always working with the expert, has to represent the interactions of every stage of the activity process and the problem solving strategy.

Representation of the interactions at each step in the activity process
The knowledge engineer and the expert have to represent the interactions of every step of the activity process with their environments. Always based on MASK [Ermine, 2002], each step proposed (figure 3) in the operational learning system is described according to its inputs, its outputs, its resources and its actors.

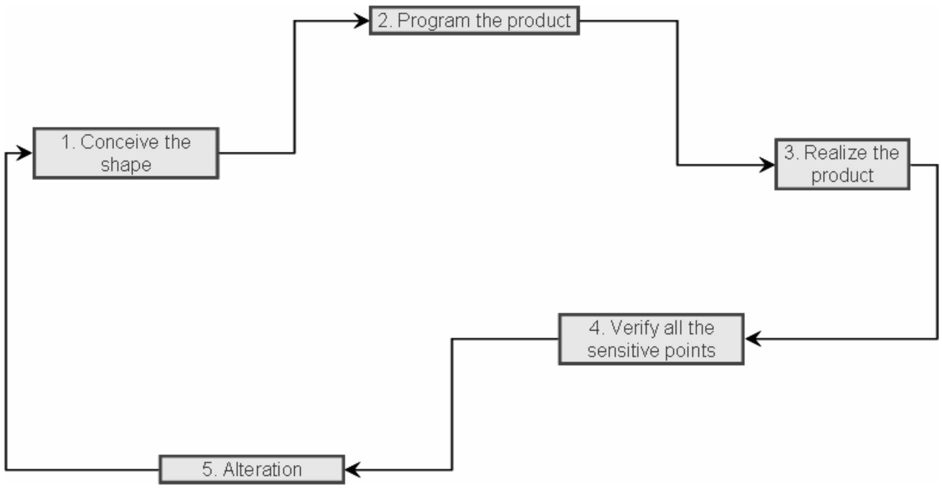


Figure 2. Representation of the activity process

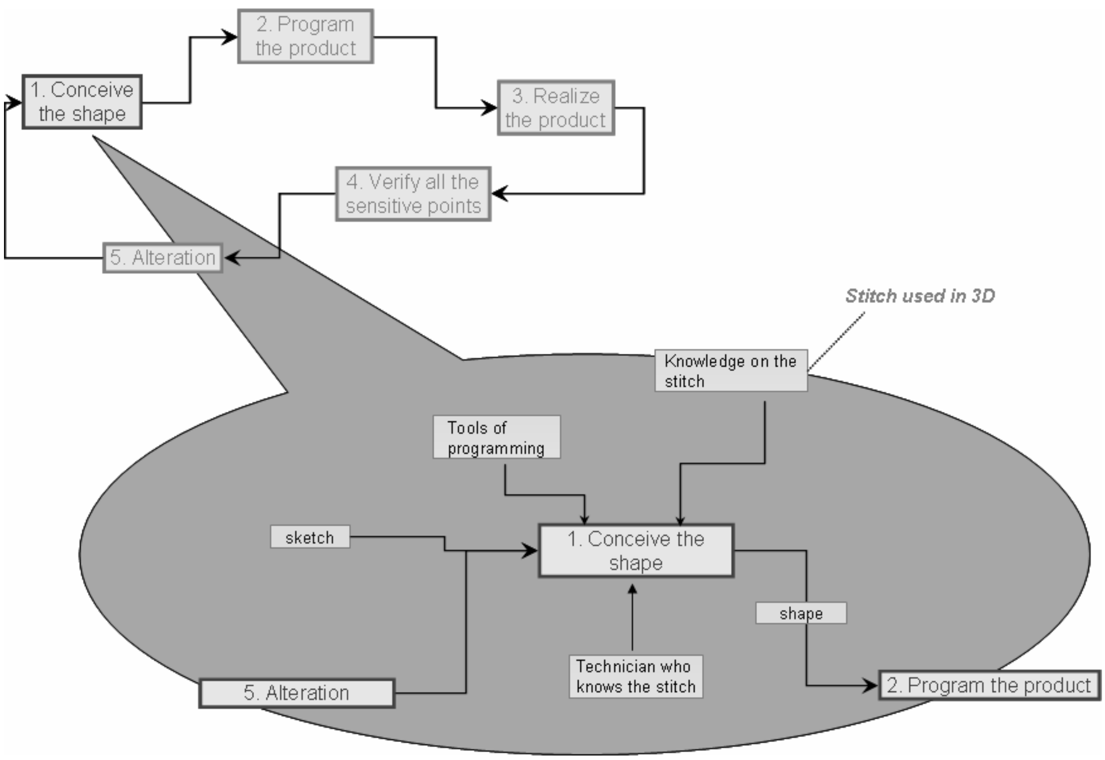


Figure 3. Representation of the interactions at each step in the activity process

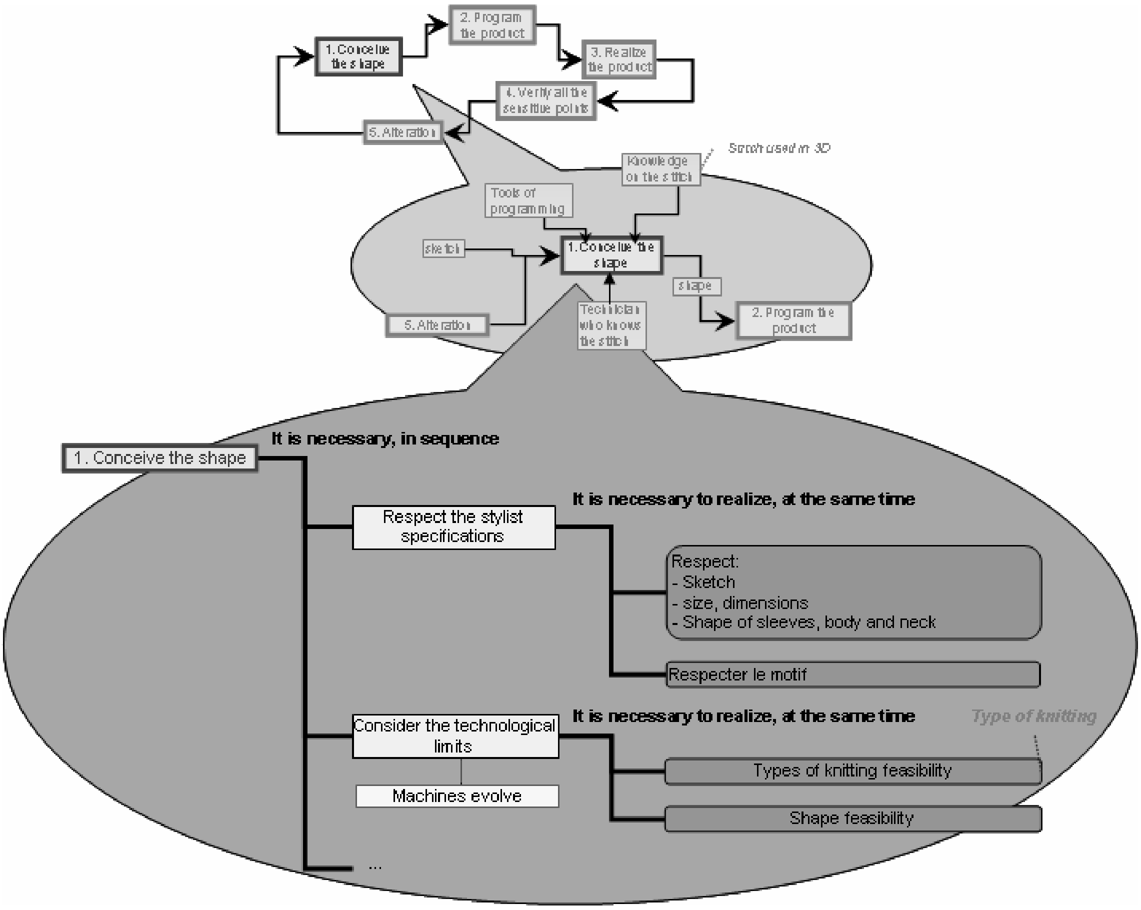


Figure 4. Representation of the problem solving strategy

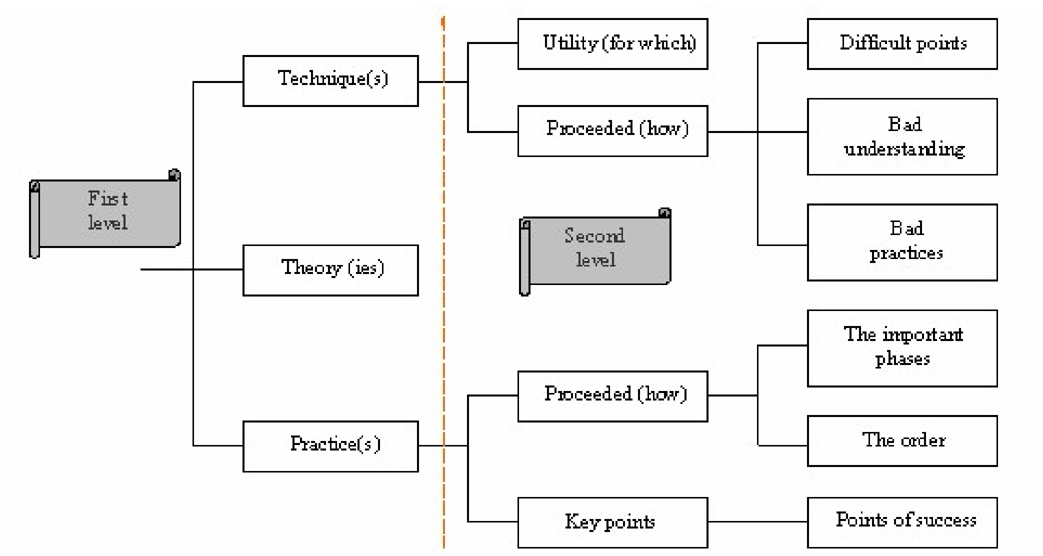


Figure 5. Guidelines tree for quizzes construction

Representation of the problem solving strategy

Every step in the activity process must be detailed, by showing how the expert behaved to reach his objectives. The problem solving strategy is generally represented by the objectives to reach and the sequence of the steps which were followed by the expert.

Every problem solving strategy (figure 4) in the operational learning system on the knitting of 3D pullover of the IFTH is represented by a list of objectives. Each objective contains a list of steps and tasks that the student has to develop.

Representation of Quizzes

Quiz, as tool of formative evaluation in our case, are built from externalization conversations with the expert. Knowledge engineer objective is to help the expert to emphasis the difficulties of learning of its know-how; specifically the difficulties of the domain, the bad understandings and the bad practices (figure 5).

A guidelines tree can help to knowledge engineer, to guide the expert to build the quiz questions (figure 5). Quiz is identified so as to advance the utility, the process and the key points of the domain.

The questions in an evaluation can be long opened answer, short opened answer, multiple choice, unique choice, true or false choice, multiple answer, crossing, among others. Quiz has to provide to student a support for every answer, as for example, by connections to related documents. A score at the end of every quiz allows helping the student to estimate its level of learning.

Representation of exercises

To allow training and getting a practical experience, the operational learning system offers the possibility of realizing exercises. On one hand the exercises allow learning in several phases of the operational knowledge, on the other hand an evaluation of the student's progress. These exercises mainly allow learning of problem solving strategies.

The learning system allows the student to realize exercises of various complexity levels. The execution of the exercises is also guided by the corresponding problem solving strategy. Indeed, this strategy is revealed to the student step by step with these objectives and the control structure at each step. The student is then invited to execute the instructions one by one.

An exercise (figure 6) in the operational learning system on the knitting of 3D pullover in the IFTH can be represented with: a statement, indications and observations, a problem solving strategy, a prototype of results (illustrations) and the result characteristics (list to tick and links with the corresponding skills).

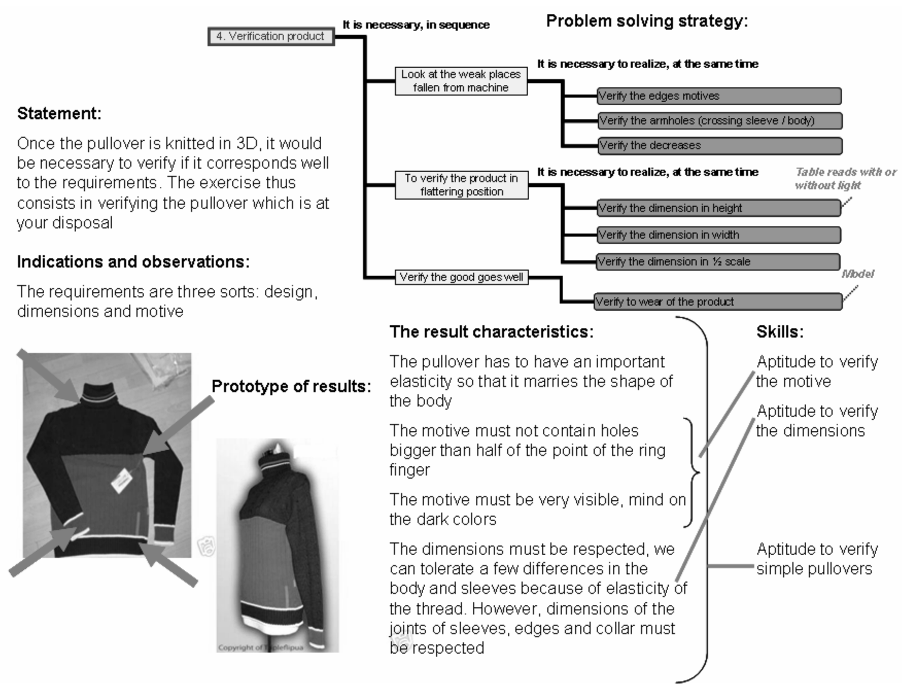


Figure 6. Representation of exercises

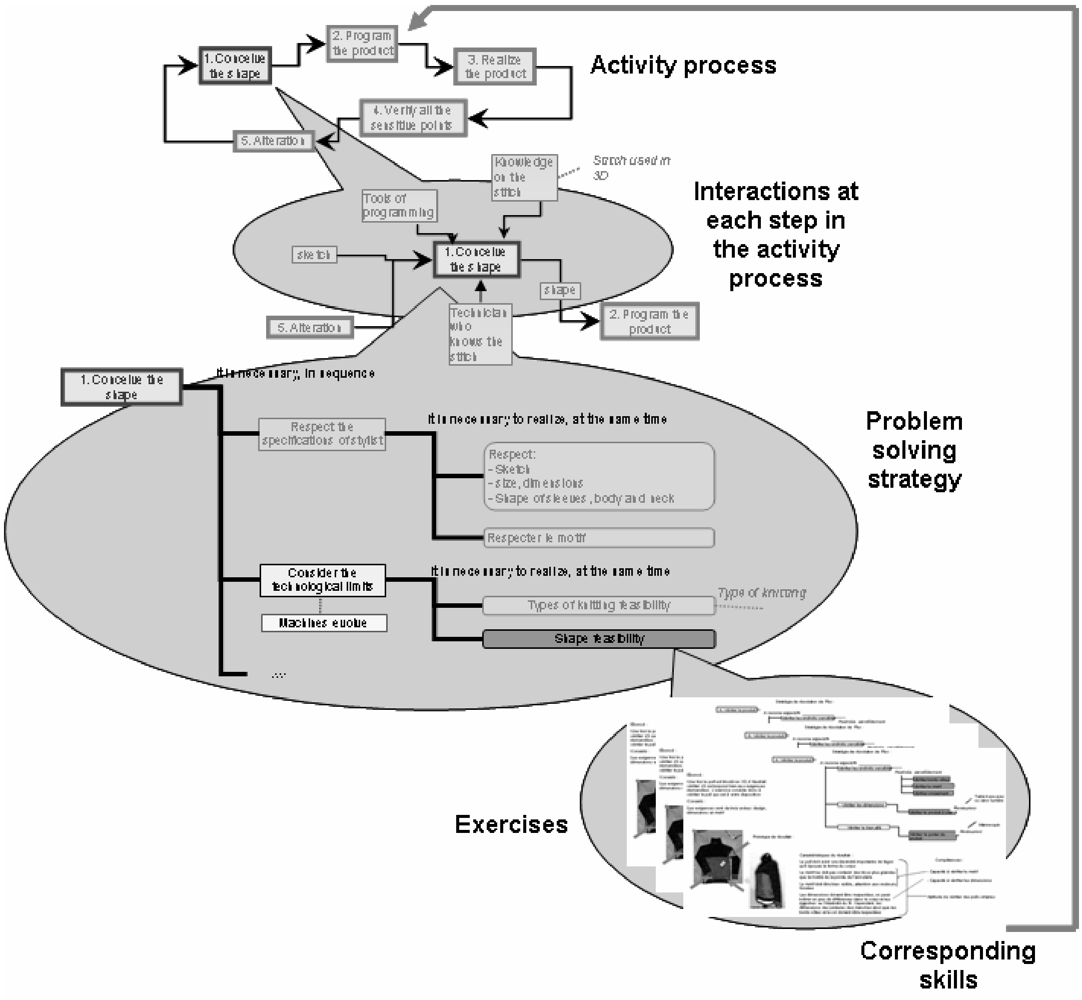


Figure 7. Process illustration of knowledge learning progress

Process of progress for the skills acquisition
The activity process can be a guide of the operational knowledge learning. It allows to supply a global view as well as to guide a student to focus on every step of the process and to learn the strategies as well as the difficulties of the activity. We show the student the global process and we invite him to explore the system step by step. For every step, the interactions with the environment are shown at first, followed by the corresponding problem solving strategy (figure 7).
The link between skills and characteristics of the exercise's results allows the progress in the acquisition of the skills relative to every activity level. More global exercises are finally presented to a student to assure a global understanding of the activity.

6 Conclusions

The knowledge management is a process which contains the capitalization of the knowledge as well as the sharing and the appropriation of this knowledge. However, the knowledge appropriation remains a subject to study. The phase of appropriation requires a particular attention because its success determines the efficiency of the organizational learning and thus the performance of the company partially.

Otherwise, educational engineering study techniques of learning and supply devices (teaching equipment, tools of evaluation and process of learning) that can be a good support for operational knowledge appropriation. We studied these techniques to supply an operational learning system based on a profession memory.

We are defining a tool to support this type of learning systems. Together with the IFTH, an application for the textile domain is being developed. The experiences on this type of domain will allow to deepen our studies and to enrich our learning systems by other techniques.

The learning system on the knitting of 3D will be integrated into a platform of training which will be integrated into the textile industry in our region. We intend to study the feedback of this type of training.

We showed the link which can exist between the knowledge appropriation and the skills acquisition. We aim at studying the representation formalisms and techniques of skills acquisition in a more detailed way.

References

[Barthès, 1996] J.-P. Barthès. ISMICK and Knowledge Management; In J.F. Schreinemakers (éd.), Knowledge Management: Organization, Competence and Methodology, In *Proceedings of* ISMICK'96, pages 9–13, Würzburg, Ergon Verlag, Rotterdam,1996.

[Castillo *et al.*, 2004] O. Castillo, N. Matta, J.-L. Ermine. De l'appropriation des connaissances vers l'acquisition des competences. In *Proceedings of* C2EI'04, Nancy, France, 2004.

[Dieng-Kuntz *et al.*, 2001] R. Dieng-Kuntz, O. Corby, F. Gandon, A. Giboin, J. Golebiowska, N. Matta, M. Ribière. Méthodes et outils pour la gestion des connaissances: une approche pluridisciplinaire du knowledge management. 2e édition, Dunod, 2001.

[Ermine, 2002] J-L. Ermine. La gestion de connaissances. Hermès sciences publications, Paris, 2002.

[Grundstein and Barthès, 1996] M. Grundstein, J.-P. Barthès. An industrial view of the process of capitalizing knowledge. In *Proceedings of* ISMICK'96, pages 258-264, Rotterdam, 1996.

[O'Leary, 1998] D.E. O'Leary. Entreprise Knowledge Management. Computer, 31, 3, pages 54-61, 1998.

[Paquette, 2002] G. Paquette. L'ingénierie pédagogique: pour construire l'apprentissage en réseau. Presses de l'université de Québec, 2002.

[Rolland, 2000] M. Rolland. Bâtir des formations professionnelles pour adultes. Editions d'Organisations, 2000.

[Steels, 1993] L. Steels. Corporate knowledge management. In *Proceedings of* ISMICK'93, pages 9–30, Compiègne, France, 1993.

[Tourtier, 1995] P.-A. Tourtier. Analyse préliminaire des métiers et de leurs interactions. Intermediate report of the project GENIE, INRIA-Dassault-Aviation, 1995.

[Van Heijst *et al.*, 1996] G. Van Heijst, R. Van der Spek, E. Kruizinga. Organizing Corporate Memories. In *Proceedings of* KAW'96, 10th Banff knowledge acquisition for knowledge-based system workshop, Banff, Canada 42-1/42-17, 1996.

AnT&CoW, a tool supporting collective interpretation of documents through annotation and indexation

Gaëlle Lortal[1], Myriam Lewkowicz[1], Amalia Todirascu-Courtier[2]
[1]Université de technologie de Troyes
ISTIT Laboratory, Tech-CICO
12, rue Marie Curie BP 2060 10010 Troyes Cedex
{lortal, lewkowicz}@utt.fr
[2]Université Marc Bloch de Strasbourg - EA1339
22, rue René Descartes 67084 Strasbourg
amalia.todirascu@umb.u-strasbg.fr

Abstract

This paper describes an Annotation Tool supporting Collaborative Work (AnT&CoW) and particularly collective interpretation of documents using annotation. In the first part, we present our methodology to design such a groupware based on a theoretical activity analysis, understanding discourse production activity as a complex writing/reading activity. Following a rhetorical discourse production theory (section 4), we model a discourse production activity and its mediatization by way of a tool (section 5). After existing annotations standards and tools have been detailed (section 6), we present our tool's requirements (section 7). AnT&CoW is following Annotea W3C standards and allows document annotating and then multi-dimensional indexing. Multi-dimensional indexing is based on a semiotic ontology represented in Topic Maps where three dimensions occur: argument, role and domain. Dimensions, mainly the domain specific dimension, are based on Natural Language Processing (NLP) techniques fitting the text up. In the last part, we present our Web-based application, its client/servers architecture and its visualization's features. Our prospects are then proposed.

1 Introduction

Nowadays, documents are a central point of interest in our organizations as many works in research show. For instance, in France, a multidisciplinary network from CNRS (National Centre for Scientific Research) works on "documents and contents: creation, indexation, navigation" (RTP-DOC). Three orientations around documents are then told apart; analyzing documents as a shape (studying structure of documents for its manipulation), as a sign (studying author's intentions when creating documents, document's intentionality), and as a medium (studying document's status in social relations) [Pédauque, R.T., 2003]. Following the second orientation, this paper intends to consider a document as a meaning-holder, which cannot be dissociated from a subject who is building or re-building it and who gives sense to it. Seeing document as a sign means that we are more interested in the creation process of a document, its interpretation, in other words in signs constituting it.

These questions are tackled here from the critical reading point of view, contrary to a reading which would not aim at producing knowledge or another text. A critical reading creates an interpretation enlightening not only the text that is read but also other texts. It can produce another text, a comment, a review, a criticism. We focus particularly on collective critical reading, which allows the building of a shared interpretation of an initial document between several participants. The drawing-up of a shared interpretation within a group takes part, according to us, of a collective sense making process [Weick, 1979]. Actually, Weick defines collective sense making in organizations as a process of collective reduction of the perceived ambiguity of a situation. By exchanging, discussing ideas, members of an organization will clarify and then share their understanding of a situation (transcribed in documents), gradually making sense.

Collective sense making in organizations from real lived situations is a theme that has been studied since the beginning of 80's. Weick's work emphasizes the sense making process, its creation and its evolution, and not the collective representation of sense. The collective sense is then not necessarily a common sense.

According to us, the collective interpretation of documents, which are the marks of the actions in the organization, will allow collective sense making. This cooperative interpretation process thus permits to take advantage of documents while letting able to overstep the setting in which the documents have been created. This process is also supporting individual identity since each participant puts his identity to the critical test, making it evolve through his/her interactions.

We propose to support this collective interpretation of documents by developing strategies for mediatized interactions around numerical documents, mostly textual. Texts' interpretation is traditionally accompanied by gloss, note, commentary, and various kinds of annotations anchored to the text itself or linking several texts or fragments of text.

We then propose to support this discursive collaboration around documents by a system allowing documents' annotation for interpretation and appropriation, objective which is not yet supported by existing annotation-based software. Actually, these tools only allow isolated annotation as textual comments, with weak indexation (date, author), hardly usable as interactions' support in a group. In fact, in a situation where we want to support a methodical texts' interpretation, textual body of comments is promoted to discourse, its context is built up by the role of the author, the semantic content, the place of the annotation into the discussion's thread. Giving this context is essential to find the design rationale of an interpretation.

Studies have been conducted at the KMI (Knowledge Media Institute) on functions of discursive comments of a document. They gave rise to the "Digital Document Discourse Environment" (D3E) [Sumner *et al.*, 2000], a web tool in which exchanging messages on a document are allowed. But, as the design of this tool has not been bound to any study of the activity of document analysis, the collaborative process of interpretation is not treated. Moreover, nothing has really been done on visualization and reuse of the exchanged messages. Actually, messages are tree-displayed and indexed according to standard attributes (date, author, title); it is as though a forum has been linked to a document. In fact, many works outline yet that online discussions are often disrupted and confused because of the numerous and frequent development of discussion threads and parallel talks. We can for example quote [Marcoccia, 2004] who stresses the phenomenon of progressive themes' digression in newsgroups, when each message in a thread introduces a theme development. The result could be a real "topic decay" [Herring, 1999].

In this paper, we first present methodological principles to design a groupware supporting active collective interpretation of documents (AnT&CoW). Then, we focus on existing works in modeling writing activities. In section 4, we present a model of discourse production stemming from rhetoric, which is adapted in section 5 to a mediatized activity. This model is the basis of AnT&CoW, which features are described in section 7, after a review of existing tools and standards for annotation in section 6. We finally present the tool architecture combining Natural Language Processing (NLP) techniques for text material processing

2 Methodological principles to design a groupware supporting collective interpretation of documents

The context of our research leads us to define new practices to support collective interpretation of digital document. Then, a classical software design process, deducing design principles from a needs analysis or an existing activity analysis, is not suitable.

The design process which we are presenting here draws its inspiration from the methodological positioning in the field of design in Educational Research by [Baker, 2000], carried on, in France, by Tchounikine [Tchounikine, 2002]. These authors distinguish models as scientific tools from models to design systems. The firsts propose a theory to understand or predict a situation or an activity; the seconds translate the firsts in models allowing design and implementation of systems supporting the situation or the activity.

However, theories from humanities usually mobilized to design groupware (activity theory, learning theory, communicative action theory…) are very difficult to use as they are. In fact, it's difficult to deduce principles of design or to adapt the definitions of these theories in a computer-mediatized framework.

Designing consists then in defining new models, with new concepts, in keeping with the theory, in order to describe an artefact supporting and marking interactions. The theory will then help us to analyze these recorded interactions.

We thus propose the following process, illustrated in Fig. 1: From a social science theory fitting to phenomena which one wishes to support/observe, use or define a *descriptive model* of these phenomena which makes the theory operational. This descriptive model allows reasoning about situations in which these phenomena would be mediatized using an information processing system. This reasoning leads to the creation of a *mediatized activity model*. This step involves researchers in humanities and social sciences responsible of the link with the description model, and computer science researchers (designers), understanding and controlling software properties. This mediatized activity model is then materialized in a *design model* describing requirements for a groupware enabling to assist interactions and also to mark them. This groupware will thus be a mean to collect corpus. This corpus, analyzed using the mobilized theory, will allow us to make evolve our comprehension of the phenomena being studied.

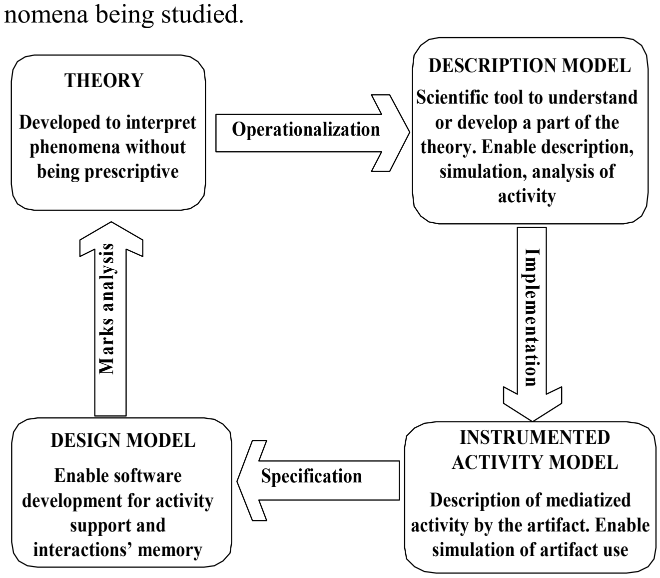


Fig. 1 – Groupware design based on a theoretical activity analysis

It seems to us that although the step of designing mediatized activity is always present while designing software, the activities of this phase are not usually explicit. It occurs as if it was possible to define design principles of an artifact supporting an activity, directly from the descriptive model of the face-to face-activity. However no one would deny that this mediatization has an impact on the activity. During this step of designing the mediatized activity the exchanges between researchers in humanities and social sciences and researchers in information and communication technologies will take place. They will then be able to build a common model reflecting the guidelines of the activity and the ways to assist this activity at the same time. This step allows the next step of design to take place. A design model will then be defined, describing the functions of the tool.
In the following section, in order to define a description model which fits our problematics of collective interpretation of documents, we present existing work on analysis of documents centered activities

3 Which theory to analyze discourse production activity?

In the field of cognitive psychology, many researchers have studied mental activities involved in writing, distinguishing text *comprehension* and text *production*.
With regard to *comprehension* models, the researches focus on memorization of text fragments, necessarily summarized. One of the most quoted model in this field is the Kintsch's constraint-satisfaction process [Kintsch, 1988]; The comprehension of text is described there as a cycle of phases of construction of a coherent mental representation of a text in the course of reading, and of phases of selection (or not) of text fragments for memorization (integration). Researches were undertaken to use this descriptive theory at constructive ends, for example for defining design principles for hypermedia documents to be easily integrated by the reader [Garlatti and Iksal, 2000]. These authors propose a guide for "good practices" in designing documents, particularly to ensure text coherence. These documents are then presented so that the reader receives help in constructing his mental model. The aim is to minimize the cognitive cost while reading the document.
Concerning *production* models, the stress is laid on editorial processes of planning, formatting and reviewing, and the control model which allows to apply these processes. The authors frequently quoted in this field are [Hayes and Flower, 1980] who proposed models of editorial strategies. Here again, this descriptive theory was used in works which gave rise to computer-supported editorial processes. In [Piolat *et al.*, 1989] a combination of three pieces of software (scripsis, scripap, scriprev) is used. Each one focuses on a process (planning, formatting, reviewing). However this work doesn't aim at proposing tools for text production within an organization, but at providing a framework for experimental study of text production.
As we presented in section 2, our approach consists in designing a groupware on the basis of the theoretical analysis of the collective activity this groupware intends to support. The descriptive models of comprehension or production offered by cognitive psychology, which we quoted above, do not appear suitable according to us for the design of a tool supporting collective interpretation of documents. In fact, they separate the memorization phase from the text formatting phase. Indeed, collective interpretation of documents mixes written activities during reading - annotations – and reading activity to produce meaning, sense. The reading/memorization phases and writing/integration are thus associated. In researches related to written didactics, reading and writing are also seen as stages of a generic activity related to the written support [Barré de Miniac, 2000]. We thus propose to use a discourse production model stemming from ancient and medieval rhetoric didactics, representing in a whole cycle both memorization and discursive production.

4 Discourse production model

Writing is the place of complex and evolutionary interactions between emotional, cognitive and linguistic factors [Barré de Miniac, 2000]. We will be interested more particularly in the cognitive factors as organizing factors of the concepts in memory and text, and in the linguistic factors as marks at the same time of a specific type of discourse and of the semantics of the document in "co-text". As an author's discourse is surrounded by social life and events, a text is surrounded by textual context making its sense.
We find these two types of factors in the rhetoric didactics. From Aristotle rhetorical theories to Hugues de St Victor's ones through Cicero or Quintilianus, discourse production is taught according to a process. Aristotelian rhetoric is focused on a final production of oral discourse (speech) without denying a memorizing phase required for any production. This phase of memorizing is better represented by rhetoric, that we will call memorial, held by thinkers quoted by [Carruthers, 1990], such as Quintilianus (the institution oratory), Cicero (De oratore, De inventione) or Tullius (Ad Herennium), and then Hugues de St Victor (Didascalicon), Fortunatianus (Artis rhetoricae libri tres) or Julius Victor (Ars rhetorica) from Middle Ages. In this approach of rhetoric, we can observe a continuum between the memorial part more "logical" or "dialectical" and the stylistic, editorial part. Rhetoric is regarded as an alliance between structuring and eloquence.
The discourse production process as recommended in this didactical context is made up of two phases: "Divisio" and "Compositio". Divisio is done while reading and consist in dividing a text in understandable units, in memorizable short segments. Compositio is the ordered combination, the suitable arrangement of "res" (conceptual or material objects) contained in the memorized segments (Fig. 2). These memorizing, Divisio, and creation phases, Compositio, are themselves divided into stages supported by the use of annotations.
The first stage of Divisio is Cogitatio. It is an individual memorial stage which consists in associating, by a conscious choice and recall, images and sections of a chrono-

logically divided content of a document in various memorial places. Textual fragments that form the text are then structured and become easily memorizable.
Collatio is the phase where textual fragments stored in several distinct places in memory are combined in a structure. In this phase connections between the various places of contents are created. A co-text is then formed by semantically binding new memorized fragments and fragments previously memorized. This phase is not specifically individual even if it structures an individual memory, insofar as this stage can be related to discursive exchanges, interactions with others.
Compositio is divided into four stages of activity evoking stages of document creation. The stage of Inventio is close to that of Collatio insofar as it is question of creating semantic links between various memorized elements, on the "res" (conceptual objects, idea) level not on the word level. An outline is formed, i.e. a set of ideas hierarchically arranged, an argumentative structure for example.

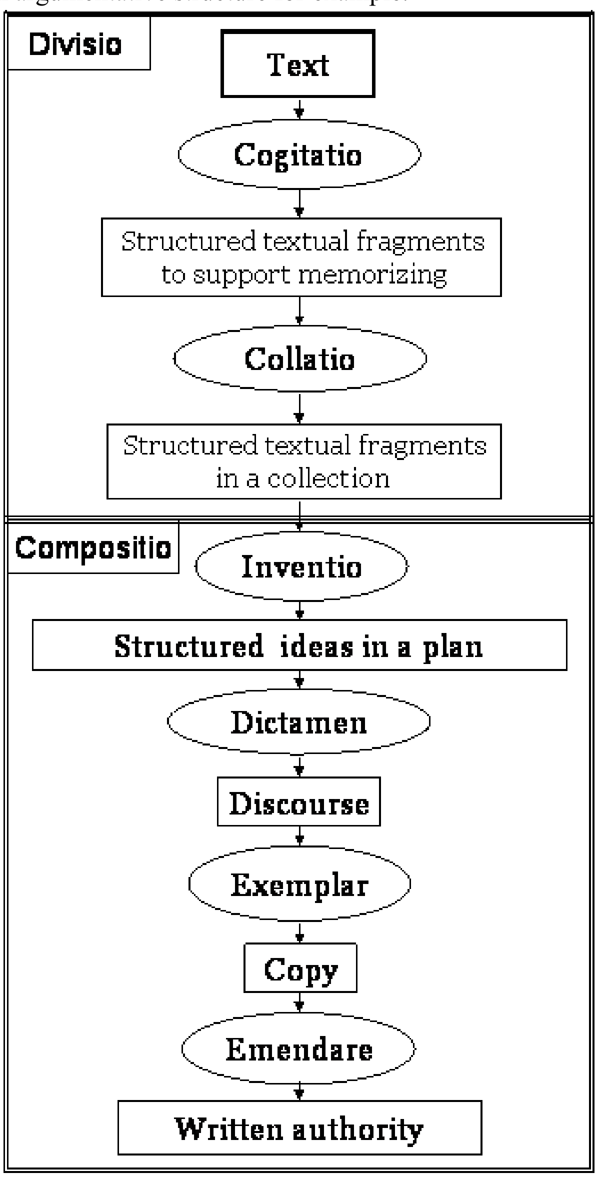


Fig. 2 – Discourse production model

The following phase will be the formatting in word of this conceptual outline. It is a traditional phase of drafting, called "Dictamen". We see with this stage the physical discourse creation, classically done on an adjustable support (a draft), where the style, the choice of the terms, therefore the textual shape of the discourse only can be modified.
The Exemplar phase consists in transforming the draft support of the discourse in a perennial support. The discourse remains strictly identical to the one found in output of the process of Dictamen.
The last phase but not the least in this succession of process is the Emendare where the final copy of the discourse is diffused and then openly commented by the addition of public comments, "notae" or arguments of an author to the original text. This phase thus makes the text become a reference text, a written document being an authority on the field.
This model represents a method of discourse production strongly supported by memory. In a computer supported collaborative work (CSCW) context, discursive creation must be supported by an adequate tool enabling storing, creating and sharing information. In order to design this tool, we first wish to model this mediatized activity of discourse production to represent the functions required for implementing in a tool.

5 Model of mediatized discourse production

We are interested here in collaborative interpreting numerical document, sense making by several participants. We will not take into account non-textual numerical documents.
The transformation of the discourse production model within a mediatized framework, enables us to define the following stages to recommend (Fig. 3). First, the text of the document is segmented to be stored in a memory as memorizable fragments.
These segments are then indexed to avoid the loss of the document structure as consistent unity. It is important to chronologically index the segments to mark the hierarchy of the various paragraphs in a text document, various words in a paragraph... This type of indexing concerns all metadata which might be automatically associated with element stored (localization, author, date...). Indexation must also be used to bind new fragments laid into the system to the conceptual set already present in the tool. We will then obtain a set of textual segments semantically bound to other textual segments. It is a process of co-textual structure creation organized by socio-cognitive as well as semantic links.
The structuring phase represents a hierarchizing process, organizing ideas according to a chronological outline. A detailed outline is defined, containing all ideas necessary to the formatting phase, the change of concepts, to words, to discourse. It is the phase where the "res" (concepts) contained in indexed textual fragments are re-used and reorganized in a new document.
The writing phase is the one where the outline is formatted in text giving a discourse as a result. This discourse is not the final objective of this activity in this vision of rhetoric, since it is then published to become an amendable object, a

writing improved by reader's feedback, themselves becoming authors in the community.
This phase when the published discourse is assessed by other members of the community is extremely important as it is allowing the validation of the Exemplar, its improvement even, and constituting a written authority, a reference discourse in the community.
Within a collective interpretation purpose, annotating a document thus consists, according to us, in following a process of formatting organized ideas in a discourse. Indeed, following the reading of a document, it comes to engage a process which enables to add an idea or an opinion structured in textual form.

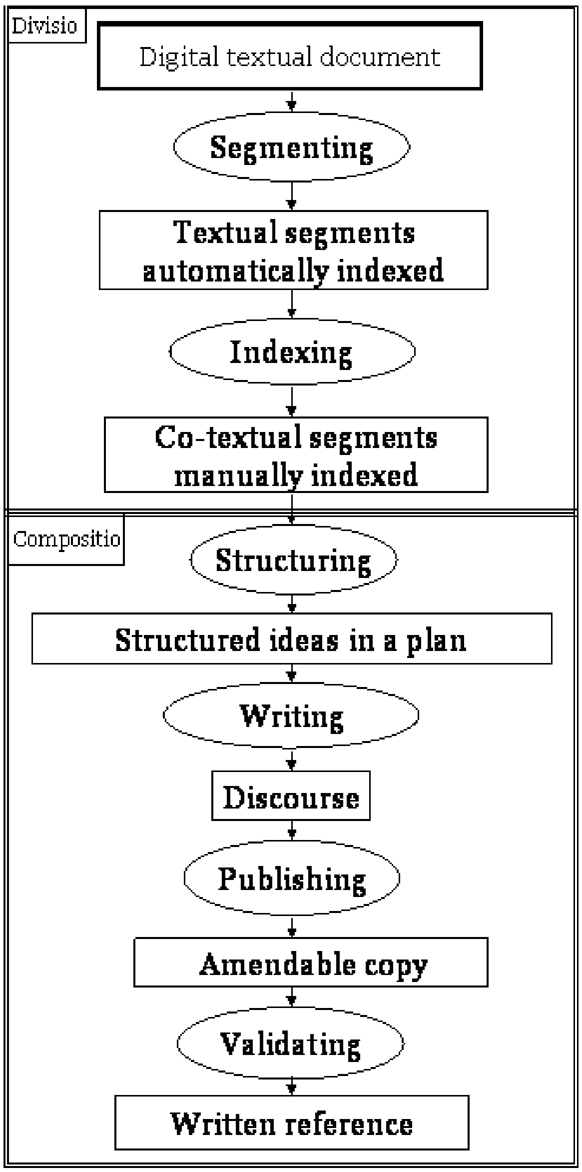


Fig. 3 – Model of mediatized discourse production

For example, in a collaborative work context, one can consider the sharing of a document in order to be commented on. After a visualization phase of the text, a reading, the text read will be segmented to allow the addition of a structured comment, of a discursive annotation. A segment will be emphasized in order to indicate the anchoring of a discursive element linked to this segment. This highlighting could be done by traditional techniques of underlining, circling, colouring segments of unsettled sizes (from a word, or a part of a word, to the paragraph, or set of separated elements). Following the segmentation and the choice of element to be annotated, an indexing phase is required, consisting in connecting segments. The tool should help the user to find semantic links between elements to structure them together and to form an organized set of textual segments according to their meaning. This meaning depends on the user's understanding. Indeed, the annotation consists in an anchor, a geographical relation, in a body, a discourse which creates its meaning amid a "co-text", but also in the whole set of textual segments stored in memory and linked to it, indexed to it by comprehensible key words, structured by and for human user. While writing this annotation, the author should organize his/her discourse to be written. This necessary step is the structuring of "rei", of concepts stored in memory, which will give rise to an outline made up of hierarchically structured arguments. The writing phase will allow constituting the body of the annotation which will be readable by a member of the discussion after publishing and thus spreading this annotation.
Just as a reference text, the annotation can be endorsed thanks to a new link brought to the latter. A reply to a comment allows taking part in the thread of discussion initiated by the first annotation.
This model of mediatized discourse production, resulting from a model of discourse production activity stemming from rhetoric, enables us to describe the requirements of a groupware assisting this type of discursive production by means of annotation.

6 Designing AnT&CoW

6.1 Existing annotations standards

As recommended through the model presented in section 5, the groupware must let users visualize a document, segment it, create various types of associations (indexing, gathering) with the various fragments, write the discourse constituting the annotation body, or publish it. The validation phase (cf. Fig. 3), optimizing collaboration through answers during the discourse, requires a specific association function as a "reply to" function to an annotation. The discursive model allows a continuous look back on the document when reading and writing, so the visualization function is predominant. The visualization is supported by the use of a plug-in into a navigator. Indeed, being an extension of a naturally used navigator and giving access to a lot of Web documents to be read, this plug-in enables visualizing simultaneously the document and the body of the annotation while writing or indexing the annotation. This annotation is captured by a "pop-up" window, then indexed to entitle its recovery after publication and creation of a set of structured documents.

In an annotation activity, several problems arise: the question of anchoring the annotation and the forms of its meta-information in the original document. These problems are tackled in the field of Semantic Web (SW), which goal is to enrich Web resources with structured descriptive information to improve their accessibility, their retrieval and the use of information. We now will describe some existing tools from this field which we can re-use and enrich in our project.

The SW identifies three types of annotations: simple metadata (modification date, author, etc.) ; annotations which we would describe as "computational" insofar as they are addressed to programs enabling them to take a profit from annotated resources [Bremer and Gertz, 2001], [Volz *et al.*, 2003], [Roussey *et al.*, 2001]; and annotations which we would describe as "social" since they are addressed to the reader, to an human user, enabling her/him to be an active Web participant.

Tools developed since the beginning of the 90's allow reviewing texts using comments or explanations, to justify decisions... In general, they consist of various elements permitting to visualize, to create, to store and to search the annotations. Annotations are defined by an anchor, some attributes and a body. They are stored on a dedicated server (annotations server), and can be classified according to their attributes, their public/private/group shared status. The annotations server contains information about the annotation localization (the document on which the annotation was created or its place in the document), its style (font, color...), its contents (text and attributes), and its function (if it is an explanation or a proposition for example). The annotations are generally tree organized. This configuration facilitates navigation in the set of annotations and their management.

These researches lead to the definition of the W3C's Annotea standard [Annotea, 2003] [Kahan *et al.*, 2001], based on a RDF annotation description [Brickley and Guha, 2004]. This standard improves collaboration through shared metadata based on Web annotations, bookmarks, and their combinations. Several annotations servers (ZAnnot, Annotea...) and annotations clients (Annozilla, Amaya...) implement the Annotea standard. The annotations server ZAnnot [Zannot, 2003] stores annotations in a RDF database. Users can interact with Zannot server by Annozilla client [Annozilla, 2004], the Mozilla navigator's plug-in, in order to search for an annotation, to create a new one or to remove another. An annotation is described by a set of metadata (its attributes defined by a RDF diagram) and a body. The RDF notation's advantage is that it is possible to personalize it, for example by adding to the annotation diagram, attributes or a set of values of attributes. This technical solution is thus interesting since it is possible to adapt the model to a need of multidimensional indexing. These dimensions supplement Annotea already existing attributes and are related to a "socio-semantic" use of the annotations in our project.

We are now going to describe and classify these existing annotations tools, and we will clarify our positioning.

6.2 Existing annotations tools

At present, several annotations clients are available, stemming from SW initiatives. Most of them adopt what we would call a "computationally-semantic" approach. This approach has, as main objective, to index Web pages more or less automatically. These tools are used for metadata creation and some are based on ontologies (and their inference engines) to support the computational annotation: OntoMat-annotizer implementing the S-CREAM framework [Handschuh *et al.*, 2002]; Melita [Dingli, 2003] or MnM [Domingue *et al.*, 2002], COHSE [Bechofer *et al.*, 2001]. KMI's Magpie [Dzbor *et al*., 2004] uses semantic annotations to support human interpretation of Web Pages. Computational annotations are geographically dependent on a part of a Web page, but they only enrich the page with concepts for automatic indexing and do not either assist cooperation or interaction between readers of a same page. In fact metadata index a page, and allow the search engines to recall a page in a better way. Finally, human users should be familiar with the ontology used to annotate the documents, and they have few possibilities to update the ontology during system use.

In our view, annotations are not only semantic annotation but discursive annotations and then more than supporting collective interpretation, they create collective interpretation.

Other annotations clients adopt a more social approach, aiming at facilitating human communication, without considering indexing features or annotation recall. In this software, these annotations can only be sorted on rudimentary metadata such as the creation date or the author: Yawas [Denoue, 2000]; CritLink [Ka-Ping, 1998]; XLibris [Price *et al.*, 1998]; etc. These annotations tools regard the annotation as a comment, a way of looking at annotation shared by some proprietary software or some plug-in application software, where the comments are neither indexed nor differentiated from the document [Windows Word comments, 2003]. The annotations are sometimes stored apart on annotations servers [Acrobat pdf, 2004] and organized in a minimalist way. However, these annotations tools do not allow connecting annotations. These tools cannot then represent a structured set of exchanges between users related to a document.

We are considering documents as mediators of discourse as KMI's D3E [Sumner *et al.*, 2000] considers. However, this tool does not allow a rich indexing of annotations, and then it will be difficult to understand the design rationale of the discussion, of a new document or even of a new concept.

Thus, even if these annotations tools support the interaction more easily than the computational annotations tools, they are not sufficient to implement our model.

We finally can classify annotations tools in two families; one concentrates on indexing Web pages, supporting their recall, while the other concentrates on human communication through comments. Following the aim to design a collaborative environment, we can deplore the lack of annotations management or co-operative work possibilities in these two families even if KMI's works are first steps in linking these two views. We thus propose to enrich them thanks to

the SW indexing techniques and to support user in her/his activity of documentary annotation. Supporting this documentary activity will help her/him working in a collaborative way. Moreover, we propose other annotations functions such as multi-anchoring (allowing connecting several fragments of documents) or the answering possibility to an annotation. This alternative approach is also the one proposed by [Bringay *et al.* 2004], in the case of an electronic Patient Record in an hospital. These authors propose a system in which the Patient Record is stored and manipulated with its annotations, as in a paper-based Record where annotations are indissociable from the original document.
In the following part, we then expose the features of an application supporting cooperation around a document, in a "Socio-Semantic Web" approach.

7 AnT&CoW requirements

Following [Zacklad et al, 2003], we define annotation as a type of located metadata, connected to another document. This unit is connected to various parameters such as time, place, participants, its public or private status, its meaning... which means that annotation is an entity made up of several parts such as its anchor (or its anchors) in a document, its attributes, and a body (the text of the annotation). We also consider that the annotation is a mark of the collaboration process which has two principal functions: planning (project management, micro-organization) and the reviewing (argumentation, annotation constituting a document body...)
Metadata suggested by Web standards (for example Annotea described above) to index annotations (name of the author, date, topic, type of annotation, etc.) are thus not sufficient for our project. In fact, with this type of index, we cannot store the organizational context (roles, profile of the participants, etc), the contextual field (specific lexicon, keywords of the field, concepts, etc), nor the type of argumentation (suggestion, opposition).
In order to allow a more subtle classification of these annotations, we thus propose to extend the collaborative annotation indexing not only by domain specific dimensions (topics), but also by a cognitive dimension thanks to an argumentative dimension (preserving the rationale of the decisions and negotiations between human participants) and an organizational dimension, using the participant's role to stress the importance of a decision.

7.1 Semiotic ontologies for multi-dimensional indexing

The three dimensions defined above are described by an ontology. From a SW point of view, ontologies are supposed to represent exhaustively the knowledge of a specific field, structuring concepts in a hierarchy by relations between them. Each concept is well defined by all its properties and the expert must thus entirely specify the relations between the concepts. However, human experts often have conflicting definitions of some concepts for which several definitions are in competition. Concurrently, specific inference mechanisms calculate the coherence and the consistency of these ontologies. Building such ontologies is a time-consuming and expensive task. Plus, on one hand, generic ontologies (EuroWordNet, DOLCE [Gangemi *et al.*, 2003]) are not adapted to domain-specific applications; they do not contain domain-specific concept definitions. On the other hand, domain-specific ontologies are not available or they are very expensive, even if their portability is increased by the use of W3C standards (OWL, RDF). Thus, it is difficult to work out a representation of the semantic contents of Web pages, even using ontologies.
To avoid this drawback, a more socio-semantics approach of the Web proposes the use of less formal ontologies, which main purpose is to help user navigating through Web pages and not to compute automatically the semantic representation of the document content. From this perspective, the concepts should be less-specified; there is no need to identify all the concepts' properties. Standards as Topic Maps (TM) (Standard ISO, [Biezunski *et al.,* 1999]) are defined for these semi-formal ontologies. TM formalism defines a network of topics covering domain-specific knowledge. Topics are defined via simple URL, so all the users share the same definition. The topics are hierarchically organised (related by "isa" relations) and associated by horizontal relations ("partOf", "used") (Fig. 4). No coherence checking mechanism is done.

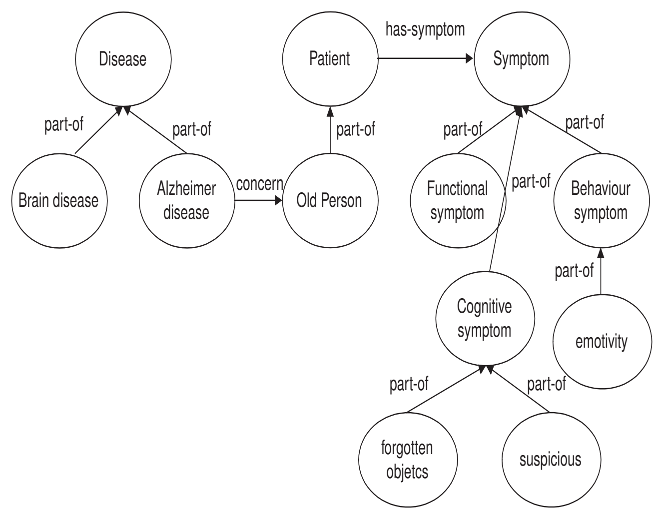


Fig. 4 – Medical domain ontology fragment in Topic Maps

While TM do not require a precise definition of concepts, and are designed to support user browsing Web pages; we adopted this formalism for representing the various dimensions of our ontology.
In our system, the organizational and argumentative dimensions are built manually. The first one is based on a social analysis of the network, and the second one is based both on a cognitive and a pragmatic analysis of interactions in the network. The domain-specific dimension requires a combination of Natural Language Processing (NLP) techniques and manual choice of terms and concepts. This ontology is

stored on an ontology server which allows an easy recall of the concepts. We focus now on the NLP techniques.

7.2 NLP tools and methods for domain contextual ontology building

Due to the low availability of domain-specific ontologies and to the fact that generic ontologies are of little use for domain specific applications, many projects aimed to use NLP techniques to extract semi-automatically terms (concept instances) [Jacquemin and Bourigault, 2003] to create term clusters (concepts) [Cimiano and al, 2004] as well as to extract relations between terms [Buitelaar and al, 2004]. The expert should name the clusters as concepts and eventually should define relations between concepts.

In our system, NLP techniques are used for two main purposes: building and maintaining the domain-specific ontology from corpora, but also for browsing and indexing annotations.

The annotation indexing can be done automatically by the tool (date, author, answered annotations codification, automatic chronological thread of discussion) or manually by the user. The annotation manual indexing phase by the user regarding to three dimensions (choice of a value representing the annotation content according to each dimension) can be tedious and we thus wish to support it thanks to NLP tools.

The *first* task, concerning ontology building is done off-line, by extracting terms from a selected corpus and by proposing a simple topic hierarchy (a term is equivalent to a topic).

Tests were carried out in the medical field (Alzeihmer's disease and memory troubles), for an Electronic Patient File (EPF) project. An EPF is a patient file created and maintained by a medical group to follow a patient and improve its cares. To be easily followed by distant members, this file is shared by means of Web interface.

It was not possible to use medical ontologies [MeSH, 2004], [UMLS, 2004] insofar as they are too generic or cover a swarms of domains (MENELAS, [Zweigenbaum and al, 1994]) far away from the application's use in the project.

For building a semi-formal ontology (structured in topics) from corpora, we identify candidate terms by using a term extractor. Among the term extractor available, we tested LIKES [Rousselot and al, 1996] which is a simple repeated segment extractor identifying sequences of words (collocation, repeated segments) occurring in the corpus. The repeated segments are potential candidate-terms, and they are organized in a tree, gathered according to their head and displayed according to their frequency of their occurrences. The candidate-terms are used to select the topics of our ontology. The outputs are filtered in order to eliminate the incorrect candidate-terms (terms finishing by a preposition, a conjunction). The majority of the candidate-terms correspond to a Head + Modifier pattern.

We carried out tests on a small medical corpus (14000 words) and obtained an approximately 100 topics ontology. The sizeable drawback of this tool remains the significant number of candidate-terms, which requires a stage of manual cleaning of the resulting hierarchy.

We developed a tool (GenTMInd), identifying hierarchical relations between terms via heuristic rules and structuring them in Topic Maps format. For example, a term matching a pattern Head + Modifier is a subconcept of the Head concept. For the moment, candidate topics should be identified among simple noun phrases (a noun phrase followed by only one prepositional phrase).

These assumptions and heuristic rules are not sufficient to identify all the hierarchical relations or all the relevant candidate-topics. User thus can manually update the ontology by adding relevant topic-keys indexing her/his annotation and by organizing them in the existing TM.

However, after a relevant corpus is gathered, we will extend the search for candidates to a set of domain-specific verbs. We will explore the context of each topic-candidate in order to identify more relations between the topics. If it is possible to find out candidate-topics frequently co-occurring (related by a syntactic relation as predicate-argument or head-modifier) in the text, it would mean that horizontal relations must be added between two candidate-topics. For example, in the context of the disease of Alzheimer, the corpus of test contains "old person", which means that relation "concern" between two topics could be added (Fig.4).

The *second* task is to help the user indexing his annotation regarding to three dimensions (other indexes like author name, date, title, are automatic), by proposing him/her a semi-automatic indexation of his/her annotation (indexes as name of the author, date or title are automatic). NLP tools scan the annotation submitted by the user, identify some relevant terms candidates and match these terms to the concepts of the ontology for each dimension. The matching process uses three resources: the indexation context, the annotation co-text and the ontology. Ontology is a vertical representation of the concepts, i.e. with paradigmatic links, while the indexation context and the annotation co-text are syntagmatic links database. The indexation context is a database storing textual contexts frequently co-occurring with the ontology topics. The annotation co-text is a database storing textual bodies of annotations and textual fragments where these are anchored (fragment of documents). Indeed, to process this mapping, we have several relations databases allowing combining paradigmatic and syntagmatic relations to improve lexical access, data recall. The mapping algorithm checks the contexts of the ontology topics and the contexts of term candidates. If similar context are found [Harris, 1988], the topic is proposed to index the candidates. The annotation tool will then propose domain specific keywords or "keysyntagms" as well as argumentative types to the user. The user will then decide if the index suggested is relevant and if s/he wishes to preserve it as metadata of her/his annotation.

By creating his/her annotation, the user decides if the annotation is anchored to one or more parts of the document or of several documents. Thus, we consider a complex annotation indexing and multi-anchoring, defining more precisely the co-text of the annotation. Once the validation is done by the user, the annotation is stored with its metadata on the annotations server.

The next step in this tool implementation is to adapt a more effective term extractor in our system, as FASTR is [Jacquemin and Tzoukermann, 1999], in order to identify the candidate-terms in the annotations bodies and to extract a concept hierarchy by the clustering techniques [Cimiano and al, 2004].
We will now present our distributed architecture and some visualization features of our annotations tool, following W3C standards and integrating NLP tools.

8 AnT&CoW: Architecture and visualization

Following the Annotea W3C standard, our client/server annotation system implements a distributed architecture (Fig.5):
The client's goal is to annotate documents (for the moment limited to annotate text or HTML pages due to format constraints), which are accessible by a Web navigator. Mainly for this reason, we chose Annozilla, a Mozilla navigator plug-in which is an Annotea client following our aim. Using XPointer, DOM standards and many functions of the Mozilla infrastructure (XPConnect, XPCom components), Annozilla offers possibilities of creating, updating and deleting annotations on a document or a part of document and gives possibilities in storing them on a local server (individual use) or a distant one (shared use).

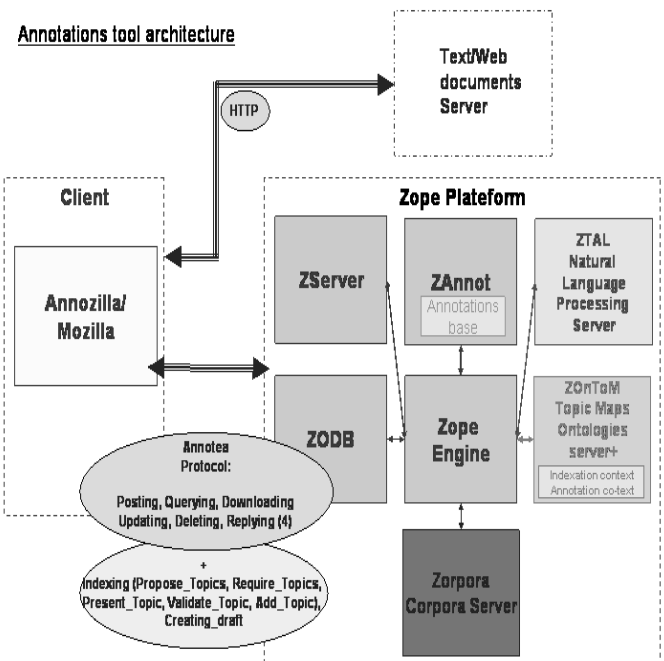


Fig.5 – AnT&CoW annotation tool Architecture

We chose a server respecting the Annotea standard, ZAnnot, developed on the Zope platform [Latteier and al, 2003] which has a Web server and several other components managing contents servers or databases. ZAnnot derives benefits from the Zope platform and manages at the same time queries sent by the Annozilla client and the reply function to an annotation.
On this platform, we encapsulate the ZTAL server for the natural language processing whose functions are defined above, the ZOnToM ontologies server represented out of TM also containing the indexation context and annotation co-text. The Zorpora server is a corpora server which contains not only the basic documents text used to constitute the domain-specific ontology dimension, but also the documents created by the project participants and eventually the authority documents shared in the project.
Since it is necessary to adapt the annotations client Annozilla for our annotation's purpose such as previously defined, we implemented the reply function from annotation to another and the indexing mechanism. To classify annotations, we extended the Annotea annotation diagram by adding metadata corresponding to our three dimensions which will be saved at the RDF format, as the other metadata and annotation bodies. For coherence reasons, our multi-dimensional Topic Map ontology is currently stored in a XTM (XML) format and is not modifiable by the user.
We provide an interface for the user allowing her/him to manage the topics of the different dimensions and to navigate through stored annotations. Navigation consists of a reading of the annotations arranged in one or more visible windows at the same time. Thus the user can, if s/he wished, display in the same document a set of annotation indexed by the same topic(s), annotation textual body and other fragments to which it is connected. (Fig.6) She/He has also possibility of recording elements gathered in only one new working document, a draft or a discussion paper shareable by the project.

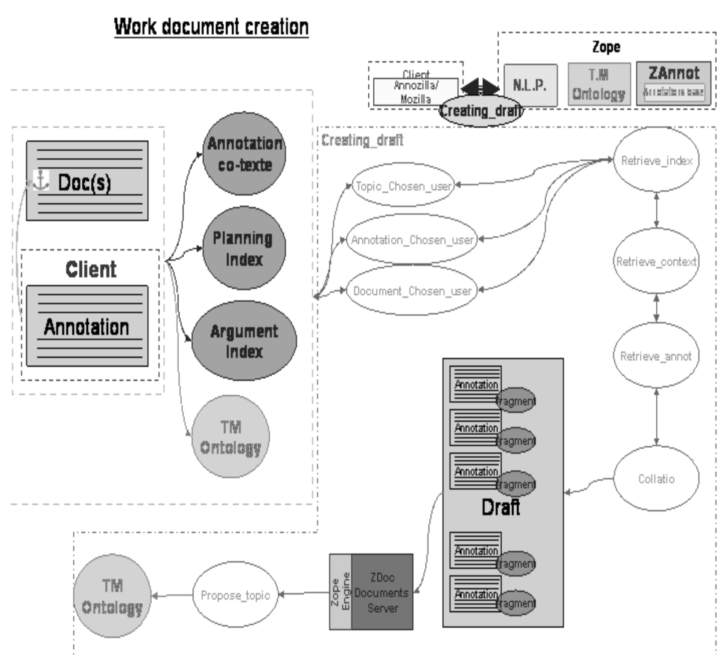


Fig.6: Work Document Creation

When a member of the project group is opening a document, s/he may open in the left side of the Web navigator main window, the Annozilla plug-in, which allows her/him to annotate as well as to retrieve and read organized annotations by means of their attributes defined above. If the author decides to create a new annotation, this annotation appears in a new window containing its body and the indexation fields in a pull-down menu as in this example with an electronic patient file (fig.7).
The next step in the tool development consists in integrating in our architecture the indexation elements, i.e. dimensions of the ontology and NLP tools; ZOnToM must be connected to the annotation server Zannot so that the TM ontology representing dimensions and the contexts/co-texts can be used for a semi-automatic indexing. The ontologies server installation in an on line process will also allow the ontology update, by way of user or of NLP tools.

rized concepts which are reorganized as an editable structure aiming at communicating about a document.
To represent this discursive annotation activity and so collective interpretation of documents, we chose a classical rhetorical model of discourse production. Adapting this model to electronic document customs allowed us to design a groupware supporting sharable annotations for document based sense making within a group: AnT&CoW. Deriving from existing annotation's standards and tools, we drew some requirements for AnT&CoW, meeting our theoretical model.
AnT&CoW is a client/server application based on a multidimensional ontology. Our tool's features are supported by Natural Language Processing tools and techniques.
A first version of this tool is in development being in keeping with an iterative design approach. This tool will allow us evaluating our hypothesis not only on discourse production model, but also on annotations status and aims.

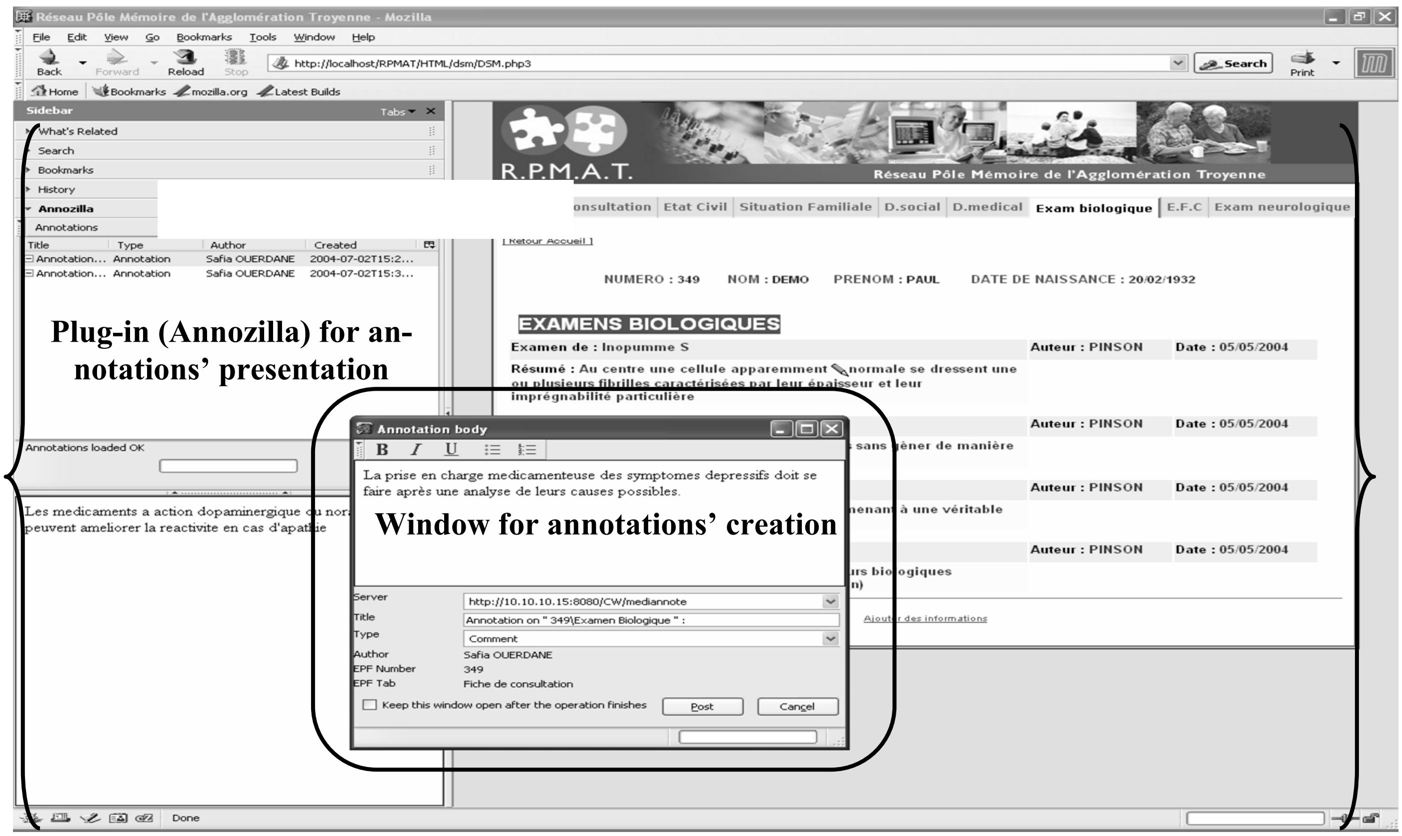


Fig.7 – AnT&CoW Interface for Electronic Patient File

9 Conclusion and prospects

The increasing number of electronic documents forces today's reader to adapt her/his practices. Traditional collective interpretation of texts by use of annotations then becomes an activity to be mediatized. Annotating is an activity mixing writing and reading and allows annotation's author to communicate with members of interest. We propose to define annotation as a kind of discourse, a structured set of memo-

Acknoledgments

This research carries a CNRS (National Center for Scientific Research)/STIC (Communication and Information Science and Technology) department funding as part of TCAN (knowledge processing, learning and new information and communication technologies) pluridisciplinary project (Mediannote project).

References

[Acrobat pdf, 2004] *Acrobat PDF*, http://www.adobe.com/support/techdocs/ac76.htm, 2004.

[Annotea, 2003] Annotea http://www.w3.org/2001/Annotea/, 2003.

[Annozilla, 2004] Annozilla, http://annozilla.mozdev.org/, 2004.

[Baker, 2000] Baker M., The roles of models in Artificial Intelligence and Education Research: a prospective view. International Journal of Artificial Intelligence in *Education Research*. Vol 11(2), p. 122-143, 2000.

[Barré de Miniac, 2000] Barré de Miniac C.*, Le rapport à l'écriture : Aspects théoriques et didactiques* coll. Savoirs mieux Ed. Septentrion Presses Universitaires, Ch. Barré de Miniac, 2000.

[Biezunski *et al.,* 1999] Biezunski, M., Bryan, M., et Newcomb, S. R., *« Topic Maps », spécification ISO/IEC 13250*, 3 Décembre 1999.

[Bremer and Gertz, 2001] Bremer J.M., and Gertz M., Web Data Indexing through External Semantic-carrying Annotations. In *11th IEEE Int'l Workshop on Research Issues on Data Engineering: Document management for data intensive business and scientific applications (RIDE-DM'2001), IEEE Computer Society*, pp. 69-76, 2001.

[Brickley and Guha, 2004] Brickley D., and Guha R.V., *Resource Description Language* - http://www.w3.org/TR/rdf-schema/, February 2004.

[Bringay *et al.* 2004] Bringay, S., Barry, C., Charlet, J., Annotations: A new type of document in the Electronic Health Record, in proceedings of *DOCAM 2004, Second International Conference on Document Research and Development in Sciences, Arts and Business*, University of California, Berkeley, October 22-24, 2004]

[Buitelaar *et al.*, 2004] Buitelaar P., Olejnik D., Hutanu M., Schutz A., Declerck T., and Sintek, M., Towards Ontology Engineering Based on Linguistic Analysis, in *Proceedings of LREC'2004*, Lisbon, ISBN 2-9517408-1-6, pp.7-11, may 2004.

[Carruthers, 1990] Carruthers M., *The Book of Memory: A Study of Memory in Medieval Culture*. New York: Cambridge University Press, 1990.

[Cimiano *et al.*, 2004] Cimiano, P. Hotho, A., and Staab S., Clustering Concept Hierarchies from Text, in *Proceedings of LREC'2004*, Lisbon, ISBN 2-9517408-1-6, pp. 1721-1724, may 2004.

[Denoue, 2000] Denoue, L., et Vignollet, L., An annotation tool for Web browsers and its applications to information retrieval, in *proceedings of RIAO 2000*, 2000.

[Dingli, 2003] Dingli A., Next Generation Annotation Interfaces for Adaptive Information Extraction. In *6th Annual Computer Linguists UK Colloquium (CLUK 0*3), January, 2003, Edinburgh, UK, 2003.

[Domingue *et al*., 2002] Domingue J.B., Lanzoni M., Motta E., Vargas-Vera M., et Ciravegna F., Mnm: Ontology driven semi-automatic or automatic support for semantic markup. In *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02),* October 2002.

[Dzbor *et al*., 2004] Dzbor, M. - Motta, E. - Domingue, J. B., Opening Up Magpie via Semantic Services. In *3rd ISWC*, November 2004, Japan

[Fayol, 1997] Fayol M., *Des idées au texte: psychologie cognitive de la production verbale, orale et écrite*. Paris: PUF, 1997.

[Gangemi *et al*., 2003] Gangemi, A., Guarino, N., Masolo, C., et Oltramari, A. Sweetening WordNet with DOLCE, *AI Magazine* 24(3): Fall 2003, 13-24, 2003.

[Garlatti and Iksal, 2000] Garlatti S., Iksal S., Méthodologie de conception de documents électroniques adaptifs sur le Web. in GAIO, M., TRUPINCIDE, E., *Document Électronique Dynamique, Actes du troisième colloque international sur le document électronique : CIDE'2000*, 2000.

[Handschuh *et al*., 2002] Handschuh, S., Staab S., et Ciravegna, F., S-cream - semi-automatic creation of metadata. In *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, October 2002.

[Harris, 1988] Harris Z., *Language and Information Columbia University Press*, New York, 1988.

[Hayes and Flower, 1980] Hayes J. R. & Flower, L. S., Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing*. Hillsdale, NJ: Lawrence Erlbaum, 1980.

[Herring, 1999] Herring, S.C., Interactional Coherence in CMC. *Journal of Computer-Mediated Communication* 4(4) : www.ascusc.org/jcmc/vol4/issue4/, 1999.

[Jacquemin and Bourigault, 2003], Jacquemin C. and Bourigault D., Term Extraction and Automatic Indexing, in Mitkov R. (ed), *The Oxford Handbook of Computational Linguistics*, Oxford University Press, pp. 599-615, 2003.

[Jacquemin and Tzoukermann, 1999], Jacquemin, C., and Tzoukermann, E., NLP for Term Variant Extraction: A Synergy of Morphology, Lexicon and Syntax. In T. Strzalkowski, editor, *Natural Language Information Retrieval*, pages 25-74, Kluwer, Boston, MA, 1999.

[Kahan *et al*., 2001] Kahan J., Koivunen M.-R., Prud'Hommeaux E., and Swick R.R., Annotea : an open RDF Infrastructure for Shared Web Annotations, *Proceedings of WWW10*, Hong-Kong, pp. 623-632, May 1-5 2001.

[Ka-Ping, 1998] Ka-Ping Y., *CritLink : Better hyperlinks for the WWW*. http://crit.org/ping/ht98.html, 1998.

[Kintsch, 1988] Kintsch W., The role of knowledge in discourse comprehension: A Construction-Integration model. *Psychological Review*, 95, 163-182, 1988.

[Latteier *et al.*, 2003] Latteier A., Pelletier M., McDonough C., and Sabaini P., *The Zope Book*, Edition 2.6. http://zope.org/Documentation/Books/ZopeBook/2_6Edition/ZopeBook-2_6.pdf, 2003.

[Marcoccia, 2004] Marcoccia M, On-line polylogues: conversation structure and participation framework in *internet newsgroups, Journals of Pragmatics*, 36 (2004) 115-145, 2004.

[MeSH, 2004] *MeSH, Medical Subject Headings*, http://disc.vjf.inserm.fr:2010/basismesh/meshv04.html, 2004

[Pédauque, 2003] Pédauque, R.T., *Document : forme, signe et médium, les re-formulations du numérique, working paper,* version 3- 8 juillet 2003, http://rtp-doc.enssib.fr, 2003.

[Piolat *et al.*, 1989], Piolat A., Farioli F., and Roussey J.-Y., La production de texte assistée par ordinateur. In G. Monteil, & M. Fayol (Eds.), *La psychologie scientifique et ses applications* (pp. 177-184). Grenoble : Presses Universitaires de Grenoble, 1989.

[Price *et al*., 1998] Price, M., Schilit, B., et Golovchinsky, G., XLibris: The active reading machine. In *proceedings of CHI'98 Human factors in computing systems,* Los Angeles, California, USA, vol.2 of Demonstrations: Dynamic Documents, pages 22-23, 1998.

[Rousselot *et al.*, 1996] Rousselot, F., Frath, P., and Oueslati, R., Extracting concepts and relations from Corpora. In *Proceedings of the Workshop on Corpus-oriented Semantic Analysis*, European Conference on Artificial Intelligence, ECAI 96, Budapest, 12 August 1996.

[Roussey *et al.*, 2001] Roussey C., Calabretto S., et Pinon J.-M., SyDoM: A Multilingual Information Retrieval System for Digital in *proc. International Conference ICCC/IFIP On Electronic Publishing (ELPUB'2001)*, Canterbury (UK), 5-7 july 2001, p. 150-164, 2001.

[Sumner *et al.*, 2000] Sumner T., Buckingham Shum S., Wright M., Bonnardel N. , Piolat A. & Chevalier A., Redesigning the peer review process : A developmental theory-in-action. In R. Dieng, A. Giboin, G. De Michelis & L. Karsenty (Eds.), *Designing cooperative systems: The use of theories and models* (pp. 19-34). Amsterdam : I.O.S. Press, 2000.

[Tchounikine 2002] Tchounikine P., Pour une ingénierie des Environnements Informatiques pour l'Apprentissage Humain. *Revue I3 Information-Interaction-Intelligence*. Vol. 2, n°1, Cepadues Editions. 2002.

[UMLS, 2004] UMLS, Knowledge Source Documentation, 2004. http://www.nlm.nih.gov/research/umls/umlsdoc.html, 2004.

[Volz *et al.*, 2003] Volz R., Oberle D., Motik B., et Staab S., KAON SERVER - A Semantic Web Management System? In: *Proceedings of the 12th World Wide Web, Alternate Tracks - Practice and Experience,* Hungary, Budapest, 2003.

[Weick, 1979] Weick K.E, *The Social Psychology of organizing*, New York, Random House, 1979.

[Windows Word comments, 2003] *Windows Word*, http://office.microsoft.com/fr-fr/assistance/HA010714941036.aspx, 2003.

[Zacklad *et al.*, 2003] Zacklad M., Lewkowicz M., Boujut J-F., Darses F., and Détienne F., Formes et gestion des annotations numériques collectives en ingénierie collaborative, *actes des journées Ingénierie des Connaissances*, Laval, 2003.

[ZAnnot, 2003], ZAnnot, http://www.zope.org/Members/Crouton/ZAnnot/, 2003.

[Zweigenbaum *et al.*, 1994] Zweigenbaum, P; et Consortium MENELAS, MENELAS : an access system for medical records using natural language. In *Computer methods and programs in Biomedicine*, 45:117-120, 1994.

Knowledge capitalization in system of equipment diagnosis and repair help

Brigitte Chebel-Morello, Ivana Rasovska, Noureddine Zerhouni
Laboratoire d'Automatique de Besançon
24 rue Alain Savary, 250 00 Besançon, France
[rasovska,bmorello,zerhouni]@ens2m.fr

Abstract

The methodology of knowledge capitalization allows to create a decision help system for diagnosis and repair of industrial plants. The objective is to create a corporate memory of an enterprise specialized in maintenance services. This memory is acquired by the knowledge capitalization process. The knowledge is "alive" and evolves in the knowledge management. The diagnostic and repair help system for maintenance operators is developed within the framework of distributed e-maintenance platform in the European project Proteus. This tool is based on the case-based reasoning technology. The principal phases of this approach are discussed and our contributions are shown.

1 Introduction

Nowadays methodologies of knowledge capitalization in enterprise has become more and more important in the industry. To capitalize knowledge means "to reuse, in a relevant way, the knowledge of a given domain previously stored and modeled, in order to perform new tasks". This knowledge is stored in a database called "corporate memory", i.e. "a structured set of knowledge related to the firm experience in a given domain". The expert knowledge should be identified, formalized and modeled in order to be retrieved, used and updated by enterprise employees [Grundstein, 1994]. The knowledge is "alive" and evolves in the knowledge management. Techniques of knowledge capitalization are based on information acquisition. This information is issued from technical documents and interviews done with the experts.

The objective of this study is to build an intelligent application for diagnosis and repair in the context of maintenance services. It is targeted to maintenance operators for their daily tasks. This decision tool is developed within the framework of the distributed e-maintenance platform shown on the fig.1. The design of a generic software architecture for web-based e-maintenance centers is the aim of the European project Proteus consisting of 16 partners from France, Germany and Belgium. The platform brings a major asset to maintenance interventions and maintenance services in general by enabling expertise via Internet directly to the user site. The web portal in the core allows different users – maintenance actors – to access to decision support for their activities. The platform integrates a number of systems and knowledge bases like CMMS (Computerized Maintenance Management System), SCADA (Supervisory Control and Data Acquisition), maintenance data bases and e-documentation and finally applications and tools for decision support.

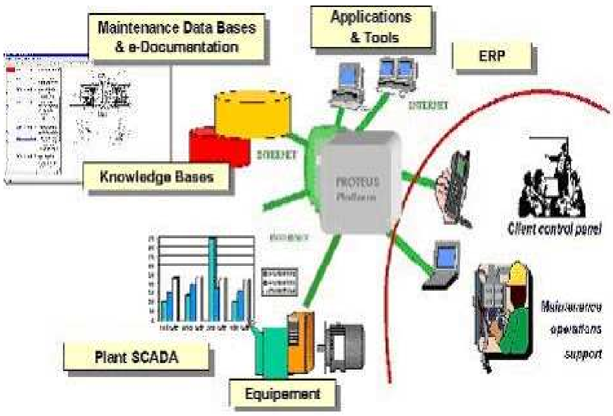


Figure 1. Architecture of Proteus platform.

The paper contains the introduction to the knowledge capitalization cycle in the section 2. Moreover, the techniques and methods of knowledge representation developed in artificial intelligence are introduced in the section 3. We choose the ontology techniques to create the maintenance domain model and we associate to this knowledge model a reasoning mechanism in order to ease its use. The case-based reasoning was chosen as the problem solving mechanism for our interactive help system and is briefly introduced. The ontology of maintenance domain is presented in the section 4. The application on the pallet transfer system SORMEL is shown with its case representation and the developed case base to test the system in the section 5. The principles and contributions in each phase of the CBR technology coupled with the ontology techniques are shown and the first results are discussed in the conclusion.

2 Knowledge capitalization in enterprise

There are several methodologies of knowledge capitalization combining the technical, human and organizational aspects which can be presented according to two approaches:

- corporate memory itself like methods REX[1], MEREX[2], CYGMA[3] that consider 6 categories of industrial knowledge with jobs' reference. Workshop FX results from work of social sciences which aim at using the actor's experience of the industrial process to create enterprise knowledge;
- models resulting from the knowledge engineering which methods are KADS[4], CommonKADS[5], KOD[6], MKSM[7] associated with its extension MASK[8], Componential framework etc., which present various conceptual models interacting to each other.

We are not interested in the industrial process in this paper so we focus on the methods of knowledge engineering. The principles are briefly pointed out in the following section.

2.1 Cycle of knowledge capitalization

The fig. 2 proposed by Michel Grundstein illustrates this process. We follow the process cycle of knowledge capitalization and we associate to each phase of this process the methods we used in the field of e-maintenance. The cycle highlights 4 phases.

Detection of information: the practices of maintenance experts were observed in their activity of industrial plant diagnosis and repair. This observation was combined with the analysis of maintenance process presented in [Rasovska *et al.*, 2004].

Storing of knowledge: the design of such a system necessitates the knowledge modeling which is declined by a representation model associated to a problem solving model. It was thus necessary to create a knowledge model of the domain shown in the next section.

Capitalization of knowledge: the capitalization is done by the platform of e-maintenance developed by project partners specialized in information technologies. The platform is used as a support for information diffusion. Web services were developed in order to relate knowledge acquisition with the access to expertise. The description of these interfaces is not studied in this paper, because this would bring nothing to the methodology of knowledge capitalization.

Update of knowledge: this step is ensured by the case-based reasoning method described later on in the paper. The access to the diagnosis service is opened to all maintenance actors, on the other hand modifications of the case base for its up to date handing-over is authorized only to the designated experts.

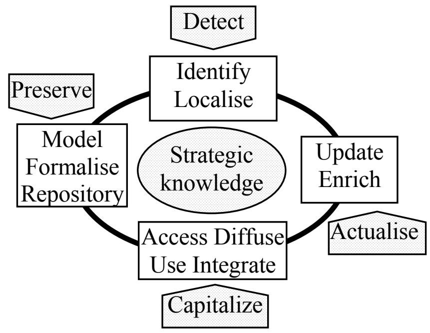


Figure 2. Cycle of knowledge capitalization [Grundstein, 1994].

The techniques and methods developed in artificial intelligence are necessary to formalize knowledge and to handle it, so the summary of them is briefly reviewed in the next section.

3 Techniques and methods of knowledge representation in AI

The most frequently used techniques of knowledge representation in knowledge capitalization are the following:

- The knowledge structuring by hypertext links between the information pages or by case bases which represents the cases with database descriptors.
- The knowledge indexation when the volume of existing documentation is huge by using linguistico-statistical techniques.
- The expert integration in the capitalization process when one is not able to model his knowledge.
- The detailed knowledge modeling used in this paper and introduced later on in the paper.

[1] REX: method of experience feedback developed for design of nuclear engines in CYGMA Cycle
[2] MEREX: developed for Renault to improve engineering process for cars
[3] CYGMA: Cycle de vie et Gestion des Métiers et Applications (Lifecycle and management of Jobs and Applications)
[4] KADS: Knowledge Acquisition and Design Structuring
[5] CommonKADS: An Advanced and Comprehensive Methodology for Integrated KBS Development
[6] KOD: Knowledge Oriented Design (Knowledge On Demand)
[7] MKSM: Method for Knowledge Systems Management
[8] MASK: Méthode d'Analyse et de Structuration des Connaissances (Method of knowledge analysis and structuration)

In our case the method of detailed knowledge modeling is proposed. This method allows to formalize expert knowledge in diagnosis and repair of industrial plants, intended for maintenance service. To handle this formalized knowledge is the task of knowledge management.

3.1 Different approaches

Methods of corporate memory construction are dissociated according to two complementary approaches of knowledge modeling:

- Bottom up approach consists in collecting verbal data from experts and gathering them to form a knowledge model.
- Downward approach "or approach directed by models" focuses on an expertise model definition in order to filter acquired knowledge and to guide effectively the acquisition process.

These two approaches were compared in [Duribreux-Cocquebert and Houriez, 2000]. We retain that the downward approach has the major advantage to separate knowledge from the field of its use. This allows to define generic components in order to be reusable in other domains. The bottom up approach presents a structured expertise model described by terminology suitable for the concrete problem which entails a lack of abstraction. Therefore our study was directed more towards the downward approach.

One of the methods of reference in this approach is KADS (Knowledge Acquisition and Design Structuring) and Common KADS. Its structure in layers uses generic problem solving methods describing the reasoning on the good abstraction level. MASK is another method based on cognitive analysis principles and on the experience feedback. It is the extension of MKMS (Management Knowledge System Modeling) proposed by CEA (Commissariat à l'Energie Atomique). It allows to represent the expertise from three complementary viewpoints: context (jobs concerned approach), know how (cognitive engineering approach) and informatics (computer engineering) [Van Craeynest et al., 2000]. These basic methods cannot take into account every particular problem type. Thus they are adapted and used in different studies.

Actually, the re-using of all methods opens perspectives to create the expertise models based on the integration of the two approaches. Duribreux-Cocquebert and Houriez [2000] propose a mixed approach combining the KADS method and the KOD method (Knowledge Oriented Design) which makes models from the core. As far as the downward methods are concerned, Talbi, et al. [2000] builds for a given problem class a methodological solution which is based on the tool OPENKADS. However, the re-using of known methods can be partial and is based on the domain elements specification or on the reasoning used in the method. As Reynaud, et al. [2000] declares: "… Applied to the domain elements, the re-using technique is based on the definition of ontology describing explicitly the domain elements. Applied to the reasoning elements, it is based on abstract descriptions of the problem solving methods.".

3.2 Knowledge representation and ontology

[Reynaud *et al*., 2000] proposes to build a model for generic problem solving system from the domain ontology specified formally. The definition of ontology is given in [Charlet *et al*., 1996]: "to make an ontology means to decide about existing individuals, concepts and properties they are characterized by and relations they are connected to each other".

Ontology refers to a set of concrete terms used to describe a certain knowledge domain and to build a representation of it. Ontology provides a knowledge representation vocabulary, often specialized to a domain or a subject matter. Knowledge representation in artificial intelligence is structural, the ontology defines concepts and instances of these concepts and might also define the goal of a specific instance. Ontological analysis clarifies the structure of knowledge and facilitates the creation of vocabulary. The knowledge can be described in a way that is language and implementation independent, and facilitate efficient knowledge sharing.

Schreiber [1994] distinguishes the domain ontology and the model ontology. The first one contains specific terms and expressions of application domain. The model ontology describes the structure imposed on domain knowledge by task and problem solving method.

A lot of ontological studies are created by lexical representation of special domain. Texts like technical documentation are resources of knowledge acquisition and are studied by scientists in the linguistic and terminology. These methods are placed in the bottom up approach.

In this context our knowledge model is proposed as the integration of expertise domain ontology. The case-based reasoning is associated to like the problem solving method.

3.3 Case-Based Reasoning

Currently developed methods focus on the expertise and are opened towards a user. They are conceived on the basis of cognitive human reasoning as Perron [2000] reminds, that is the goal of case-based reasoning methods. These methods use similar previous cases to solve a new problem. This is one of the mechanisms of analogical reasoning where the episodic memory organisation is used as shows Kolodner [1991].

During the last years the case-based reasoning has begun to play a significant role in the knowledge management. This approach is close to the human reasoning because it uses similar cases to make a decision. Moreover, there is a dynamic aspect of knowledge capitalization included in the permanent knowledge evolution. This represents a difficulty for many methods except the case-based reasoning technology that solve this problem in its evolution cycle. The case-based reasoning implements a knowledge base made up of cases containing the experience of already solved problems where one can seek cases similar to the problem to be solved. This case base or knowledge base will make part of the emergent corporate memory based on the ontology techniques introduced above.

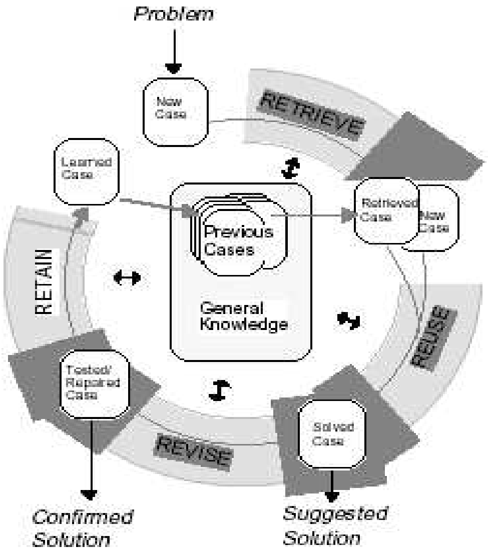


Figure 3. The CBR cycle [Aamodt and Plaza, 1994].

Aamodt and Plaza [1994] present the case-based reasoning as a problem-solving paradigm and propose four principal phases as shows the fig. 3. The retrieve phase find the most similar case or cases according to the similarity between the request and previously experienced cases stored in the case base. The reuse phase uses solutions of the similar cases in order to solve the new problem. The differences between the reminded case and the new case are taken into account and the old solution is adapted to the new situation. The phase "revise" of the proposed solution evaluates the proposed solution in the real world. And the retain phase store a new case in the case base. An other very important task in the CBR cycle is the case representation and acquisition which needs knowledge representation techniques.

The cases are stored and organized according to well defined criteria making it possible to find them effectively [Fuchs, 1997]. Moreover, the acquisition of a new case makes it possible to make evolve the knowledge. The CBR feasibility for the decision-making aid for operators in industrial supervision was shown in the study of the decision making process [Mille, 1995]. In this paper the phase revise and retain are not studied. Just the first three steps are described.

[Lenz, *et al.*, 1998] distinguishes two different points of view on CBR applications: the domain and the task type. This study is situated in mechanical engineering domain and has the diagnosis as the task type. The CBR support tool is designed to the maintenance domain. The goal is to decide about the diagnosis in diagnostic process and to propose the repair operation. Hence, in the next section the model of diagnostic and repair help system applied to maintenance domain is proposed.

The current research in this domain is focused on detailed knowledge representation and in this view, Althoff [2001] considers the CBR as the technology of knowledge-based system implementation. The knowledge is stored in the form of cases in a case base which can thus be regarded as a knowledge base. The cases in the case base are created by representation and detailed modeling of domain knowledge. We propose the articulation of ontology techniques with the creation of case in the case-based reasoning. The case vocabulary joins the domain ontology and we can use this for the case base. The knowledge items in the ontology are relied by relations as "composed-of, is, has-for-instance" etc. We form these concepts in according to case structure and we create the case base for our CBR system.

[Bergmann *et al.*, 2001] distinguish 3 different types of CBR systems with regard to their case representation:

- textual – the cases are in a free text form.
- structural – the cases are in a form of lists of questions and answers (for each case the questions can be different).
- conversational - the cases are in a form of lists of attributes which are specified progressively and the all the characteristics are known like the domain model.

The knowledge capitalization, issued from the knowledge engineering, and the case-based reasoning as one of the techniques from artificial intelligence, join each other. We thus worked out the structural case-based reasoning system to create an interactive system of knowledge capitalization in e-maintenance.

4 Ontology of maintenance domain

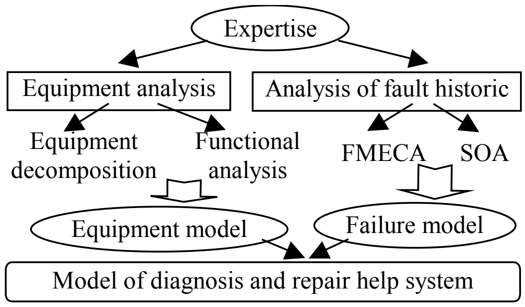


Figure 4. Equipment expertise.

To make a domain ontology means to model this domain, i.e. to define objects, to describe them and to structure the gathered descriptions. This implies to precise the description language and the acquisition system. During the realisation of application the informal description of objects should be transformed in the formal one [Bachimont, 2000]. Bézivin [2000] stressed that meta models as they are used by OMG (Object Management Group) answer well this constraint. OMG approved the notation UML which permits to specify, create, visualise and store the system objects in the form of diagrams comprehensible for both users and developers [Rumbaugh et al., 1999]. In particular, class diagram allows to represent ontology of domain knowledge.

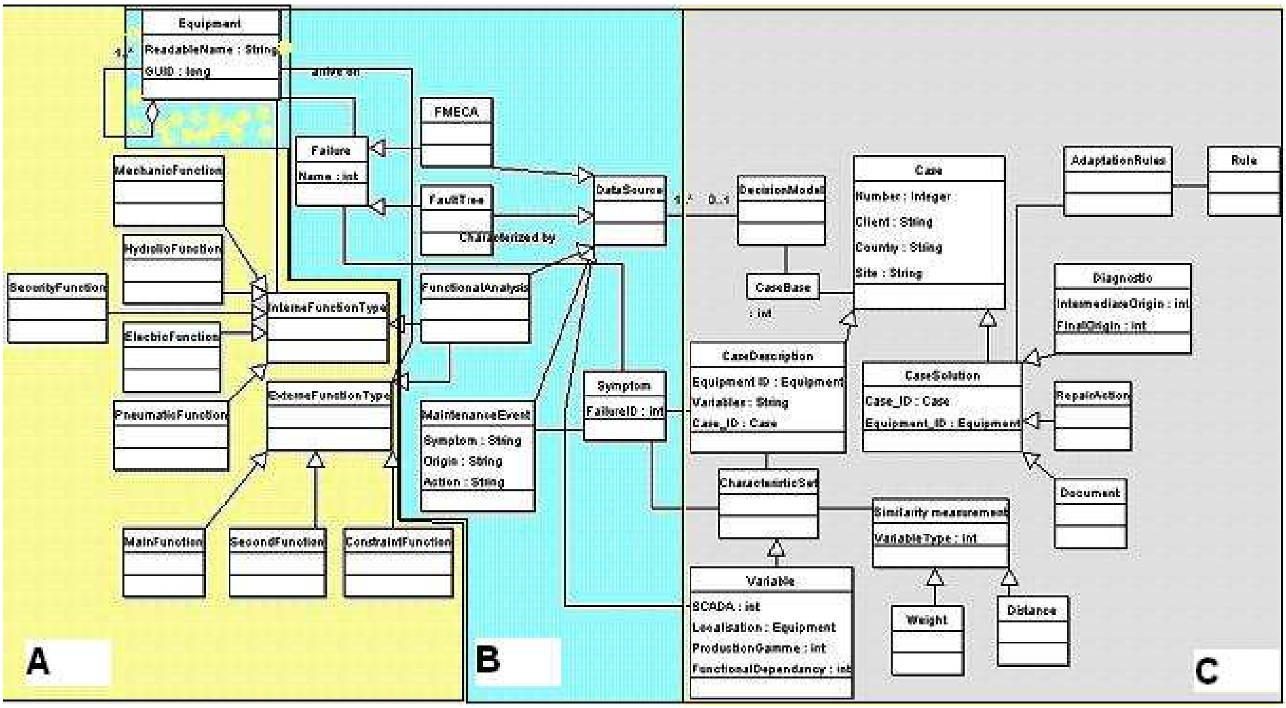


Figure 5. General maintenance ontology of diagnosis and repair help system.

The ontology of expertise domain in maintenance used for decision support system is developed from the maintenance process analysis, reliability concepts, the analysis of equipment to be maintained and the expert reasoning and practice. It is developed in the ontology editor Protégé[9] [http://protege.stanford.edu]. The diagnosis is based on two complementary analyses, the first of the equipment and the second related to the fault history recorded in the CMMS (Computerized Maintenance Management System) shown in the fig. 4. It is the methodology for creating an ontology. This expertise representation on the equipment will be translated in the form of class diagram. The purpose of functional analysis is the hierarchical equipment model. The equipment decomposition analyses equipments parts. The method FMECA (Failure Modes and Effects Criticality Analysis) identifies failure modes of equipment parts and its criticality. Finally SOA system identifies the symptom, the origin and the action of a failure. The case representation and modeling in the computerized decision-making system are based on the equipment and failure models.

The proposed data model is based on the previous analysis of general maintenance process presented by [Rasovska, et al., 2004]. This paper focuses on the diagnostic and repair tasks and proposes an UML class diagram of CBR support. In the fig. 5 the UML class diagram is presented. It describes, on the one hand, the equipment, its decomposition and functionality on fig. 5A that serves later to the ontology of case descriptors. The equipment analysis is made by reliability tools that professionals apply on equipments in order to assure their maintenance and that are shown on fig. 5B. These two parts make part of the domain ontology. On the other hand, the diagram describes the structure of diagnosis and repair help system based on the equipment analysis presented in the domain ontology (see fig. 5C). The CBR concepts form the model ontology.

The equipment is decomposed in the form of a tree structure. Each equipment has the function specification described by the classes *InternalFunctionType* and *ExternalFunctionType*. The first one specifies type of equipment function like mechanical, hydraulic etc. The second one represents hierarchy and dependencies between each function of the equipment. The equipment

[9] Protégé is an integrated software tool to construct domain ontology, to customize data entry forms and to enter data as well as to develop knowledge-based systems.

is linked with the class *Failure*. This is identified in the FMECA and makes part of the *FaultTree*. In the fault tree a breakdown is characterized by a *Symptom* which is the description of this breakdown caused by another failure called origin.

The decision help system is described by the class *DecisionModel*. The case-based reasoning stores the cases (*Case*) in the *CaseBase*. The case is composed of two different parts, namely the description of the case (*CaseDescription*) and its solution (*CaseSolution*). The description is associated with the *CharacteristicSet* containing the variables (*Variable*) which characterize the symptom of the failure (*Symptom*). The *CharacteristicSet* is linked by the *SimilarityMeasurement* based on *Distance* and *Weight* to find the similar cases in the case base. The diagnosis consists in describing the symptom by variables which allows identification of failure origin and thus determination of *RepairAction* and suitable technical *Document* for the operation. The class of rules *AdaptationRules* is introduced to adapt the solutions of old cases to the new ones. Each class of this diagram can be instantiated in order to create objects and consequently cases as. A case is represented by an instantiation of the class diagram [Ruet, 2002].

5 Application

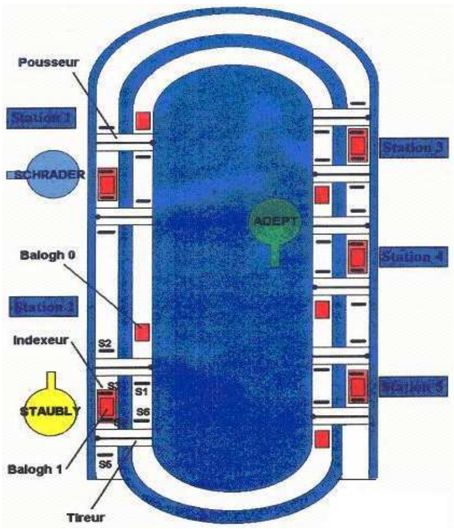


Figure 6. The transfer system SORMEL.

The application of the decision support system was done on the pallet transfer system shown in fig. 6. It represents a flexible production system. It is composed of five robotized working stations which are served by a transfer system of pallets organized into double rings (internal and external). Each station is equipped with pneumatic actuators (pushers, pullers and indexers) and electric actuators (stopper) as well as a certain number of inductive sensors (proximity sensors).

An inductive read/write module allows to identify and locate each pallet and to provide information relative to required operation in a concrete station. The displacement of the pallets is ensured by friction on belts which are involved by electric motors. Each pallet has a magnetic label that is used like embarked memory. This memory can be read in each working station thanks to magnetic read/write modules (Balogh) and allows the memorizing of the product assembly sequence. These labels thus enable to determine the pallet path through the system.

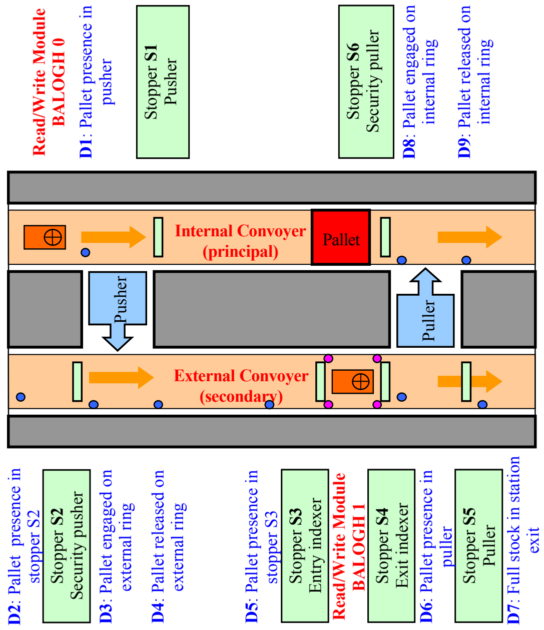


Figure 7. The working station of SORMEL.

The working station is described in more details in fig. 7. The pallets are conveyed on the interior ring which allows the transit between the various stations. When the pallet should be handled by a robot in the concrete working station (information read on the label of the pallet), the latter is deviated on the external ring where the concerned working station is. The working station is situated on the external ring and contains pneumatic and electric actuators (puller, pusher, indexer, stopper) and inductive sensors.

5.1 Architecture of decision support system

The architecture of proposed decision support system for diagnosis and repair within the e-maintenance platform is illustrated in fig. 8. The web portal of CBR tool for users is connected by Proteus with the CBR algorithm module and the web services (developed under Java and Python). This module is connected with the case base and description procedures developed in Protégé. The description procedures formalise dynamically the suitable questions in order to work out the description of the new problem (target case). The questions are asked to an operator, or to other modules integrated in the platform.

The CBR module uses the ontology of Belfort site and the generic ontology of maintenance.

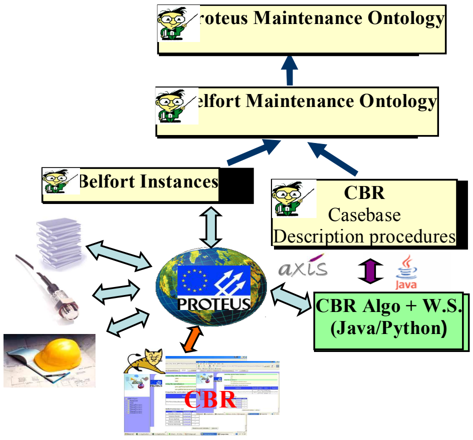


Figure 8. Architecture of decision support system.

Once the maintenance intervention request with the basic failure information has been received, maintenance operator asks for help from the diagnosis and repair decision support system. The system works out a new problem description by questions to the operator and by automatic collection of information and values coming from sensors from integrated systems in the platform, namely monitoring system SCADA and CMMS. A new problem is matched against cases in the case base and the most similar cases are retrieved. The solution suggested by these cases is adapted (reused) for the new situation and tested for success.

5.2 Case elaboration

The case representation and case acquisition are essential development components in the creation of CBR applications. The phase of case acquisition proves to be a significant aspect of knowledge engineering. The case development consists in facilitating the problem description in order to allow the search of a case whose solution will be most easily adaptable. The general method lies on completion or filtration of problem description which is based on domain knowledge. So that the eventual incomplete description is deduced and the weighting of descriptors is done in function of identified dependencies between new problem's descriptors and the searched solution's ones.

A case is a description of problem solving episode. In general, it is the association of some problem and its solution. There is a number of different theories on case representation but the most often used one is structured in a list of descriptors that take the form of complex objects. The case representation requires to list the various system components and to characterize them. It is based on the domain ontology developed in Protégé and shown on the fig. 9 and on the model ontology of CBR system shown on the fig. 10. The ontology of CBR system manipulates the domain ontology and use it in order to create the cases in the case base.

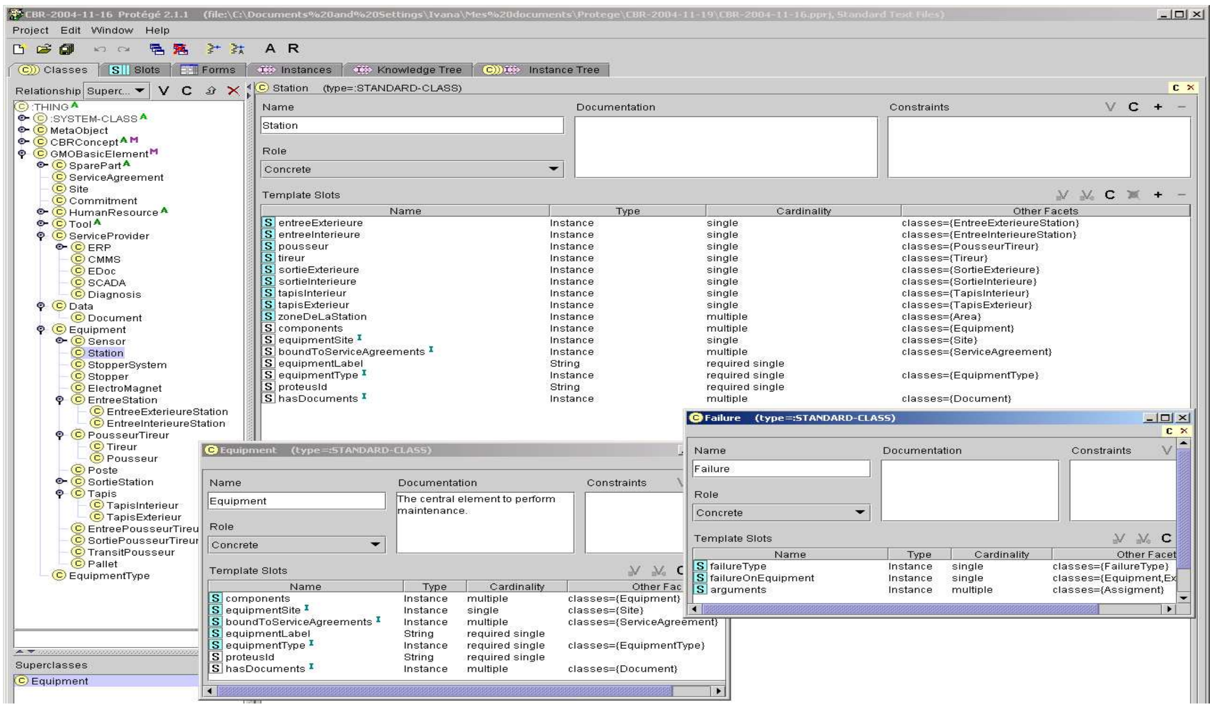


Figure 9. Domain ontology in Protégé.

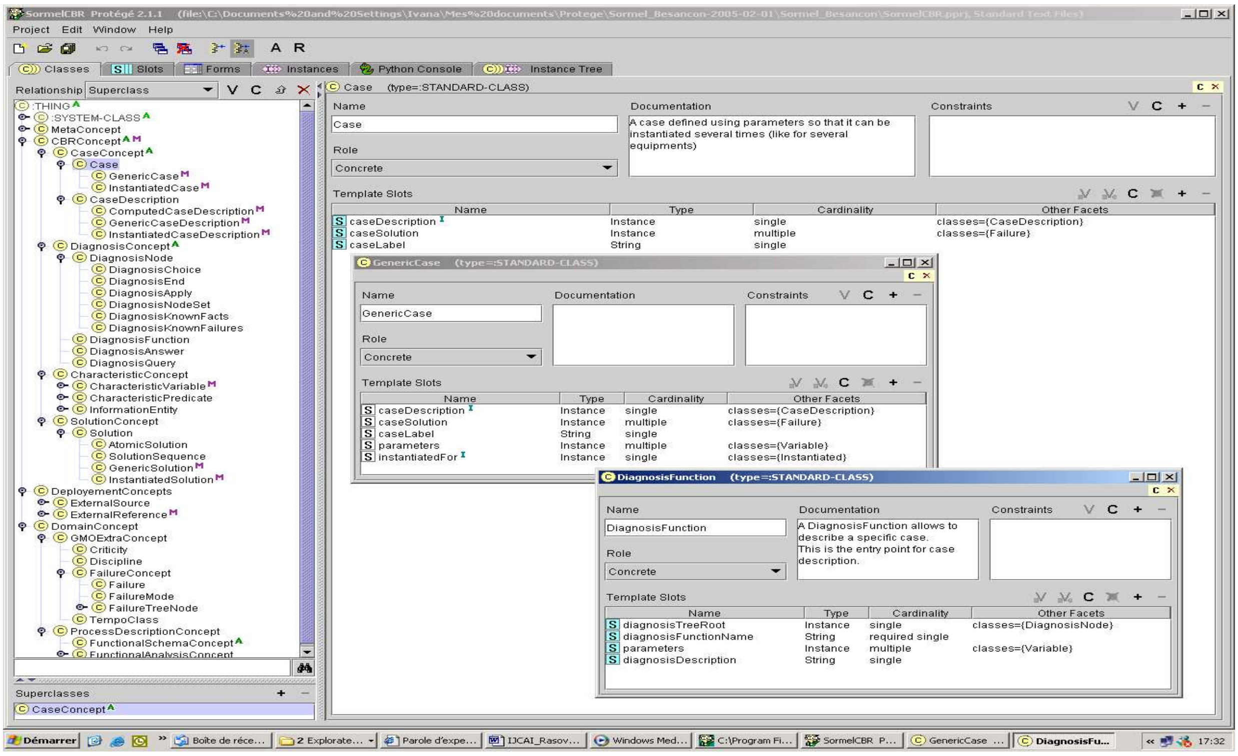


Figure 10. Model ontology of CBR system in Protégé.

5.3 Case representation

Case (type 1)	**Attribute : value**
Description Symptom : problem with	 Transfer [alimentation, grafcet, transfer]
Context identification localization	Localization_set: station [station, belt, simple turn, double turn] Localization_zone: pusher [puller, pusher, indexer, conv-int, conv-ext] Localization_sub-zone: entry [entry, pusher,exit]
Context's attribute: state	Sensor D1: 1 [0(indicates no pallet), 1(indicate pallet)] Balogh 0: 1 [1(pallet enters into the station), 0(pallet doesn't enter)] Stopper S1: 0 [0(stopper on the top), 1(stopper in the bottom)] Pusher: does not return to its position [push, does not return to its position, does not push]
Solution Functional equipment mode	 Symptom: problem with transfer means blocked pallet Localization of this pallet : station.pusher.entry D1: good function (OK) Balogh 0: pallet does not enter into the station (OK) S1: good function (OK) Pusher broken up
Action	Jack of pusher except service

Figure 11. Case representation.

The context of the case definition in the case base is the transfer system SORMEL. The case characteristics are issued from components of different nature such as sensors, controllers and control (command) units. To each component its state and an failure mode is associated. In the case description on fig. 11 there are states of sensors, actuators and pallets. Moreover the problem is localized from comparison of different components' states. The case solution raises from the functional mode of component where the problem is located.

The example of a case is presented at fig. 11.

The case is elaborated from the symptoms description characterizing the problem nature. During the acquisition of a new problem description one specifies:

- context (system, subsystem, component) by locating the failure,
- components of this context and their states (equipment and its value is listed).

The problem solution summarizes components identified in the context with their operating (functional, failure) modes. This leads to the identification of the failed one and to the repair action associated to the proposal of the operator skills for this intervention, required spare parts, required tools and suitable technical documentation. The case representation is object oriented.

From the localization of zones, sub-zones etc. the components are identified (sensors, actuators, Balogh etc). For this the failure detection rules are applied, created in studying the case base:
IF ((pallet is present) AND (sensor does not signal)) THEN [failing sensor OR pallet in a bad direction]
Once the generic rules identified, the case base will be composed of a minimum of generic cases representing a more consequent case unit.
In the proposed system, the case acquisition is done by filling out a form, during the new problem description. This questionnaire follows the tree decomposition of cases from the case base, and the closed questions presented depend on the failure localization. The recording of a new case in the case base is done during retaining step after its revision.

5.4 Retrieve and reuse

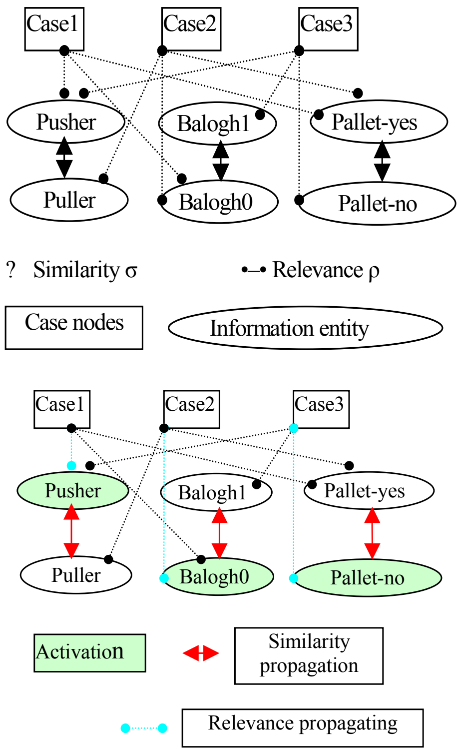


Figure 12. The case retrieval nets.

The ontology was established like a tool of knowledge sharing for various actors of the e-maintenance platform and is at the origin of the case base for decision support tool based on CBR technology. The ontology takes the form of a net to which the principle of case retrieval nets (CRN) presented in [Lenz, 1999] is applied. This structure enables to make the base available to other systems providing necessary information like SCADA, CMMS etc. The basic knowledge in CRN is an information entity represented as terms in the ontological structure to which we give acceptable values. A case is a set of these information entities (IE) and the case memory is a net with nodes for the IEs observed in the domain and additional nodes denoting the particular cases. IE nodes may be connected by similarity arcs and a case node is reachable from its constituting IE nodes via relevance arcs (see figure 12). Different degrees of similarity and relevance may be expressed by varying arc weights.
Given this structure, case retrieval is performed by

- activating the IEs given in the query
- propagating activation according to similarity through the net of IEs
- and collecting the achieved activation in the associated case nodes.

For more details see [Lenz, 1999]. Thanks to the CRN retrieval, a case is retrieved in a tree structure. This avoids to put the same questions on the failure several times in the different steps.
The similarities between the variables are done by the simple comparison of their values. Two points of view on the similarity can be taken into account: the similarity within the description of case and the similarity within the solution of case. Similarity measures are adapted to the object oriented case representation. The path of pallet in the transfer system is taken into account in the case comparison. The similarity $Sim(O_1,O_2)$ represents the global similarity between two case descriptions O_1 and O_2

$$Sim(O_1,O_2)=\sum_{i=1}^{p}\omega_i sim_i(o_1,o_2)\text{, where}$$

ω_i is the weight of attribute I, p is the number of attributes and sim_i is the local similarity calculated for the common class of two representations of attribute i

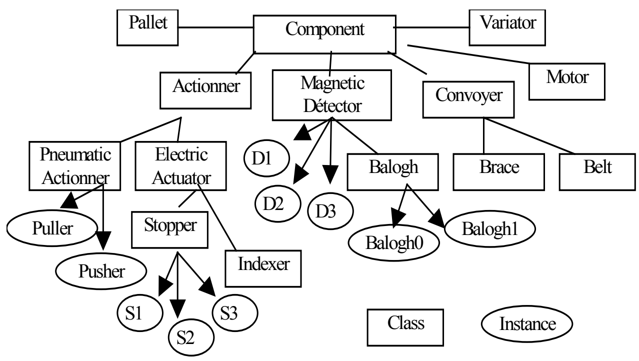


Figure 13. Ontology of case descriptors.

For two objects o_1 and o_2 the similarity is calculated while going up to the first common concept of these objects and by comparing the slots/attributes common on this level. For example, one object relates to the pusher. The new object corresponds to the puller and in the descriptors hierarchy (ontology of case descriptors on fig. 13) one can see that the two objects belong to the class actuator. The generalized class actuator will lead to the solution for the new problem. The ontology of case descriptors is issued

from the domain ontology and serves for the retrieve and reuse phase of the case-based reasoning cycle.
The reuse phase consists in re-using similar explanations to substitute suggestions for solutions by other elements chosen according to differences to reduce. In the reuse phase, the components hierarchy will be used in order to generalize the decision rules. The transfer system consists of 5 identical stations; it is thus possible to build generic classes to adapt the solutions for each particular station. In the work of Lenz on case retrieval nets, the adaptability of cases is not taken into account in the retrieve phase. In this proposition the descriptors hierarchy is used to replace a given descriptor by an other one of the same family and to apply actions associated to it. The adaptation strategy is introduced based on the adaptation operators designed to the diagnosis and repair domain. An adaptation operator is applied to a characteristic attribute of a case solution. Three types of possible adaptation are distinguished:

- insertion of an action;
- suppression of an action;
- modification of action parameters or modification of action.

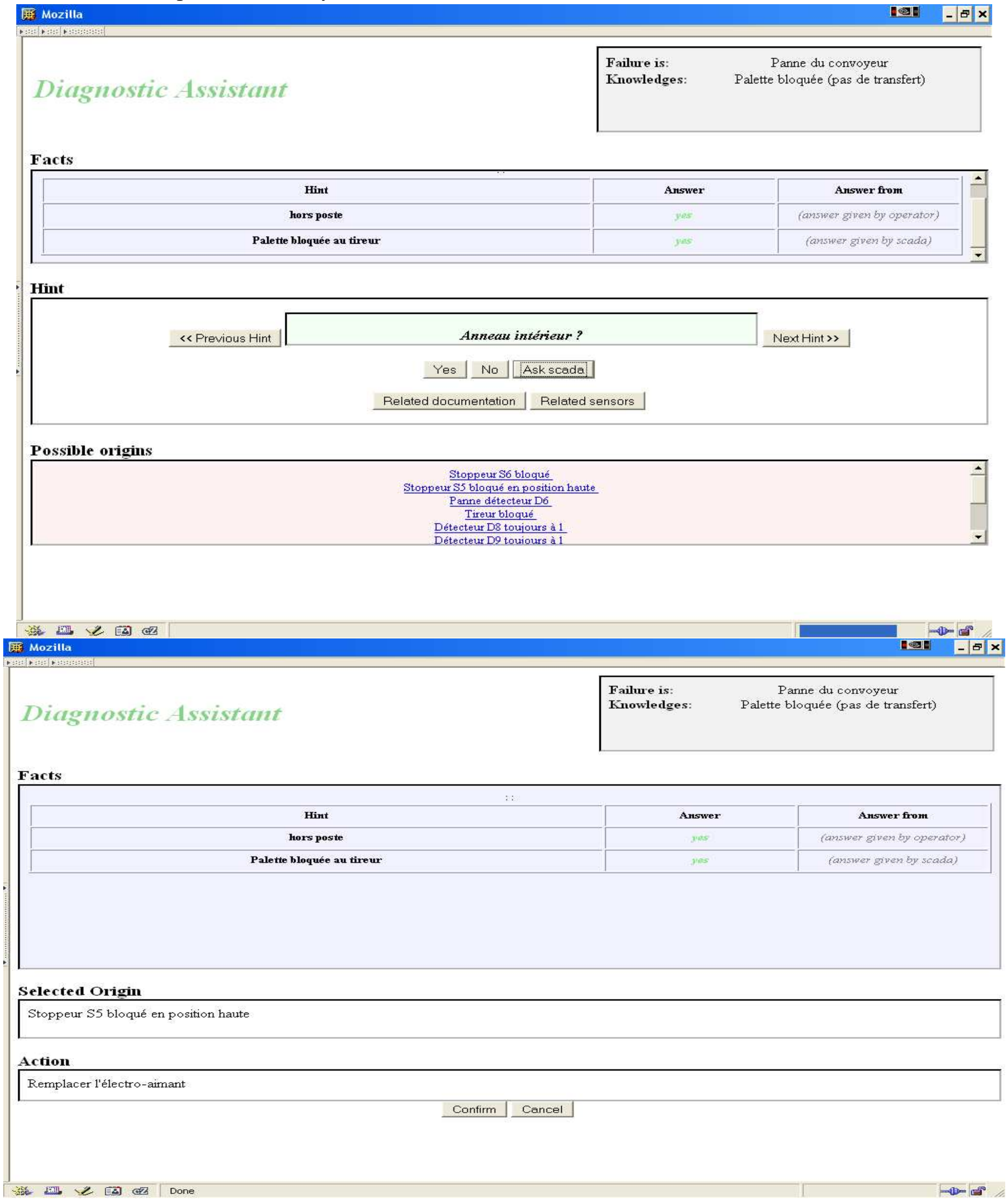


Figure 14. Example of print screens for diagnostic assistance.

5.5 User interface

One of the reproaches to the expert and knowledge based systems is that they are not adapted to users who are not specialist in the informatics and therefore the systems are not exploited as they should. In order to answer all the potential user needs, this tool intended to function in e-maintenance is regarded as an interface of the web service. In the fig. 14 examples of print screens are shown. The programming language to implement the web service of CBR is Java. User interface is developed in html.

Conclusion

We have proposed the knowledge capitalisation method in order to create a corporate memory in the enterprise specialised in maintenance services. It is the method of knowledge modelling which purpose is to formalise the expertise knowledge and to develop tools for this knowledge manipulation like other methods CommonKADS, MKMS and MASK.

In the last decade, there have been more and more artificial intelligence methods and applications used to improve industrial process operations and production. Among the most popular, there are case-based reasoning techniques applied in the fields where human operator experiences are demanded, such as fault diagnosis, repair help systems, management decision support systems or logistic support systems. The CBR systems make part of actual generation of expert systems. CBR systems are able of cooperating with human users and so the quality of support given and the manner of its presentation are important issues. For the implementation of decision support systems, it is necessary to cover the design process, the matching of requirements and needs for proposed systems. The process of equipment maintenance handles a huge volume data that is not always complete that requires up to date handing-over. The incremental decision help tools are to be envisaged. Process operation support systems require a CBR method that can represent system dynamics and fault propagation phenomena. Nevertheless, there is some problem in using the CBR system. Until today the systems are often isolated and they cannot cooperate with other systems. There is no aspect of standardisation in the creation of case and case vocabulary. On the other side the ontology techniques aims at the systematic creation and storage of knowledge assets based on the characterization of knowledge items. So the CBR system is completed by using the ontology techniques which are based on the formal description and standardisation but have no reasoning mechanism allowing to use existing knowledge. Thus the case representation join the creation of domain ontology and the cases as knowledge items are reused and exploited by case-based reasoning mechanism. The knowledge items are easily handled and with the case retrieval nets approach can be manipulated in the CBR cycle.

The CBR is designed as an interactive system; it can deal with the knowledge of experts. This knowledge goes behind the standard solutions of text books [Lenz et al., 1998]. The implementation of appropriate CBR support is possible only with the handling of incomplete information. A case connects information which have appeared together in a problem solving process. Actually the case base contains about twenty cases. This has allowed to test the case retrieval. These tests will follow up. The limits and the constraints of the proposed methodology application are in the domain modeling which is after all inevitable for the implementation of intelligent application system.

This study has focused on the ontology construction of the CBR system and its application to the real industrial platform. The system that use case-based reasoning techniques is designed to support human decision-making, learning and action. This hypothesis is supported by a number of already realised and successfully implemented projects of similar specialisation.

Acknowledgements

This study was funded by PROTEUS project programme sponsored by the French Ministry of the Economy, Finance and Industry and the Federal Ministry of Education and Research of Germany under the label of European Commission Initiative ITEA.

References

[Aamodt and Plaza, 1994] A. Aamodt et E. Plaza. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications, 1994, 7(i): pp 39-59.

[Althoff, 2001] K.D. Althoff. Case-Based Reasoning. In: S.K. Chang (Ed.), Handbook on Software Engineering and Knowledge Management, pp 549-588, 2001.

[Bachimont, 2000] B. Bachimont. Engagement sémantique et engagement ontologique. In J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (Eds), *Ingénierie de connaissances*. Paris, France : Editions Eyrolles, 2000, pp. 305-323.

[Bergmann, 2001] R. Bergmann. Highlights of the European INRECA Projects. In Aha & Watson (Eds.). Proceedings of the 4th International Conference on Case-Based Reasoning, ICCBR, Vancouver, Canada, 2001.

[Bézivin, 2000] J. Bézivin. De la programmation par objets à la modélisation par ontologies. In J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (Eds), *Ingénierie de connaissances*. Paris, France : Editions Eyrolles, 2000, pp. 209-222.

[Charlet *et al.,* 1996] J. Charlet, B. Bachimont, J. Bouaud and P. Zweigenbaum. Ontologie et réutilisabilité:

expérience et discussion. Acquisition et ingénierie des connaissances : tendances actuelles, 1996.

[Duribreux-Cocquebert and Houriez, 2000] M. Duribreux-Cocquebert and B. Houriez. Application industrielle d'une approche mixte de modélisation des connaissances. In J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (Eds), *Ingénierie de connaissances*. Paris, France : Editions Eyrolles, 2000, p.25.

[Fuchs, 1997] B. Fuchs. Représentation des connaissances pour le raisonnement à partir de cas : Le système ROCADE. Thesis, Laboratoire Image Signal Acoustique de CPE Lyon, 1997.

[Grundstein, 1994] M. Grundstein. Développer un système à base de connaissances : un effort de coopération pour construire en commun un objet inconnu. CP2I, Paris, 1994.

[Kolodner, 1991] J.L. Kolodner. Improving Human Decision Making through Case-Based Decision Aiding in AI Magazine, 1991, pp 52-68.

[Lenz *et al.,* 1998] M. Lenz, B. Bartsch-Spörl, H.D. Burkhard and S. Wess. Case-Based Reasoning Technology: From Foundations to Aplications, Lecture Notes in AI, Springer, 1998.

[Lenz, 1999] M. Lenz. Case Retrieval Nets as a Model for Building Flexible Information Systems. Thèse à l'Université de Humboldt, 1999.

[Mille, 1995] A. Mille. Raisonnement basé sur l'expérience pour coopérer à la prise de décision, un nouveau paradigme en supervision industrielle. Thèse à l'Université de Saint-Etienne, 1995.

[Perron, 2000] L. Perron. SIRIUS : une approche du savoir centrée sur l'utilisateur. In J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (Eds), *Ingénierie de connaissances*. Paris, France : Editions Eyrolles, 2000, pp. 481-498.

[Rasovska *et al.,* 2004] I. Rasovska, B. Chebel-Morello and N. Zerhouni. A conceptual model of maintenance process in unified modelling language. In Proceedings of INCOM'2004 - 11th IFAC Symposium on Information Control Problems in Manufacturing, Salvador-Bahia, Brasil, 2004.

[Reynaud and Tort, 2000] Ch. Reynaud and F. Tort. Diriger la réutilisation de composants à l'aide d'ontologies. In J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (Eds), *Ingénierie de connaissances*. Paris, France : Editions Eyrolles, 2000, p.59.

[Ruet, 2002] M. Ruet. Capitalisation et réutilisation d'expériences dans un contexte multiacteur. Thèse au Laboratoire Génie de Production de l'Ecole Nationale d'Ingénieurs de Tarbes, 2002.

[Rumbaugh *et al.,* 1999] J. Rumbaugh, L. Jacobson, G. Booch. The unified modeling language reference manual. Reading, Addison Wesley, 1999.

[Shreiber *et al.,*1994] A. Schreiber, A. Terpstra, P. Magn and M. Van Veelzen (1994). Analysing and Implementing VT Using CommonKads. In Proceedings of the 8th Banf Knowledge Acquisition for Knowledge-Based Systems Workshop, 1994.

[Talbi and Laublet, 2000] N. Talbi et P. Laublet. Construction d'un modèle générique d'expertise à partir d'un exemple. In J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (Eds), *Ingénierie de connaissances*. Paris, France : Editions Eyrolles, 2000, p.43-58.

[Van Craeynest *et al.*, 2000] J.M. Van Craeynest, J.-L. Ermine and Ch. Chagnot. Capitalisation des connaissances dans le cadre d'un transfert industriel. In J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (Eds), *Ingénierie de connaissances*. Paris, France : Editions Eyrolles, 2000, p.465.

Cooperative Work in a Health Care Network Through a Virtual Staff

Marek Ruzicka, Rose Dieng-Kuntz, David Minier, Olivier Corby
INRIA, ACACIA Project
2004 route des Lucioles, B.P. 93
06902 Sophia Antipolis Cedex
{Marek.Ruzicka, Rose.Dieng, Olivier.Corby}@sophia.inria.fr

Abstract

The paper presents a software tool called Virtual Staff which enables members of a healthcare network to visualize their collective reasoning: in order to diagnose the pathology of the patient (according to the symptoms expressed by the patient, the observations or analyses of the doctor and the already known health problems of this patient), or in order to determine the best possible therapeutic procedures. This Virtual Staff aims at offering to the users a service of support to cooperative reasoning, during the phases of elaboration of diagnosis or therapeutic decision and a service of constitution of a memory of decisions of the community constituted by the members of the care network. Another goal of Virtual Staff is to prepare repository of patient cases which will be reused for decision support in new cases.

Keywords: ontologies, knowledge management, cooperative work, decision-making support, QOC, SOAP, health network.

1 Context : Needs in a Care Network

The collaborative project *"Ligne de Vie"* (*Life Line*), aims at developing a knowledge management (KM) tool for a care network. Specialised in a particular domain or in a specific pathology, a care network is a health network gathering all the actors intervening in the care or follow-up processes. The objective of the network is to ease (a) communication and collaboration among these actors in spite of their physical distance, (b) the regular follow-up of the patient et (c) the respect of best practices inside the network.

The ideas proposed by the industrial partner for launching the project Ligne de Vie stem from fifteen years of experiences in health domain and from contacts with a network dedicated to diabetes: Nautilus offers a software for management of electronic medical record, Episodus, relying on a problem-oriented vision of the patient records and articulated around the notion of « Life Line » enabling to represent the life of the patient from his/her birth till his/her death with all the health problems encountered by this patient. As long as the patient still suffers from a problem, this problem remains open. When the patient gets completely cured from it, the problem is closed. Figure 1 shows an example of such life line of a patient.

This article focuses on a software tool called *Virtual Staff* which enables network members to *visualize their collective reasoning*: in order to diagnose the pathology of the patient (according to the symptoms expressed by the patient, the observations or analyses of the doctor and the already known health problems of this patient), or in order to determine the best possible therapeutic procedures.

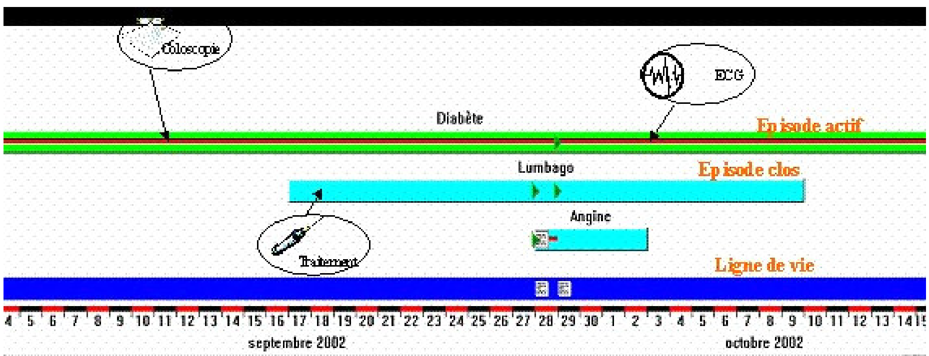


Fig. 1. Episodus representation of life-line

This Virtual Staff should offer to the users *a service of support to cooperative reasoning*, during the phases of elaboration of diagnosis or therapeutic decision and *a service of constitution of an organisational memory* – the memory of decisions of the community constituted by the members of the care network.

In the next session, we will explain the objectives of the Virtual Staff (VS). Then, we will describe the models of SOAP and QOC graphs on which the VS relies and we will present the architecture and functions of the VS, before illustrating cooperation the VS enables for the members of the care network. Finally we will discuss the interest of the software for KM in a health-centred community and then conclude.

2 Objectives of the Virtual Staff

In the hospital, the unity of location and of time allows the doctors to meet as a staff in order to discuss about the decisions to take. In a care network, the Virtual Staff aims to be a collaborative work supporting tool, allowing the real time update and history of therapeutic decisions. As an *electronic board* where each one can note information readable by the other members of the team, it constitutes a discussion support that may be synchronous (if the participants take part to the discussion at the same time or in the same place) or asynchronous (if each one accesses it at the moment appropriate to him/her). Starting from the patient's health problems, the members of the team will formulate diagnostic hypotheses and proposals for a treatment. Via this Virtual Staff, the care team will connect the various elements of the patient record useful for the discussion, and thus will converge in an asynchronous way towards the definition of new health problems and of new therapeutic actions. The formulation of diagnostic hypotheses is a priori reserved to the medical actors, whereas the discussion on the treatment could sometimes imply non medical professionals (for example, a welfare officer could emit arguments against the choice of a heavy treatment incompatible with housing conditions of the patient).

We suppose that many cases will be created *asynchronously* by *multiple users*. This requires visualizing and storing all information on the patient case, including multi-user and temporal aspects. Another objective of Virtual Staff is to collect solved cases in a special *repository* and to reuse them for retrieving past patient cases similar to a new patient case to solve. This kind of reasoning is similar to case-based reasoning [Moussavi, 1999].

3 Models of SOAP and QOC graphs

3.1 Weed's SOAP Model

In Virtual Staff, the dependencies between the various diagnostic and therapeutic hypotheses can be represented through a graph using the concepts defined in the Nautilus ontology. The doctor will reason by linking the health problems to the symptoms, the clinical signs and the observations in order to propose health care procedures.

The Virtual Staff can thus rest on the SOAP model (Subjective, Objective, Assessment, Plan) used by the medical community [Weed, 1971]. In this model:

- the S nodes describe current symptoms and clinical signs of the patient,
- the O nodes describe analyses or observations of the physician,
- the A nodes correspond to the diseases or health problems of the patient,
- and the P nodes correspond to the procedures or action plans set up in order to solve the health problems.

This SOAP model is used in the medical community to structure a patient record. Therefore, its use to structure the doctor's reasoning - that relies on the same concepts - seems natural.

3.2 QOC Model (Question-Options-Criteria)

Sometimes, the doctor may need to visualize all the possible solutions and the arguments in their favour or against them. The QOC model (Question Options Criteria) [Maclean *et al.*, 1991], used by CSCW community for support to decision-making or for design rationale in a project, can then be useful. In this model, a question Q corresponds to a problem to solve. To solve the question Q, several Options are thought out, with, for each option, the criteria in its favour and the criteria against it: each option is thus connected positively or negatively to criteria. The QOC graph is reduced to a tree if no criterion is linked to several options.

Two types of questions are possible for the Virtual Staff:

- *Diagnosis of pathology* (i.e. find the right A in the SOAP model): Which pathology explains the clinical signs of the patient?
- *Search of a prescription* (i.e. find the right P in the SOAP model): Which action plan will enable to treat the diagnosed pathology?

In the Virtual Staff, among the criteria to be satisfied, there are the patient's symptoms and the doctor's observations: for a question about the patient's pathology, each possible option will be linked by a positive influence link to the symptoms and observations compatible with this option, and by a negative link to the symptoms or observations rather incompatible with this pathology. The criteria will thus consist of S or O nodes of the SOAP model but they may also sometimes correspond to A or P nodes, if some diseases are incompatible or if some health care procedures are exclusive.

For support to decision on a treatment to cure the diagnosed pathology, the options will be the possible treatments, each one connected by a positive link to the criteria encouraging to choose it and by a negative link to the criteria inciting to reject it.

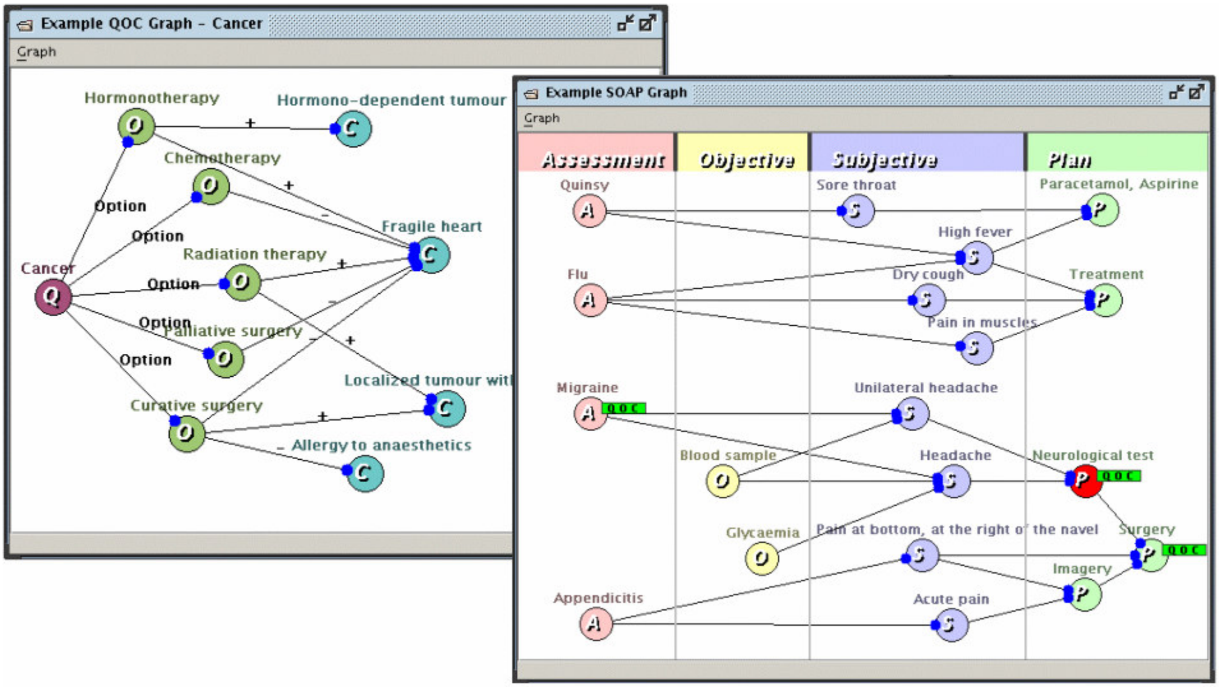


Fig. 2. Example of SOAP and QOC graphs

3.3 Combination of graphs in Virtual Staff

In the Virtual Staff, we combine both models: SOAP to visualize the medical record and QOC in phase of decision to choose between pathologies or between action plans. Each patient case consists of exactly one SOAP graph and any number of QOC graphs. SOAP graph is divided into four zones according to the four SOAP types of nodes (Assessment, Objective, Subjective and Plan) and for better clarity, nodes of same type are clustered in the corresponding zone.

Any *A* or *P* node could have an associated QOC graph. In Virtual Staff we distinguish between open and closed QOC graphs. An open QOC graph corresponds to a support for choosing a decision not yet taken, while a closed QOC graph corresponds to a decision already taken. Figure 2. shows basic user interface of both graph types.

4 Architecture and Functions of the Virtual Staff

Virtual Staff (VS) is not a single stand-alone application. It is not either only a graphical editor or visualization tool of SOAP and QOC graphs. Besides VS, there are three other important components involved in preparation and reasoning of patient cases. First of all there is the Nautilus Database of electronic patient records (EPRs) serving as source of initial medical data for new patient cases. To acquire consistent terminology and structure in graphs, nodes and arcs are connected with terms taken from ontology. First, the Nautilus DB had been transformed into RDF(S) [Lassila and Swick, 1999] encoded ontology. Most of nodes and arcs in Virtual Staff are therefore associated with concept and relation types from this ontology. This also ensures compatibility of concepts used in both Nautilus DB/Episodus and Virtual Staff. Finally we need some tool for searching terms in the ontology. We choose CORESE search engine which enables information retrieval based on RDF(S) semantic annotations on documents or on persons [Corby and Faron-Zucker, 2002]: CORESE can thus easily cooperate with Nautilus ontology. Whole context is described on fig. 3.

4.1 Nautilus DB of electronic patient records

Episodus, software of our industrial partner, offers management of electronic medical records stored in the Nautilus DB. Through Episodus, we could see the whole "life line" of given patient including evolution of all symptoms, diseases, treatments and other health care procedures (see fig. 1). It is useful to start a consultation with patient not with an "empty graph", but with a graph containing all mentioned items which are still active at the time of the consultation.

Hence we developed an exchange format for sending initial information about patient to Virtual Staff and possibly for later update of data in Nautilus DB according to acquired results. This format is written in XML and is fully compatible with Virtual Staff and the industrial partner is currently developing interface for Episodus to enable communication through this exchange format.

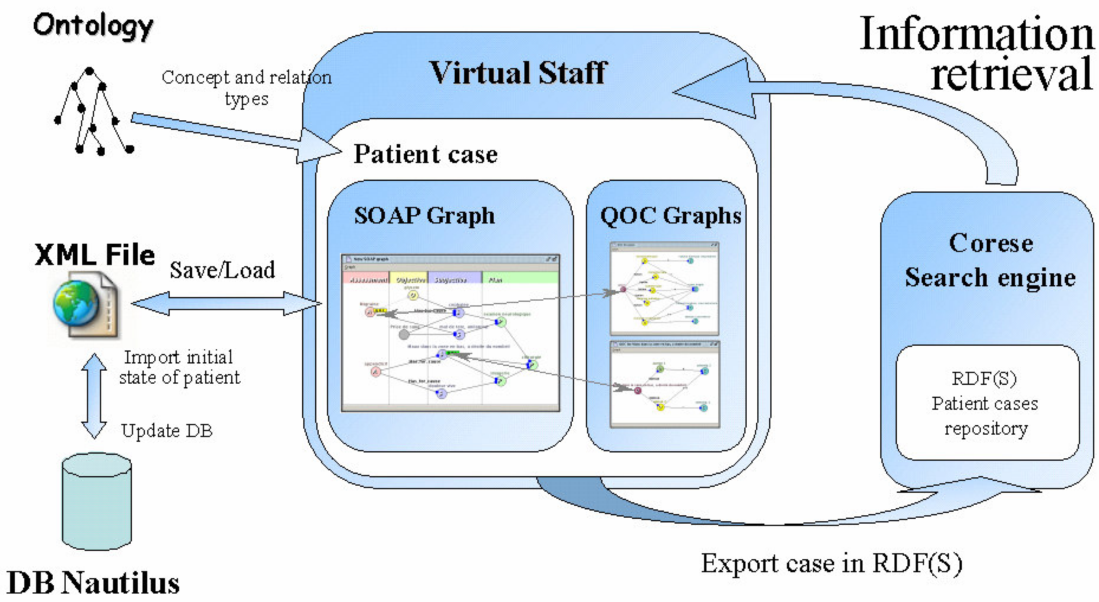


Fig. 3. Overview of Virtual Staff and its associated components

When a patient consults his/her doctor for new symptoms, the physician will create an instance of Virtual Staff. The system will then initialise a SOAP graph with all currently open pathologies and prescriptions for this patient, based on available information in XML file imported from Episodus. The initial A and P nodes automatically added in the graph correspond to the pathologies and the health care procedures already existing and open in the patient's life line.

4.2 Extended Nautilus Ontology

Extended Nautilus ontology was obtained from heterogeneous information sources (a database, a textual corpus, a classification). As it is represented in RDF(S), it can be directly accessed and queried by Corese Search engine (this was impossible with Nautilus DB conceptual schema). Advantage of this ontology is that graph nodes can be associated with concept types compatible with Nautilus DB. Without this compatibility it would not be possible to process initial import. Fig. 4 shows an extract of this ontology.

Just as nodes are associated with concept types, arcs should be associated with relation types. But as this ontology did not offer these relation types, it was necessary to enrich it with a minimum set of relations: *option* between question and option; *positive criteria*, *negative criteria* etc. between option and criterion, for QOC graphs. For SOAP graphs, some propositions should be refined after testing of Virtual Staff by doctors.

Fig. 4. Extract of Extended Nautilus Ontology

This also points to flexibility of Virtual Staff. Everything is taken from the ontology and no concept/relation type is hard-coded, so it is possible to switch to another ontology represented in RDF(S) without modifying source code.

4.3 Virtual Staff

From the user point of view, Virtual Staff is a graphical editor of SOAP and QOC graphs, with possible connection of nodes and arcs with the ontology (through Corese search engine) and data interchange with Nautilus DB of EPR(s) (through XML interchange format).

Fig. 5 gives a general overview of Virtual Staff user interface. Main (right) part of window represents electronic board for creating and updating graphs. Left part gives us information about selected graph and node/arc, about all involved participants and about previous contributions. Selection of particular participant/contribution will enhance appropriate nodes with *color rings*. These *color rings* depend on to the contribution author, its role or the last modification time. This enables quick overview of which part of graph was created by which participant, at what time etc. Color enhancement could help new participant to better and faster understand patient case [Benbasat *et al.*, 1986].

After brief examination of single case using these enhancements, the doctor should have sufficient information about case evolution and about reasoning of other participants.

It is possible to represent SOAP and QOC graphs through conceptual graphs [Sowa, 1984], built by using the concepts and relations of the Nautilus ontology. Due to the correspondence between conceptual graphs and RDF(S) language [Corby *el al.*, 2000 and 2004; Corby and Faron-Zucker, 2002], they can also be represented in RDF(S). Virtual Staff therefore implements another save function, this time using RDF(S) format. Through saving cases in RDF(S) we can obtain repository of closed/solved patient cases which could be later queried by Corese search engine or analysed by statistics and data-mining methods.

To reach platform-independence Virtual Staff is completely implemented in JAVA. Connection to Nautilus DB or other database of EPR(s) is optional, but to run Virtual Staff, the user needs to rely on an ontology represented in RDF(S).

4.4 CORESE

CORESE (COnceptual REsource Search Engine) [Corby and Faron-Zucker, 2002] is a semantic search engine dedicated to RDF(S) and is widely used within VS. Since the extended Nautilus ontology is represented in RDF(S) we can use Corese for finding requested concept and relation types.

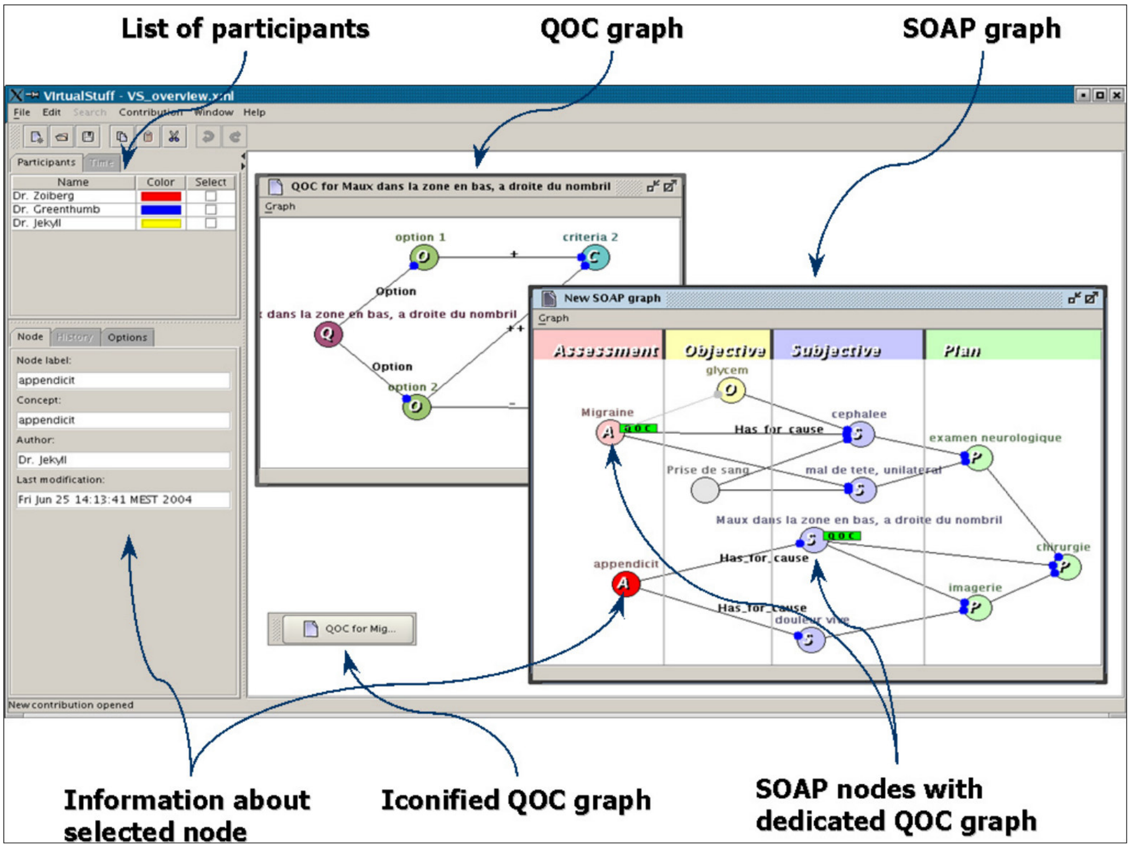


Fig. 5. Virtual Staff user interface

While adding or updating nodes in already existing part of the graph, the system can propose a list of possible concepts to help the user to build the SOAP and QOC graphs.

Another possible usage of Corese is querying of patient cases repository. We can query through Corese query language [Corby *et al.,* 2004] questions like “which positive criteria are in favour of radiotherapy option in QOC graphs for cancer” or “what is typical relation for two given concept types from Nautilus ontology”. Answers to these questions could help the user in two ways. Firstly, they can serve as kind of a remainder, so that (s)he does not forget any option and/or criteria in QOC graphs and second, they can enable case-based reasoning while filling empty parts of graphs or making the QOC decision.

4.5 Data interchange between components using XML format

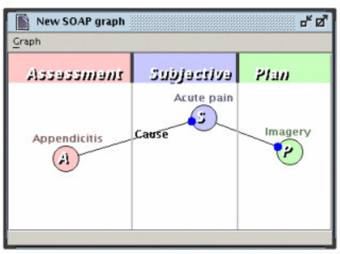
One of our assumptions is that any change in the graph must have a responsible participant. Each time a participant (any doctor or other medical person) opens new or already existing case, Virtual Staff automatically creates new *“contribution”* for this participant. Everything performed during this contribution (adding/changing nodes and relations) is stored in an XML file using *delta function* and for tractability reasons; it cannot be directly altered later.

For example one doctor adds couple of nodes in SOAP graph and then closes his/her contribution. Then another doctor in a following contribution can modify these nodes and/or delete them, but Virtual Staff will remember both states, before and after modification. On the other hand, Virtual Staff cannot manage changes made within one contribution (e.g. when a doctor adds a new node and then removes it, nothing is stored in the delta function; if (s)he altered the name of some node several times, only the last value is stored etc.).

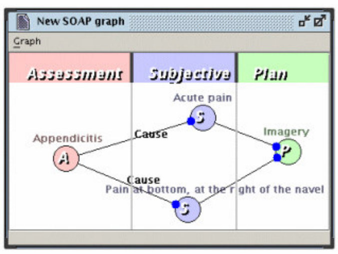
Fig. 6 presents two states of SOAP graph made by two consecutive contributions and shows the corresponding parts of XML file (for saving space we skip definition of relations between nodes). In the first part an initial state of SOAP graph is defined, this means that every node and relation is stored, whereas in the second part there are only new nodes and modifications to existing nodes. This approach has several advantages. The user can easily find who is responsible for any part of any graph at any time, can backtrack to previous contributions and see evolution of the patient’s case and thanks to delta function, (s)he can also see whether some parts of a graph were deleted and/or changed by different participants. For example from fig. 6, (s)he knows how the graph looks like and even more (s)he can see that Dr. Jekyll does not agree with Dr. Zoidberg in a particular part of the case. So it offers tractability on possible divergences between opinions of several doctors, with their respective arguments.

5. Cooperation using the Virtual Staff

Some patient cases are solved by a single doctor, but other complex cases need cooperation of several doctors having different medical specialties (in particular, in networks dedicated to heavy pathologies). Cooperation itself could be divided by time in *synchronous* and *asynchronous* and by place in *face-to-face* and *distant*. In ideal world, all cases will be solved in synchronous face-to-face sessions, but in practice doctors have to overcome “gaps” between places and time. While designing Virtual Staff, we had to take into account multi-user aspect with respect to all possible cooperation types.




```xml
<Contribution open="2004-07-26T17:00:02" close="2004-07-26T17:04:44" participant="Dr. Zoidberg">
  <SOAPGraph title="New SOAP graph">
    <NodeList>
      <Node ID="n0" label="Appendicitis" x="67" y="124">
        <Concept ID="PAPPA1"/>
      </Node>
      <Node ID="n1" label="Acute pain" x="229" y="76">
        <Concept ID="SDOUL1"/>
      </Node>
      <Node ID="n2" label="Imagery" x="330" y="116">
        <Concept ID="GIMAG1"/>
      </Node>
    </NodeList>
    <RelationList> <!--  skipped, similar as Nodes --> </RelationList>
  </SOAPGraph>
</Contribution>
```




```xml
<Contribution open="2004-07-26T17:00:02" close="2004-07-26T17:06:49" participant="Dr. Jekyll">
  <SOAPGraph>
    <NodeList>
      <Node ID="n1">
        <Concept ID="SD0311"/>
      </Node>
      <Node ID="n3" label="Pain at bottom, at the right of the navel" x="211" y="184">
        <Concept ID="SDOUL1"/>
      </Node>
    </NodeList>
    <RelationList> <!--  skipped, similar as Nodes --> </RelationList>
  </SOAPGraph>
</Contribution>
```

Fig. 6. XML format for storing cases using delta function

Today it is easy to share XML or RDF(S) files, so the problem with distance between places is not the crucial one. No software tool can be sophisticated enough to fully replace natural collective discussion, but especially in medical domain, where there are many specialists and competence fields, it is difficult to get all required doctors in the same place at the same time. In the Virtual Staff, it is possible to edit a patient's case asynchronously and members of health network can append their part of "knowledge" in the case at any time. Moreover Virtual Staff provides for new users valuable information about patient case history.

It is understandable that each specialist could have a little bit different opinion on the same patient's case. After the patient's visit to the general practitioner, the doctor could for example prepare QOC graph for some new symptoms to diagnose. From his/her point of view (s)he makes a general model of decision process using this graph according to his/her medical specialty. Then, if another doctor e.g. cardiologist examines the graph from viewpoint of his/her medical specialty, the graph can be enriched with new options and criteria and possibly with a new decision of QOC. This process can continue iteratively for other relevant medical specialties.

If every participant is obliged to introduce his/her medical specialty on the VS (we call it *role of participant*) and as far as all graph states and decisions are saved using the delta function, knowledge contained in graphs can be analyzed from much more complex view. Searching history of patient case enables to see not only patient state for a given time, but it also brings us information about reasoning of all participants. In QOC graphs, the user can find out answers to questions like "*why and when particular treatment started*", "*which criteria were taken in account when treatment started*" etc. Each doctor could see which diseases and health care procedures were active during decision making and so on. If treatment failed for some reason and should be replaced (through a new QOC decision), then it is possible to list through all previous treatments and situations when these treatments were applied, and then to decide new treatment or to continue the older one (e.g. if new criteria recommending this treatment are given). So, Virtual Staff helps to visualize progress of all contributions and clearly shows reasoning of different participants.

6. Discussion

From organizational viewpoint, the organization constituted by a health care network can be considered as a virtual enterprise, with a rather informal structure, and its members constitute a community gathered by a common objective (i.e. offer the best health care and follow-up for the patients), each member having also more specific objectives due to his/her profession (e.g. doctor vs. nurse vs. social worker) and to his/medical specialty. The kind of cooperation in this organization may also depend on the kind of network: some networks are dedicated to a heavy pathology (e.g. diabetes, oncology, etc), and will gather members from different professions and different specialties, while other networks will rather gather the same kind of health centers or of professionals (e.g. hospital professional network, liberal practitioner network, pharmacist network, nurse network, etc) and others will be dedicated to a type of patient (new born, old people). The kind of interactions will of course differ according to the type of network. Our approach, based on support to cooperative reasoning, seems useful for a healthcare network dedicated to a heavy pathology. In [Bardram and Sølvkjær, 1996], four types of collaborative aspects in clinical research are emphasized: communication, sharing different records and material, planning of collaboration and collaborative problem solving: our approach tries to tackle at least patient record sharing and collaborative problem aspects.

From cognitive viewpoint, the members of the network build themselves a mental representation of the patient's case. The graphs handled in the Virtual Staff aim at enabling a participant to visualize partly this representation, to share it with other participants and to make it evolve through cooperation with other participants. SOAP model seems relevant for medical reasoning and QOC for representing diagnosis and therapeutic decisions since QOC model is known as useful for design rationale of a design project: patient's health care and follow-up can be considered as a therapeutic project to which some members of the health care network will part in.

From technological viewpoint, in addition to SOAP and QOC graphs, we mainly rely on semantic Web technologies:

- ontologies for representing the concepts shared by the network members,
- RDF(S) for representation of ontology and annotations on the patient's record,
- Corese semantic search engine for querying the ontology and RDF annotations (in particular on one or several virtual staffs concerning one or several patients),
- XML for interchange between the virtual staff and the Episodus software describing the life time of the patient.

Our approach can be seen as tackling KM from several viewpoints:

- *Organizational memory:* we build a memory of a specific community constituted by the health care network members: these members may be individual or organizations (e.g; hospital, health centre). More precisely, we build a memory of health projects about patients: each life line of a patient is considered as the trace of a medical project, with events, phases, actors playing a role in this project. Tractability of the project decision rationale is tackled by QOC graphs.
- *Support to cooperative work:* our approach follows the suggestions of [Pratt *et al.*, 2004] for taking inspiration of CSCW for heath care support. We try to offer, in a longer-term, shared understanding, informed participation and social creativity, as in the vision of Fischer [Fischer, 2000]. And according to us, support

to cooperative work is clearly an approach for knowledge management that also has the advantage to be user-centered.

- *Case-Based reasoning*: even though we do not use classic Case-Based Reasoning techniques [Moussavi, 1999], querying through Corese on past patient cases and past VS sessions in order to have suggestions for a new patient case aims at the same objective as case-based reasoning.

The first version of the Virtual Staff was validated by the industrial partner, from two viewpoints: its functions and its interfaces. This evaluation led to several improvements of the graphical interface of the virtual staff: the new interface - illustrated by the figures through the paper - helps the user to really follow the reasoning guided by the SOAP model (each part of the graph is dedicated to nodes expressing either an Assessment or an Objective or a Subjective or a Plan) or by the QOC model (Question → Option → Criterion). We will also ask actual physicians to handle the Virtual Staff in order to evaluate it from the viewpoint of a real end-user.

The Nautilus ontology and the Virtual Staff were validated by Nautilus society. Notice that, even though the Virtual Staff was implemented with the Nautilus ontology (for collaboration reasons), it would be possible to adapt the Virtual Staff to another medical ontology such as UMLS meta-thesaurus [Pisanelli *et al.*, 1998].

As IBIS method is close to QOC method, we can compare the Virtual Staff to gIBIS [Conklin and Begeman, 1988] that already relies on graph visualization and offers argumentation on decisions. But our originality is to combine both SOAP graphs and QOC graphs and to rely on a medical ontology for building and handling these graphs.

We must notice that so far the Virtual Staff has not yet been used in real practice. Further experiences will enable to determine whether such a tool can be accepted by medical professionals and avoids the criticisms against previous tools a la gIBIS. Another cooperative tool for a healthcare network is WebOnColl [Chronaki *et al.*, 1997], a web-based medical collaborative environment relying on user profiles and virtual workspaces.

7. Conclusions

In this paper, we presented software tool called Virtual Staff enabling members of a health care network to create, modify, and consult alone or together: SOAP graphs describing the links between diagnostic and therapeutic hypotheses, symptoms and observations, and QOC graphs for support to decision-making. The nodes of both kinds of graphs are typed by the concepts of the Nautilus ontology. Such a combination of these SOAP and QOC models with an ontology is original and illustrates the interest of an ontology to help the user to visualize a reasoning or a decision-making process.

Virtual Staff is also designed for multi-user cooperative work with possible backtracking through evolution of cases. This enables to see how knowledge contained in graphs was modelled and how reasoning and decision making were processed. Solved and closed cases are stored in RDF(S) repository and they can be analysed to get new knowledge (to improve ontology etc.) or reused in case-based reasoning on new cases.

As a further work, the industrial partner will proceed to end-user focused evaluation by physicians taking part in an actual network (probably in diabetes).

Acknowledgements: We thank very much Philippe Ameline, the author of the Nautilus database who launched the project “Ligne de Vie”, had the initial idea of the Virtual Staff and validated our implementation of the Virtual Staff. We also deeply thank the “Ministère de la Jeunesse, de l’Éducation Nationale et de la Recherche” for funding the project “Ligne de Vie”.

8. References

[Bardram and Sølvkjær, 1996] Bardram, J. E. and Sølvkjær, M.: Computer Supported Cooperative Work in Clinical Practice. 13th Int. Congress of the European Federation for Medical Informatics, 1996.

[Benbasat *et al.*, 1986] Benbasat, I., Dexter, A., Todd, P.: An Experimental Program Investigating Color-Enhanced and Graphical Information Presentation: An Integration of the Findings, Communications of the ACM, 29:1094-1105, 1986.

[Chronaki *et al.*, 1997] Chronaki, C.E., Katehakis, D.G., Zabulis, X.C., Tsiknakis, M., Orphanoudakis, S.C.: WebOnCOLL: medical collaboration in regional healthcare networks. IEEE Transactions on Technology in Biomedicine, vol. 1, n. 4, 257-269, December 1997.

[Conklin and Begeman, 1988] Conklin, J., Begeman, M.: gIBIS: A Hypertext Tool for Exploratory Policy Discussion, Communications of the ACM, 140-152, 1988

[Corby *el al.*, 2000] Corby, O., Dieng, R., Hébert, C.: A Conceptual Graph Model for W3C Resource Description Framework. ICCS’2000, Springer-Verlag, LNAI 1867, Darmstadt, 2000, p. 468-482.

[Corby and Faron-Zucker, 2002] Corby, O., Faron-Zucker, C.: Corese : A Corporate Semantic Web Engine. In Workshop on Real World RDF and Semantic Web Applications - 11th International World Wide Web Conference 2002 Hawaii. May 2002. http://paul.rutgers.edu/~kashyap/workshop.html

[Corby *et al.,* 2004] Corby, O., Dieng-Kuntz, R., Faron-Zucker, C.: Querying the Semantic Web with the CORESE Search Engine. Proc. of ECAI’2004, Valencia, 2004.

[Fischer, 2000] Fischer, G.: Shared Understanding, Informed Participation and Social Creativity: Objectives for the Next Generation of Collaborative Systems, COOP'2000, IOS Press, p. 3-16.

[Lassila and Swick, 1999] Lassila, O., Swick R. R.: Resource Description Framework (RDF) Model and Syntax Specification. W3C Recomm., 22 Feb. 1999, http://www.w3.org/tr/rec-rdf-syntax/

[Maclean *et al.*, 1991] Maclean, A., Young, R., Bellotti, V., Moran T.: Questions, Options, and Criteria: Elements of a Design Rationale for User Interfaces. IJHCI, 6(3/4):201-250. 1991.

[Moussavi, 1999] Moussavi, M. : A Case-Based Approach to Knowledge Management. In Aha, D. ed, Proc. of AAAI'99 Workshop on Exploring Synergies of Knowledge Management and Case-Based Reasoning, July 1999, Orlando, AAAI Press Techn. Report WS-99-10.

[Pisanelli *et al.*, 1998] Pisanelli, D. M., Gangemi, A., Steve, G.: An Ontological Analysis of the UMLS Metathesaurus. Journal of the American Medical Informatics Association, 5:810-814, 1998.

[Pratt *et al.*, 2004] Pratt, W., Reddy, M., McDonald, D., Tarczy-Hornoch, P., and Gennari, J. Incorporating Ideas from Computer-Supported Cooperative Work. Journal of Biomedical Informatics, 37/2. pp 128-137. 2004.

[Sowa, 1984] Sowa, J. F.: Conceptual Structures: Information Processing In Mind and Machine, Addison-Wesley, 1984.

[Weed, 1971] Weed, L. D.: The Problem Oriented Record as a Basic Tool in Medical Education, Patient Care and Clinical Research. Ann Clin Res 3(3):131-134. 1971.