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 present, in this article, a "topic map" system applied to 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, collective semantics "in progress".
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 to use, to organize and to 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.

To achieve this goal, Research and Development especially in Knowledge Engineering should not concentrate all their efforts on automated systems, forgetting the human cognitive and social aspects of the shared meaning. It is necessary to use computers and the Web as supports for individuals and groups Knowledge Management activities. So people will be able to improve their capacity to comprehend information cognitively and socially within their activities, to create collectively knowledge and to share complex semantics originating from the current real world.

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 the moment, the experience is limited to the French speaking OSS community, but we consider including other languages).

The concrete application of this study is visible on the Web[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,

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

the topic map will 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 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. A community may use a topic map or a viewpoints map and decide to build such a map collectively, in order to help members to organize their tasks, to retrieve information resources, to classify skills, etc. In many cases, the goal or the result is to strengthen the group as a community of practice, and to favor a shared culture across members' activities.

Especially in business activities with expertise, communities are permanently inventing. The categories, the knowledge and the meaning of the words change. Neologisms and new classification schemes appear and disappear in a continuous manner within business lexicons. Communities Business notions are not always unanimous from one expert to another one, depending on different skills, opinions or roles in the processes. 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 we shall propose it further, 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 do use the low-level open standards (XML, RDF) of the semantic Web "Cake" proposed bay 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 Communi-

ties 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. 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 to overcome these conflicts by communication.

Building shared meaning from several viewpoints, with participation of experts or community members, is especially crucial in the situations where these underlying semantic resources to be elicited and maintained are very voluminous, evolutionary and even contentious (e.g. metadata of competitors together in a e-Marketplace). In such cases, the communities need S2W applications to organize themselves their activities of co-construction by themselves, *i.e.* adapt roles and internal services in order to bootstrap, build and maintain the semantic structures they need. In order to co-construct such artifacts in a continuous manner, in the flow of the activity, they have to be helped by well-adapted S2W tools and by accompanying methodologies, including the State-of-the-Art of CSCW tools. So the users remain active co-builders during the whole lifecycle of the S2W application.

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 Knowledge Acquisition point of view, a particular bottleneck exists for building and maintaining formal computable domain ontologies, especially in large domains with frequent changes such as 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*".

With the S2W is 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:

- In many cases, the construction of semantic structures can neither lie on a single individual nor be made *a priori*. In these cases, a co-construction among several actors is inevitable. It is also possible to imagine a co-

construction "by doing" which would occurs as a permanent process. This process would be a key issue for the Knowledge Management in the organization.

- The actors need to organize themselves for benefiting from their differences and their complementarities, and to resolve potential conflicts among antagonist points of view. So it is important to offer an interaction framework which takes into account the diversity of vocabularies and conceptual systems from the actors in their context.
- 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"). Variations of these three main KBM-roles are possible (for example, "editor-translator", if the KBM is a multilingual one, "chief-editor" in case of a hierarchical organization of the map validation process, etc.).

In the "commerce" situations taken in a broad sense, buyers and sellers speak different "sub-languages". Each one brings complementary expertise to "co-construct" the catalogues. For example, in the context of the proposed Yellow Pages artifact for Open Source Software (see §4), idioms are different in cases of the software developer, the service company, the software user, the economic analyst, the jurist, etc. Each of them has a different viewpoint on the considered software entity – though some persons could belong to several roles or sub-communities, speak several "sub-languages" and carry out translations between them. This point is a major stake for the success, and sometimes the survival, of the organizations today.

Since 2001, when we initiated this "Knowledge-Based Marketplaces" concept (KBM) at the Tech-CICO Laboratory [Cahier and Zacklad, 2002], we have initiated and stud-

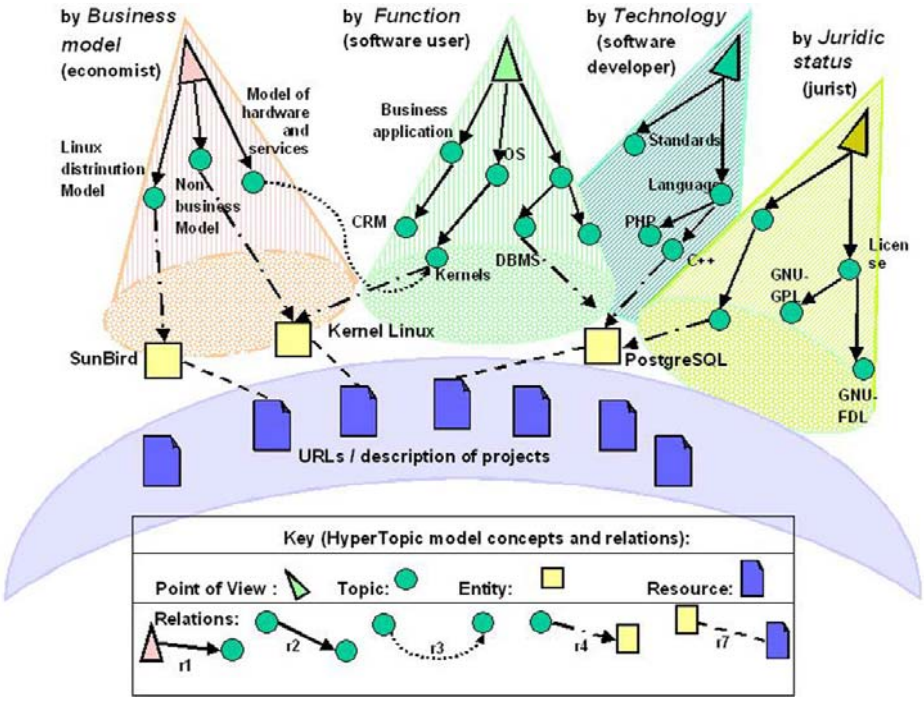

Figure 1 - Multiple points of view in a socio semantic web semi-formal ontology, according to HyperTopic model (exemple of Yeposs application)

ied several real-size KBM experiments and applications in intranet or extranet contexts. These applications are located in various fields, but they always link actors who are suppliers or buyers of "resources", in the wide sense. "Entities" classified in these KBM systems are products or services, R&D projects, knowledge or human resources to solve problems, skills [Cahier *et al,* 2004], etc. And presently KBM application deals with software projects, in the case of the OSS application detailed in §4. In a KBM actors use and construct the topic map as a pivot to manage and retrieve the information describing the entities of the collection. Exciting aspects in these concrete KBM experiments were:

- to improve the discussion and cooperation abilities in the community,
- to alleviate the back-breaking work of fine-grained classification by sharing it between more numerous and more competent actors,
- and to detect the possibilities for new knowledge to emerge from the collaboration processes.

In addition, in the case of the end of 2004 initiated OSS application that we describe below, we try to go beyond all our preceding firewall-confined KBM applications, to enlarge the field of experiment:

- for the first time, towards a large-scale and open usage, proposed to a mature community intensively using cooperative tools on the Web,
- for the first time also, with the ability given to the actors to debate at a fine-grained level, by mean of multiple forums threads, and thus to discuss separately each Point of view or each Topic on the map.

3 Methodology, models and tools

3.1 General methodology

A general guideline to build S2W applications is to give to 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 in 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 for methodology a pragmatic framework. In particular, in his domain, a community must face a continuous growth of the information. 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.

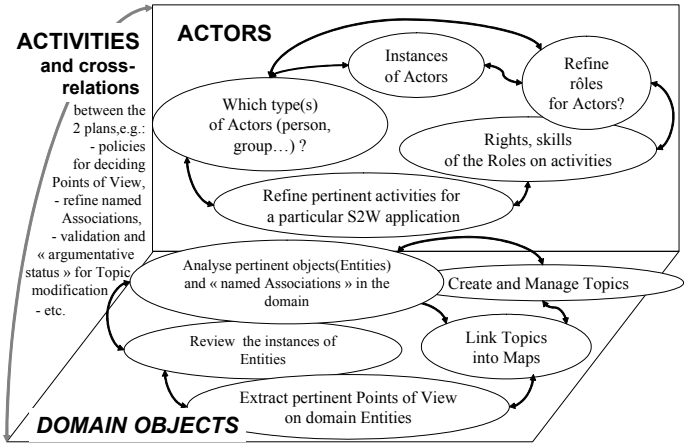

Figure 2 - S2W applications requires a threefold analysis

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, brain storming, 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.

In such a collective design problem solving situations, 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 the goal, the co-building of 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. In a first approach, HyperTopic concepts and relations can be seen, as the elements of the key of 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.

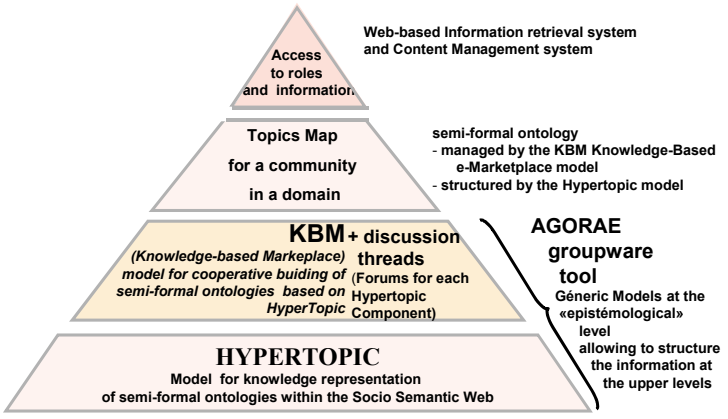

Figure 3 - Architecture of a socio semantic Web application based on HyperTopic and KBM models

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.

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 [TMO, 2001] 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.

According to the TMO standard, Topics are not concepts but simple or complex linguistic expressions expressing "a subject we speak about". We also use this definition for 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 TMO standard, in particular to facilitate their handling. For that, in addition of 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 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.

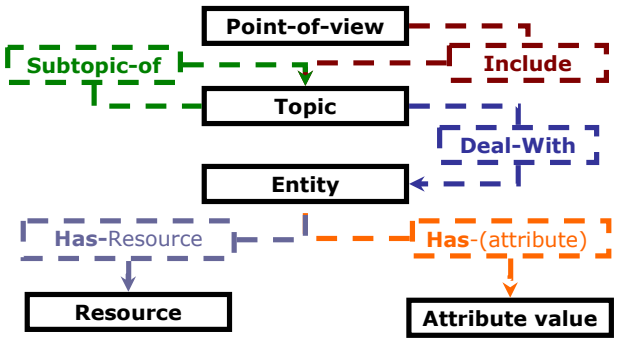

Figure 4a - HyperTopic Basis (associations are dashed)

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 others approaches like FacetMap [XFML, 2002] in which the "facets" translate couples of (attribute, value) attached to the considered objects, without proposing the explicit possibility to gather facets in points of view meaningful for the actors, as HyperTopic proposes.

In HyperTopic, a point of view is a point of vision on an 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.

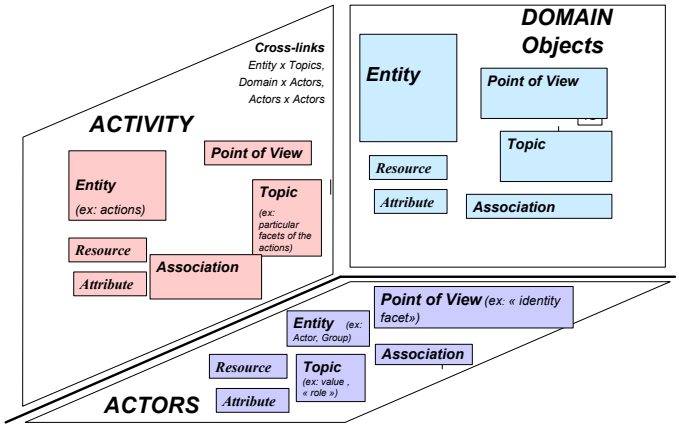

Figure 4b - The HyperTopic Model (v2), emphasizing the 3 dimensions of the socio semantic Web

In its second version, HyperTopic allows specification and seeks of what are the business objects, and what changes in these objects by the time, 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.

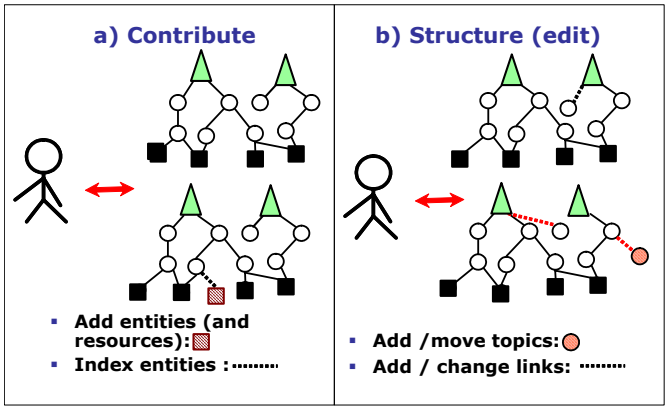

Figure 5 - Contributor (a) and semantic editor (b) actions with HyperTopic

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 has today main features needed to evoked profiles (initial design and normal use in "cruising" mode) of KBM-like applications: 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 actors. 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 declare 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.

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

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 order to conduct to a real participation, indeed loyalty, 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:

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

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 elements the other members will put as well in the system.

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

[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.

open source software, users are often in the same time developers and secondhand 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.

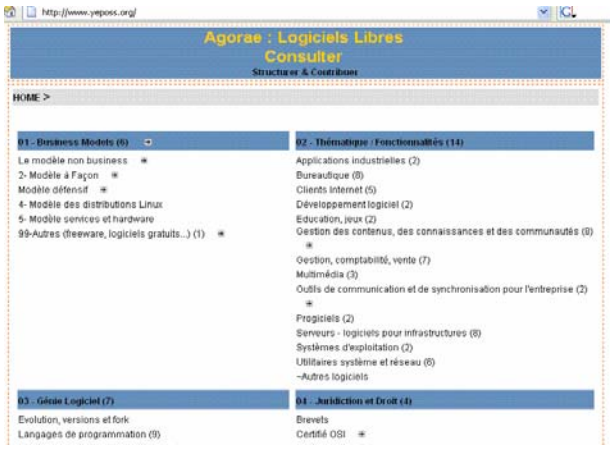

Figure 7a – OSS application, home page reading

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.

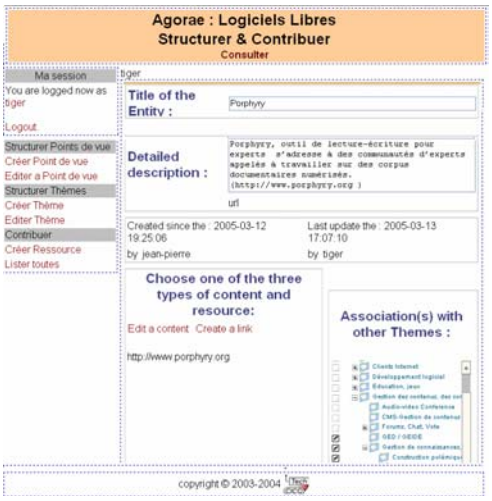

Figure 7c – OSS application, contributing

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.

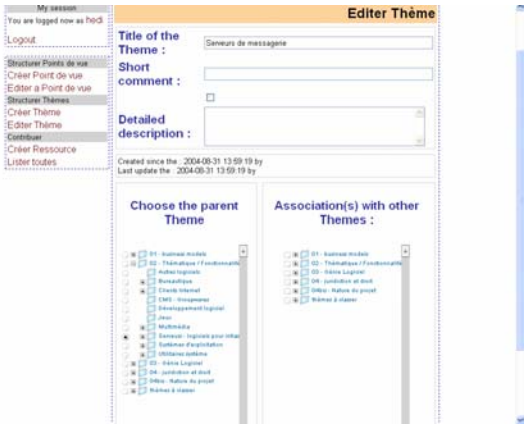

Figure 7d - OSS application, reviewing

5 Future work

At the current stage of the project, we have not definitively fulfilled our goal to "map" the OSS industry. But 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 AGORAE application.

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 runtime" of a shared meaning, and better understand the conditions that may smooth 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.

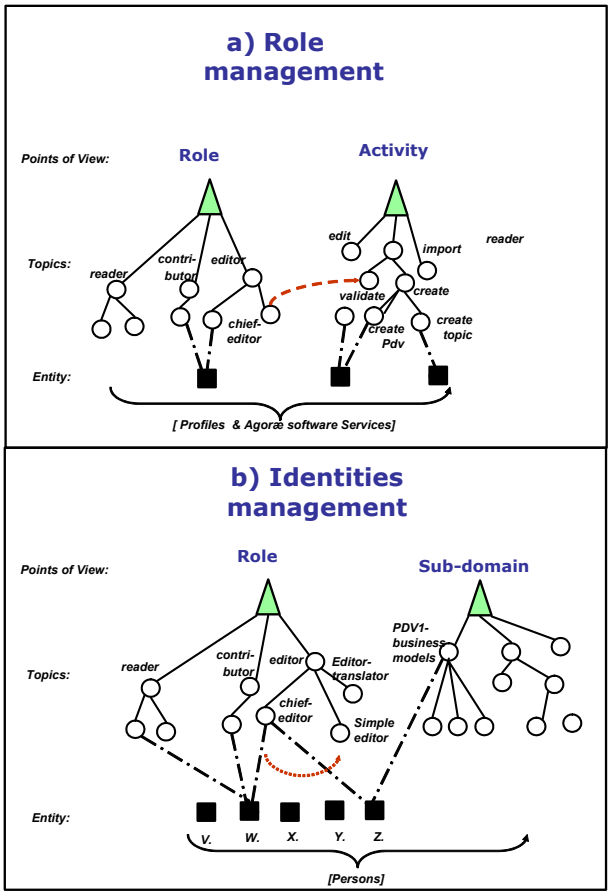

Figure 8 –Flexible edition of KBM roles (a) , of identities and rights (b), based on HyperTopic.

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 encourage 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 stay 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 observed it 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.

The HyperTopic model, that we propose as a foundation for the S2W, could be helpful at the methodological level, to take in account the relation between the growth of the information and the different knowledge of multiple actors to interpret and classify it. For example, in the OSS application, the detailed topics of the "Business model" Point of View represents at a given moment a revisable theory, which through the system is explained by the mean of categories, topics, and examples, transmitted to novices and new community members, discussed and improved at a fine-grained level by the actors themselves, etc.

According to us, the co-construction of S2W applications using co-built Topic Maps is a crucial issue for Knowledge Management. This problematic is interesting as well for its Knowledge aspects (Knowledge engineering, CSCW) as for its management dimension (organizational issues, incentives problems…). The challenge is now to find and evaluate new tools for a collective knowledge management based on S2W

approach. We have to find new ways to create a community dynamics with people interacting only through a Web site. This dynamic is essential in the process of a real self organization in a perspective of creating such S2W applications.

The experiment is still in progress so we cannot give definitive answers, but fist results are following our hypothesis, that it is possible to co-create lively semantics on the Web from virtual communities. According to us, these results will contribute to give to S2W an operational perspective in many and diverse areas. Like the Web standards, which have permit in a few years to increase strongly the number of people accessing to information, we would hope that semantic Web standards, and particularly those for S2W, will increase strongly the number of people actively contributing to knowledge building and sharing, from a simple browser. We also hope that socio semantic Web will increase the granularity and the quality of the knowledge maps lively managed by this way.

References

[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.

[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". extended 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

[TMO 01] TopicMaps.Org Specification, 2001. http://www.topicmaps.org]

[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. submission to IC'2005 Ingénierie des Connaissances, Nice, june 2005