A WEB-BASED ANNOTATION SYSTEM FOR IMPROVING COOPERATION IN A CARE NETWORK

MYRIAM LEWKOWICZ, GAËLLE LORTAL, AMALIA TODIRASCU, MANUEL ZACKLAD, MOHAMED-FOUED SRITI*

University of Technology of Troyes / ISTIT lab/Tech-CICO team
12, rue Marie Curie . BP 2060 . 10010 Troyes Cedex - France
*E-mail : {surname.name}@utt.fr, *sritimoh@utt.fr*

Coming from the needs of a care network for a cooperating system enabling its members to share information and argue cases, this paper presents a work in progress which aim is to propose a Web-Based system using annotation for cooperation through an electronic patient file (EPF). We first describe the EPF, and then we focus on our socio-semantic web positioning, on existing web standards for annotation, and on requirements for the EPF annotation tool. Finally, we present a distributed architecture for this web-based application.

1 Introduction

A distributed care network is an organizational device enabling a group of people to work together in order to improve care on a particular disease. The one studied here was created in order to deal more efficiently with the memory disorders of elderly people living in a restricted geographical area. It brought together various actors, such as health professionals (family doctors, neurologists, geriatricians, speech therapists, nurses, etc.) and social workers. Generalist and specialist doctors are working together on cases of Alzheimer's disease during sometimes long-way off meetings. In order to complete these synchronous phases of cooperation by asynchronous phases, a technical device has been added to the organizational one : a web-based system playing the role of a boundary object [1,2,3] between all the members of the network : an Electronic Patient File (EPF).

In this paper, we first describe the existing EPF and its lacks, and then we focus on our socio-semantic web positioning, on existing web standards for annotation, and on requirements for the EPF annotation tool. Finally, we present a distributed architecture for this web-based application

2 Electronic Patient File : A Web-Based System for Information Sharing in a Care Network

The EPF takes into consideration static medical observations and activities, as well as dynamic, structured, creative pieces of information. It is structured on the medical SOAP format [5], which contains : Subjective data; new Objective data; Assessment of the new data and new Plans that are determined by the new data. Then, our EPF contains Symptoms

and clinical signs of a patient, Observations and analysis of doctors about a case, Assessment of the disease within a case, and Procedures or action plans for resolving health problems. The EPF is included in a public web site and is structured by 8 tabs pointing at dynamic page : Summary form (basic civil information, contact), Civil status record, Familial status, Social file, Medical file, Medical examination (Biological tests, Neurological tests, Cognitive tests). These data are shared by participants from specialized areas : medical (general practitioner, specialist...), social (social worker...), and administrative (secretary...). Each category is allowed to consult certain types of data and to act on certain types of data.

The EPF helps participants to work together and share data, but participants expressed the need to enlarge EPF, in order to permit less structured exchanges. Thus, analysing actors' exchanges while face-to-face meetings led to the definition of needs in annotating. Doctors need for example to discuss on analysis results, on primary diagnosis, on decisions like putting the patient in hospital leaving him/her at home... In order to permit these discussions, and then improve the cooperation into the network, we propose to add to the EPF a web-based annotation system. This way, annotations not only help in structuring and in improving information access as in classical Semantic Web approach, but also in structuring social exchanges in a Socio-semantic web orientation [4]. In order to permit these informal discussions through all the professionals of the care network, we propose to add a web-based annotation system to the EPF.

3 Proposition of a socio-semantic Web application for Annotation

3.1 Socio-Semantic Web approach

Socio-Semantic Web (2SW), as defined by Zacklad [4], is an extension of a Cognitivo-Semantic Web proposed by Caussanel [6]. The Cognitivo-Semantic Web aimed to add to Semantic Web researches the cognitive aspects of a user making a request on the web : initial design of the request, updating of the request according to the increasing knowledge amount, and the result evaluation. In order to take in account these cognitive activities, we recommended the use of more simple knowledge representation languages, as Topic Maps, rather than using formal logics. While extending this perspective, the Socio-Semantic Web (2SW) is defined beside a 'Social Web' which corresponds to all the web applications dedicated to the increase of a mutual awareness in distant interactions (newsgroups, chat, instant messenger...). The 2SW aims at supporting more structured design activities, in which interactions are based on shared documents or information by a group, pursuing common objectives, at least for a while. Considering these objectives, 2SW has to

contribute to the building of a structured representation, both of the domain and of the group.

Cooperation spaces proposed by the 2SW have then to supply communicational and documentary functions, beside the groupware's specific functionalities. But, contrary to these applications which were based on proprietary environments, 2SW is following the Open philosophy of Semantic Web. This open and distributed nature of Semantic Web, enlarged both with Cognitivo-Semantic Web and Socio-Semantic Web, seems then to offer good design principles for our application.

3.2 Existing Web standards for Annotation

Our aim to allow a user to add some comments on a patient file questions us on the anchoring and the form of this meta-information which we want to add to the original file. This problematics is tackled in the Semantic Web field of research. We are now going to describe existing tools in this field, which we could reuse and enrich for our project.

Semantic Web identifies three annotation types : simple meta-data (modification date, author...), annotations which we will call computational because they address software agents and they allow them to exploit resources in a better way [7,8,9], and annotations which we will call cognitive because they address the reader, allowing him/her to be an active actor of the Web. Applications aiming to enrich online documents, to go to a collaborative Web are not new; several tools have been developed since the early 90. These tools are generally made of several elements which permit to visualize, create, store, and search annotations, each annotation being defined by an anchor, attributes, and a body. Annotations are stored on a dedicated server and can be sorted according to their attributes, can be public or private, can be shared by a defined group. These servers contain information on the localization, the style, the content, the function of the annotations.

All these researches lead to define the W3C standard Annotea [10,11], which is based on a RDF description of annotations. The 'ZAnnot' [12] annotations' server stores the annotations in a RDF database, and the users can interact with this server via Annozilla client [13] to search an annotation, to create a new one, or to delete another. An annotation is described as a set of meta-data (its attributes), and a body. The meta-data are defined by a RDF schema and give information as the author of the annotation, its date of creation, its type (comment, question, correction, etc.), the original document, the part of document which is annotated. The advantage of this RDF notation is that it is possible to personalize it by adding new attributes and a set of values of these attributes to the schema of an annotation. This

technical solution is then interesting for our project, because, as we will expose in Sec. 3.3, we wish to index the annotations according to several points of view. These points of view are different from existing Annotea attributes, and are related to a cognitive use of the annotations in the care network.

We are now going to present requirements for a Socio-Semantic web-based application supporting cooperation around the EPF.

3.3 Requirements for a Socio-Semantic web-based annotation tool

With Zacklad [14], we define annotation as a located meta-data, bound to a document. An annotation is a track of a collaboration process, which has two main functions : Planning (project management, micro-organization), and Reviewing. Meta-data proposed by existing web standards (as Annotea which we described above) for indexing annotations (Author name, Date, Subject, Annotation Type), are not sufficient to take into account the role of annotation in this collaborative process. In fact, with this kind of index, we miss the organizational context (roles, profile of the participants, etc.), the domain context (professional keywords, domain-specific terms, concepts), and argument types (proposition, opposition, etc.). In order to allow a more subtle information retrieval, we then propose to extend annotation indexing with regard to domain-specific point of view, but also to add social dimension of indexing : the arguing point of view, as well as the organizational point of view, using the actors' role to increase the importance of the decision.

Using Ontologies for Indexing Annotations

Each of the three points of view defined above is described by an ontology. From a Semantic Web perspective, ontologies should represent exhaustively domain-specific knowledge structured as a hierarchy of concepts and relations between concepts. Each concept is well defined by its properties, so the expert should completely specify these relations. Generally speaking, ontologies' coherence and consistency should be computed automatically, by specific inference mechanisms. Building such ontologies is a time-consuming and expensive task. Plus, on one hand, generic ontologies (EuroWordNet, DOLCE [15]) 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). To avoid this drawback, the Socio-Semantic Web proposes the use of less formal ontologies, which main purpose is to help the 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 [16] (TM) ISO norm are defined for these semi-formal ontologies. TM formalism defines a network of topics covering domain-specific knowledge. Topics' definitions 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. (1)). No coherence checking mechanism is used by 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 point-of-view ontologies.

In our system, the micro-organizational and argumentative ontologies 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 ontologies require a combination of Natural Language Processing (NLP) techniques and manual choices of terms and concepts. We focus now on these NLP techniques.

Using NLP for building domain-specific ontology

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) [17], to create term clusters (concepts) [18], as well as to extract relations between terms [19]. 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 first task is done off-line, by extracting terms from a selected corpus and by proposing a simple topic hierarchy (a term is equivalent as a topic). The second task is to help the user indexing his annotation regarding the three points of view (other indexes like author name, date, title, are automatic), by proposing him/her a semi-automatic indexation of his/her annotation.

For the first task, starting from text to create a semi-formal ontology from scratch, needs to use a term extractor. We use a simple repeated segment extractor, LIKES [20], to identify sequences of words occurring several times in the corpora. The repeated segments are potential candidate terms and they are tree-organized in a tree. Most of the candidate terms match a Head+Modifier pattern. We developed a tool (GenTMInd), identifying hierarchical relations between terms via heuristic rules and structuring them in Topic Maps. This way, a term matching a pattern Head + Modifier is a subconcept of Head concept. For the moment, candidate topics should be identified among simple noun phrases (a noun group followed by only one prepositional phrase). For the second task, proposing terms (keywords) to users for indexing annotation means that NLP tools should be able to scan a short text, and match it to the ontology concepts, for each viewpoint. For that point, we want to use a matching algorithm between the text, body of the annotation, and a part of a document or the Topic Maps built by the system. The next step to be implemented is to adapt a more efficient term extractor, as FASTR [21] is to identify candidate terms in annotation's body, and to extract concept hierarchy by adapting the clustering techniques [18].

We are now going to present the distributed architecture of our Web-based annotation system, integrating W3C standards, and NLP tools.

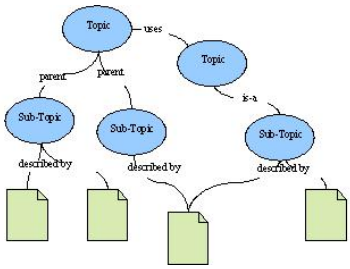

FIG. 1: Topic Map formalism

3.4 Architecture

According to Annotea W3C standard, the architecture of our annotation system is a distributed client/server architecture (Fig. (2)) :

- Client part : As we mentioned before, we need to annotate the EPF pages, which are accessible via a Web browser. Annozilla, a plug-in for Mozilla Navigator, is an Annotea client that permits to reach this aim. Using XPointer and DOM standards and a lot of Mozilla Framework facilities (XPConnect and XPCom components, see McFarlane [22]for details), Annozilla gives the user the ability to create, update and remove annotation on a document or on a part of the document, and gives the possibility to store it in a local (individual use) or in a remote server (shared use).

- Server part : we have chosen the Annotea client ZAnnot developed under Zope Platform [23] which owns an Object DB and a Web server and some of other components (ZClass, Zope Products...). ZAnnot as a Zope product takes advantage of the Zope Platform and handles all annotations queries sent by the client Annozilla ; in addition it handles and manages the reply services.

The particularity of our approach is that we have to annotate dynamic (PHP) Web pages. To achieve this, we use the EPF pages to exchange some hidden parameters between EPF Server and ZAnnot like the EPF identifier. After we adapted Annozilla to annotate the EPF, we implemented the reply ability between annotations and the indexing function. To classify annotations, we extended the Annotea Annotation Scheme, by adding some indexing metadata (the three viewpoints) which will be saved in RDF format with the other annotations metadata and body. For coherence reasons, the Topic Maps of the three points of view are actually stored also in RDF format and they are not modifiable by user.

We provide users an interface allowing him/her to manage the topics in the different points of view and browse the annotations' list through it. When a member of the care network will open an EPF, he/she will be able to open on the left-side of the screen the Annozilla plug-in which permits both to annotate and to retrieve past annotations, sorted according to the indexes which we have defined above. If he/she decides to create a new annotation, this annotation appears in a new window containing its body and the fields for indexation.

The next step is to involve the Ontology server and NLP Tool, and thus, the Topic Maps will be stored in the Ontology server and will be updated by the user

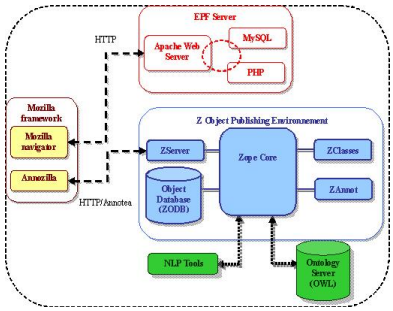

FIG. 2: Web-based Annotation System Architecture

or by the NLP tool.

4 Conclusion

In this article, we first described the needs of members of a care network to co-operate around information laid in an Electronic Patient File. In a medical context where data is strongly coded and fixed, annotation seems to be a natural way of allowing arguing and collaborative work. We have then presented a web-based tool to support collaboration through annotation. We propose to take into account the working context of participants by indexing annotations according to domain, micro-organization structure, and argument process enabling the emerging of new solutions. Our annotation application uses the W3C Annotea Standard and NLP tools to create semi-formal ontologies which will permit to help users to index semi-automatically their annotation. The project is currently under development, and an experiment in the care network is planned to validate the annotation's model.

Références

1. Star, S. and Griesemer, J.Institutional ecology, "translations" and boundary objects : amateurs and professionals in Berkeley's museum of vertebrate zoology, 1907-1939, Social Studies of Science, 19, 387-4201989.
2. Wenger, E.Communities of Practice - Learning, Meaning and Identity, Cambridge University Press, Cambridge 1998.
3. Bannon, L.,Kuutti, K. Shifting expertises on organisational memory : from storage to active remembering, in IEEE HICSS 96, Vol. III, Information Systems - Collaboration Systems and Technology IEEE Computer Society Press, Hawaii, pp. 156-1671996.
4. Zacklad M. Barbaud X. Vers une application du Web Socio Sémantique pour la réalisation d'un système d'information destiné aux réseaux de santé, In Second séminaire francophone du Web Sémantique Médical, Rouen, 2004.
5. Weed L.D., The Problem Oriented Record as a Basic Tool in Medical Education, Patient Care and Clinical research. Ann Clin Res 3(3) :131-1341971

6. Caussanel J.,Cahier J-P.,Zacklad M., Charlet J. Les Topic Maps sont-ils un bon candidat pour l'ingénierie du Web sémantique. In Bruno Bachimont, editor, Actes des 6èmes Journées Ingénierie des Connaissances, pages 233- 52, Rouen, France, 28-30 mai 2002.
7. Bremer J.M., 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.
8. Volz R., Oberle D., Motik B., 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.
9. Roussey C., Calabretto S., PINON J.-M., SyDoM : A Multilingual Information Retrieval System for Digital Librairies in proc. International Conference ICCC/IFIP On Electronic Publishing (ELPUB'2001), Canterbury (UK), 5-7 july 2001, p. 150-164, 2001
10. Kahan J., Koivunen M.-R., Prud'Hommeaux E., Swick R.R. Annotea : an open RDF Infrastructure for Shared Web Annotations, Proceedings of WWW10, May 1-5 2001, Hong-Kong, pp. 623-632 2001.
11. http ://www.w3.org/2001/Annotea/
12. http ://www.zope.org/Members/Crouton/ZAnnot/
13. http ://annozilla.mozdev.org/
14. Zacklad, M., Lewkowicz M., Boujut, J-F., Darses, F., Détienne, F., Formes et gestion des annotations numériques collectives en ingénierie collaborative, actes des journées Ingénierie des Connaissances 2003, Laval 2003.
15. Gangemi A., Guarino N., Masolo C., Oltramari, A. Sweetening WordNet with DOLCE, AI Magazine 24(3) : Fall 2003, 13-24 2003.
16. M. Biezunski, M. Bryan, S. R. Newcomb, " Topic Maps ", ISO/IEC 13250 specification, 3 Decembre 1999.
17. Jacquemin C.,Bourigault D., Term Extraction and Automatic Indexing, in Mitkov R. (ed), The Oxford Handbook of Computational Linguistics, Oxford University Press, 2003, pp. 599-615 2003.
18. Cimiano, P. Hotho, A., Staab S. Clustering Concept Hierarchies from Text, in Proceedings of LREC'2004, Lisbon, 25-28 mai 2004, ISBN 2-9517408-1-6, pp. 1721-1724 2004.
19. Buitelaar P., Olejnik D., Hutanu M., Schutz A., Declerck T., Sintek, M. Towards Ontology Engineering Based on Linguistic Analysis, in Proceedings of LREC'2004, Lisbon, 25-28 mai 2004, ISBN 2-9517408-1-6, pp.7-11 2004.
20. Rousselot,F., Frath, P., 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.
21. 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.
22. N. McFarlane, " Rapid Application Development with Mozilla ", Bruce Perens' Open Source Series,2003.
23. A. Latteier, M. Pelletier, C. McDonough, P. Sabaini, "The Zope Book", 2.6 Edition, 2003.