

An annotation service for e-library: enhancing collaborative reading

Gaëlle Lortal¹, Myriam Lewkowicz¹, Amalia Todirascu²

¹ICD, University of Technologie of Troyes, 10010 Cedex, France (lortal@utt.fr, lewkowicz@utt.fr)

²LILPA, Marc Bloch University, 67084 Strasbourg Cedex, France (amalia.todirascu@umb.u-strasbg.fr)

ABSTRACT

Libraries are places to share and to exchange information. As well, e-libraries can facilitate information access to digital documents and then improve knowledge exchanges and social interaction between members of communities on the Web. This article presents AnT&CoW (Annotation Tool for Collaborative Work), an annotation plug-in. Its design is based on scenarios. AnT&CoW supports discussions between users of an e-library; it enables document and annotation indexation according to a community classification schemes. User interactions might result in new document creation, so the tool also supports document creation.

Keywords: Library and information science, annotation, expertise sharing, SW technologies

1. INTRODUCTION

Digital libraries facilitate information access and storage, available for a large community of users. Digital documents should be easily retrieved and archived, so large number of researches focuses on the development of efficient information retrieval techniques and optimized data storage. As stated by [21], these issues are only some aspects of the information environment. Libraries are also places for users to share and to exchange knowledge about library contents. New services are then required: communication between members, knowledge structuring and organization, contextual archiving [2] and contextualizing information exchanged around the documents. Our issue concerns the definition of e-services to enhance document reading and to permit users to share and exchange knowledge.

Texts comments are traditionally exchanged by annotation on hardcopy of texts [22]. Even if several systems provide digital documents annotation functions, they do not propose global annotation support. They support document access through metadata or they support unstructured member' discussions through posts. Our tool provides functions as Web documents annotation and annotation structuring in discussion thread by means of semi-automatic indexing.

In this article, we first present the annotation role for Knowledge Management (KM) in e-library. In the next section, we focus on the annotation role in emerging communities. Finally, we present our scenario based annotation tool for documentation and expertise sharing and re-use.

2. ANNOTATION USE FOR KNOWLEDGE MANAGEMENT IN E-LIBRARY

Annotations are active elements of document creation and become fragments of a written document. From [9], we can define several types of annotation: (1) The gloss which is a fragment of text explaining a part of a document, (2) the underlining mark pointing parts of a document, (3) the note paraphrasing the main point of a document, (4) the comment bringing new ideas out, and

(5) the discursive comment which is an arguing and organised comment built cooperatively by exchanges among authors. These annotations are firmly bound to a document or to a topic. From a Semantic Web (SW) perspective, [6] annotation enriches the document in order to improve automatic document indexing and retrieval [15]. Annotation is then used for Electronic Document Management (EDM), document content processing, document structuring, services interoperability, and for some specific types of cooperation (as in [19]'s scenarii of cooperation).

As well, annotation is an activity enabling interpretation of a document and common interpretation shared with others users. A document is considered as poor as it only contains data or information. In our framework, annotating a document means enriching it by linking an advice and innovative ideas [1]. When annotating, a user contextualizes document with a social interpretation [20]; as he/she relates an annotation to a document, he/she enriches it with traces of his/her interpretation rising from his/her experience.

Annotating also contributes to the representation of an event in the memory of each user. Individual knowledge is then created while writing and exchanging. By means of annotation, users enrich documents with individual knowledge built during reading. This knowledge is built cooperatively and shared by other members through exchanges of interpretation. SW techniques for EDM are then not suitable to merge knowledge and social experience.

Capturing annotations and structuring of annotation becomes then crucial to enable users to exchange and to understand others' interpretation. We then designed a tool supporting the annotation activity which also helps users to take part in a community by supporting interpretation exchanges between them.

3. EMERGING COMMUNITIES AND ANNOTATIONS

3.1. Collaborative exchanges between users

Annotation considered as a text fragment bound to a document or a topic supports exchanges among users. This is a very ancient activity coming from a medieval practice of text interpretation; annotations enable expertise sharing through hard copies diffused around the world. Nowadays, in a computer supported activity, as in a newsgroup or a blog, annotations could be seen as posts bound to a discussion thread. They are structured around a topic by means of rough metadata (author, time, reply to) and supported by Social Web tools [17] [10]. Several annotations tools are available to annotate documents in order to share opinions and beliefs [11], [24], [18]. These tools support the emergence of new communities (newbies, specialists) sharing their expertise and the expertise reuse.

Annotations are also well-known means for collaborative writing [5], [28]. Annotations contain several users interpretations. Once collected, they represent a basis to write a new document. The draft created by the collection of these annotations is a multi-authored text involving a community of authors.

Annotations are then means to involve several users and create a community. However, to understand and reuse these annotations, users need to keep the context of their creation by structuring them on the basis of other annotations and stored documents.

To support community emergence, our tool should enable shared interpretation through exchanges, but also support collaborative annotation structuring and indexing according to a shared classification.

3.2. Classification elaboration and sharing

Community members are linked together by a topic of interest, but also implicitly by the “sociolect” [23] (their “jargon”) they develop when they communicate and exchange information. A sociolect is the way a social group is speaking i.e. the lexicon it shares, the connotations it shares. This sociolect enables them to identify themselves as being a part of the same community. To classify their annotations, documents and topics, community members will use their shared vocabulary. This shared language is working as a “folksonomy”, a social classification [25]. Folksonomies permit to trace knowledge creation for instance in looking at the indexes which are newly created for a document collection. Some annotation tools are based on ontologies, which enable document structuring and categorization, as recommended by the Semantic Web [6]. We can quote among others [12], [13], [14], [16], [27] where annotations are used to index and enable search engines to retrieve documents. But ontologies are heavy and set once only (or nearly). They hardly permit new concepts and neology in classification as folksonomies. Folksonomies or sociolectal classification could strengthen the community.

In fact, this sociolect enables to grasp community concepts because the community is defining a classification scheme (CS) which will help them to

structure the concepts they handle [26]. Following a CS, users can index and retrieve annotations or documents according to a topic but also to an argument (to understand the benefits or drawbacks, to know pro or cons).

In a digital library with a large collection of documents, building classification resources is a tedious and time-costly activity. This acknowledgement leads us to suggest the use in our annotation tool of a Natural Language Processing (NLP) module enabling the semi-automatic extraction of terms and relations resources from texts. This NLP module will also help a user indexing her/his annotation.

Using semi-automatic classification, community members will then be able to elaborate and share their own classification based on their specific vocabulary. Such a classification is easily shared by newcomers. They easily understand the common vocabulary as they can find examples of use in messages let about documents. That is especially true when topics are non-specific ones and vocabulary is composed of everyday words. We are now going to illustrate that by an annotating scenario for e-library which is the basis of the development of our plug-in.

4. A SCENARIO BASED ANNOTATION TOOL FOR DOCUMENTATION AND EXPERTISE SHARING AND RE-USE

4.1. A scenario of an annotation service

We consider the following scenario between two researchers (Sandy and Leon) searching for documentation using the IEEE digital library and a librarian (Julia) of this library.

Sandy is looking for information about distributed data query. She is trying to find information in the IEEE Intelligent Systems magazine as she found that this magazine covers systems designers and information manager domains. She finds an interesting paper and from this one she tries to find other documents on the same topics but she cannot. Then, she decides to let a message asking for related publication. Julia finds this annotation and decides to answer it giving several other magazines compatible with Sandy’s needs. Magazines are accessed according to indexes set by librarians. Sandy has then some more papers, but it is still not subtle enough. Leon, who is a senior researcher in distributed data sharing through ontologies find Sandy’s annotation on a paper he is downloading, and decides to answer it giving an author’s name (Fig.1.). Sandy finds this answer and thanks Leon who proposes further help on the subject. They start collaborate in exchanging opinions and advices on data sharing, linking documents in the digital library via their annotations. They also index these annotations by their own terms and link them to a discussion thread. Then, they strengthen a topic in adding links as well as their social network about data sharing in discussing. They also look to the

future considering that their exchanges on the topic could be gathered on a multi-authored paper.

In a digital library framework, annotations should enable user to create a network among documents and other users. A user needs then to know what other users think about documents, to read annotations let about a document, but also to give advice to other readers (like recommendations for instance), and then be entitled to

annotate a document or answer a comment. These needs can be summarized as followed:

- Annotation creation (document creation, comment editing)
- Classification elaboration of shared knowledge
- Indexation of a document or of a fragment of document

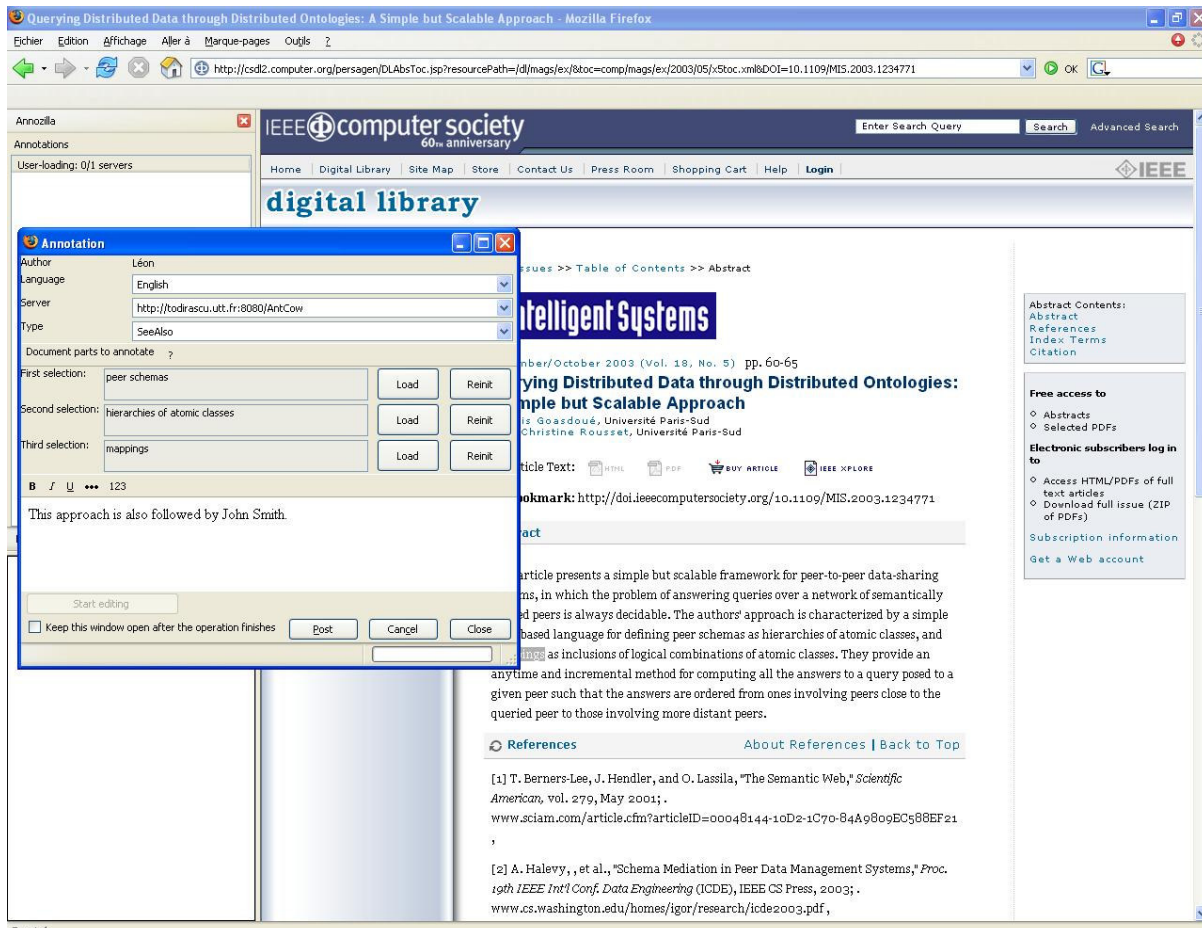


Fig.1. Multi-anchoring annotation through AnT&CoW plug-in interface

4.2. Requirements for an annotation e-service

Having described the different purposes of annotations in previous section, we can identify functions to support collective annotations of documents in an e-library. We can consider three main functionality families: communicating, browsing, and interpreting:

1. Communicating: Supporting interpretation means handling annotations as creating fragments of discourse and enabling discourse by creating threads of annotations. Functions as selection of document fragments (highlighting, circling...), anchoring discourse fragments to documents or other annotations (answering, multi-anchoring...), are then necessary.

2. Browsing: Once created, an annotation should be indexed to be easily retrieved. Browsing is based on annotation indexing. Indexation allows structuring annotations in browsable knowledge map as Topic Maps formalism [7]. To index subtly these fragments, the user should be involved. But to support the user in this time-expensive task, we suggest using Natural Language Processing tools to propose user domain specific terms and the annotation arguing type.
3. Interpreting: Thirdly, users should be able to create new documents which gather ideas emerging from collective brainstorming and exchanges around a document. Our tool should contain a gathering functionality allowing creation of a new document to work on.

4.3. An annotation plug-in for e-service

AnT&CoW [22], an annotation plug-in which is still under development, roughly implements these functionalities. We are now focusing on implementing the visualization of the classification. AnT&CoW re-uses Annozilla [4], an open-source annotation plug-in for Mozilla-Firefox browser, which follows W3C Annotea protocol [3]. When the user launches Annozilla, the plug-in appears as a frame on the left hand side of the screen (Fig.1). We improved Annozilla in order to facilitate communication. Annotations posted by several members are stored on a server and can be classified according to several viewpoints defined by community members. Annotations can be anchored by a user on a document, creating links between several fragments. The sense of the link between the original text and the annotation is explained by the argument written by the annotation's author in the annotation body. This body could also be annotated and then anchor an annotation. Multi-anchoring and replying is part of the genesis of a new document. Actually, our tool enables multi-anchoring only on one document at a time, but we will go beyond technical limits in order to enable multi-anchoring among several documents. From the note paraphrasing an external document, new ideas are brought out and form a discursive comment. The cooperative structuring of a discursive comment during discussion around a document can be merged into a new document: AnT&CoW allows the composition of a document representing a collective interpretation by gathering all the annotations related to the same topic. Gathering these fragments is done manually by choosing fragments or automatically after a keyword selection. This new document is chronologically arranged and enables visualization of argumentation by showing annotations, their indexes, and their links.

AnT&CoW is still under development but already implements basic collaborative functionalities allowing collective activity awareness [8] and collective knowledge creation through communication and indexation means.

5. CONCLUSION

When going to a library, registered members are used to speak to librarians as well as to other members. When logging on to a digital library collection, users should be able to interact with other users in order to improve their knowledge about the document collection and also to broaden their social and knowledge networks. We assume that encapsulating an annotation service as AnT&CoW in a digital library enables creating communities as well as managing knowledge. Communication facilities enable creation of a shared language used to index members production and enable them to retrieve documents. This classification is

adaptable and has the granularity needed by the community.

The next step in our iterative development is to visualize the classification in order to permit users to add and link concept they consider crucial for indexing, even if it does not appears yet in any documents.

ACKNOWLEDGEMENTS

This research carries a co-funding, from the CNRS (National Center for Scientific Research)/STIC (Communication and Information Science and Technology) department as part of TCAN (knowledge processing, learning and new information and communication technologies) pluridisciplinary project (Mediannote project) and from the European Social Fund and Champagne-Ardenne regional council.

REFERENCES

- [1] Ackerman M. and Halverson C., (2004), Sharing Expertise: The Next Step for Knowledge Management in *Social Capital and Information Technology*, Huysman M. and Wulf V. (ed)
- [2] Amar M., Hufschmitt B., Cotten J-P and Varet M.M., (2003), *Documentation et philosophie. 2 : À propos de l'indexation discursive Autour des travaux de Muriel Amar* Presses universitaires franc-comtoises, 2003 Série Philex
- [3] Annotea, (2001) <http://www.w3.org/2001/Annotea/>
- [4] Annozilla, (2004) <http://annozilla.mozdev.org/>
- [5] Bentley, R., Horstmann, T., Sikkil, K. and Trevor, J., (1995), Supporting Collaborative Information Sharing with the World Wide Web: The BSCW Shared Workspace System, in *The World Wide Web Journal: Proceedings of the 4th International WWW Conference, Issue 1, December 1995, pp 63-74*. O'Reilly
- [6] Berners-Lee T., Hendler J., and Lassila O., (2001) The Semantic Web, *Scientific American*, 2001
- [7] Biezunski, M., Bryan, M., and Newcomb, S. R., (1999) *Topic Maps*, spécification ISO/IEC 13250, 3 Décembre 1999.
- [8] Carroll J., Neale D., Isenhour P., Rosson M, McCrickard D., Notification and awareness: synchronizing task-oriented collaborative activity, in *International Journal of Human-Computer Studies*, v.58 n.5, p.605-632, 2003,
- [9] De Libera A., (2000), *La philosophie médiévale*, Paris, PUF (coll. Que sais-je ? 1044), 4e éd.
- [10] Delacroix J., *Les Wikis* M2 éditions, 2005
- [11] Denoue, L., and Vignollet, L., (2000) An annotation tool for Web browsers and its applications to information retrieval, in *proceedings of RIAO 2000*
- [12] Dingli A., (2003) Next Generation Annotation Interfaces for Adaptive Information Extraction. In *6th Annual Computer Linguists UK Colloquium (CLUK 03)*, January, 2003, Edinburgh, UK

- [13] Domingue J.B., Lanzoni M., Motta E., Vargas-Vera M., and Ciravegna F., (2002), Mnm: Ontology driven semi-automatic or automatic support for semantic markup. In *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, October.
- [14] Dzbor, M., Motta, E. and Domingue, J. B. (2004), Opening Up Magpie via Semantic Services. In *3rd ISWC*, November 2004, Japan
- [15] Handschuh S. and Staab S. (2003) Annotating of the Shallow and the Deep Web, in *Annotation for the semantic Web*, S. Handschuh, S. Staab, IOS Press p.25-45
- [16] Handschuh, S., Staab S., and Ciravegna, F., (2002), S-cream - semi-automatic creation of metadata. In *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, October 2002
- [17] Jordan K., Hauser J., and Foster S., (2003) The augmented Social Network : Building identity and trust into the next-generation Internet, in *firstmonday vol.8*, N.8, august 4th 2003 http://www.firstmonday.org/issues/issue8_8/jordan/
- [18] Ka-Ping Y. (1998), *CritLink : Better hyperlinks for the WWW*. <http://crit.org/ping/ht98.html>, 1998.
- [19] Koivunen, M-R, and Swick R.R. (2003) Collaboration through annotations in the Semantic Web in *Annotation for the semantic Web*, S. Handschuh, S. Staab, IOS Press p 46-60.
- [20] Laclavik M., Gatial E., Balogh Z., Habala O., Nguyen G., and Hluchy L., (2005), Experience Management Based on Text Notes (EMBT), in *Proc. of eChallenges 2005 Conference, Innovation and the Knowledge Economy, Volume 2, Part 1: Issues, Applications, Case Studies*; Edited by Paul Cunningham and Miriam Cunningham; IOS Press, pp.261-268.
- [21] Lagozei C., Krafft D.B., Payettei S., and Jesurogaii S., (2005), What Is a Digital Library Anymore, Anyway? Beyond Search and Access in the NSDL in *D-Lib Magazine*, November 2005, Vol. 11 N°11, ISSN 1082-9873 <http://www.dlib.org/dlib/november05/lagoze/11lagoze.html>
- [22] Lortal G, Lewkowicz M., and Todirascu-Courtier A., (2005), AnT&CoW, a tool supporting collective interpretation of documents through annotation and indexation, in Dieng R., & Matta N., in *Proceedings of KMOM-IJCAI Workshop* p.43-54.
- [23] Marcellesi, J.-B. and Gardin B. (1974), *Introduction à la sociolinguistique : la linguistique sociale*, Paris, Larousse, coll. « Langue et langage »
- [24] Price, M., Schilit, B., and Golovchinsky, G. (1998), 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.
- [25] Shirky, C., (2005), *Ontology is Overrated: Categories, Links, and Tags* http://www.shirky.com/writings/ontology_overrated.html
- [26] Simone C. and Sarini M., Adaptability of Classification Schemes in Cooperation: What does it mean?, in *Proceedings of ECSCW 2001*, 19-38, 2001.
- [27] Sumner T., Buckingham Shum S., Wright M., Bonnardel N., Piolat A. and Chevalier A. (2000), Re-designing 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.
- [28] Sun C., Jia X., Zhang Y., and Yang Y., (1997), REDUCE: a prototypical cooperative editing system, in *Proceedings of the 7th International Conference on Human-Computer Interaction* , pp.89-92, San Francisco, USA, Aug. 24-30, 1997