

# Annotations: a way to capture experience

Gaëlle LORTAL<sup>1</sup>, Myriam LEWKOWICZ<sup>1</sup>, Amalia TODIRASCU-COURTIER<sup>2</sup>

<sup>1</sup> CNRS – ICD Tech-CICO, Université de Technologie de Troyes,  
12, rue Marie Curie 10010 Troyes Cedex, France  
{lortal, lewkowicz}@utt.fr

<sup>2</sup> Linguistic, Languages and Speech - EA 1339, University Marc Bloch,  
22, rue Descartes 67084 Strasbourg Cedex, France  
amalia.todirascu@umb.u-strasbg.fr

**Abstract.** This paper describes a cooperative tool used by a mechanical engineering team carrying out an asynchronous and distributed work. The tool supports communication phases as well as document and product creation phases by means of annotations. Memorizing all the annotations and creating documents from these annotations enable the team members to capture their experience. Annotation is here defined as a continuum fulfilling several purposes from communication and argumentation to indexation.

## 1 Introduction

Distributed collective practices involve multiple actors sharing multiple documents. During a project, members handle documents such as texts, plans, models, specifications. During asynchronous phases of work, they exchange comments, they modify documents and they can negotiate some parts of new documents being created. As they have these activities of interpreting and reviewing documents, they build up their experience. To support these activities, annotation is an interesting tool. Annotation enables communication and argumentation, and might facilitate retrieval of past experiences acquired during a project. We assume that annotation enables the team members to capture the experience developed in asynchronous work. In an experience management framework, we aim at capturing and collecting experience through annotation gathering [27].

Annotations are traces of experience linked to a document. Once captured and structured according to a classification scheme, they can easily be retrieved by project members [22]. Annotations also contain member's expertise and enrich the document with traces of experience. The enriched document is then a matter for Knowledge Management (KM) techniques. In sharing their expertise via annotations, project members create new intellectual capital [8]. This social capital is built by shared interpretations of documents. We can support project members during asynchronous phases of work by a groupware capturing traces of experience.

We expose here the behaviour of a mechanical engineering team carrying out an asynchronous and distributed work, from preliminary design phase to production phase. Members of this team discuss design details based on project documents such

as description and plans, and they send newly-created documents to the others. On the basis of a distributed collective practices analysis, we describe a scenario used to design a computer-based annotation tool. We classified activities of our scenario, and for each, we describe a function that can support it and how we aim at implementing this function or reuse an existing annotation tools function. Our tool, which is capturing the experience, enables activity awareness [14] and collective sense making [28].

## **2 Annotation to capture experience**

Annotations are active elements of document creation and parts 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 out new ideas, and (5) the discursive comment which is an arguing and organised comment built cooperatively on exchanges among authors. These fragments are firmly linked to a document or to a topic. This relational feature of annotation is the one allowing Semantic Web (SW) [6] to consider annotation as enriching the document in order to improve automatic document indexing and retrieval [16]. Annotation is then used for Electronic Document Management (EDM), document content processing, structuring a document, services interoperability, and for some specific cooperation types (as in [21]'s scenarii of cooperation).

Annotation is also an activity enabling document interpretation and interpretation-sharing with other project members. The document is considered as poor, since it only contains data and information. In our framework, annotating a document means enriching it by the advices and innovative ideas of a member, i.e. a project context [1]. As a member links an annotation to a document, he/she enriches it with traces of his/her interpretation rising from his/her experience.

Annotating also helps forming representation of an event. Individual experience is then built while writing and exchanging. By means of annotation, project members built their experience and let traces of their expertise within project documents. They enrich them with individual knowledge built during the project. This knowledge is built cooperatively and shared by other members through exchanges of interpretation. SW techniques for EDM are not then subtle enough to manage knowledge and social experience.

Capturing annotations and tracing experience through annotation then becomes crucial in memorizing an experience built cooperatively during asynchronous phases of work. As we aim at memorizing this experience, we designed a tool-supporting annotation activity in order to assist collective sense making [28]. This tool is based on the observation of a mechanical engineering project presented in the following section.

### 3 Scenario-based requirements

The scenario is within the context of a project involving an association committee (AC) (mainly plane pilots), and a mechanical engineering team (divided into a Design Team -DT- and a technical team -TT-). The aim is to reuse a car-engine as an aero-engine. Team members participate in the project during their spare time. So, despite members' co-location, the work is asynchronous. Each DT member develops a part of the engine. Mechanical engineering design is a highly collaborative field and consequently, the design team needs perpetual feedback from TT (technicians and suppliers). DT then relays feedbacks from the AC managing expenses. Drawings and documents represent discussion basis between TT and DT about technical feasibility and about tool or material availability. TT mainly manages communication with suppliers. Within the DT, communication is widespread even if information should always be forwarded to the group supervisor. The only medium available in this team to mediate this asynchronous written communication is email. Email is used as a comment coming with a document as in these following descriptions:

1. A member creates a first draft with computer-aided design tools, each member then discusses this draft, asking for complement or verifying calculus. To do so, team members write down their ideas and questions as comments anchored to the draft(s). Each member puts his/her annotation on any documents, but a subtle micro-organization structure within the team involves some ethical rules. After the revision and the updating process, detailed plans are then communicated to the TT in order to begin production phases. TT receives drawing plans and their comments, which explain how the DT made a decision.
2. These annotations are often merged as new documents following an explaining purpose (for example by listing and explaining in an email, all the modifications which are visible in the document).
3. Annotations are identified as elements of solutions, of negotiations and of discussions, helping design process retrieval. As our teams need to share and manage an important number of documents, an adapted classification should be set up, based on several pieces of information: author, date, content, aim, recipient, which part of the document/plan it comments, etc. Since it is an innovative project, they can only partly define an a priori classification for documents to be produced. Their classification is possibly extended as work progresses. They try to organize documents and annotations by storing their mail by date, title and author. Here they need a tool to support their classification and share their storage.

This situation description shows us that indexation and communication functionalities are crucial to understand on which basis a solution has been adopted. It is also crucial for experience building. Annotation is used for communication, indexation and is a part of the experience elaborated during the project. We are now going to look at existing annotation tools which could provide necessary functions to support these different activities.

### 3.1 Supporting Communication

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

Annotation for communication is partly supported by the SoW (Social Web) approach [19], enabling mediated communication by comments. Newsgroup, blog, or wiki [10] enable on-line “polylogue” [24] (dialogue among more than two speakers) by publishing messages related to a discussion thread. Conversation fragments are not yet structured, which leads to topic digression [24] and decay phenomena [18]. Annotation as a comment is supported by annotation clients ([11], [20], [25]) which sort annotation on rudimentary metadata (creation date, author). Annotation could be stored apart on annotation servers, so, differentiated from the document [29] or not [2]. However, these tools do not allow connecting annotations nor structuring exchanges between users about a document. D3E [26] also considers documents as discourse media but does not allow a rich indexation of annotations. These tools do not focus on tracing the design rationale underlying the discussion enabling an *a posteriori* understanding of exchanges which took place.

### 3.2 Supporting Indexation

Once created, an annotation should be indexed so as to be easily retrieved. Browsing is based on annotation indexing. Indexation allows structuring annotations in a browsable knowledge map as Topic Maps formalism allows [7].

Several annotation clients [23] are available from Semantic Web initiatives (SW). “Semantic annotation” of the SW is “computational annotation”. An objective of SW is to index Web pages, and allow search engines a better information recall structuring annotation by underlying ontologies ([17], [12], [13]). These annotations lightly support readers of a page to cooperate or to interact. Magpie [15] uses annotations to support human interpretation and enables multiple viewpoints indexation. These tools enable additional annotation to a document, which permits the DT to explain the TT when and on which document a decision was made. But these annotations only help users to structure or to share her/his understanding of the text. To index subtly these fragments, users should be involved. But to support users in this time-consuming task, we suggest using Natural Language Processing tools to offer users domain specific terms and annotation arguing types.

As we have already claimed, we do not only need annotations to index but also to negotiate. Our purpose is to support document interpretation or recall of an existing interpretation but also to support the creation of new ideas (from collective interpretation). So, annotation is crucial to index and then to capture experience, but we also need a tool enabling exchanges via annotations i.e. enabling discourse around a document.

### 3.3 Supporting Interpretation

Thirdly, users should be able to create new documents which gather together 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.

DT needs to create documents collaboratively i.e. to re-use, re-structure and rewrite existing comments to build a draft on which members can work or to build a *final* document. As we have shown above, two families of annotation tools are available: one focuses on Web pages indexation, while the other focuses on human communication through comments. We can deplore the lack of annotation management or the poverty of cooperative functions in these two families. [26], [15] are the first steps in linking these two points of view. We thus propose to design a tool combining functions supporting SoW and SW activities (i.e. to comment, to answer, to organize).

## 4. Ant&Cow

AnT&CoW [23], an annotation tool which is still under development, roughly implements these functionalities. It re-uses Annozilla [4], an open-source annotation plug-in for the Mozilla-Firefox browser, which follows W3C Annotea protocol [3]. We improved Annozilla in order to facilitate communication and collecting experience. We consider that, in so doing, we are improving activity awareness and expertise-sharing. When a user launches Annozilla, it appears as a frame on the left of the screen. Annotations posted by several members are stored on a server and can be classified according to several viewpoints defined by the group. The user chooses a part of a document to anchor his/her annotations, creating links between several fragments. The link is explained by the argument written by the annotation's author in the annotation body. This body could also be annotated and then it can anchor an annotation. Multi-anchoring and replying could be the genesis of a new document. From the note paraphrasing an external document, new ideas are brought out, and form a discursive comment. The cooperative structuring of this discursive comment during discussion around a document can form a new document. For the moment, our tool enables multi-anchoring only on one document, but we hope to go beyond technical limits in order to enable multi-anchoring among several documents. AnT&CoW allows also the composition of a document representing a collective interpretation. It is called "generate draft" as the new document built is a gathering of linked fragments. Gathering these fragments is done manually by choosing fragments or automatically after a keyword selection. Gathering different fragments written by several authors, means gathering members' viewpoints in one document. This new document is chronologically arranged and represents visualization of experience by showing annotations, their indexes, and their links (see rings in fig.3).

AnT&CoW is still under development but already implements basic collaborative functionalities allowing collective activity awareness, collective sense making and expertise-sharing through communication and indexation means.

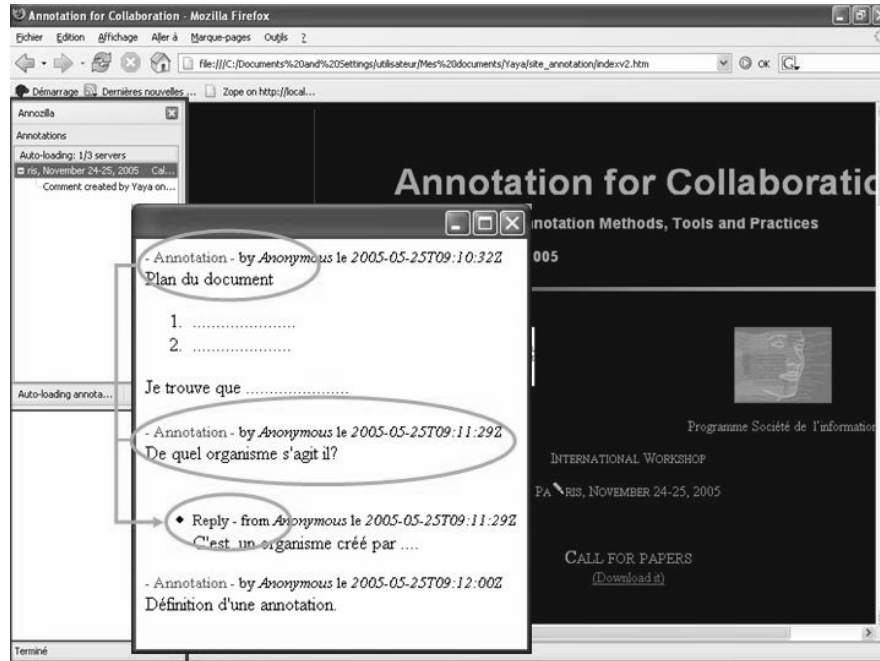


Fig.3 AnT&CoW generating draft functionality

## 5. Conclusion

We have defined an annotation as an interpretation enabling project members to share their expertise. Annotation is an experience contextualisation which enriches the document content with knowledge. This knowledge is built cooperatively among members during asynchronous work phases and should be collected and stored in order to be shared and re-used. To understand this activity in order to design a tool-supporting annotation activity, we studied a mechanical engineering project. We brought to light three main families of functionalities, communicating, browsing and interpreting. Interpreting enables annotation gathering in order to create a new document. This document represents an interpretation, or in other words a discussion with its arguments. Annotation is then a central element to trace the individual experience as well as the project experience in asynchronous or distributed projects. In order to support this activity, an annotation tool is being developed. It enables users to annotate existing documents and to produce new documents based on annotations which have been correctly indexed.

Now, we aim at improving this indexation following [1]'s approach. We try to combine repositories with social network by means of visualization of experience following three viewpoints; a social-cognitive dimension (the shared language of the

group), a relational dimension (negotiation and argumentation) and a structural dimension (micro-organisation roles and actors). These viewpoints are supported by several technical means and among them a fine-grained indexation based on Natural Language Processing tools.

## Acknowledgements

This research carries a funding from the National Center for Scientific Research /Communication and Information Science and Technology department as part of TCAN 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. Acrobat PDF, (2004) <http://www.adobe.com/support/techdocs/ac76.htm>
3. Annotea, (2001) <http://www.w3.org/2001/Annotea/>
4. Annozilla, (2004) <http://annozilla.mozdev.org/>
5. Bergmann R., (2002), Experience Management: Foundations, Development Methodology, and Internet-Based Applications (Lecture Notes in Artificial Intelligence)
6. Berners-Lee T., Hendler J., et Lassila O. The Semantic Web, *Scientific American*, 2001
7. Biezunski, M., Bryan, M., et Newcomb, S. R., (1999) « *Topic Maps* », spécification ISO/IEC 13250, 3 Décembre 1999.
8. Bresnan M., Edelman L., Newell S., Scarbrough H. and Swan J., (2004), The Impact of Social Capital on Project-Based Learning, in *Social Capital and Information Technology*, Huysman M. and Wulf V. (ed),
9. De Libera A., (2000), *La philosophie médiévale*, Paris, PUF (« Que sais-je ? » 1044), 4<sup>e</sup> éd.
10. Delacroix J., *Les Wikis* M2 éditions, 2005
11. Denoue, L., et 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 *6<sup>th</sup> Annual Computer Linguists UK Colloquium (CLUK 03)*, January, 2003, Edinburgh, UK
13. Domingue J.B., Lanzoni M., Motta E., Vargas-Vera M., et Ciravegna F., (2002), Mnm: Ontology driven semi-automatic or automatic support for semantic markup. In *13<sup>th</sup> International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, October.
14. Dourish, Paul, & Belotti, Victoria. (1992). Awareness and coordination in shared workspaces. *Computer Supported Cooperative Work 92 Proceedings*, (pp. 107-115) New York: Association for Computing Machinery.
15. Dzbor, M. - Motta, E. - Domingue, J. B. (2004), Opening Up Magpie via Semantic Services. In *3<sup>rd</sup> ISWC*, November 2004, Japan
16. 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

17. Handschuh, S., Staab S., et Ciravegna, F., (2002), S-cream - semi-automatic creation of metadata. In *13<sup>th</sup> International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, October 2002
18. Herring, S. (1999). Coherence in CMC. *Journal of Computer-Mediated Communication*, 4(4). Retrieved January 12, 2005, from <http://jcmc.indiana.edu/vol4/issue4/herring.html>
19. Jordan K., Hauser J., Foster S., (2003) The augmented Social Network : Building identity and trust into the next-generation Internet, in *firstmonday* vol.8, N.8, august 4<sup>th</sup> 2003 [http://www.firstmonday.org/issues/issue8\\_8/jordan/](http://www.firstmonday.org/issues/issue8_8/jordan/)
20. Ka-Ping Y. (1998), *CritLink : Better hyperlinks for the WWW*. <http://crit.org/ping/ht98.html>, 1998.
21. 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.
22. Laclavik M., Gatial E., Balogh Z., Habala O., Nguyen G., Hluchy L., (2005), Experience Management Based on Text Notes (EMBET), in Proc. of eChallenges 2005 Conference, Slovenia, 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.
23. Lortal G., Lewkowicz M., Todirascu-Courtier A. (2005a). AnT&CoW, a tool supporting collective interpretation of documents through annotation and indexation, in Dieng R., & Matta N., *Proceedings of KMOM-IJCAI Workshop* p.43-54.
24. Marcoccia, Michel. On-line Polylogues : conversation structure and participation framework in Internet Newsgroups. *Journal Of Pragmatics*, 2004, vol.36 n°1, p.115-145.
25. Price, M., Schilit, B., et 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.
26. Sumner T., Buckingham Shum S., Wright M., Bonnardel N., Piolat A. & 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.
27. Sun Z., Finnie G., (2005) Experience Management In Knowledge Management, In: Khosla R, Howlett RJ, Jain LC (eds) Knowledge-Based Intelligent Information and Engineering Systems: 9th Intl Conf KES2005, Melbourne, Australia, September 14-16, 2005, Proc, Part I, LNAI 3681, Springer-Verlag, Berlin Heidelberg 2005, pp. 979-986
28. Weick, K.E., (1979) *The Social Psychology of organizing*, New York, Random House.
29. Windows Word (2003), <http://office.microsoft.com/fr-fr/assistance/HA010714941036.aspx>