Semantic Annotation Tools for Learning Material

Faical Azouaou[1], Weiqin Chen[2] and Cyrille Desmoulins[1]

[1] CLIPS-IMAG & Joseph Fourier University
BP 53, 38041 Grenoble cedex, France
{Faical.Azouaou, Cyrille.Desmoulins}@imag.fr

[2] Department of Information Science and Media Studies, University of Bergen
PostBox 7800, 5020 Bergen, Norway
weiqin.chen@infomedia.uib.no

Abstract. This paper aims at providing the specification for semantic annotation tools for e-learning. From the specific requirements of annotating learning material, we categorize and evaluate the existing annotation tools, mainly general purpose ones. We illustrate two research prototypes of annotation tools we developed, and evaluate to what extend the specific requirements of annotating learning material are reached by these research prototypes.

1 Introduction

Many uses of annotations and metadata on learning material have been described, in ecological use reports or in research project, in various contexts and various roles [1, 2]. However before people or software agents can use them, such annotations of learning material have to be created, automatically or manually. Currently few tools exist dedicated to this particular task of annotating learning material.

This paper[1] aims at explaining the specificity of annotating learning material and providing specifications for automated and manual annotation tools for e-learning. To come to such specification, we start from two different viewpoints. The first is the specific requirements that the e-learning context brings for annotation tools. The second is a review of existing annotation tools, mainly general purpose ones. As most of these tools are quite similar, we analyse their characteristic properties and categorise them to three most important factors as regard to semantic web and e-learning. We then evaluate the strength and weakness of each category regarding the requirements we have specified for annotating learning material.

We further illustrate on two examples of annotation tools we have developed in France and in Norway. We demonstrate how it is possible to define the functionalities of annotation tools for a specific use taking into account our requirements and adapting functionalities of general purpose tools of the same category.

[1] This work has been partially funded by the Kaleidoscope network of Excellence. EU IST Technology Enhanced Learning (TEL) project 507838.

2 Requirements for e-learning annotation tools

General annotation tools usually provide domain-independent annotation supports. They are designed to fulfill the general requirements such as ease of use, efficiency, etc.[3]. However, these tools do not take into considerations of special requirements for special domains. For example, in the context of e-learning, the annotation of learning material has different requirements. Below we list the requirements for e-learning annotations tools:

Usefulness: takes into account teaching/learning context
- 1.1. Teaching/learning domain (topics to be taught).
- 1.2. Teaching/learning objectives and the addressee of the annotation.
- 1.3. Teaching/learning activities (exercise, lab work, lesson, field studies, etc.).

Shareability: enables teaching/learning actors to communicate through annotation.
- 2.1. With an explicit semantic related to the teaching/learning context.
- 2.2. By complying with e-learning standards (LOM, IMS-LD, etc.).
- 2.3. By the means of the visual form of the annotation are used to.
- 2.4. By enabling to share annotation with others in the same e-learning context

Usability
- 3.1. Annotation made manually does not disturb teaching/learning activities
- 3.2. Annotators are put in their usual teaching/learning context while annotating.

3 Characterizing and evaluating existing annotation tools

In this section we discuss the different definitions of annotation in various contexts. We further review existing tools, mainly general purpose ones, based on the requirements of annotating learning material.

As there are many annotation tools and most of them are quite similar, our method is to extract properties characterising them. Focusing on the three most important ones as regard to semantic web and e-learning, we obtain a reduced number of categories on which we can situate each annotation tool.

3.1 Annotation definition and properties of annotation tools

According to the Merriam-Webster on-line dictionary [4], an annotation is *a note added by way of comment or explanation or the act of annotating.* This definition, as many definitions from research literature, specifies that an annotation is both an object added to a document and the activity that produces this object. This twofold view on annotation is also reflected in the formal definition we present hereafter.

Euzenat [5] formalized semantic annotation in the context of the Semantic Web. From two sets of objects, documents and formal representations, two functions can be created: a function from document to formal representations, called *annotation* and a function from formal representations to documents called *index*. The corresponding

activities are annotation and indexing. So, we can also formalize non-semantic annotation as a function from documents to non-formal representation, and the activity to create this function.

To extract properties characterising annotation tools, we studied the annotation activity and what characterises it. We established that the annotation activity on a computer depends on three main factors:

- The author of the annotation (the annotator).
- The addressee of the annotation (the user of the annotation).
- The fact that the annotation is semantic or not (see previous section).

These three factors provided us four properties of annotation tools:

Automatic versus manual annotation: Annotating is the process that creates a function from a document to a representation, formal or not formal, creating such a function involves three sub-processes. To choose a document or a part of document to be annotated (source); to choose the element of representation that is the result of the function (target) and finally to define the properties of the function itself. Consequently, automatic annotation means that the three annotation sub-processes are performed automatically by a software agent; manual annotation means that they are performed by a human agent, even if he/she uses software tools for that and semi-automatic annotation means that the human agent is helped by the software tools to perform at least one of the three annotation sub-processes.

Cognitive versus non cognitive annotation: Two properties describe the annotation addressee. The first is the cognitive aspect of the annotation, representing whether annotation can be handled by human, in this case, annotation has a visible shape, we call it "cognitive annotation" [6].

Computational versus non computational: The second aspect describing the annotation addressee is whether the annotation is aimed to be used by a software agent (computational) or no (non computational).

Semantic versus non semantic annotation: The third factor characterising annotation activity on a computer is the fact that it has an explicit semantics for the computer, and not only for the human that created it or handle it.

3.3 Evaluating annotation tools

The three main factors provided us with four dimensions to group and evaluate annotation tools:

- The author: automatic, manual or semi-automatic annotation.
- The addressee: cognitive versus non cognitive annotation and computational versus non computational annotation
- Computational semantics: explicit semantics for the computer.

The combination of these four dimensions makes a table of 24 cells. Each annotation tool can be categorized by one of the cells.

A second table shows in each cell of the table to what extends each of the three requirements in Section 2 are reached. "R" indicates realized requirements and "P" possibly realized requirements, which means that users could use the tool to somehow reach the requirements although the tool does not realize the requirement.

Table 1. Existing annotation tools by categories

Semantics	Author / Adressee	Manual	Semi-automatic	Automatic
Non semantic annotation	Cognitive and non computational annotation	Imarkup, Acrobat , Web-Notes , CoNote, WebAnn, Epost		Google's ToolBar
	Non cognitive and computational annotation	Manual index in libraries	MyAlbum Annotate	Google search engine
	Cognitive and computational annotation	Knowledge Pump, Xlibris		Cached Google Links
Semantic annotation	Cognitive and non computational annotation	Annotea + Amaya, Yawas [7], ThirdVoice Mark-Up		
	Non cognitive and computational annotation	Edutella, OntOmat, SHOE, HTML-A, WebKB, Karina		AeroDAML
	Cognitive and computational annotation	Mangrove, SMORE	MnM,Melita, Teknowledge, IMAT	KIM, MnM, Magpie, COHSE

Table 2. Evaluation of existing tools based on requirements for annotating learning material

Semantics	Author / Adressee	Manual	Semi-automatic	Automatic
Non semantic annotation	Cognitive and non computational annotation	P: 2.3 R: 3.1		R: 3.1
	Non cognitive and computational annotation	Nothing	P: 2.2 R: 3.1	Nothing
	Cognitive and computational annotation	R: 2.4, 3.1		Nothing
Semantic annotation	Cognitive and non computational annotation	P: 1.1 1.2 1.3 2.1 2.2 2.4 3.2		
	Non cognitive and computational annotation	P: 1.1 1.2 1.3 2.1 3.2 R or P: 2.2, 2.4		P: 1.1 1.2 1.3 2.1 2.4 3.2 R or P: 2.2
	Cognitive and computational annotation	P: 1.1 1.2 1.3 2.1 2.2 2.4 3.2	P: 1.1, 1.2, 1.3, 2.2 2.4 R or P: 2.1, 3.1	P: 1.1, 1.2, 1.3, 2.2 2.4 R or P: 2.1, 3.1

This evaluation table points out the following interesting results:

- All the non semantic cognitive tools realize the 3.1 requirement (does not disturb the activity) but it is not the case for semantic tools.
- Some non cognitive computational semantic tools already use the e-learning standards (mainly LOM).
- Very few other e-learning requirements are currently respected but some could be reached with an adaptation of semantic tools: usefulness, shareability and usability concerning teaching context (2.1, 2.4, 3.2).
- The 2.1 and 3.1 requirements are yet respected by some tools that provide annotation with ontologies of teaching topics.

4 Research annotation tools

As we explained in section 3, annotation tools depend on the use of the annotation, both the creating use (the means provided to the annotator) and the annotation end-user (its addressee). Therefore to specify an annotation tool dedicated to e-learning means to clarify to which one of the 18 tools categories it belongs to, describe the specificities of the learning context and specify the functionalities provided by the tool to its users. We illustrate this method with two research tools dedicated to annotating learning material.

4.1. MemoNote

MemoNote is an annotation tool developed at the CLIPS laboratory (Grenoble). Although many learning and training activities are now supported by e-learning systems, users have usually no means (or very poor means) to manage the note of events and knowledge they want to memorize during these activities and to retrieve in the future. The MemoNote project aims at formalizing and implementing computerized external memories made of notes added directly and voluntary on the training material by its user.

MemoNote is cognitive, semantic and manual or semi-automatic annotation tool. It enables the user to annotate pedagogical documents. For a specific teaching activity, MemoNote can adapt the user's context by selecting a set of ontologies which describes the users, the teaching domain, the pedagogical activities (content, location, time) and the pedagogical objectives.

This ability to change its context with a set of ontologies makes MemoNote both a generic tool, which can be used in every context, and a specific one, once the context is fixed by ontologies.

The user has two annotation means:

- Manual annotation. The user himself/herself must define the three facets.
- Semi-automatic annotation. The annotator defines the source of the annotation by selecting an annotation tool and the annotation anchor. An annotation pattern is attached to each tool enabling MemoNote to deduce partly or entirely the semantic and episodic facets

The user interface in both cases is the same. It has three main parts. The first part is a reader (reading software) embedding MemoNote annotation tools. It provides reading facilities quite similar to paper ones. In this reading interface, the user can choose an annotation tool (for example red underlining) and put it on the document surface (on the touch screen). The second part is the annotation interface where the user can define (or not) each semantic fields (addressee, objective, content, importance and confidence). The third part is the ontology browsing interface. For each attribute, the user want to define, this interface pops up until the ontological value of the field is fixed. For some entirely automatic patterns, the interface for annotation and ontology browsing does not open and fields are filled in automatically.

4.2. AnnForum

AnnForum is an annotation tool developed at the University of Bergen (Norway) to support the annotation and reuse of collaborative knowledge building forum as new learning resources. According to [5], annotation is not always productive if it hasn't been designed in close relation to its use, it will produce limited benefits. AnnForum is a computational, cognitive, semantic, manual and semi-automatic annotation tool.

FLE3 [8] is a web-based groupware for computer supported collaborative learning (CSCL), which is used in a university course INFO281 (Introductory Artificial Intelligence). It is based on progressive inquiry learning, a type of activity where students engage in a research-like process by posting messages to categories (problem, hypothesis, scientific material, etc). There is a large amount of messages posted in each semester on FLE3. With AnnForum, by reusing FLE3 as new learning resources, future students can benefit from former students' knowledge and experiences.

A conceptual domain model is used in AnnForum to describe the domain concepts (Artificial Intelligence) and the relationships among them, which collectively describe the domain space. Once the conceptual domain model is available, annotations can be created by the teachers linking previous knowledge building to elements of this model. To support teachers in creating such annotations, we designed a keyword recognizer and an algorithm to determine the relevance of a message to a concept in the domain model. The keyword recognizer identifies the occurrence of the topics, including their names and variants of the names in the domain model. Relevance is determined using an algorithm that applies a weight to the keywords in the messages. The annotation of the messages from the system is then shown to the teacher who can add or remove the related topics on the interface and then elaborates the annotation manually. The semantic information added into the forum enables the reusing facility to detect messages and teaching material from the previous knowledge building which are relevant to current discussion topics and present them to the students.

4.3 Evaluation of Memo Note and AnnForum against the requirements

MemoNote tool as a cognitive, non computational and semantic annotation tools, respects the 2.2 requirement using OWL to represent ontologies and RDF to represent annotations. The 1.1, 2.1 and 3.2 requirements that were potential for general purpose tools are respected by MemoNote. This is made using the set of ontologies defining the annotation context and from which values are taken to specify the content of an annotation. Currently the 2.4 requirement is not yet implemented. We have started to formalize how a group could share manual annotations and create a collective manual annotation.

AnnForum allows teachers to create a domain model (1.1). This model has an explicit semantic network to support the annotation (2.1). Teachers can make annotation with teaching objectives in mind (1.2). The annotation can be created manually or semi-automatically. The annotation process does not disturb the teaching activities (3.1). By adding a learning model that complies with LOM or IMS-LD, it is possible to use AnnForum to annotate teaching/learning activities (1.3). Since the annotation is based on explicit semantics, it is also possible for teachers to share the annotations.

Conclusion

In this paper we first presented the specific requirements of annotating learning material. Based on these requirements, we categorized and evaluated the existing annotation tools. We have also presented two annotation tools which are under development particularly for learning material.

Although each of these annotation tools fulfills some of the requirements for learning material annotation, there are still some problems that need further investigation.

First, the requirements we presented might not cover all the requirements. Teachers, learners, and other actors may have their own needs when they annotate learning material. We should look more into the special requirements from different parties.

Second, in order to take into consideration teaching/learning context, annotation tools should be able to combine domain and teaching/learning ontologies.

Finally the categorization we provided is a first mean, for a particular use, to illustrate what new directions research should be followed. The tools respecting the most of the requirements are computational, cognitive and semantic, meaning that the promising direction could be that the user can let the software compute inferences for him. It means to make MemoNote also a computational tool with which the annotator would be able to create annotations that will automatically remind him/her annotations at a certain time in the future, depending on its current learning/teaching task. For AnnForum, the emphasis will be on the semantic use of e-learning standards (LOM, IMS-LD) in order to be able to annotate teaching/learning activities and support the share and reuse of the annotations.

References

1. Marshall C: Toward an ecology of hypertext annotation, in ACM Hypertext. Pittsburgh, PA, 1998, pp 40-49
2. Marshall C, Price M, Golovchinsky G, Schilit B: Designing e-Books for Legal Research, in Joint IEEE and ACM Conference on Digital Libraries (JCDL01). Roanoke, Virginia, 2001
3. Handschuh S, Staab S: Annotating of the Shallow and the Deep Web, in Annotation for the Semantic Web. Edited by (Eds.) SH-sSS. Amsterdam, IOS Press, 2003
4. Webster M: Merriam Webster: On-line dictionnary, 2004
5. Euzenat J: Eight Questions about Semantic Web Annotations. IEEE INTELLIGENT SYSTEMS 2002; 17(2):55-62
6. Caussanel J, Cahier J-P, Zacklad M, Charlet J: Cognitive Interactions in the Semantic Web, in SemantivWeb. Hawai, 2002
7. Denoue L, Vignollet L: An annotation tool for Web browsers and its applications to information retrieval, in Content-Based Multimedia Information Access (RIAO 2000). Paris, France, 2000
8. Hakkarainen K, Muukkonen H, Lakkala M: Collaborative technologies for facilitating progressive inquiry: Future Learning Environment Tools, in The 3rd International Conference on Computer Support for Collaborative Learning. Palo Alto, California., 1999, pp 406-415.