Semantic annotation for the teacher: models for a computerized memory tool
Faiçal Azouaou, Cyrille Desmoulins

To cite this version:

HAL Id: hal-00190399
https://telelearn.archives-ouvertes.fr/hal-00190399
Submitted on 23 Nov 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract. This article aims to propose a model of semantic annotation dedicated to the teacher. This model must adapt to the teacher’s activity specificity, who needs to master both a pedagogical and domain expertise. In this paper, we analyze the particularity of the teacher’s annotation language that enables the teacher to express his own expertise. First, we identify the concepts of this annotation language used by the community of teachers. Then we propose a conceptual model of this language based on ontologies. We use these ontologies to propose an annotation model (MemoNote) in order to enable the teachers annotating using their own language. In order to check the validity of this model (ontologies and annotation language) in representing the teachers’ semantics, we describe the results of an investigation done with chemistry teachers. Finally we specify the external and internal representations of the annotation tool for the implantation.

Introduction

The teacher during his activities handles various teaching documents (designing, reading, reviewing... etc). At the same time, he needs to memorize ideas and corrections or to plan actions to be made. This memorization is often materialized by annotations that the teacher puts on these documents.

Whereas the teacher nowadays uses more and more software tools to handle teaching documents in a digital format, annotation is still often made on paper, which requires the teacher to first print his documents, and implies an extra effort. For example, the result of the IMAT european project [1] pointed out the need for teachers to be able to annotate directly on the digital document and to manage a software memory of his activity. Thus, it appears necessary to propose a software annotation tool enabling teachers to express and clarify their feedbacks directly on the digital teaching documents.

As in any community of practice [2], [3], the teachers’ community uses a specific shared language to annotate. This language is a set of common forms and concepts. Consequently, the teacher needs a dedicated tool of annotation that integrates this language.

The object of this article is indeed to identify the teacher’s annotation language to be used as a basis for defining a tool of semantic annotation dedicated to him. The assumption of our research is that the memory resulting from this tool could enable the teacher to improve the effectiveness of his teaching and support its activity through the remembrance it provides.

The article is organized as follows. In the first section, we specify the meaning of a semantic annotation and we explain why this semantic is important. In order to provide semantics to teacher’s annotation, we characterize in section two the teacher’s annotation language. Starting from this characterization, we define in the third section the basic concepts of this teacher’s annotations language, represented with dedicated ontologies. We then use
these ontologies and a generic annotation model to propose a conceptual model for a dedicated teacher annotation language. We explain in section five, how we can use this ontology based model to improve the annotation’s retrieval. In order to check the validity of this conceptual model (ontologies and annotation model) in representing the teachers’ practices, we present in section six, the results of an investigation done with chemistry teachers. In the last section, we describe the implementation of this conceptual model in a teacher dedicated annotation software tool in terms of external and internal representations.

1. What is a semantic annotation?

Before studying the annotation language, we need to specify precisely what an annotation is in general and what a semantic annotation is in particular. Some authors provide informal definition of an annotation, varying upon the research field Human Computer Interaction (HCI[4], the cognitive science [5], and the digital libraries and document retrieval field [6]. To sum up, all these informal definitions agree 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 here.

Euzenat [7] formalized semantic annotation in the context of the Semantic Web. From two sets of objects, documents and formal representations, two functions can exist: a function from document to formal representations, called *annotation* and a function from formal representations to documents called *index*. Usually, these two functions are created at the same time during an activity called annotation or indexing. The Semantic Web aims to provide annotations on web documents with an explicit semantics for the computer, and not only for the human that created it or handle it. A Semantic Web will create an extension of the current World Wide Web, in which information is given well-defined meaning, so machines become much better able to process and "understand" the data that they merely display at present [8].

Marshall [15] makes a distinction between explicit and implicit annotations. An explicit annotation is an annotation that other readers can understand and interpret. At the opposite, an implicit annotation is “telegraphic, incomplete and tacit”. These annotations “pose interpretive difficulties for anyone other than the original annotator”. It is the case of highlighted text, of a cryptic asterisk without comment, etc. Thus, if every annotation has a semantic, it remains mostly implicit. Annotation semantics is implicitly carried, for a given annotator, by the shape used. For example, a given annotator used to underline in red the parts which he considers as important. The lack of explicit annotation semantics makes it difficult for other people to interpret them. A useful annotation tool should enable an annotator to explicit the semantic of his annotations.

2. Teacher’s annotation language for authoring pedagogical documents

As in any community of practice [2]; [3] teachers use a specific shared community language to annotate. This language is a set of common annotation forms and objectives. The annotation’s objective is considered as a central point by most of the authors, because it defines the semantics of the annotation.

Whereas we did not found any result about teacher’s annotations objectives, some authors studied the various objectives an annotation could carry in general. Marshall studied university students’ annotations and extracted the following objectives [2]: procedural signalling for future attention, place marking and aiding memory, problem-working, interpretation, tracing progress through difficult narrative and incidental reflection about the material circumstances of reading. This study also points out the fact that an annotation can serve several objectives at the same time.
Veron [9] and Huart [10] took up the objectives of Virbel [11] on active reading at the BNF (Bibliothèque Nationale de France). They identified four families of annotation goals: classifying (organising into a hierarchy, contextualising); adding information (reformulating, commenting, documenting); planning (scheduling, indirect annotating); correlating.

Finally, Mille [12] studied the didactic annotation of a text exam. She identified several annotation goals, grouped in two main categories: understanding the document and finding information.

A dedicated teacher’s annotation tool should enable teachers to express their own expertise using their own annotation language as they are used to do on paper. To identify this language (teacher’s annotation objectives), we first study the nature of his expertise. We consider, as [13] that the teacher’s annotation is a language that references his self-expertise: while annotating the teacher is in fact transforming his implicit knowledge into an explicit form.

Teaching expertise has many facets, according to [14], he uses during his activity. The teacher organizes the subject to be taught (domain) in several lessons, and each lesson may include several topics that are combined into learning objectives. For each topic, the teacher defines appropriate pedagogical presentations and activities in order that the learners reach their learning objectives. They use different teaching documents. The teacher, then, should have two kinds of expertise:

1. **Pedagogical expertise**: knowledge about organizing the lesson, evaluating learners, designing pedagogical activities, asking good questions, etc.
2. **Domain expertise**: declarative or procedural knowledge of the domain to be taught.

Some annotations concern only the document itself: the teacher annotates to memorize elements about the design, the structure of the document; for instance; the teacher annotates to correct a syntax error, to move a paragraph, or to add a picture.

Consequently, the teacher’s annotations express his objectives relating to three different levels: pedagogy, domain and document.

- **The pedagogy level**: The teacher organizes the content to be taught in several lessons; he adapts the content to the learners’ context. For each lesson, he designs different activities that help learners building their own knowledge. The teacher also decides to ask appropriate questions to learners and adapts the different lessons to the feedbacks and questions of the learners [14]. All these teacher’s activities mean a high level of pedagogy expertise, that the teacher can memorize using annotation.

- **The domain level**: The domain level covers the knowledge specified in the content of the lessons. This knowledge can be rather general knowledge like “including/understanding the theory of relativity” or more precise one like “knowing the capitals of the countries of Europe”. It can also be declarative knowledge such as “knowing the great cities of the world”, or procedural one such as: “to know how to carry out an experiment of oxidation in chemistry’s lab”. We situate in this level the teaching activities specified in the program or by the teacher himself.

- **The document level**: The document level concerns what is related to the document itself; in particular it concerns two of its structure:
  - **Physical structure**: the document presentation, its typographical characteristics: font, colour, size, grease…
  - **Logical structure**: the role and the nature of each segment in a document: title, subtitle, paragraph, etc
The teacher augments his own memory using annotations about these three levels. The result of the memorisation is the objective of the annotation. For instance, during the design of the pedagogical document, the teacher annotates that he should review in the future some exercises’ results (domain expertise), or add some definitions for the learners (pedagogical expertise).

3. Ontology based conceptual model for annotation semantics

In the previous section, we have classified the teacher annotation’s semantics into three levels: pedagogy, domain to be taught and document. To model these objectives, we choose to use ontologies [15]. They formalise the concepts shared by a community and their relationship (hierarchy, metonymy, etc.) by providing a precise and explicit semantic. They define the scope of the set of “concepts” handled by the annotator and they also enable the annotations to be shared among people using the same ontologies.

To design these ontologies, for the field of chemistry (1st year of university), we have used the literature (for pedagogy and document levels); teaching and learning documents (for domain level).

3.1 Pedagogy annotation objectives ontology

The pedagogy ontology concerns the annotations that the teacher makes to memorize elements of the pedagogy level.

<table>
<thead>
<tr>
<th>To memorize...</th>
<th>To memorize...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Non significant objective.</td>
<td>4. learning objectives not ambitious enough</td>
</tr>
<tr>
<td>1.1. To ignore.</td>
<td>4.1. Compared to the students</td>
</tr>
<tr>
<td>1.2. To work if there remains time.</td>
<td>4.2. Concept already comparable by the students</td>
</tr>
<tr>
<td>1.3. To reduce</td>
<td>4.3. Not enough of concepts</td>
</tr>
<tr>
<td>2. Significant objective</td>
<td>4.4. Too low constraints on the situation</td>
</tr>
<tr>
<td>2.1. To deepen</td>
<td>4.5. Too much time in the meeting</td>
</tr>
<tr>
<td>2.2. To illustrate</td>
<td>4.6. Too low material constraints.</td>
</tr>
<tr>
<td>2.3. To evaluate</td>
<td>5. Non relevant learning objective</td>
</tr>
<tr>
<td>3. Badly elaborated objective</td>
<td>5.1. Not part of the program</td>
</tr>
<tr>
<td>3.1. Measurable objective not assessed</td>
<td>5.2. Already represented in another objective</td>
</tr>
<tr>
<td>3.2. undefined situation of training</td>
<td>6. Bad content of the text (spot of reading learning).</td>
</tr>
<tr>
<td>3.3. Non objective evaluation</td>
<td></td>
</tr>
<tr>
<td>3.4. Non operable objective</td>
<td></td>
</tr>
</tbody>
</table>

3.2 Domain annotation objectives ontology

This second ontology concerns teacher’s annotations relating to the domain level. The teacher is more an expert within teaching and pedagogy than the domain he teaches, thus, he needs to annotate elements of this domain to not forget them. So, a novice teacher of database can annotate the SQL’s syntax of a specific database management system (DBMS), especially if this system is different from that he taught the previous year. This domain ontology depends on the topic to be taught. The domain of our study is the chemistry program (1st year of university).
### 3.3 Document annotation objectives ontology

Finally, the last ontology concerns annotations’ objectives relating to the logical and physical structures of the document: titles, paragraphs, font, colours, etc. The teacher annotates to memorize different improvements and changes to do on the document. These annotations have an effect on the reading of the document (increased comfort of reading, better structured document, corrected errors...).

#### Table 3 Ontology of annotation’s objectives on the document level

<table>
<thead>
<tr>
<th>1. To restructure</th>
<th>4. To create a relation between two passages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1. To give a title</td>
<td>4.1. Relation presentation /detail</td>
</tr>
<tr>
<td>1.2. To treat on a hierarchical basis</td>
<td>4.2. Relation presentation / explanation</td>
</tr>
<tr>
<td>1.3. To synthesize</td>
<td>4.3. Relation definition / explanation</td>
</tr>
<tr>
<td>1.4. To reformulate.</td>
<td>5. To review</td>
</tr>
<tr>
<td>2. To add a personal remark</td>
<td>5.1. textual error</td>
</tr>
<tr>
<td>2.1. To criticize</td>
<td>5.2. incomplete illustration / table</td>
</tr>
<tr>
<td>2.2. To express a related idea</td>
<td>5.3. missing illustration / table</td>
</tr>
<tr>
<td>2.3. To develop</td>
<td>5.4. missing index/glossary</td>
</tr>
<tr>
<td>2.4. To express its own comprehension</td>
<td>5.5. incorrect assertion</td>
</tr>
<tr>
<td>2.5. To add an example</td>
<td>5.6. ambiguous content</td>
</tr>
<tr>
<td>2.5.1. To solve a problem</td>
<td>5.7. an indefinite abbreviation</td>
</tr>
<tr>
<td>2.5.2. To explain textually</td>
<td>5.8. document’s structure</td>
</tr>
<tr>
<td>2.6. To refer to another document</td>
<td>5.9. To remove a passage</td>
</tr>
<tr>
<td>3. To categorize</td>
<td>5.10. To reformulate a passage</td>
</tr>
<tr>
<td>3.1. By importance’s value</td>
<td>5.11. To add a passage</td>
</tr>
<tr>
<td>3.2. By predetermined type</td>
<td>5.12. To plan an action</td>
</tr>
<tr>
<td>3.2.1. Theorem</td>
<td>5.13. To support the attention</td>
</tr>
<tr>
<td>3.2.2. Definition</td>
<td>6. to spot</td>
</tr>
<tr>
<td>3.3. By personal type</td>
<td></td>
</tr>
<tr>
<td>3.4. By content’s similarity</td>
<td></td>
</tr>
</tbody>
</table>

### 3.4 The relation between the three types of objectives

When the teacher annotates with an objective relating to the knowledge level, this annotation has often an effect on the two other levels (domain and document). Indeed, there is dependency between the three levels. Each learning topic of the domain is adapted to learners using pedagogy and then transcribed on teaching and learning documents.

We illustrate this dependency using an example: a teacher prepares his chemistry lab; he decides to plan an assessment (pedagogical element) during the lab. In order not to forget to do this assessment, he annotates his document with a comment. Before to go to the chemistry lab, he reminds with the help of this annotation, to add an assessment to the activity described by the document. Consequently, the teacher modifies the document (document level) by adding...
the assessment questions (pedagogy level), which relate to a specific knowledge objective (domain level).

4. Ontology based annotation model

In order to define a teacher dedicated annotation model, we use the ontologies specified in the previous section and a generic annotation model.

In [16], we presented a generic pedagogical annotation model (MemoNote) composed of three parts:

1. **The tangible part**: represents the visible part of the annotation (the form, the anchor...). It is constituted of the following attributes: physical anchor (URL + location in the document), visual form and the syntactic anchor (annotated content).

2. **The episodical part**: describes the context of the annotation (author, date, location...). It is made up of the following attributes: Author, Date, location, Activity and context of memorizing.

3. **The semantic part**: express the meaning the author gives to the annotation. This semantics is represented primarily by an objective attribute. It is made up of the following attributes: objective, content, importance, confidence, recipient, activity and remembrance’s context.

Among these three annotation parts, the semantic one is the most significant. Indeed the author of the annotation annotates for a given objective which is often implicit in the annotation form. The annotation is required to understand and re-use this annotation. Consequently, the loss of this implicit semantics makes the annotation useless.

In order to propose a dedicated teacher annotation model, we modify the generic model by adding elements relying to the teacher activity. In particular, we change the episodical and the semantic parts:

1. **The tangible part**: same as previous.

2. **The episodical part**: the teacher annotation model should specify the teacher annotation context: the type of the domain to be taught (chemistry, mathematics...), the type of activity (exercise, lab, course...), the phase of the teaching (before the course, during the course (with the learners). These data will be used by the teacher to retrieve annotations he has done in a particular context (for instance: last month during the lab).

The **semantic part**: we divide the semantic part into three categories which correspond to the three levels: pedagogy, domain and document. The annotation’s objective (the main attribute of the semantic part) takes its value in one of the three ontologies specified in the previous section.

5. Semantic indexing using ontologies

As specified by Euzenat [7], each annotation action define an inverse function which is indexing. Concretely speaking, while teacher is annotating a document using our ontology-based model, he/she is simultaneously indexing it with concepts from the three ontologies of objectives. The result of the annotation process is a semantically annotated document that is indexed with ontological concepts instead of simple keywords. But unlike the semantic web indexing which aims to describe the objective content of resources, our annotation model enables the teacher to index documents using his own viewpoint using subjective annotations.
Ontology indexing provides several advantages while retrieving documents or annotations[17]. It helps users to select queries criteria thanks to the ontology guidance. For instance, in order to review and correct the domain-level errors on all his pedagogical documents, the teacher will define a search criteria by choosing in the document’s ontology the concept “to memorise possible errors”, which will displays him all the documents with annotations “to memorise possible errors”. Consequently, the teacher will no longer use key words but only select a concept (or several) in one ontology (or more) and be guided by the ontologies hierarchy. In this way, the user interface’s usability is improved.

6. Investigation

In order to check the validity of this model (ontologies and annotation language) to represent the teachers’ practices, we describe in this section, the results of an investigation done with chemistry teachers.

We realized six interviews with chemistry teachers. We wanted to extract the semantics of the annotations they add on their teaching documents for chemistry’s lab. First, during a six-month period, each teacher annotated his chemistry’s lab document. We organise individual interview with each teacher. They were asked to bring their documents on which the colours of annotations were deferred. Then on the basis of these documents we ask the teachers to explain the semantics of each annotation using our ontologies.

This exploratory investigation provided several results. First, it partially validated our model based on three levels: i.e. we could verify that the teachers (chemistry in our investigation) really annotate their teaching documents according to objectives relating to the three levels. Then, this investigation provides us elements to make up the three ontologies by adding new concepts that were lacking to the teachers during the interview. For instance:

- Document related ontology.
  - to spot
- Domain related ontology.
  - to plan changes
  - material problem
  - time problem
  - irrelevant passages
- Pedagogy related ontology.
  - Mediocre text’s content (useless pedagogically).

This investigation confirmed our assumption (described in section 3.4) about the existence of dependencies between the three ontologies. Indeed, to express the objective of a given annotation, teachers often indicated an objective taken in several ontologies. These dependencies are in single-directed: pedagogy to domain, domain to document. For example if the teacher annotates to ignore a given objective, this will have an effect on the pedagogical activity defined for this objective, which does not have any more reason to exist on the document and consequently will be removed from it.

On the other hand this dependency does not exist in the other direction: if the teacher annotates at the document level (to correct a misspelling for example), this will not have an effect on the knowledge or pedagogy levels.

7. Implantation’s details

To implement our teacher’s annotation model, we re-use our generic annotation tool called “MemoNote”. MemoNote enables users to manage the note of events and knowledge they
want to memorize during their pedagogical activities (teaching or learning) and to retrieve them in the future. The MemoNote project aims at formalizing and implementing computerized external memories made of notes added directly and voluntarily on the training material by its user. It covers memorization and remembering tools, for individual and groups, mainly for teachers and learners. The MemoNote annotation tool represents the memorization part of the project. It is currently dedicated to personal annotation but not to teacher. The first mock-up of the tool has been implemented on TabletPC computers for pencils based annotations, extending the MobiPocket reader software[18]. A quite similar web based mock-up has been developed too, mainly to provide retrieval and synchronisation functionalities.

7.1 MemoNote background

MemoNote enables the user to annotate pedagogical documents. It is not dedicated to a special field of teaching or a specific type of activity. For a specific teaching activity, MemoNote can adapt the user’s context by selecting a set of ontologies. This set of ontologies describes the users, the teaching domain, the pedagogical activities (content, location, time) and the annotation’s 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. This formalizes the results of ecological studies on annotation of teaching material [19] demonstrating that the annotation process is rather generic whereas the annotation content (forms, objectives, etc.) depends upon the learning/teaching context.

7.2 External representations of the conceptual model.

While reading a document in its pedagogical activity, the MemoNote user annotates this document by:

- Defining the source of the annotation (tangible part): anchor: where it is located on the document, Visual form: the shape and color the annotation takes on the document.
- Defining the target of the annotation (semantic part): addressee, objective (at document, domain and pedagogical levels), content.
- Defining the annotation link itself (episodic part): annotator, date, location, teaching context,

The user interface in both cases is the same (figure 1). It has three main parts. The first part is a reader (reading software) embedding MemoNote annotation tools. In the first mock-up, this reader is MobiPocket [18]. 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) shown on the top the TabletPC on figure 3. 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. The main pattern type is a pattern where there is only one ontology to fix and a subpart of the ontology is selected by the pattern; and a pattern where some fields remains to be defined from scratch.

7.3 Internal representations for the conceptual model

Representing the semantics of annotation with ontologies is the same idea that in the semantic Web [8] approach (the main difference is that we use it for subjective annotation). We can then use the same languages as the semantic web to represent annotations and ontologies.

First, to represent and store the annotations we use RDF (Resource Description Framework)[20]. RDF is an infrastructure that enables the encoding, exchange and reuse of structured metadata. RDF is an application of XML that imposes needed structural constraints to provide unambiguous methods of expressing semantics. RDF additionally provides a means for publishing both human-readable and machine-processable vocabularies designed to encourage the reuse and extension of metadata semantics among disparate information communities [21].

Then, for representation of the different ontologies, we use OWL [22], otherwise known as the Web Ontology Language. OWL provides a language for defining structured ontologies that provide rich integration and interoperability of data. It uses both the URIs for naming and the description framework provided by RDF. OWL builds on RDF and RDF schema, adding more vocabulary for describing properties and classes as well as relationships between classes [23].

The use of these two semantic web standards guarantees the capacity to share annotations between different teachers, even if these annotations are made using different tools, because the two languages represent a unified data exchange format. These two languages also offer us the possibility to publish the annotated documents on the web; thus they can be indexed using the annotations by the new semantic web search engines like SWOOGLE [24] or they can also be processed by different web agents.

8. Conclusion

In this paper, we have demonstrated that the teaching activity is particular, because the teacher has both an expertise in the domain that he teaches and an expertise in pedagogy. When annotating, teacher expresses an annotation’s objective which is related to the concepts of these two expertises. In addition, the teacher can annotate the physical and logical structure of the document itself. Thus, this teacher’s annotation language enables the teacher to explicit objectives that belong to three different levels: pedagogy, domain and document. These objectives represent the annotation’s semantics that remains mostly implicit. The lack of explicit annotation semantics makes it difficult to reuse the annotations.

We model this shared language using ontologies. We propose then an ontology based annotation model dedicated to the teacher. This model enables the teacher to explicit his annotation’s objective using the three levels of semantics (pedagogy, domain and document). Using this model the teacher can retrieve his annotations easily guided by the ontologies.
We carried out a first validation by making an exploratory study with chemistry’s teachers which enabled us to confirm our assumption and to supplement and correct our three ontologies of objectives. Lastly, we presented an implementation of this model using languages borrowed from the semantic web researches.

Our exploratory study relates to the particular case of chemistry. We need to check out in what extent our annotation model is generalizable to the other teaching’s disciplines and the way of rapidly extracting annotation ontologies for a given domain.

References