A flexible model for the collaborative annotation of digitized literary works

dh2012.uni-hamburg.de /conference/programme/abstracts/a-flexible-model-for-the-collaborative-annotation-of-digitized-literary-works.1.html

XML

Gayoso-Cabada, Joaquin, Universidad Complutense de Madrid, Spain, gayoxo@gmail.com
Ruiz, Cesar, Universidad Complutense de Madrid, Spain, cruiz85@gmail.com
Pablo-Nuñez, Luis, Universidad Complutense de Madrid, Spain, lpnunez@filol.ucm.es
Sarasa-Cabezuelo, Antonio, Universidad Complutense de Madrid, Spain, asarasa@fdi.ucm.es
Goicoechea-de-Jorge, Maria, Universidad Complutense de Madrid, Spain, mgoico@filol.ucm.es
Sanz-Cabrerizo, Amelia, Universidad Complutense de Madrid, Spain, amsanz@filol.ucm.es
Sierra-Rodriguez, Jose-Luis, Universidad Complutense de Madrid, Spain, jlsierra@fdi.ucm.es

Introduction

The Complutense University has been one of the first European universities that has collaborated with Google's project[1] by putting on the Web 100,000 volumes from its ancient fund. However scholars notice that these digitized texts are often of no much use to professors-researchers-students in literature unless additional tools are provided, to enhance the educational and research value of this material. In particular, the ability of making annotations on these texts has been largely recognized as a basic mean of adding value to this kind of digitized resources (Rios da Rocha et al. 2009). In this paper we present the annotation model used in @Note 1.0, a system developed at UCM funded by the Google's 2010 Digital Humanities Award program.

@Note 1.0 allows us to retrieve digitized works from Google Books collection and add annotations to enrich the texts with research and learning purposes: critical editions, reading activities, e-learning tasks, etc. One of the main features of @Note annotation model, which distinguishes it from similar approaches (Azouaou & Desmoulins 2006; Bechhofer at al. 2002; Koivunen 2005; Rios da Rocha et al. 2009; Schroeter et al. 2006; Tazi et al. 2003), is to promote the collaborative creation of annotation schemas by communities of researchers, teachers and students, and the use of these schemas in the definition of annotation activities on literary works. It results in a very flexible and adaptive model, able to be used by many different communities of experts in literature defending different critical literary theories and for different annotation tasks. In this paper we present this annotation model.

The @Note Annotation Model

Structure of the model

The structure of the @Note annotation model is summarized in the UML class diagram (Booch et al. 2005) of Fig. 1. In this model:

- *Annotation management communities* are groups of *annotation managers*, experts in literature (teachers, researchers, etc) who act as administrators to create activities, to select works and to organize activity groups.
- *Annotation communities*, in their turn, are groups of *annotators*, students / pupils interested in literature who perform proposed annotation activities.
- Each a*nnotation activity* comprises (i) a *digitized work*, (ii) a *metalevel-oriented annotation schema*, (iii) a *work-oriented annotation schema*.
- In this context, the *works* are the literary texts that can be annotated during the annotation activities. *Annotations*, in their turn, are characterized by: (i) an *annotation anchor* (the region of the work to which the annotation refers), (ii) an *annotation content* (a free rich-text piece that actually configure the

annotation), (iii) a set of *annotation types* (semantic qualifiers for annotations) chosen from the annotation schemas attached to the annotation activity (at least one from the metalevel-oriented annotation schema).
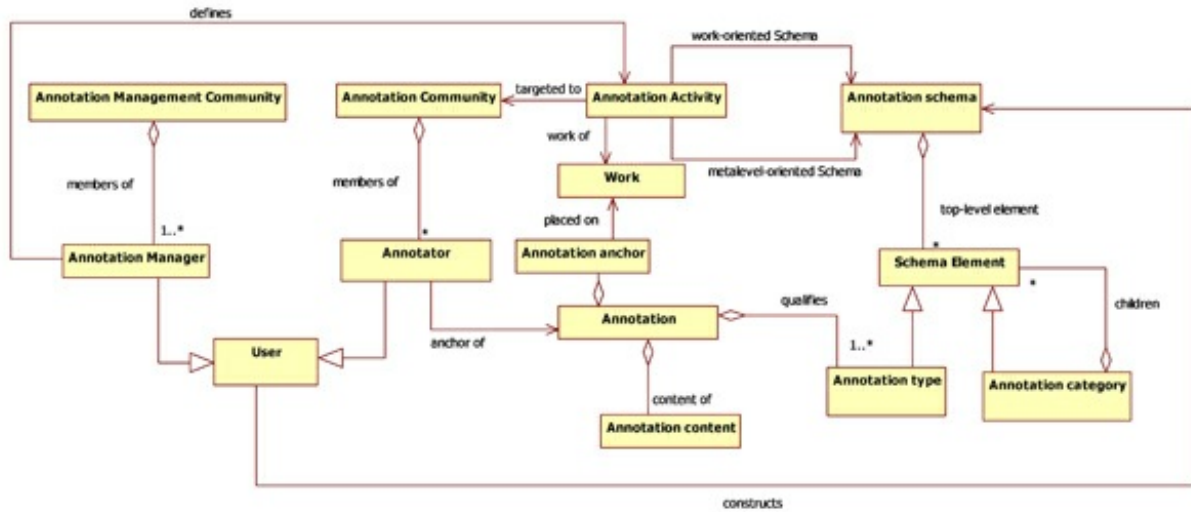

Figure 1: @Note information model

- The *annotation schemas* are explicit formalization of the types of annotations that can be carried out on works. In @Note, annotation schemas are hierarchies formed by annotation types and *annotation categories* (sets of annotation types and/or others, more specifically, annotation categories). In their turn, they can be *metalevel-oriented annotation schemas* (schemas which usually comprise concepts concerning particular literary theories around which the annotation activities are articulated), or *work-oriented annotation schemas* (schemas that capture aspects relative to the relationships between annotations and their anchors). While schemas of the first type are created by annotation managers, schemas of the second type are usually created by annotators.

In the context of the annotation management community, an annotation schema can be *public* or *private*. A private schema is only accessible for the annotation manager who created it. On the contrary, a public schema is accessible for all the annotation managers. Annotation managers have unlimited privileges on all the schemas to which they can access (i.e., they can create new annotation types and categories, they can blend two different types/categories in a single one, they have renaming and erasing privileges, etc), with the exception of modifying the *public/private* character (it only can be done by the schema's creator). In addition, when annotation managers create annotation activities, they only can choose those schemas to which they have access grants. Concerning annotators, they can add new types and categories to the book-oriented annotation schema, but they can't perform any other modification.

Figure 2: Example of rules governing the annotation process (informally described using natural language)

The annotation process

The @Note annotation process governs how to create the different types of information elements envisioned in the annotation model. For this purpose, @Note introduces a set of rules governing aspects like information visibility, creation and modification privileges of annotations and annotation schemas, etc. Although, by lack of space, these rules will not be detailed here, in Fig. 2 we include an example concerning an informal description of some of the rules governing the management of annotation schemas.

Annotation browsing and recovering

Annotation schemas in @Note are seen as T-*boxes* of description logic theories (Brachman & Levesque 2004). For instance, Fig 3a shows, edited in @Note, a fragment of the annotation schema used at UCM in an English Literature introductory course, while Fig 3b despites the description logic's counterpart. This simple interpretation is still powerful-enough to enable powerful *annotation browsing* and *annotation recovering* behavior. Indeed:

(a)

- Analysis
  - Structure
    - Structure_Type
      - Narration
      - Use_Of_Frames
    - Plot
    - Setting
- Criticism
  - Bibliographical
  - Psychological
  - Cultural

(b)

Structure $\subseteq$ Analysis
Structure_Type $\subseteq$ Structure
Narration $\subseteq$ Structure_Type
Use_Of_Frames $\subseteq$ Structure_Type
Plot $\subseteq$ Structure
Setting $\subseteq$ Structure
Bibliographical $\subseteq$ Criticism
Psychological $\subseteq$ Criticism
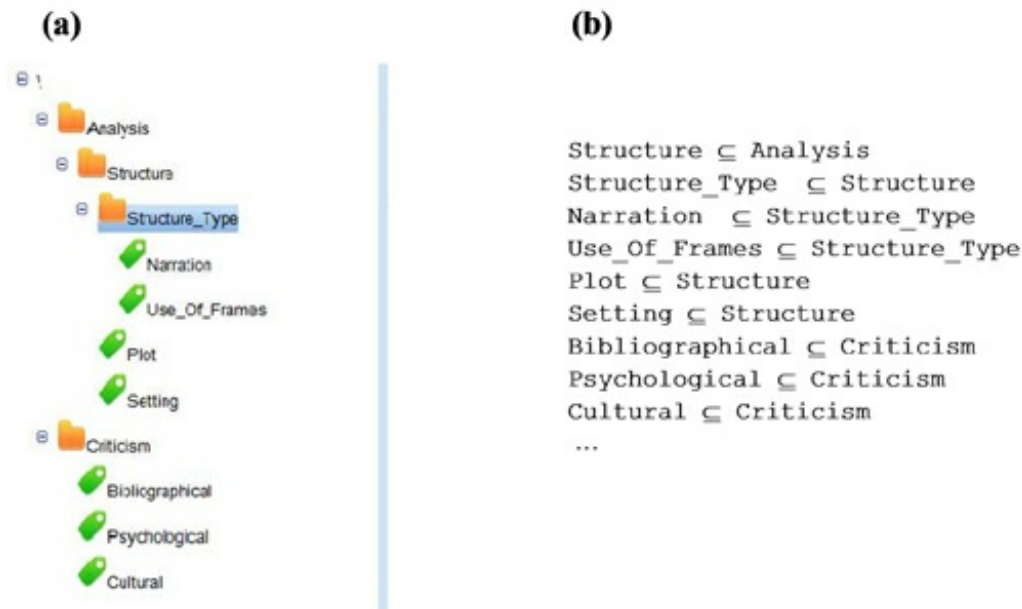Cultural $\subseteq$ Criticism
...

Figure 3: (a) A fragment of annotation schema (b) description logic counterpart

- Annotations can be browsed using annotation schemas, in a similar way to a folder explorer in a filesystem. In each step, there is a category or a type selected, and the user will *see* all the annotations entailed by such a selected element.

- Queries consist of arbitrary Boolean formulae involving annotation types and categories, being the outcomes the annotations entailed by such formulae.

In both cases, since entailment will be performed according to the description logic principles, the process will take into account the *is-a* relationship made explicit by the annotation schema.

Some technical details

The system has been entirely developed using Google technologies for the development of Rich-Internet Applications (RIAs) (Fraternali et al. 2010): GWT in the client side and the Google App Engine's facilities in the server side (Unruh 2010). Fig. 4 shows some snapshots of the system. The current version runs on the fully free-access books integrated in Google Books, and, in particular, on the UCM-Google collection. The works retrieval is achieved by the use of the Google Books API through REST (Richardson & Ruby 2007), and then presented to clients in an asynchronous way to keep them responsive to their events.
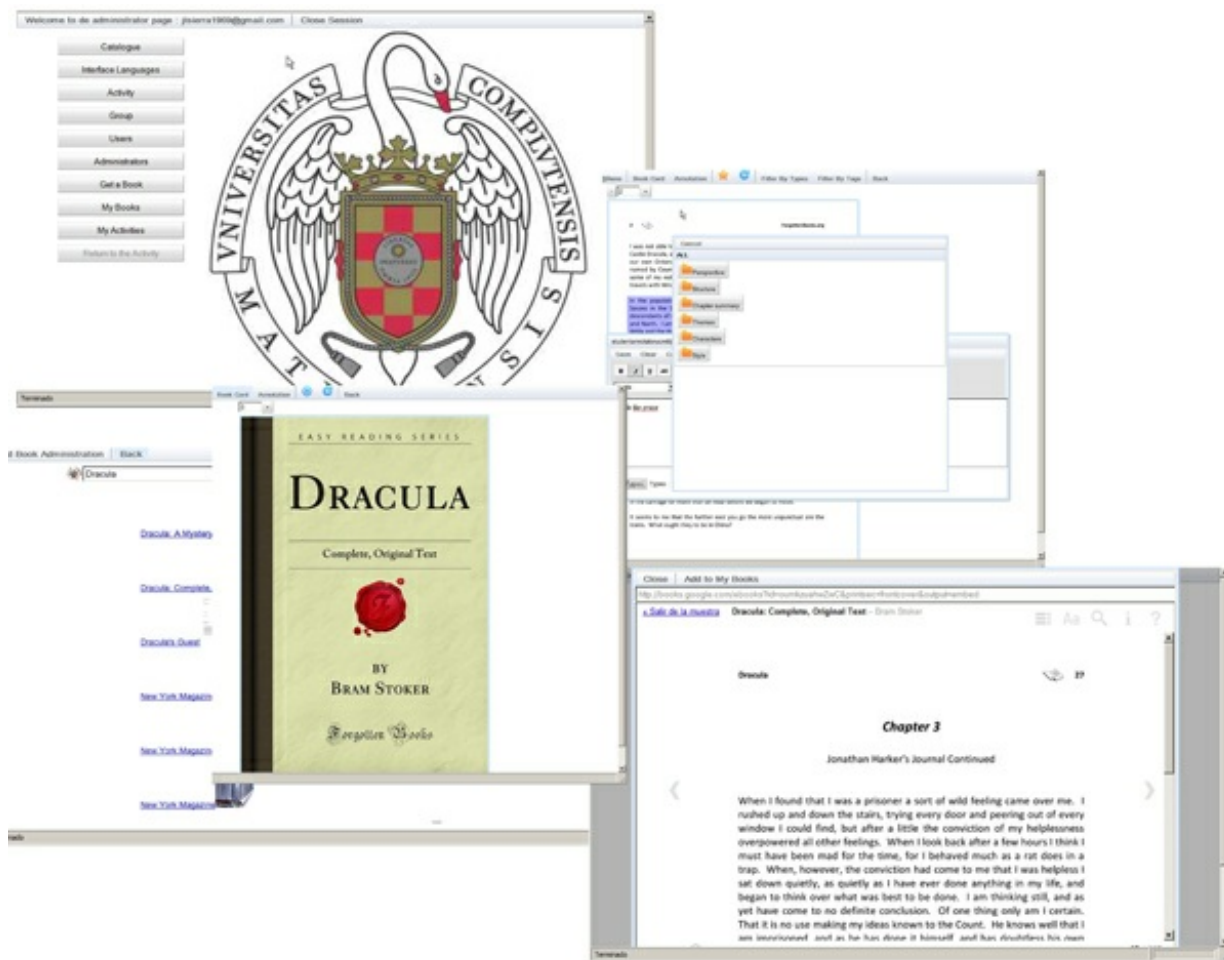
Figure 4: Some snapshots of @Note

Conclusions and Future Works

@Note promotes a fully collaborative annotation process, in which not only literary works are collaboratively annotated, but also annotation schemas are collaboratively created. The @Note system has been evaluated at UCM by several researchers and students in literature. They highlighted the flexibility of the annotation model, and, in particular, the ability to create and share annotation schemas tailored according to different critical perspectives and annotation activities. Additionally, they appreciate a sufficient expressive power from a browsing and recovering point of view. They also remarked the educational potential of the tool, although some advanced features could add some conceptual difficulties for students.

Currently we are working to adapt the annotation tool in order to facilitate its connection to a repository of learning objects, so as to allow the storage of literary texts' annotations as learning objects (Polsani 2003), and to make possible the recovery of those annotations and move about them according to the associated metadata. Thus, we are developing a communal working space for the creation of written compositions in different traditions and languages. We are also experimenting with the students' capability for developing their own catalogues, annotating the literary texts according to them and reusing their annotations in the production of critical essays. Additionally, we are working on connecting our system with other digital libraries (in particular, with Hathi Trust[2]). Finally, we are planning to address interoperability issues, in order to enable the interchange of annotations according to some of the emerging standards proposed by the digital humanities community (e.g., OAC[3]).

Acknowledgements

References

Azouaou, F., and C. Desmoulins (2006). MemoNote, a context-aware annotation tool for teachers. *Proceedings of the 7th International Conference on Information Technology Based Higher Education and Training* ITHET'06, Sidney, Australia, July 2006.

Bechhofer, S., L. Carr, C. Goble, S. Kampa, and T. Miles-Board (2002). The Semantics of Semantic Annotation. *Proceedings of the First International Conference on Ontologies, DataBases, and Applications of Semantics for Large Scale Information Systems* ODBASE'02, Irvine, CA, USA, October 2002.

Booch, G., J. Rumbaugh, and I. Jacobson (2005). *The Unified Modeling Language User Guide* (*2nd Edition*). Boston: Adisson-Wesley.

Brachman, R., and H. Levesque (2004). *Knowledge Representation and Reasoning*. Amsterdam: Morgan-Kauffmann.

Fraternali, P., R. Gustavo, and F. Sánchez-Figueroa (2010). Rich Internet Applications. IEEE *Internet Computing* 14(3): 9-12.

Koivunen, M.-R. (2005). Annotea and Semantic Web Supported Collaboration. *Proceedings of the UserSWeb Workshop – 2nd European Semantic Web Conference*, Heraklion, Grece, June 2005.

Polsani, P. (2003). Use and abuse of reusable learning objects. *Journal of Digital Information* 3(4).

Richardson, L., and S. Ruby (2007). *Restful web services*. Beijing: O'Reilly.

Rios da Rocha, T., R. Willrich, R. Fileto, and S. Tazi (2009). Supporting Collaborative Learning Activities with a Digital Library and Annotations. *Proceedings of the 9th IFIP World Conference on Computers in Education* WCCE'09, Bento Gonçalves, Brazil, July 2009.

Schroeter, R., J. Hunter, J. Guerin, I. Khan, I., and M. A. Henderson (2006). Synchronous Multimedia Annotation System for Secure Collaboratories. *Proceedings of the Second IEEE International Conference on e-Science and Grid Computing* e-Science '06, Amsterdam, Netherlands, December 2006.

Tazi, S., Y. Al-Tawki, and K. Drira K. (2003). Editing pedagogical intentions for document reuse. *Proceedings of the 4th IEEE Technology Based Higher Education and Training* ITHET'03, Marrakesh, Morocco, July 7-9, 2003.

Unruh, A. (2010). *Google App Engine Java and GWT Application Development* . Birmingham: Packt Publishing.