

Online Authoring System with Semi-automatic Annotation and Visual Knowledge Representation

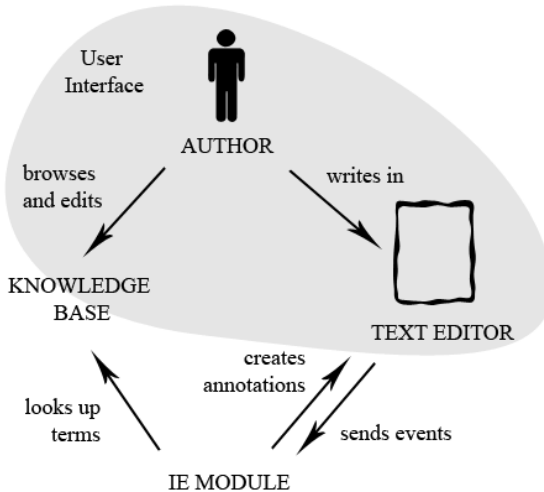
Yasen Kiprova and Ivan Koychev¹

Faculty of Mathematics and Informatics, University of Sofia, Bulgaria
yasen.kiprova@gmail.com, koychev@fmi.uni-sofia.bg

With the recent development of technology e-books are becoming more and more popular. Some of them already incorporate rich multimedia and hyperlinks, adding references to different resources. On the other hand, the Semantic Web initiative aims to develop technologies that allow explicit knowledge to be incorporated in the web content. The SmartBook project aims to develop a framework that will allow the e-books to be augmented, employing Semantic Web technologies [4]. The paper presents a work in progress, which aims to create a book authoring environment that will enable integration of knowledge and data in e-books.

Several systems approached the problem of manual and semi-automatic annotation. [e.g. 2, 5]. We envisage two ways in which the approaches in these systems can be further developed and adopted for our purpose: better usability and better information extraction. The system architecture contains three main components: web interface, knowledge base (KB) and information extraction (IE) module (fig. 1).

Our main goal is to provide a practical user interface, which stimulates authors to include semantic data into their work. This includes the following aspects:



- Suggestions and auto-completion of KB instances while typing;
- Browsing and modifying the KB;
- Modifying annotations and adding new ones.

For the purpose of browsing and modifying the KB we will customize the Javascript Infoviz Toolkit (<http://thejit.org/>). In the first prototype of our system the ontology would not be modifiable through the interface.

Fig. 1. System workflow and interaction between modules.

¹ Also associated with Institute of Mathematics and Informatics - BAS.

Our second goal is a useful interface for readers. It will benefit from the integrated semantic annotation. Further it can create his/her personal extension by adding annotations and by editing the KB. We are going to use the graph visualization of the KB, since we believe this is the most intuitive one.

For the purpose of helping authors with automatic annotation, we are adopting the KIM platform [3], customized for online annotation while typing. The KIM platform, integrates modern information extraction tools with the scalable semantic repository OWLIM [3]. It also provides ontology driven interface for browsing through the set of processed documents. Our system will use the KIM IE module which steps on GATE [1] and relies heavily on lookups – terms will be recognized in the text if they exist in one of the terminology dictionaries. Once recognized (manually or by the IE module) in text, terms will be added to the semantic repository as instances. This way the KB of our system will be updated constantly by the users, which will result in more complete dictionaries and better information extraction.

The *authors' interface* will consist of two parallel screens, one for writing text and one for viewing and editing the KB. When a sentence is written in the text editor, asynchronous callbacks will tell the IE module to annotate it. The mentioned annotations will be highlighted and their instance URIs will be sent to the KB editor. The graph visualization engine will browse the KB for the mentioned instances and will adjust towards them. Authors will see the mentioned instances highlighted in the graph screen and will be able to follow relations away from what is mentioned. They will also be able to manually mark a piece of text and annotate it, providing an instance from the KB.

Readers will be able to browse the KB through the graph visualization tool and will have limited editing permissions. Clicking on a mentioned instance will direct them to the respective place in the text.

Acknowledgements. This research is supported by the SmartBook project, subsidized by the Bulgarian National Science Fund, under Grant D002-111/15.12.2008.

References

1. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In: Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia (2002)
2. Ciravegna, F., Dingli, A., Petrelli, D., Wilks, Y.: User-System Cooperation in Document Annotation Based on Information Extraction. In: A. Gomez-Perez, R. Benjamins (eds.) EKAW 2002, Spain, LNCS-LNAI, vol. 2473, 122-137
3. Kiryakov, A., Simov, K., Dimitrov, M.: OntoMap: portal for upper-level ontologies. In: Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS 2001), pp 47-58. ACM, 2001
4. Koychev, I., Dicheva, D., Nikolov, R.: SmartBook: Semantics Inside. To appear in *Serdica Journal of Computing*, 2010
5. Lanfranchi, V., Ciravegna, F., Petrelli, D.: Semantic Web-based Document: Editing and Browsing in *AktiveDoc*, In: Proceedings of the 2nd European Semantic Web Conference, Heraklion, Greece, May 29-June 1, 2005
6. Popov, B., Kiryakov, A., Ognyanoff, D., Manov, D., Kirilov, A.: KIM – a semantic platform for information extraction and retrieval. *Nat. Lang. Eng.* 10, 375-392 (2004)