System Description: KAT an Annotation Tool for STEM Documents

Mircea Alex Dumitru, Deyan Ginev, Michael Kohlhase, Vlad Merticariu,
Stefan Mirea, and Tom Wiesing

Computer Science
Jacobs University Bremen
http://kwarc.info

Abstract. We presents the KAT system, browser-based annotation tool
for linguistic/semantic annotations in structured (XHTML5) documents.
As it is parametric in the annotation ontology and represents annotations
as RDF, it can easily be integrated into RDF-based corpus management
systems; we present an integration into the CorTeX system.

1 Introduction

Current natural language understanding systems do not work particularly well
on mathematical and technical documents as they cannot deal with formulae,
diagrams, and the special, technical vocabularies and discourse conventions of
such documents. To retrain existing tools and evaluate new ones specifically
developed for STEM documents, we need to establish manually annotated doc-
ument corpora. Unfortunately, even the annotation tools used in computational
linguistics do not work well with mathematical documents, as they assume plain
texts.

2 System Architecture and Realization

3 Conclusion

[1] [2]
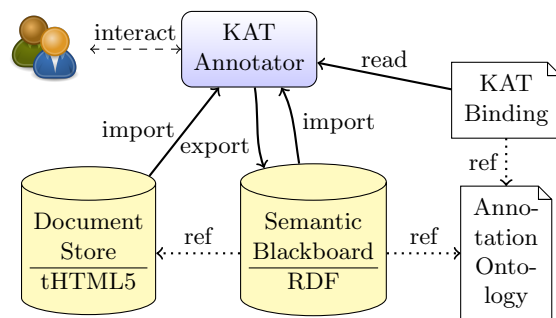
EdN:1
EdN:2

---

[1] EDNOTE: MK@MK: say something
[2] EDNOTE: Acknowledgements

Fig. 1. The KAT System Architecture