



TextAnnotator

A web-based annotation suite for texts

Giuseppe Abrami, Alexander Mehler, Manuel Stoeckel

Goethe University Frankfurt | Text Technology Lab



20. - 24. July 2020
Virtual

Introduction

- ▶ The annotation of natural language texts and their use is addressed in many projects in the digital humanities.
- ▶ This not only involves the generation of training data, but also the correction of errors by automatic preprocessing.
- ▶ Many annotation tools usually only allow simple annotation of text with simple visual annotation support.
- ▶ However, this limited use of annotation tools shows a gap in the large field of digital humanities that can be closed by the so-called **TextAnnotator** (Helfrich et al. 2018).
- ▶ Annotation of multi-tokens, named entities, temporal, argumentative, prepositional and rhetorical structures as well as annotation using knowledge resources.

Features

- ▶ Browser-based annotations of natural language text
- ▶ UIMA-based annotations
- ▶ Preprocessing with the help of the *TextImager* (Hemati, Uslu, and Mehler 2016)
- ▶ Platform independent and cross-technology utilization through WebSockets.
- ▶ User- and group-related access permissions on projects and resources.
- ▶ Collaborative and simultaneous annotations based on different annotation views (AV).
- ▶ Instantaneous evaluation of the annotation quality, by using Inter-Annotator-Agreement.

In medias res

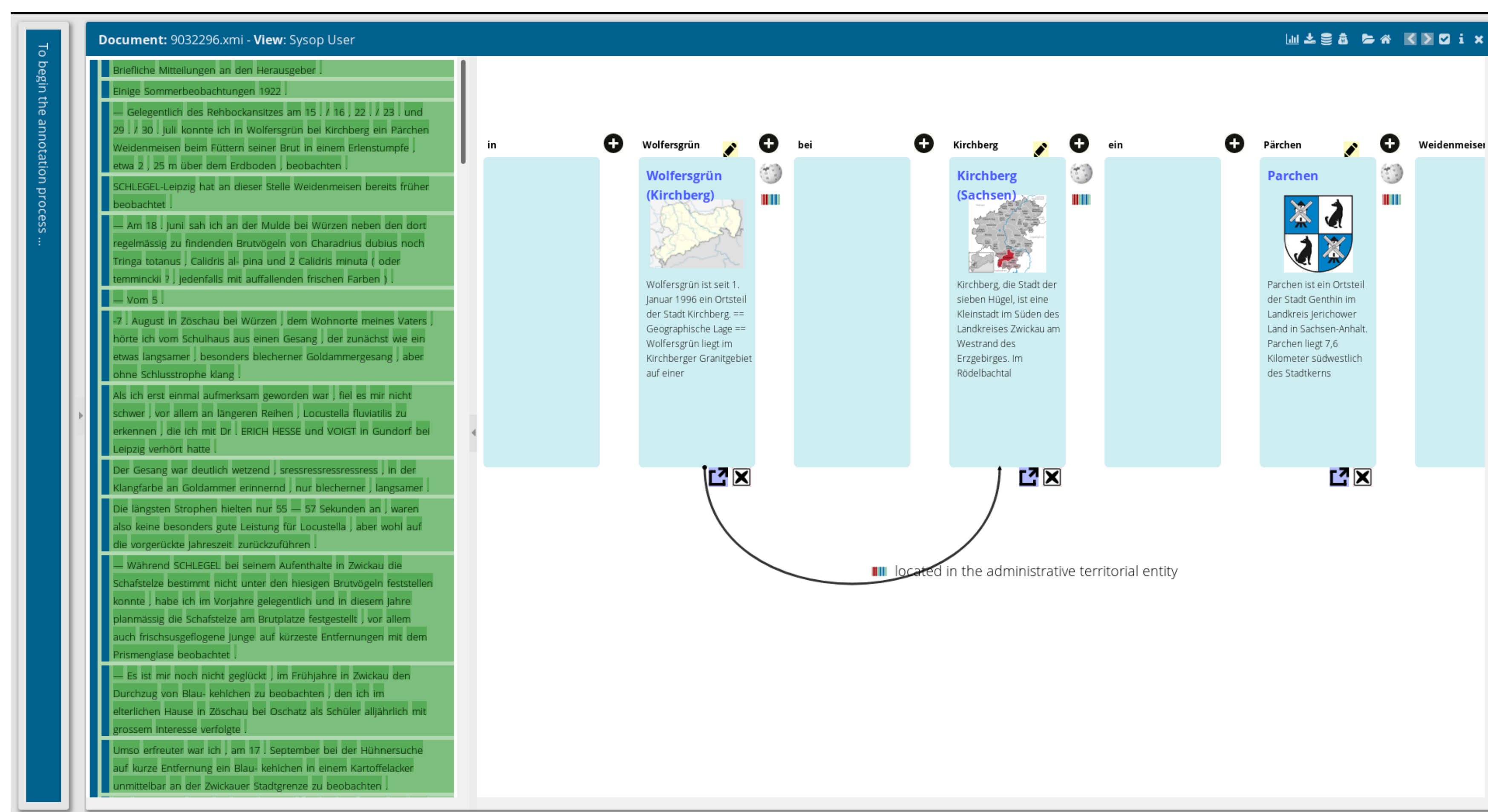


Figure 1: Extract from an annotation session and the use of *KnowledgeBaseLinker*. The individual tokens can be linked to knowledge resources or the entries can be modified. In this scenario the texts were already automatically preprocessed by the *TextImager* and the edge connecting two entities shows an implicit relation which was interpreted based on the Wikidata entries of the respective assignments to the knowledge database Wikipedia.

Annotation Views

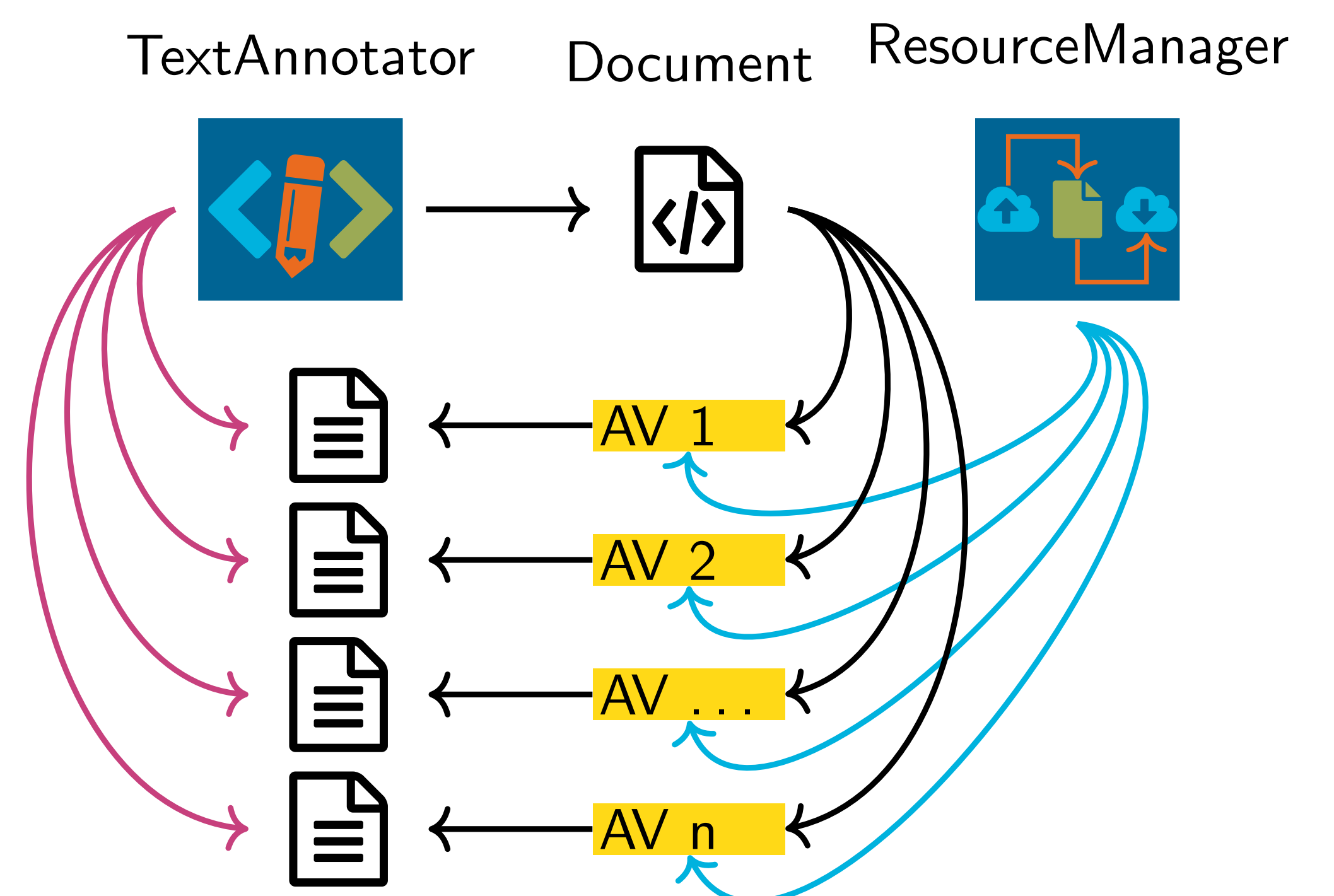


Figure 2: Schematic diagram of the use of annotation views (AV). *TextAnnotator* has access to the documents, which contain annotation views that are accessible to the user. By this assignment the *TextAnnotator* uses the respective annotations in the individual views for annotation through implemented tools.

Inter-Annotator-Agreement

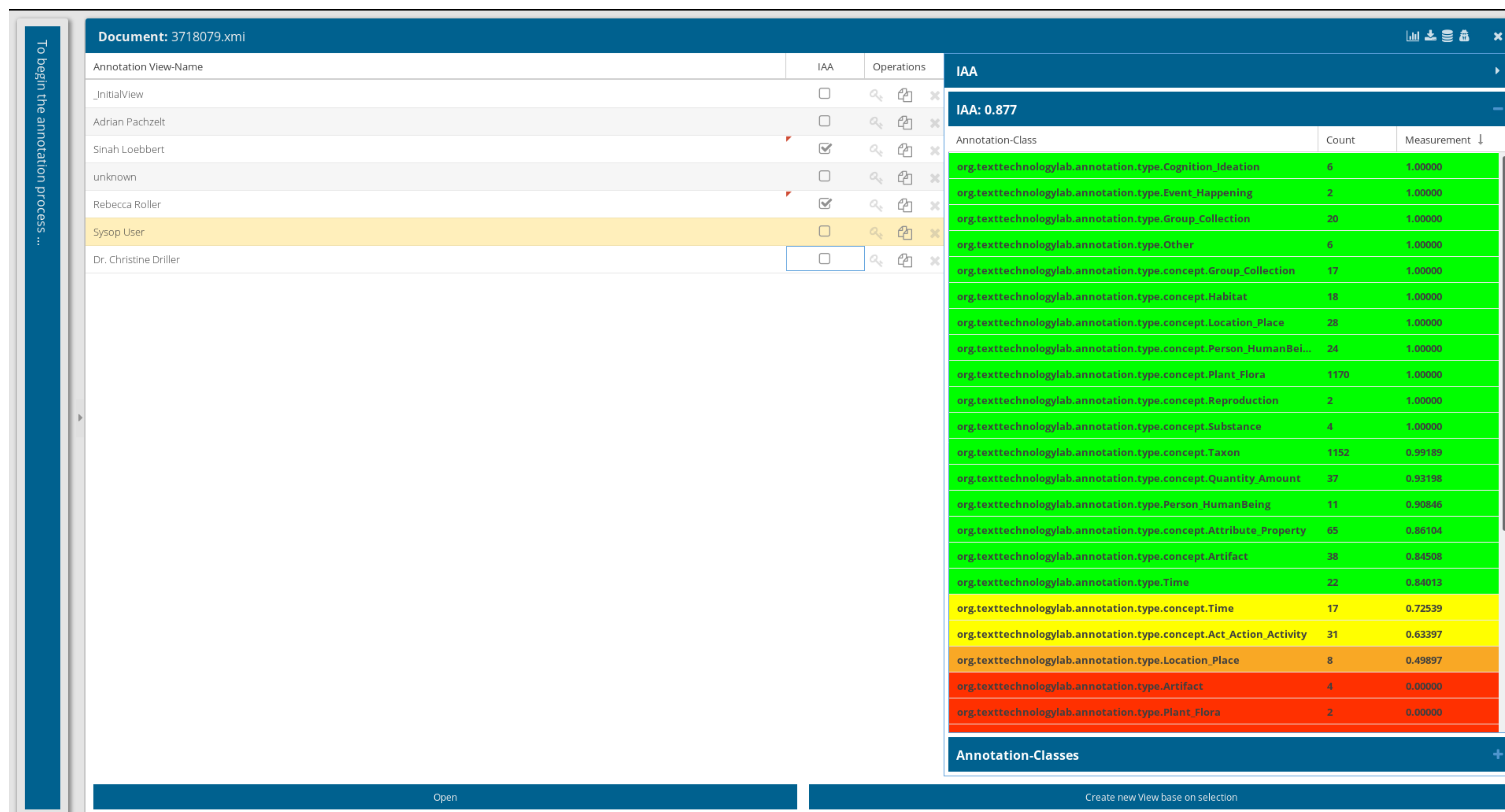


Figure 3: An open document is shown in the *TextAnnotator*. In the left area the annotation views can be seen. User views carry the name of their respective users, all other views are annotation views that can be authorized by the user. On the right, the IAA value of the document is visualized based on the selected annotation views (left) and the previously selected annotation classes. The basis for agreement of the annotations can be modified arbitrarily (selected views, classes) and is calculated directly. At the same time, the agreement is highlighted in different colors.

References

Abrami, Giuseppe and Alexander Mehler (2018). "A UIMA Database Interface for Managing NLP-related Text Annotations." In: *Proceedings of the 11th edition of the Language Resources and Evaluation Conference, May 7 - 12 LREC 2018*. Miyazaki, Japan.

Abrami, Giuseppe, Alexander Mehler, Philipp Helfrich, and Elias Rieb (2018). "TextAnnotator: A Browser-based Framework for Annotating Textual Data in Digital Humanities." In: *Proceedings of the Digital Humanities Austria 2018*. Salzburg, Austria.

Abrami, Giuseppe, Alexander Mehler, Andy Lücking, Elias Rieb, and Philipp Helfrich (2019). "TextAnnotator: A flexible framework for semantic annotations." In: *Proceedings of the Fifteenth Joint ACL - ISO Workshop on Interoperable Semantic Annotation (ISA-15)*. ISA-15, Gothenburg, Sweden.

Castilho, Richard Eckart de and Iryna Gurevych (Aug. 2014). "A broad-coverage collection of portable NLP components for building shareable analysis pipelines." In: *Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT*. Dublin, Ireland: Association for Computational Linguistics and Dublin City University, pp. 1-11.

Castilho, Richard Eckart de, Eva Mújdrizca-Maydt, Seid Muhie Yimam, Silvana Hartmann, Iryna Gurevych, Anette Frank, and Chris Biemann (Dec. 2016). "A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures." In: *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*. Osaka, Japan: The COLING 2016 Organizing Committee, pp. 76-84.

Ferrucci, David, Adam Lally, Karin Verspoor, and Eric Nyberg (2009). *Unstructured Information Management Architecture (UIMA) Version 1.0*. OASIS Standard.

Gerdes, Kim (Aug. 2013). "Collaborative Dependency Annotation." In: *Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013)*. Prague, Czech Republic: Charles University in Prague, Matfyzpress, Prague, Czech Republic, pp. 88-97.

Gleim, Rüdiger, Alexander Mehler, and Alexandra Ernst (2012). "SOA implementation of the eHumanities Desktop." In: *Proceedings of the Workshop on Service-oriented Architectures (SOAs) for the Humanities: Solutions and Impacts, Digital Humanities 2012*. Hamburg, Germany.

Helfrich, Philipp, Elias Rieb, Giuseppe Abrami, Andy Lücking, and Alexander Mehler (2018). "TreeAnnotator: Versatile Visual Annotation of Hierarchical Text Relations." In: *Proceedings of the 11th edition of the Language Resources and Evaluation Conference, May 7 - 12 LREC 2018*. Miyazaki, Japan.

Hemati, Wahed, Tolga Uslu, and Alexander Mehler (2016). "TextImager: a Distributed UIMA-based System for NLP." In: *Proceedings of the COLING 2016 System Demonstrations*. Federated Conference on Computer Science and Information Systems, Osaka, Japan.

Hinrichs, Erhard, Marie Hinrichs, and Thomas Zastrow (2010). "WebLicht: Web-based LRT Services for German." In: *Proceedings of the ACL 2010 System Demonstrations*. ACLDemos '10. Uppsala, Sweden: Association for Computational Linguistics, pp. 25-29.

Klie, Jan-Christoph, Michael Bugert, Beto Boulosa, Richard Eckart de Castilho, and Iryna Gurevych (2018). "The INCEPTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation." In: *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, pp. 5-9.

Meyer, Christian M., Margot Mieskes, Christian Slab, and Iryna Gurevych (Aug. 2014). "DKPro Agreement: An Open-Source Java Library for Measuring Inter-Rater Agreement." In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations*. Dublin, Ireland: Dublin City University and Association for Computational Linguistics, pp. 105-109.

Stenetorp, Pontus, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii (2012). "BRAT: A Web-based Tool for NLP-assisted Text Annotation." In: *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*. EAACL '12. Avignon, France: Association for Computational Linguistics, pp. 102-107.

TextAnnotator

www.textannotator.texttechnologylab.org

