Continuous Annotation Quality Control, Support for Hierarchically Structured Label Sets and Long-Segment Annotation with WAT-SL 2.0

Christina Lohr, Johannes Kiesel, Stephanie Luther, Johannes Hellrich, Benno Stein, Udo Hahn

1 Jena University Language & Information Engineering Lab (JULIE Lab), Friedrich-Schiller-Universität Jena, Germany
2 Faculty of Media, Bauhaus-Universität Weimar, Germany

This work was partially supported by the Deutsche Forschungsgemeinschaft (DFG) under grants HA 2097/8-1 within the STAKI2 project (Semantic Text Analysis for Quality-controlled Extraction of Clinical Phenotype Information within the Framework of Healthcare Integrated Biobanking) and the German Federal Ministry of Education and Research (BMBF) within the SMITH project under grant 01ZZ1803G.

Web Annotation Tool for Segment Labeling – Basic Design

• Annotation of long text segments
• Web-based front-end with Java back-end

Four easy setup steps
1. Selection of TXT files
2. Definition of annotation categories
3. Configuration of annotator accounts
4. Tuning colors of annotated categories

Novel Features in WAT-SL 2.0

Annotation Monitoring & Quality Control
• Overview of task and segment completion
• Admin can login as any annotator
• Monitoring time spent on each task
• Computing IAA values (Krippendorff’s α)

Annotation Usability
• Hierarchically structured labels
• Keyboard shortcuts

Clinical Annotation Project
• 1K German clinical documents
• 170K Text segments
• 8 Medical students
• 2 Annotation supervisors
• 4 Annotation iterations
• 6-21 Labels per round

Based on feedback from annotators WAT-SL 2.0 was implemented

Export formats
• CSV files
• ANN files (BRAT-compatible)

Download WAT-SL 2.0 with example annotation data
https://github.com/webis-de/wat

This work was partially supported by the Deutsche Forschungsgemeinschaft (DFG) under grants HA 2097/8-1 within the STAKI2 project (Semantic Text Analysis for Quality-controlled Extraction of Clinical Phenotype Information within the Framework of Healthcare Integrated Biobanking) and the German Federal Ministry of Education and Research (BMBF) within the SMITH project under grant 01ZZ1803G.