WebAnno: a flexible, web-based annotation tool for CLARIN

Richard Eckart de Castilho, Chris Biemann, Iryna Gurevych, Seid Muhie Yimam

This work is licensed under a Attribution-NonCommercial-ShareAlike 4.0 International.
If you are interested in using this material under different conditions, please contact us.
WebAnno – an annotation tool for text

- **Team tool**
  - Allows a distributed team of annotators to work on a corpus
  - Supports different roles within the team (e.g. user / manager)

- **Flexible**
  - Multi-layer annotation with configurable annotation layers
  - Different annotation modes including correction and learning modes

- **Web-based**
  - Available to annotators everywhere, no installation effort
  - All configuration performed through the web interface

- **Platform independent**
  - Platform independent Java-based application

- **Open source**
  - Allows the community to participate
WebAnno – an annotation tool for CLARIN

- Developed based on the requirements of CLARIN F-AG 7…

- … but also used beyond F-AG 7

- … used and recognized beyond CLARIN
  - Search “WebAnno” on Google Scholar
  - See our public users mailing list

- WebAnno is the first annotation tool to supporting WebLicht TCF
  - Worked with TCF developers to improve TCF support updating files!

- WebAnno team is constantly in touch with the community
  - Visit [http://webanno.googlecode.com](http://webanno.googlecode.com) after the talk to participate in our survey!
Annotation examples

Part-of-Speech & syntactic dependencies

Named entities

AC Milan chief executive Adriano Galliani believes that Mario Balotelli is destined for greatness.

Co-reference

One day, a wolf sees her and is overcome by the urge to eat her.

As she is going to deliver the food to her grandmother, the wolf approaches her and asks her where
Main Menu

- **Annotation**
  - Annotate texts from scratch

- **Correction**
  - Review and correct previously annotated documents

- **Automation**
  - Employ integrated machine learning capabilities

- **Curation**
  - Compare annotations from different annotators and merge them

- **Monitoring**
  - Assign workload to annotators and monitor their progress

- **Projects**
  - Create new projects

- **Manage users**
  - Create new user accounts
Workflow of a WebAnno project
Curation

highlight sentences with disagreement

curator’s editor

display annotators color-coded agreement

demo-anno-en/PosDep.tcf

showing 1-5 of 7 sentences

1 Ms. Haag plays Elianti.
Rolls-Royce Motor Cars Inc. said it
expects its U.S. sales to remain steady
at about 1,200 cars in 1990.

3 The luxury auto maker last year sold
1,214 cars in the U.S.

BELL INDUSTRIES Inc. increased its
quarterly to 10 cents from seven cents
a share.

5 The new rate will be payable Feb. 15.
Built-in layers vs. custom layers

- **WebAnno offers various built-in annotation layers**
  - User can immediately start annotating
  - Only linguistic layers
  - Layer semantics are known

- **Custom layers allow WebAnno to be adapted to unforeseen tasks**
  - Adapt to non-linguistic annotation tasks
  - Adapt to unforeseen linguistic annotation tasks
  - Layer semantics are unknown

- **Import/export of annotated data**
  - Layers with known semantics convert from/to many formats (TCF, CoNLL, …)
  - Layers with unknown semantics convert from/to generic formats (XMI, …)
Layer types

- Existing built-in layers were generalized into three layer types

- **Span layer** – POS, lemma, named entity, …

- **Relation layer** – Syntactic dependencies, …
  - Attaches to span annotations
  - Directed, reversible arcs

- **Chain layer** – Co-reference chains, …
  - Undirected arcs

- **Layers can be further customized using “behaviours”**
  - Character-based or token-based
  - Single/multiple token
  - Crossing of sentence boundaries
  - Stacking
Custom layer examples

Semantic predicates and arguments (span/relation)

Person (span) / Relationship (relation)
Custom layer examples

Semantic predicates and arguments (span/relation)

Person (span) / Relationship (relation)
Custom layer configuration

Layers

Control behavior

Features

Controlled vocabulary
Integrated machine-learning

- Annotating data from scratch is more work than correcting
- WebAnno learns from pre-annotated data and makes suggestions
- Accept suggestions with a single click
- Correct suggestions to improve training data
Example: Chunking

- Externally pre-annotated secondary data
  - Part-of-Speech training data
  - POS-tagger model
  - POS tagged text

- Data annotated in WebAnno
  - Chunk-annotated documents

- Externally pre-annotated primary data
  - Chunk training data
  - Chunker model
  - Chunk suggestions
# Automation configuration

<table>
<thead>
<tr>
<th>Details</th>
<th>Users</th>
<th>Documents</th>
<th>Layers</th>
<th>Tagsets</th>
<th>Guidelines</th>
<th>Export/Import</th>
<th>Automation</th>
</tr>
</thead>
</table>

**Select automation layer**

- [Chunk] value

- **Start automation**

**Target layer** | **TAB-SEP target** | **Other layers** | **TAB-SEP feature**

**Import new documents**

- **Format**: TAB-SEP
- **Files**: Choose Files, no files selected

**Documents**

- test_chunk.txt
- train_chunk.txt

---

**Training data example**

- Rockwell B-NP
- International I-NP
- Corp. I-NP
- 's B-NP
- Tulsa I-NP
- unit I-NP
- said B-VP
- it B-NP
- signed B-VP
- a B-NP
- tentative I-NP
- agreement I-NP
- extending B-VP
- its B-NP

**Primary training data**

**Secondary training data**
Deploy WebAnno as you need it

click to start

webanno-standalone.jar

personal workstation

migrate projects

on-premise group server

to come…

cloud-based group server

CLARIN infrastructure service
Where we want to go from here…

- **Extend the scope of WebAnno**
  - Support for slot-based annotation layers (semantic annotations)
  - Tagset constraints
  - Support for more built-in linguistic layers

- **Improve continuously based on user feedback**
  - More efficient annotation interface
  - Support for additional corpus formats
  - … your feedback?

- **Deploy as a CLARIN infrastructure service**
  - CLARIN AAI support
  - Reduce administrative overhead for operators
  - Self-service for project managers
WebAnno Community Survey

This survey is designed to help us understand how people use WebAnno or plan to use WebAnno. So if you are not already a user of WebAnno, we still appreciate your feedback. In this case, please answer the questions such as they reflect your intended use. If you find any of the questions not to be applicable, please simply leave it empty.

Thank you for participating in our survey!

http://webanno.googlecode.com

Visit me in the demo session!