Visualization, Search, and Error Analysis for Coreference Annotations
Markus Gärtner, Anders Björkelund, Gregor Thiele, Wolfgang Seeker and Jonas Kuhn
Institut für Maschinelle Sprachverarbeitung, University of Stuttgart
firstname.lastname@ims.uni-stuttgart.de

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG) via Project D8 of the SFB 732 and by the Bundesministerium für Bildung und Forschung (BMBF) via project No. 01UG1120F, CLARIN-D center Stuttgart.

EXPLORATION VIEWS

Text View
- Textual representation with color markup
- Customizable formatting and text properties
- Filter out singletons or clusters that are of no interest

Entity Grid View
- Tabular view inspired by Barzilay and Lapata’s Entity Grid [1]
- Lists entities as columns and sentences as rows
- Cell contents customizable via Label Patterns
- Graphical error summary when used for error analysis

Tree View
- Displays clusters as subtrees of a virtual document root
- Node and edge text customizable via Label Patterns
- Export for several formats (svg, png, xml,...)

Both the entity grid and tree views support Label Patterns to customize text content. Those patterns are strings that define the format according to which a mention will be displayed. They allow various properties of a mention to be used as label text (e.g. $form$ extracts the full surface form of a mention, whereas #form# would only extract the surface form of the head word of a mention). Switching between exploration views requires only a simple click and selections or filtering of mentions can be preserved when switching to another view.

ERROR ANALYSIS

Trees representing (a) system output, (b) correct annotation, and (c) a merged version of the other two with highlighting of their individual differences (i.e. the errors made in the prediction)

- Assign multiple stand-off coreference annotations per data set
- Select up to two annotations for comparison in the exploration views
- Fine-grained analysis with 5 different error types (false positive, false negative, foreign antecedent, invented antecedent, invalid cluster root)
- Specialized visualizations available for grid and tree views
- Experimental prototype of a quantitative error breakdown (needs further work)

SEARCH

Example search query for cataphoric pronouns and corresponding result overview
- Interfaces with the built-in search engine of ICARUS [2]
- Enables searches on sets of documents
- Express queries graphically or in plain text
- Inspect search results with any of the 3 available exploration views

SUMMARY
- Multiple exploration views for coreference annotations
- Interactive graphical interface integrated in ICARUS [2]
- Fine-grained comparison of different data sets
- Supports various levels of user expertise
- Highly customizable visualization
- Java-based, platform independent, requires no installation

The latest version can be found here: http://www.ims.uni-stuttgart.de/data/icarus.html

REFERENCES