CUTE – CRETA Un-/Shared Task on Entity References

Call for submissions

We invite for contributions to an shared/unshared task workshop, to be held during DHd 2017, Bern, Switzerland or shortly thereafter in Stuttgart, Germany[1]. While *shared tasks* are used as a direct benchmark for different systems/approaches/methods on a clearly defined and evaluated task, *unshared tasks* are open to various kinds of contributions, based on a common data set. Shared and unshared tasks in the fields of digital humanities are a promising way of fostering collaboration and interaction between Humanities scholars and Computer Science researchers.

Specifically, we invite scholars, researchers and scientists to collaboratively work on a heterogeneous, German-language corpus that has been annotated with entity references (see below). The corpus comprises the following texts:

- One speech from four different debates from the German national parliament (Bundestag): S. Leutheuser-Schnarrenberger on Oct. 28, 1999; A. Merkel on Dec. 16, 2004; A. Ulrich on Nov. 15, 2007 and A. Karl on March 17, 2011
- Letters from Goethe's *The Sorrows of Young Werther* (1787) from May 4th to June 16
- The segment titled *Zur Theorie des Kunstwerks* from Adornos *Ästhetische Theorie*
- books 3 to 6 of Wolfram's von Eschenbach *Parzival*

Each text (and genre) has its own characteristics. Nevertheless, all annotators followed the same uniform annotation guidelines which will also be released and discussed.

We invite contributions to one of the following tasks:

1. **Automatic entity reference detection**: Experiments on automatically predicting annotations on unseen texts, using rule-based or statistical systems

2. **Visualising entity references in text**: Visualisation options for the (interactive) exploration of the existing or new entity reference annotations

3. **Annotation Analysis**: Qualitative or quantitative analysis of the existing annotations or annotation guidelines, or annotation experiments on the applicability of the guidelines to new texts

4. **Freestyle**: Anything goes

---

[1]Depending on the acceptance of the unshared task as a pre-conference workshop at DHd 2017.

Contributions to task 1 will be evaluated quantitatively and competitively (shared task). Contributions to tasks 2 to 4 will be evaluated qualitatively by the organisation committee (unshared task). Technical details to data formats and evaluation will be available from the workshop web page: http://www.creta.uni-stuttgart.de/index.php/cute.

The corpus and the annotation scheme have been selected with a number of research questions in mind originating from different fields of humanities and social sciences. The submissions are invited to take these questions into account (or get inspired by them), although it is not a submission requirement to pertain to one of these questions.

- Entity networks

  - How connected and dense are networks based on person mentions?
  - Are central persons, protagonists identifiable?
  - Are abstract concepts defined, or are they introduced (and used) in certain constellations, as Adorno claims to do?

- How to compare entity mention structures or networks across different texts (that might be written in different languages)?

  - How different are the character relations in the German *Parzival* and its French original or other Arthurian romances?
  - Can we identify the Werther-triangle (Werther, Lotte, Albert) in Werther-adaptations, even if the names might be different?

- Are entities mentioned in re-occurring contexts or context categories?

  - How relevant are national, trans- or international organisations in the context of parliamentary debates on certain policy areas?
  - Do character mentions (always/often/never) appear in a given thematic context?

- Are there interactions between entities, within the same or across different types?

  - Which political parties tend to refer to which organisations in political debates?
  - Do some characters (only/always) appear at certain locations?

- Can we observe changes related to some of the above aspects over time?

  - Are some (political) organisations more present in parliamentary discourse before and after, e.g., the German reunification?
  - How does the context of character mentions change over the course of a single text?

Submission information

Submissions should be in the form of abstracts of approximately 1,000-1,500 words length. All abstracts should be in PDF format. Details on the submission procedure will be published on the homepage.

In order to evaluate the annotation quality in **task 1**, we will provide further text excerpts (from the four text works mentioned). These then need to be (automatically) annotated within few days and will be evaluated based on manual annotations (by the organisers). Technical details about the format of the annotations data to send in will be available from the workshop homepage.

Important Dates

Task 1

- December 1: test data for task 1
- December 5: submission of annotated test data
- December 10: notification about eavaluation results
- December 15: submission deadline for abstract

Tasks 2–4

- December 15: Submission deadline
- December 19: Notification

Workshop

February 13 or 14 (Bern) or March (Stuttgart), 2017

Organisation and Contact

Contact

Nils Reiter, Gerhard Kremer
e-mail: cute@ims.uni-stuttgart.de

Program Committee

- Nils Reiter, Natural Language Processing
- André Blessing, Natural Language Processing
- Nora Echelmeyer, Medieval Studies

- Steffen Koch, Visualisation
- Gerhard Kremer, Natural Language Processing
- Sandra Murr, Modern German Literature
- Maximilian Overbeck, Social Sciences
- Axel Pichler, Philosophy/Literary Studies

Organisation Committee

- Martin Baumann, Visualisation
- Manuel Braun, Medieval Studies
- Thomas Ertl, Visualisation
- Dominik Gerstorfer, Philosophie
- Markus John, Visualisation
- Cathleen Kantner, Social Sciences
- Evgeny Kim, Natural Language Processing
- Roman Klinger, Natural Language Processing
- Jonas Kuhn, Natural Language Processing
- Thu Le, Natural Language Processing
- Catrin Misselhorn, Philosophy
- Sarah Schulz, Natural Language Processing
- Sebastian Padó, Natural Language Processing
- Thomas Rainsford, Romance Studies
- Sandra Richter, Modern German Literature
- Achim Stein, Romance Studies
- Gabriel Viehhauser-Mery, Digital Humanities
- Claus Zittel, Literary Studies/Philosophy