

CUTE – CRETA Un-/Shared Task zu Entitätenreferenzen

Aufruf zu Beiträgen

Webseite: <http://www.creta.uni-stuttgart.de/index.php/de/cute/>

Wir laden dazu ein, Beiträge zu einem *shared/unshared Task* Workshop einzureichen, der im Rahmen der DHd 2017 in Bern (CH) oder kurz danach in Stuttgart (DE) stattfindet¹. Im Gegensatz zu *shared tasks*, bei denen die Performanz verschiedener Systeme/Ansätze/Methoden direkt anhand einer klar definierten und quantitativ evaluierten Aufgabe verglichen wird, sind *unshared tasks* offen für verschiedenartige Beiträge, die auf einer gemeinsamen Datensammlung basieren. Shared und Unshared Tasks in den Digital Humanities sind ein vielversprechender Weg, Kollaboration und Interaktion zwischen Geistes-, Sozial- und ComputerwissenschaftlerInnen zu fördern und zu pflegen.

Konkret rufen wir dazu auf, gemeinsam an einem heterogenen Korpus zu arbeiten, in dem Entitätenreferenzen annotiert wurden (siehe unten). Das Korpus besteht aus:

- Jeweils einer Rede aus insgesamt vier Parlamentsdebatten des Deutschen Bundestags (S. Leutheuser-Schnarrenberger am 28.10.99, A. Merkel am 16.12.04, A. Ulrich am 15.11.07 und A. Karl am 17.03.11)
- Briefen aus Goethes *Die Leiden des jungen Werther* (Ausgabe von 1787) vom 4. Mai bis einschließlich 16. Juni
- Dem Abschnitt *Zur Theorie des Kunstwerks* aus Adornos *Ästhetischer Theorie*
- Den Büchern 3 bis 6 von Wolframs von Eschenbach *Parzival* (mittelhochdeutsch)

Auch wenn jede Textsorte ihre eigenen Besonderheiten hat, wurden alle nach einheitlichen Annotationsrichtlinien annotiert, die wir ebenfalls veröffentlichen und zur Diskussion stellen möchten.

Wir laden ein, Beiträge zu folgenden Aufgaben einzureichen:

1. **Automatische Erkennung von Entitätenreferenzen:** Experimente zum automatischen Erkennen von Annotationen auf noch nicht annotierten Texten, mit regelbasierten und/oder statistischen Systemen
2. **Visualisierung von Entitätenreferenzen im Text:** Visualisierungsmöglichkeiten zur (interaktiven) Exploration der vorhandenen oder neuer Annotationen

¹Das hängt davon ab, ob der “Unshared Task” als vorangehender Workshop bei der DHd 2017 angenommen wird.

3. **Annotationsanalyse:** Qualitative oder quantitative Analyse der vorhandenen Annotationen oder der Annotationsrichtlinien; Annotationsexperimente zur Anwendbarkeit der Richtlinien auf neue Texte
4. **Freestyle:** Kreative Ideen, die sich nicht direkt einem der obigen Tasks zuordnen lassen

Beiträge zu Aufgabe 1 werden quantitativ evaluiert und im Wettbewerb mit den Evaluationsergebnissen der anderen Beiträge verglichen (Shared Task). Beiträge für die Aufgaben 2 bis 4 werden vom Organisationskomitee qualitativ evaluiert (Unshared Task). Technische Details zu den Datenformaten und zur Evaluation sind auf der Webseite abrufbar: <http://www.creta.uni-stuttgart.de/index.php/cute>.

Das Korpus sowie das Annotationsschema wurden vor dem Hintergrund von Fragestellungen aus Geistes- und Sozialwissenschaften zusammengestellt bzw. entwickelt. Die Einreichenden können diese Fragen miteinbeziehen (oder sich von ihnen inspirieren lassen), auch wenn das keine Voraussetzung für die Teilnahme ist.

- Entitäten-Netzwerke
 - Wie sind unterschiedliche Personen innerhalb eines bestimmten Textkorpus verbunden?
 - Sind zentrale Figuren identifizierbar?
 - Würden in Theodor W. Adornos Texten abstrakte Konzepte – wie Adorno es selbst forderte – hinreichend definiert bzw. in bestimmten Konstellationen eingeführt (und benutzt)?
- Wie können Strukturen oder Netzwerke der Entitäten-Erwähnungen textübergreifend verglichen werden (auch unter Einbezug unterschiedlicher Sprachen)?
 - Wie unterscheiden sich beispielsweise die Figurenbeziehungen des deutschen *Parzival* von seiner französischen Vorlage oder von anderen Artusromanen?
 - Ist die Werther-Dreiecksbeziehung (Werther, Lotte, Albert) in Werther-Adaptionen identifizierbar, auch wenn die Namen anders lauten?
- Tauchen Entitäten in sich wiederholenden Kontexten oder Kontextkategorien auf?
 - Wie relevant sind nationale, trans- oder internationale Organisationen in Parlamentsdebatten zu bestimmten Politikbereichen, wie z.B. der deutschen Außenpolitik?
 - Tauchen bestimmte Figuren-Erwähnungen (immer/oft/nie) in einem gesuchten thematischen Kontext auf?
- Kommt es zu Interaktionen zwischen Entitäten desselben Typs bzw. typübergreifend?
 - Welche politischen Parteien neigen in politischen Debatten dazu, sich auf welche Organisationen zu beziehen?
 - Erscheinen manche Figuren (nur/immer) an bestimmten Orten?

- Sind zu manchen der oben genannten Aspekte Veränderungen im Zeitverlauf feststellbar?
 - Sind manche (politische) Organisationen im parlamentarischen Diskurs vor oder nach einem Ereignis stärker präsent, z.B. im Zuge der Deutschen Wiedervereinigung?
 - Wie ändert sich der Kontext von Figuren-Erwähnungen innerhalb eines einzelnen Texts?

Details zu den Einreichungen

Beiträge sollen in Form von Kurzbeschreibungen im Umfang von ca. 1.000 bis 1.500 Wörtern eingereicht werden. Die eingereichte Kurzbeschreibung ist im PDF-Format einzureichen. Details dazu werden auf der Workshop-Homepage veröffentlicht.

Zur Evaluation der Annotationsqualität in **Task 1** werden wir weitere Textauschnitte (aus den vier genannten Werken) bereitstellen, die innerhalb von wenigen Tagen von den verschiedenen Systemen der Beitragenden annotiert werden sollen. Wir werden diese im Anschluss anhand von manuellen Annotationen evaluieren. Technische Details zu den Datenformaten werden auf der Workshop-Webseite veröffentlicht.

Wichtige Termine

Task 1

- 1. Dezember: Testdaten für Task 1
- 5. Dezember: Abgabe der annotierten Testdaten
- 10. Dezember: Benachrichtigung über Evaluationsergebnisse
- 15. Dezember: Einreichungsfrist Abstract

Tasks 2–4

- 15. Dezember: Einreichungsfrist Abstract
- 19. Dezember: Benachrichtigung über Annahme

Workshop

13. oder 14. Februar (Bern) oder März (Stuttgart) 2017

Organisation und Kontakt

Kontakt

Nils Reiter, Gerhard Kremer
E-Mail: cute@ims.uni-stuttgart.de

Organisationskomitee

- Nils Reiter, Computerlinguistik
- André Blessing, Computerlinguistik
- Nora Echelmeyer, Literaturwissenschaft (Germanistische Mediävistik)
- Gerhard Kremer, Computerlinguistik
- Steffen Koch, Visualisierung
- Sandra Murr, Literaturwissenschaft (Neuere Deutsche Literatur)
- Maximilian Overbeck, Sozialwissenschaften
- Axel Pichler, Philosophie/Literaturwissenschaft

Programmkomitee

- Martin Baumann, Visualisierung
- Manuel Braun, Literaturwissenschaft (Germanistische Mediävistik)
- Thomas Ertl, Visualisierung
- Dominik Gerstorfer, Philosophie
- Markus John, Visualisierung
- Cathleen Kantner, Sozialwissenschaften
- Evgeny Kim, Computerlinguistik
- Roman Klinger, Computerlinguistik
- Jonas Kuhn, Computerlinguistik
- Thu Le, Computerlinguistik
- Catrin Misselhorn, Philosophie
- Sarah Schulz, Computerlinguistik
- Sebastian Padó, Computerlinguistik
- Thomas Rainsford, Romanistik
- Sandra Richter, Literaturwissenschaft (Neuere Deutsche Literatur)
- Achim Stein, Romanistik
- Gabriel Viehhauser-Mery, Digital Humanities
- Claus Zittel, Literaturwissenschaft/Philosophie