



Universität Stuttgart

Interaktive Annotation von Entitäten in CRETA

Dr. Axel Pichler
Maximilian Overbeck

CRETA 

CENTER FOR REFLECTED TEXT ANALYTICS

- I. Die erste CRETA-Phase: Annotation von Entitätenreferenzen

- II. Illustration des Annotationsworkflows
 - Plenarprotokolle des deutschen Bundestages
 - Theodor W. Adornos *Ästhetische Theorie*

- III. Kritische Reflexion: Möglichkeiten und Grenzen in der Generierung eines *universell* applizierbaren Annotationsschemas

I. Ausgangssituation in der ersten CRETA-Phase

Annotation von Entitätenreferenzen

- Warum Entitäten?
 - Einfache Konzepte
 - Geringe Kontextabhängigkeit für eine valide Bestimmung von Referenzen auf Entitäten
 - Möglichkeit einer validen, automatischen Erkennung von Referenzen auf Entitäten
 - Generierung großer Datenmengen

I. Ausgangssituation in der ersten CRETA-Phase

Annotation von Entitätenreferenzen

- Allgemeines Kategorienschema
- Entitäten = »spezifische, durch Benennung unterscheidbare Objekte in einer echten oder fiktiven Welt«
- Annotation v. *Personen, Organisationen, Orten, Ereignissen, Werken, abstrakten Konzepten*
 - Eigennamen: z.B. »der Warschauer Pakt«
 - Gattungsnamen: z.B. »die 300 Tibeter, die vergangenen Mittwoch protestierten«
 - Referenzausdrücke als maximale Nominalphrasen
 - Ausschluss von Personalpronomina: »Wir«, »Er«, »Sie« usw.

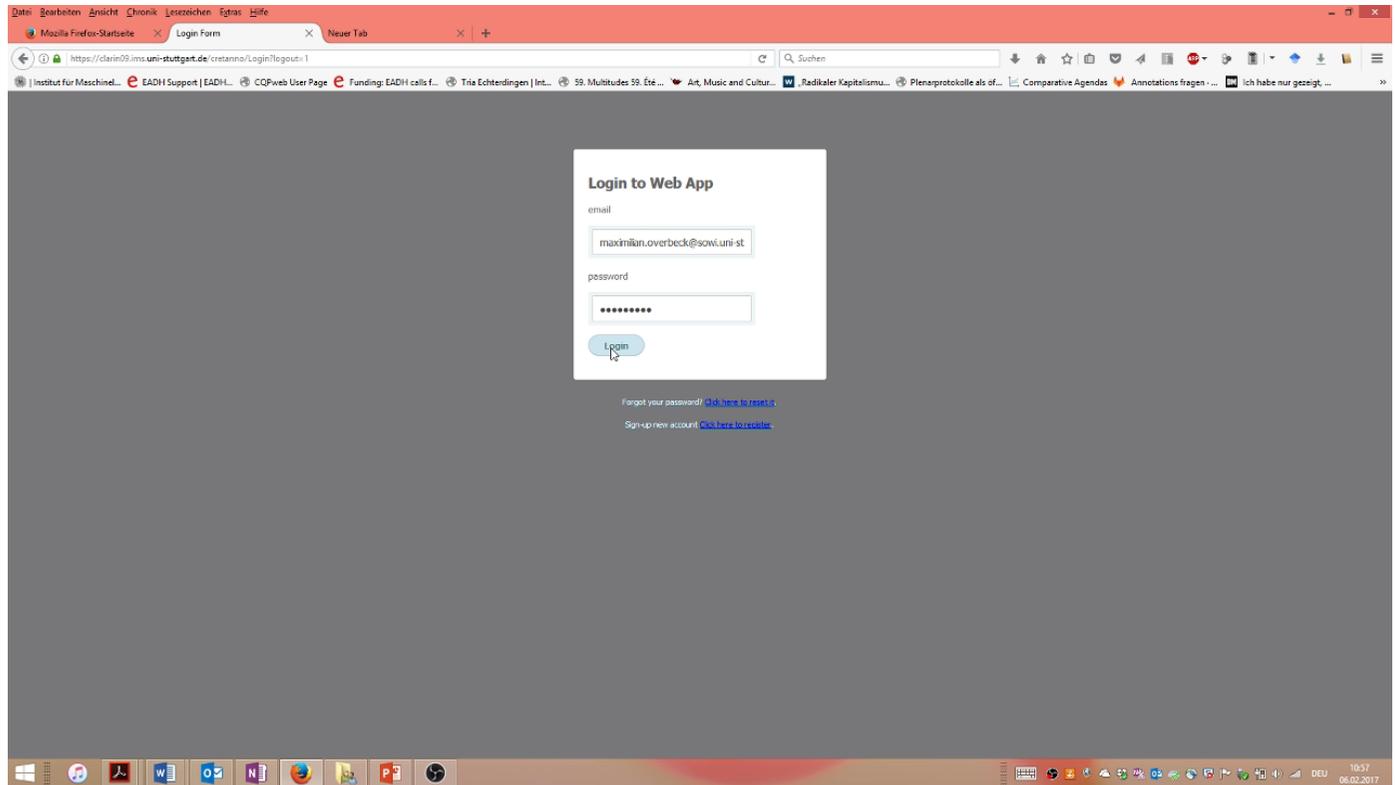
II. Workflow

Sozialwissenschaftlicher Schwerpunkt

- Entwicklung einer fachspezifischen Fragestellung
- Anwendung der CRETA-Richtlinien auf das Testmaterial
- Erstellen von fachspezifischen Zusatzrichtlinien
- Annotieren
- Wöchentlicher Austausch
- Festlegen eines Goldstandards

II. Workflow

Von der Annotation in CRETanno...



II. Workflow

... hin zum Festlegen eines Gold-Standards.

The screenshot shows a web browser window with the URL https://clarin09.ims.uni-stuttgart.de/cretaanno/GetArticle?doc_id=68&renderer_type=2. The page displays a document with line numbers on the left and text on the right. The text is in German and appears to be a historical document. The browser's address bar shows the URL, and the page title is "CRETAAnno". The browser's taskbar at the bottom shows various application icons and the system clock indicating 11:07 on 06.02.2017.

1 Ist zwelvet herzen nächgebür.
daz muoz der sêle werden sôr.
gesmachet unde gezieret
ist, swâ sich parmeret
5 unverzaget mannes muot,
als agestlern varve tuot.
der mac dennoch wesen geil
wand an im sint beidiu teil,
des himels und der helle:
10 der unstaete geselle
hât die swarzen varve gar,
und wirt och nâch der vinstler var:
sô habet sich an die blanken
der mit staeten gedanken.
15 diz vliegende bîspel
ist tumben luten gar ze snel,
sine mugens niht erdenken:
wand ez kan vor in wenken
rehte alsam ein scheleic hase.
20 zin anderhalb ame glase
geleichet, und des blinden troum,
die gebent antlûzes roum,
och me mit staete niht gesin
dirre triebe ihle schîn.
25 er machet kurze vroude altdâr,
wer routet mich dâ nie kein hâr
gewuots, inne an mîner hant?
der hât vil nâhe gratte erkant
sprich ich gein den vorhten och,
daz gleichet mîner wilze doch.

2 wil ich triue vinden
aldâ si kan verswinden,
als viur in dem brunnin
und daz tou von der sunnen?
5 ocht erkant ich nie sô wîsen man,
ern mohte gerne kûnde hân,
weilher sture dislu maere gemt
und waz si guoter lère vernt,
dar an si niemer des verzagent,
10 beidiu si vliehent unde jagent.

II. Workflow

Typische Auslöser für Annotationsdivergenzen

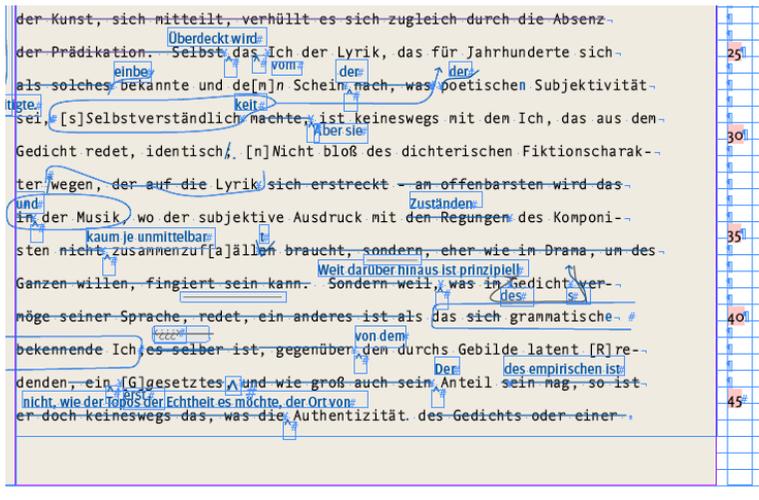
- Falsche Anwendung der Annotationsrichtlinien
- Ambiguitäten in den Annotationsrichtlinien
- Ambige Textstellen (*vgl. Gius & Jacke 2016*)
 - Metonymie: Die Deutschen [ORG] zwingen den Griechen [ORG] einen harten Sparkurs auf.
 - Multireferentialität: Europa steht vor einer großen Herausforderung

II. Workflow

Adornos *Ästhetische Theorie*

Ausgangssituation:

- Ziel 1: Hybride Textedition von Adornos *Ästhetischer Theorie*



- Ziel 2: Analyse der *Ästhetischen Theorie* mit Hilfe von DH-Tools

II. Workflow

Adornos *Ästhetische Theorie*

Adaption der CRETA-Annotationsrichtlinien in Hinblick auf Adornos *Ästhetische Theorie*:

- Überprüfung des heuristischen Potentials der für CRETA festgelegten Entitätskategorien.
- **Fazit:** Nur eine der sechs Kategorien (Personen/Figuren, Orte, Werke, Organisationen, Ereignisse, abstrakte Konzepte) ist von hoher interpretatorischer Relevanz: die sog. *abstrakten Konzepte*. Im Falle der *Ästhetischen Theorie* handelt es sich bei den *abstrakten Konzepten* mehrheitlich um Begriffe.

II. Workflow

Adornos *Ästhetische Theorie*

Bestimmung der heuristisch relevanten Begriffe:

- **Problem:** Eine genaue Bestimmung dessen, was als textinterpretatorisch relevantes Konzept zu erachten ist, ist schwierig. Zahlreiche der in Adornos *Ästhetischer Theorie* relevanten Konzepte entsprechen nicht der gängigen philosophischen Bestimmung eines Begriffs. Das gilt insbesondere für Ausdrücke aus dem Kontext der Kunstinterpretation.
- **Lösung:** Annotation zentraler Begriffe nach dem Begriffsregister der Leseausgabe.

III. Kritische Reflexion

Vorteile

- Quantifizierbarkeit/große Textmenge
- Empirische Validierung von Theorien
- Möglichkeit der digitalen Weiterverarbeitung der Daten
- Möglichkeit der Anknüpfbarkeit für ähnliche Fragestellungen

III. Kritische Reflexion

Probleme

- Grundproblem: Theoriebeladenheit der Annotationsrichtlinien und -resultate
- Die Entitätenannotation folgt nicht aus einer explizit am Gegenstand entwickelten Fragestellung, sondern aus der Grundintention eine allgemein verfügbare und weiterverwertbare Datenmenge zu generieren.
- Die Bestimmung der Annotationsrichtlinien folgt bereits bestimmten theoretischen Vorannahmen, die im Zuge der Adaption der Annotationsrichtlinien in Hinblick auf die jeweiligen Textkorpora noch zunehmen.

III. Kritische Reflexion

Konsequenzen

- Problem 1: Genese der Daten muss stets mit den Daten zur Verfügung gestellt werden. Bedarf potentiell einer »Übersetzung der Daten«.
- Problem 2: Daten selbst sind bereits Resultat bestimmter Vorannahmen und dementsprechend nicht für Fragestellungen, die mit diesen Vorannahmen nicht übereinstimmen, von Relevanz.

Vielen Dank!



Maximilian Overbeck

Institut für Sozialwissenschaften

Abt. III: Internationale
Beziehungen und Europäische
Integration



Axel Pichler

Stuttgart Research Centre for Text
Studies