

Koreferenz mit BART

Spezifikationsvortrag zum Softwareprojekt im Sommersemester
2014

Julian Baumann, Xenia Kühling, Sebastian Ruder

27. Mai 2014

Inhalt

- 1 Einführung
- 2 BART
- 3 Stanford Sieves
- 4 Module und Aufgaben
 - Module
 - Aufgaben
- 5 Zeitplan
- 6 Softwarespezifikation
- 7 Quellen

evtl Koreferenz

■ Erklärung Koreferenz

BART

- Beautiful Anaphora Resolution Toolkit
- Entstanden im Projekt
Exploiting Lexical and Encyclopedic Resources For Entity Disambiguation im John Hopkins Summer Workshop 2007
- System für automatische Koreferenzresolution
- Weiterentwicklungen im Rahmen von shared tasks, für verschiedene Sprachen

Wie funktioniert BART

- Modularer Aufbau:
- Vorverarbeitungsphase
- Extraktion NP- Kandidaten, NP- Merkmale, Kandidatenpaare

Wie funktioniert BART

- Resolution mit Soon Algorithmus
- Kandidatenpaare werden paarweise anhand ihrer Merkmale verglichen
- Ergebnisse

Aufgabe

- Resolution in BART mit System der Stanford-NLP-Gruppe
- Bestes Ergebnis bei CoNLL-2011 shared task, obwohl hauptsächlich regelbasiert

efregrgoiu

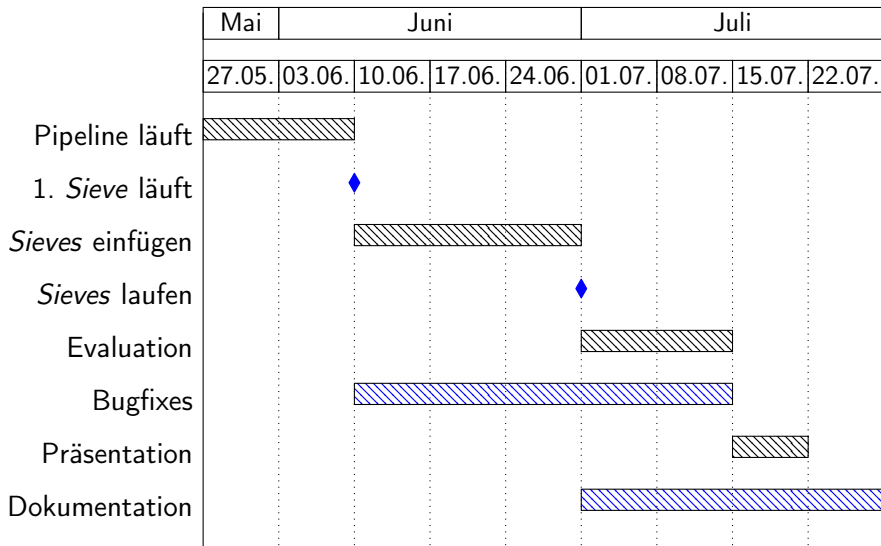
Aufgabenverteilung

bis 10.06.: Aufteilung der Pipeline

- *DiscourseEntity*: Julian Baumann
- *Sieve & StringMatchSieve*: Xenia Kühling
- *SieveDecoder*: Sebastian Ruder

ab 10.06.: Aufteilung der Sieves

- *RelaxedStringMatchSieve*, *PreciseConstructsSieve*,
(*SpeakerIdentificationSieve*)
- *StrictHeadMatch[ABC]Sieve*, *RelaxedHeadMatch*
- *ProperHeadNounMatch*, *PronounMatch*



Softwarespezifikation

- Datenformate
- Interfaces
- Datenstrukturen
- benötigte Ressourcen, verwendete Algorithmen