Projektgruppe 2

C. Ahrens, A. Beyer, C. Michels

Fachbereich II – Computerlinguistik und Digital Humanities

Universität Trier

9. Februar 2015

Übersicht

Übersicht

Was ist Barth 2.0?

Barth

- Barth
    - Barth
- Barth
- Barth

Was ist Reconcile?

Reconcile

- Reconcile
    - Reconcile
        - Reconcile
- Reconcile
- Reconcile

Was ist DCoref?

DCoref

- deterministisches Modul der Stanford CoreNLP
  - hängt von anderen Modulen ab (pos, lemma, ner, parse)
  - diese Module müssen in der Pipeline vorher angewendet werden
- arbeitet mit Sieben
  - erste Stufe: bevorzugt Recall (detection)
  - zweite Stufe: Siebe bevorzugen Precision
- Post-Processing: mehr Precision (?)

Übersicht

Suche nach dem besten Ansatz

Vergleichsgegenstände

- Reconcile
- DCoref

Vergleichspunkte

- Precision
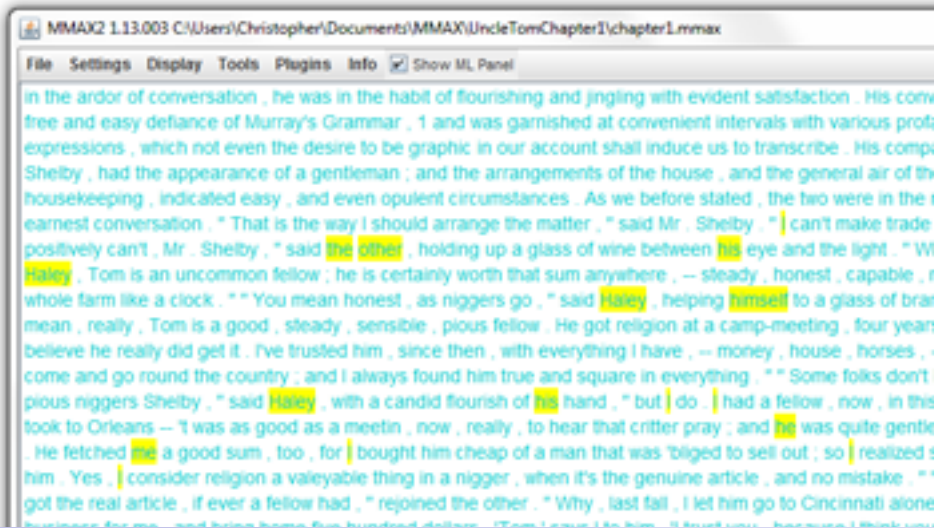- Recall
- FScore

Suche nach dem besten Ansatz

Werkzeug: MMAX

- Schwierigkeiten bei der Verwendung
    - Vergleich war entsprechend auch schwierig

Testdaten: Uncle Tom's Cabin

- erste Hälfte des ersten Kapitels, Reintext
    - bis zu dem Wort *Wilberforce*
    - eindeutige Grenze

"Version Space (Angepasst)". *Wikimedia Commons.*

Formatieren: Reconcile

Reconcile

- Formatieren
    - Formatieren
    - Formatieren

Formatieren: DCoref

DCoref

- Formatieren
    - Formatieren
    - Formatieren

Indizieren: Reconcile

Indizieren
- Indizieren
    - Indizieren

Indizieren: DCoref
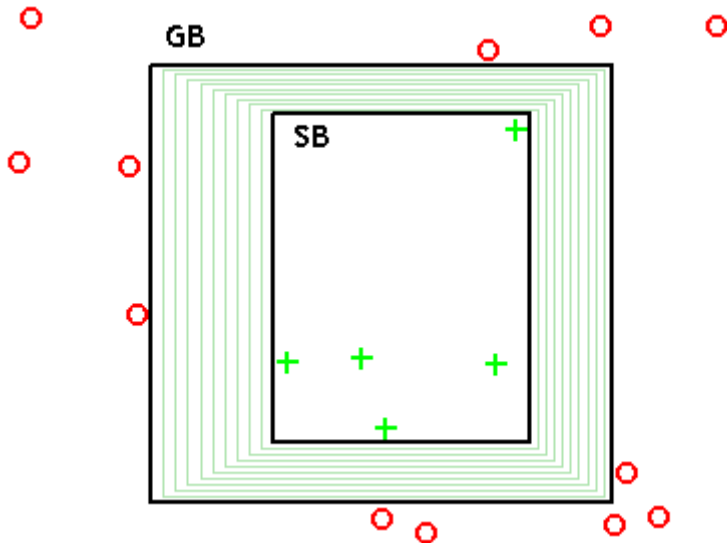
Indizieren

- Indizieren
  - Indizieren

Übersicht
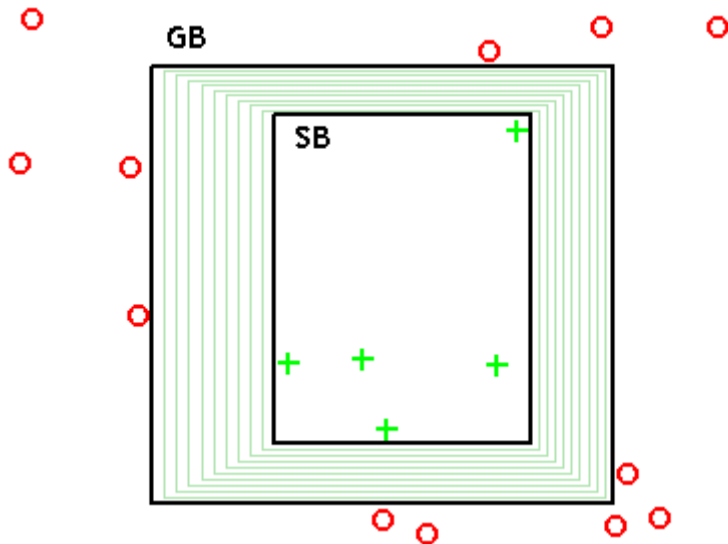
"Version Space (Angepasst)". *Wikimedia Commons.*

"Version Space (Angepasst)". *Wikimedia Commons.*

Zeit für Fragen

Gibt es noch offene Fragen oder Dinge, die unklar geblieben sind?

Zum Schluss

Vielen Dank für eure Aufmerksamkeit!