

Ausblick I: nächste Schritte

Nächste Schritte:

- ▶ erweitern der Metadaten-Datenbank mit Metadaten:
 - ▶ aus bestehenden XML files
 - ▶ aus zusätzlichen Quellen (z.B. Sitzverteilungen)
- ▶ Überführen der Strukturierten Daten im XML-Format in einen Document Store
- ▶ Analyse der Daten:
 - ▶ Clustering (Top-Down, Bottom-up)
 - ▶ LDA (Latend Dirichlet Allocation)

Ausblick II: Ziele Vorlesungszeit

Ziele bis Ende Vorlesungszeit:

- ▶ XML-Daten in Document-Store ablegen
- ▶ Metadaten-Bank mit weiteren Metadaten erweitern
- ▶ analysieren der Daten mittels mind. zwei Clustering-Verfahren,
- ▶ Cluster mit wahrscheinlich gleichen Schreiber (aber nicht Redner) finden und darstellen

Ausblick III: Ziele Semester

Ziele bis Ende Semester:

- ▶ Analyse mittels LDA
- ▶ Visualisierung der Ergebnisse der LDA-Analysen
- ▶ weitere Cluster-Verfahren nutzen
- ▶ alle (sinnvollen) Ergebnisse vereinen und Darstellen
- ▶ Untersuchung warum manche Analysen fehlerhafte/schlechte Ergebnisse lieferten

Ausblick IV: Erweiterbarkeit

Erweiterbarkeit wenn uns Zeit bleibt:

- ▶ zusätzlichen Metadaten-Quellen auffinden und in die bestehende Metadaten-Datenbank überführen
- ▶ weitere Metadaten erzeugen zB. Pos-tagging, n-gramme und Kookkurrenzen
- ▶ Analyse-Verfahren erweitern
 - ▶ andere Cluster-Algorithmen
 - ▶ LDA mit anderen Parametern
 - ▶ LDA mit anderen Features
- ▶ Visualisieren der Ergebnisse

Ausblick V: Einschränkungen I

Einschränkungen wenn wir nicht alle Ziele schaffen:

- ▶ weniger Analyseverfahren nutzen, zB. nur ein Clusteringverfahren
- ▶ Datenbereinigung verkürzen
- ▶ weniger Metadata als feature nutzen

Ausblick VI: Einschränkungen II

Wichtigste Ziele:

- ▶ Daten in Datenbank strukturiert ablegen
- ▶ Clustering-Verfahren auf unsere Daten Anwenden
- ▶ Interpretation der Ergebnisse

neue Ziele:

- ▶ vermutlich nicht