



Abschlussprojekt

Im Verzeichnis `daten/countries` finden sich Wikipedia-Artikel über Länder. Jeder Landestext hat Abschnitte, die durch „== Abschnitt ==“ überschrieben sind. Erzeugen Sie 20 Satzbeispiele als „Dokumente“. Die Beispiele sollen aus mindestens 5 Ländern und aus mindestens 5 verschiedenen Abschnitten stammen.

Beispiel: 20 Sätze, aus 10 verschiedenen Ländern, davon gehört jeder Satz zu einem der Abschnitte „History“, „Climate“, „Economy“, „Tourism“, „Energy“.

Für 20 diese Beispielsätze sind mindestens 5 Beispielfragen zu entwickeln, um ZWEI der folgenden Funktionalitäten einer semantischen Suche zu vergleichen bzw. zu demonstrieren:

- a) Query Expansion durch Synonyme
 - b) Query Expansion mit word2vect
 - c) Passage Retrieval mit sBERT
 - d) Pseudo Relevance Feedback
-
1. Testen sie die gewählten Techniken mit den erzeugten Sätzen und Fragen.
Ermitteln sie als einfache Metrik, wieviel % der Queries das am ehesten korrekte Dokument als erste Antwort enthalten
 2. Präsentieren Sie in einer Powerpoint die Ergebnisse ihrer Tests und diskutieren Sie anhand der Ergebnisse die gewählten Techniken.

Inhalt der Abschlusspräsentation:

- Welche Daten wurden herangezogen (Länder, Abschnitte)
- Was waren die Fragen
- Welche 2 Methoden
- Vergleich der Ergebnisse
- Diskussion der Ergebnisse