Aantekeningen 10-6-2022

Vragen:

* Structuur: wat is literatuur en wat is resultaat?
* Discussie over het niet begrijpen van idf van de tekst waarover het beschrijft; principe veel data leidt tot begrip;
* Resultaten qua balancing POS vs. genormaliseerde idf: is dit interessant?

Vs. word length?

* Hoe balancing: unigram als deel van meer complexe term
  + Analyse kijken naar wat wordt geannoteerd?
  + Overlappend?
  + Per zin; welke termen moet ik annoteren; welke tokens
  + Per zin scores geven;
  + Concept extractie kan je ook discussie schrijven;
  + Mag ook met de hand erbij
  + Leg uit
  + Was token centraal complex in deel

Introductie: wat gaan we doen?

Motivatie

* Hoofdvraag; deelvraag
* Deelvraag 1: Wat weten we over x?
* Deelvraag 2: Benadering/approach? Hoe onderzoek je?
  + Welke data? Hoe evolueren?
  + Evoluatie/ CLEF achtergrond
  + Best ingewikkeld
  + Abstracts als proxy
* Resultaten
* Wat kunnen we hieruit krijgen?
  + Gaat niet alleen uit lexicale complexiteit, maar ook centrale termen zijn belangrijk
* Terugroepen op de hoofdvraag; deelvragen terugroepen;
* Belang aanroepen
* Let op analyse vs. voorspelling; mag er beide in terugkomen maar scheidt het
* Misschien juist analyse over POS in plaats van voorspellend

Conclusie: Het gaat niet alleen om lexicale complexiteit; het is een andere vraag!

* Discussie eventueel: wat is een concept? Koningin Maxima of Koningin Maxima der Oranje Nassau etc.

Indeling per alinea:

Beginnen: wat is de vraag die we hier stellen?

Eind: wat had de lezer mee moeten krijgen uit dit stuk tekst?

NLP

Concepten heel erg biased door de query

Ir book.com

Search Engines Information Retrieval in Practice

Justin zobel: writing for computer science