

Vortrag 2: Joseph Birkner „Ranking with neural network derived document vectors“. Betreuer: Wenpeng Yin

In dieser Arbeit geht es darum, eine Suchmaschine zu entwickeln, die intellektuelle Vorschläge für Onlinekurse liefert.. Dafür wird ein Korpus mit den Kursbeschreibungen genutzt, das im Rahmen von einer vorigen Bachelorarbeit am CIS erstellt wurde.

Das Thema liegt somit in dem Bereich von IR. Man benötigt also eine Suchanfrage und eine Datenbank mit den Dokumenten und Ranking Algorithmen , um die treffendste Ergebnis zu liefern (ubiquitous vertical search). Dafür werden auch die Metadaten benutzt, die die Information über den User spezifizieren (history search, Alter, Geschlecht etc.).

Es besteht aus zwei Unterbereichen: Repräsentationsoptimierung - DSSM (Deep semantic structured matching) und Passende optimierung - DRMR (Deep relevance matching model).

Das Ziel der Arbeit ist Optimierung der Dokumentationsdarstellung nämlich Entwicklung des System, die eine gute Dokumentendarstellung für eine effiziente Ranking erzeugen kann.

Traditionell benutzt man in IR TF-IDF Maß zur Beurteilung der Relevanz von Termen in Dokumenten. Die Nachteile davon sind aber, dass die Reihenfolge der Wörter ignoriert wird und dass man eine Annahme macht, dass die Wörter unabhängig sind.

Josephs Lösung ist den semantischen Raum des Dokument mit Hilfe von Word2Vec für und DocVec zu kreieren.

Die Aufgaben sind die Dokumentenvektoren (embeddings) zu generieren (30 Dimensional), trainiert mit 1200 Kursbeschreibungen, LSTM Modelle zu trainieren mit unterschiedlichen Daten (vermutlich Wikipedia Korpus, Konferenzen) und das Ganze zu evaluieren.