

## Semantische Suche von juristischen Dokumenten mittels Word Embedding und Netzwerkanalyse

Eine neue Art der Suche nach juristischen Dokumenten zu eruiieren und einen lauffähigen Prototyp zu bauen, ist das Ziel dieser Bachelorarbeit.

Es sollte möglich sein, einen Sachverhalt oder Gerichtsentscheid in die Suche einzugeben. Mithilfe der Netzwerkanalyse oder des Word Embedding-Verfahrens werden dem Benutzer der Suche ähnliche juristische Dokumente angezeigt.

Für den Prototyp wurden circa 100'000 deutsche Gerichtsentscheide vorselektiert. Die Dokumente wurden bereinigt und für das Information Retrieval mit Elasticsearch indexiert. Es wurde eine Netzwerkanalyse mit NetworkX auf den Gerichtsentscheiden durchgeführt. Mithilfe von PageRank und einem eigens für diese Bachelorarbeit implementierten Algorithmus konnten die Gerichtsentscheide nach ihrer Relevanz gewichtet werden.

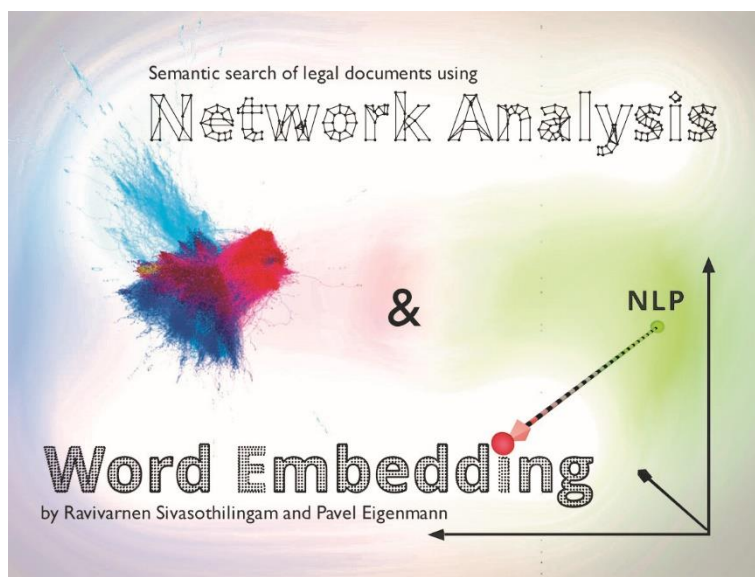
Um eine Suche nach ähnlichen Dokumenten mit einem neu verfassten Sachverhalt möglich zu machen, wurde Word Embedding angewendet. Für die Auswertung der unterschiedlichen Verfahren wurden von einem Juristen mehrere Testfälle erstellt. Ein Testfall beinhaltet die Eingabe eines Sachverhalts oder Gerichtsentscheids und die Ausgabe der zehn ähnlichsten Gerichtsentscheide. Mit dem Kendall-Score konnten die generierten Ranglisten mit den Testfällen verglichen werden. Das fastText-Verfahren mit Skipgram und 300 Dimensionen erreichte das höchste Ergebnis.

Diese Bachelorarbeit wurde in Zusammenarbeit mit dem Wirtschaftspartner Weblaw AG durchgeführt.



Diplomierende  
Pavel Eigenmann  
Ravivarnen Sivasothilingam

Dozent  
Mark Cieliebak



Dieses Bild zeigt links ein Netzwerk über alle Gerichtsentscheide, wobei die Farben unterschiedliche Gerichte darstellen. Rechts wird ein Vektor gezeigt, welcher das Word Embedding illustriert.