

Projektidee NLP

SS2021

Christian Gunzelmann, Diane Mittermair

Fragestellungen

- Textquelle: Heise Newsticker Kategorie Wirtschaft
- Was gibt es für Themen?
- Wie haben sich die Themen im Laufe der Jahre verändert?
- Gibt es hier interessante Auffälligkeiten?
- Gibt es Themen welche mehr diskutiert werden als andere?
- Wie können Auffälligkeiten mit Geschehnissen in Verbindung gebracht werden?
- Wie wurde über ein Thema x Jahre zuvor berichtet?

Idee

- URL: [Heise Online Newsticker Wirtschaft](#) (Zusatz: IT, Wissen)
- Fokus soll auf die sich im Verlaufe der Zeit ändernde Themen und Schwerpunkte sein
- Datenextraktion und -aufbereitung
- Erstellen von Clustern zu bestimmten Zeiträumen (1 Jahr gleitend von 2004 bis 2020)
 - z.B. K-Means
 - Schwerpunkt- und Themenänderungen sichtbar machen und analysieren
 - Visualisierung des Zeitverlaufs durch sich verändernde Wordcloud
- Welche Themen erzeugen mehr Diskussionen? Gibt es Veränderungen in den Interessen der Leser? (Kommentare mit einbeziehen)
- Finden von ähnlichen Artikel aus vergangenen Jahren