

Source Code zur Bachelorarbeit

Hier befinden sich jeglicher Source Code, der zu der während der Erstellung der Bachelorarbeit verwendet wurde. Dabei ist das Projekt inhaltlich entsprechend der Arbeit gegliedert.

Nachfolgend ist eine Beschreibung aller Inhalte dieses Ordners.

Ordner: data_selection

Zeitliche Selektion (Kapitel 3.3.3.1) - date_selection.py: Auswahl der Artikel basierend auf den Veröffentlichungsdatum

Inhaltliche Selektion (Kapitel 3.3.3.2) - topicmodelling_eval.py: Evaluierung der Topic Modelling Modelle - topicmodelling_train.py: Training verschiedener Topic Modelling Modelle mit Grid Search - topicmodelling_selection.py: Selektion der Artikel anhand der identifizierten Topics

Zudem befinden sich hier auch alle trainierten Modell im **SubOrdner: Ida_models** und eine csv mit den Ergebnissen der Evaluierung.

Ordner: sentiment

Sentiment Analyse (Kapitel 3.3.4) -evaluation.py: Bestimmung des Median Absolute Error zwischen manuell bestimmten Tonalität und automatisch bestimmten - labeling_process.py: Manueller Labeling Process (Commandline Interface) -selection.py: Zufällige Auswahl der Artikel für die vorabdurchgeführte Stichprobe - calc_sentiment.py: Bestimmung der Tonalität und speichern in der Datenbank -evaluation_other_approaches.py: Bestimmung der Tonalität der alternativen Ansätze und Erstellung des Plots

Ordner: translation

Standardisierung (Kapitel 3.3.2) -translate.py: Übersetzen des Textes der deutschen Artikel -translate_headlines.py: Übersetzen der Überschriften der deutschen Artikel (relevant für ein Ansatz der Tonalitätsbestimmung)

Ordner: webscraper

Webscraper (Kapitel 3.3.1.2) -Hier sind alle Webscraper enthalten. Zudem auch während dem Scraping erzeugt Logs zum nachvollziehen, ob der Scrape erfolgreich geklappt hat

Ordner: outdated

Hierin sind Ansätze die anfangen wurden zu implementieren aber nicht in die Arbeit geschafft haben. Darunter fällt ein Ansatz zur Datenselektion mit Doc2Vec und der Scrape der Quelle Twitter.

Ordner: plots

Hierin sind alle Skripte für die Erstellung der Plots. Zudem im **Subordner: images** auch die Plots abgelegt als png.

Ordner: resources

Hierin sind alle Resources (Skripte & Files) enthalten, die oft in anderen Skripten verwendet werden. -database.py: Regelung des Datenflusses mit der Datenbank (mongodb) -keys.py: Bereitstellung der verwendeten Keys (API und Passwörter), tatsächliche Key-File allerdings nicht in der Abgabe aus Datenschutzgründen. - logger.py: Implementierung eines Loggers -preprocessing.py: verschiedenste Preprocessing funktionen

Zusätzlich sind hier die Antworten des Survey zur Interpretation der Wordclouds (für besser verständliche Plots)

File: requirements.txt

-liste aller verwendeten Python Libraries, die gemäß pip install -r requirements.txt installiert werden können

File: .gitignore

-Spezifikation aller Files die nicht von Git versioniert werden sollen