

Datenmodell

Das vorliegende Modell beschreibt die aus unserer Sicht vorliegenden fachlichen Daten. Diese sind grob in drei Kategorien zu unterteilen: Rohdaten, verarbeitete Daten, sowie applikationsspezifische Daten. Da das Ziel zunächst ein Prototyp ist fließt dieses in die Entwicklung des Modells mit ein.

Die Rohdaten sind die in der Datenextraktion gesammelten Daten aus den Social Media, sowie potenziell weitere Informationsquellen (z.B. Pegelstände, Trenderkennung etc.).

Die Gesamthierarchie der Daten setzt sich daher auf verschiedenen Ebenen zusammen:

Grundsätzlich besitzen Daten einen eindeutigen Identifikator und eine Iteration, welche es uns ermöglicht die Datensätze grob nach Erfassungsdatum zu selektieren.

Daten aus den Sozialen Medien besitzen im Modell zusätzlich einen Autor, dieser ist nur ein Text welcher einen Namen darstellt (diese Einschränkung ist nötig um nicht direkt personenbezogene Daten zu sammeln). Des Weiteren besitzt jeder Datensatz der sozialen Medien ein spezifisches Veröffentlichungsdatum.

Da sich unser Prototyp zunächst auf Twitter als Plattform bezieht nutzen wir im ersten Schritt nur Tweets als Informationsquelle. Diese enthalten im Datenmodell den eigentlichen Inhalt, die Anzahl der aktuellen Retweets und Follower (des Nutzers), sowie die Geolocation und einen vom Absender definierten Wohnort, sowie eine Referenz auf einen möglichen ursprünglichen Tweet. Diese Angaben sind allerdings nur dann vorhanden wenn sie vom Tweetverfasser zur Verfügung gestellt wurden. Darüber hinaus verfügt jeder Tweet über beliebig viele Hashtags, welche ihm zugeordnet sind.

Die verarbeiteten Daten sind zunächst eine reine Kopie der Rohdaten und werden durch die Analysegruppe durch gewonnene Informationen erweitert. Für das erste Anwendungsszenario soll daher zunächst jedem Tweet eine Kategorie zugewiesen werden. Diese soll grob den Inhalt des Tweets wiedergeben. Weitere Informationen sind ebenfalls geplant.

Die dritte Kategorie, welche im Modell genannt wird gehört zur eigentlichen Applikation. Diese enthält ein Dashboard mit Informationen welche in beliebig vielen Widgets dargestellt werden. Darüber hinaus sind für die Applikation Nutzer samt eines rudimentären Authentifikationsverfahrens anhand einer E-Mail Adresse und eines Passworts vorgesehen.

Für den bisherigen Prototypen fallen alle genannten Informationen auf einer Ebene zusammen, sodass die Datenbanktabelle dementsprechend gestaltet ist.