

Aufspüren von Technologietrends

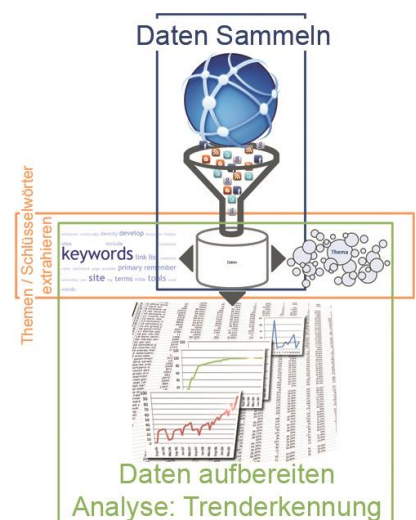
Eric Schmidt, CEO von Google, schätzt die Datenmenge im Internet auf grob 5 Millionen Terabytes. Seit sich das Internet zum Mitmach-Web weiterentwickelt hat, fließen die Informationen noch schneller als zuvor. Es geht so weit, dass einige Informationen nur im Internet erhältlich sind. Unser Trendscouting-System nutzt diesen grossen Datenpool, um Technologietrends zu finden. Zur Beobachtung von Entwicklungstendenzen haben wir Indikatoren definiert und in das System eingebaut. Es misst die Anzahl Artikel aus RSS-Feeds, die zu einem Thema geschrieben wurden und überprüft die Popularität auf Facebook, indem es die Like-Button-Klicks sammelt. Anhand der charakteristischen Begriffe, die ein bestimmtes Thema auszeichnen, wird auf Google Trends überprüft, ob die Google Nutzer vermehrt nach diesen Begriffen gesucht haben. Aus den Artikeln erkennt das System automatisch neue Themen und wichtige Schlüsselbegriffe, die mit den Indikatoren als Grundlage für die Trenderkennung dienen. Dafür werden diverse Information Retrieval-Techniken eingesetzt. Der k-Means-Algorithmus wird für das Clustering der gesammelten Texte eingesetzt. Für die Klassifizierung von Dokumenten kommt der Naive-Bayes-Algorithmus zum Einsatz und die Schlüsselwörter werden mittels TF-IDF ermittelt. Zur Trenderkennung baut das System für einige Indikatoren Zeitreihen und berechnet den linearen Trend. Eine weitere Analyse sucht Hotspots in der Datenkollektion.

Anschliessend werden die spannendsten Resultate dem Benutzer auf einem Dashboard präsentiert. Die interessanten Themen kann er dann markieren und über längere Zeit beobachten. Innerhalb von zwei Monaten wurden zu 99 Newsfeeds 120'000 Artikel gesammelt und über 2 Mio. Einträge für Facebook-Likes-Klicks in der Datenbank generiert. Innert kurzer Zeit wurden Tausende von Schlüsselwörtern und Themen erzeugt. Wir haben ein System gebaut, dass innert kurzer Zeit eine enorm grosse Datenbasis für diverse Analysen erstellt.



Diplomierende
Fabian Geiger
Thierry Musy

Dozierende
Gerold Baudinot
Martin Braschler



Diese Grafik zeigt auf, dass unsere Arbeit in die Bereiche Datensammlung, Themen- und Schlüsselwort-extraktion sowie Trendanalyse aufgliedert ist.