

## Information Discovery - Finden, nicht suchen!

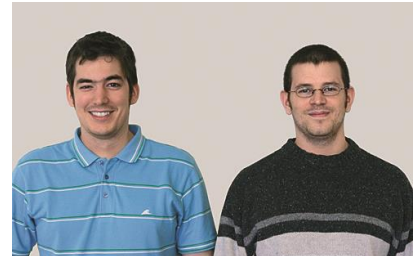
Wie kann man etwas finden, ohne danach zu suchen? Mit dieser und ähnlichen Fragen beschäftigt sich der Schwerpunkt Business Information Engineering des Instituts für angewandte Informationstechnologie (INiT) der ZHAW seit mehr als zwei Jahren. Auf Basis von vier Prototypen aus früheren Diplom- und Projektarbeiten entwarfen wir in dieser Bachelorarbeit ein System, das bisherige Ideen und Ansätze unter einem Dach vereinen soll.

Das System Information Discovery System verarbeitet einen kontinuierlichen Strom von Informationen, indem in regelmässigen kurzen Intervallen über vordefinierte Newsfeeds Nachrichten aus dem Internet ins System eingelesen werden. Mittels verschiedener Information Retrieval-Techniken wie z.B. Stemming, Indexing, Clustering und Filterung werden die Daten aufbereitet, um interessante Informationen zu extrahieren. So ist das System in der Lage, selbstständig aktuelle Themen zu entdecken und verschiedene Nachrichten einem oder mehreren Themen zuzuordnen. Ebenfalls kann die Reputation eines entdeckten Themas und der dazugehörigen Nachrichten bestimmt werden.

Anhand einer grafischen Benutzeroberfläche wird dem Benutzer ersichtlich, welche neuen Themen entdeckt wurden und wie diese gewichtet sind. Die Beschriftung des Themas variiert in ihrer Schriftgrösse nach Anzahl der Nachrichten. Ebenso wird bei der Bewertung des Themas verfahren, indem die Schrift bei negativen Themen rot und bei positiven grün eingefärbt wird. Über ein gestapeltes Flächendiagramm kann der Benutzer die Entwicklung der Anzahl Nachrichten herauslesen und wie sich die Reputation aus den Nachrichten zusammensetzt.

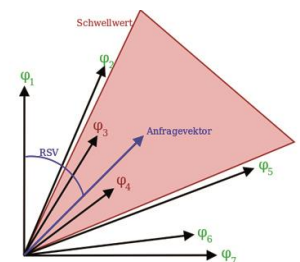
Dem Benutzer ist es natürlich auch möglich die automatische Bewertung des Systems für einen Newsartikel zu überschreiben. Diese Anpassung hat wiederum Einfluss auf spätere Bewertungen, da das System aus den Korrekturen des Benutzers lernt. Zu einem Thema gehörende Newsartikel oder auch das Thema selbst kann vom Benutzer gelöscht werden. Über eine Art Favoritenliste ist es auch möglich Themen langfristig mitzuverfolgen.

Mit dem entwickelten Information Discovery System erhalten Benutzer ein neuartiges Information Retrieval-Werkzeug, mit denen sie neue Themen entdecken, deren Reputation bestimmen lassen und diese langfristig verfolgen können.

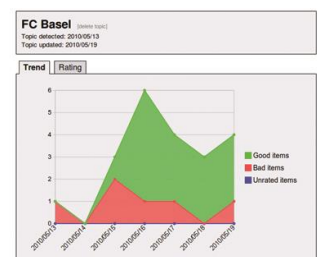


Diplomierende  
Remo Häberling  
Stefan Münch

Dozent  
Martin Braschler



Dies ist eine vereinfachte Darstellung des mathematischen Vektorraummodells. Liegt ein Dokumentenvektor innerhalb des roten Bereichs, wird dieses Dokument vom System negativ bewertet.



Aus diesem Ausschnitt aus dem Webfrontend kann der Trend eines ausgewählten Themas beurteilt werden. Dabei werden die Anzahl der erschienenen Dokumente am entsprechenden Tag dargestellt.