Mehr Qualität durch semantische Analyse: Retresco entwickelt neue Software für das Archiv der Süddeutschen Zeitung

- Machine-Learning auf Basis der Retresco-eigenen semantischen Textanalyse
- Optimierung archivarischer Arbeitsprozesse durch themengenaue Klassifikation und Visualisierung in einem Wissensnetz

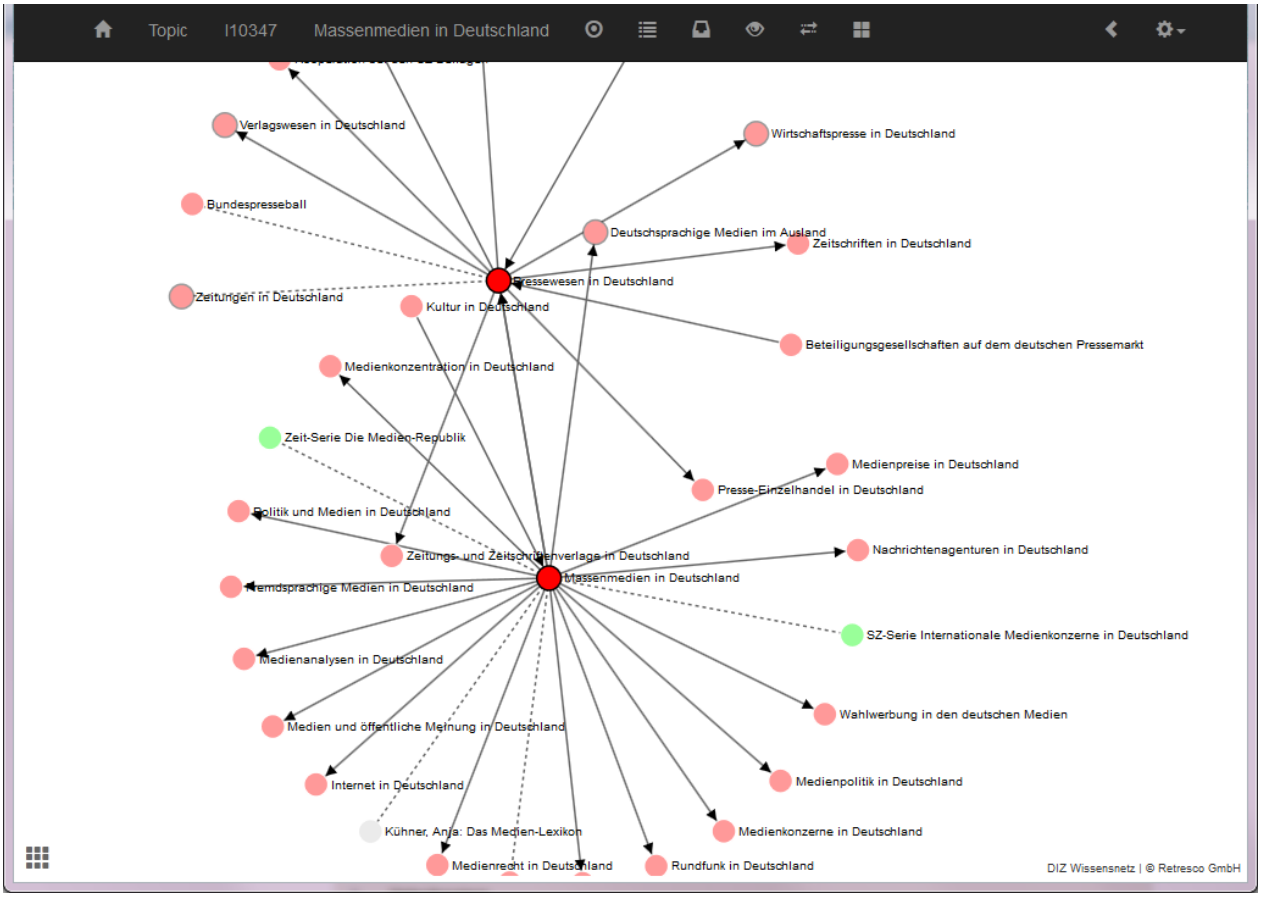
Berlin, den 06.11.2014 – Archive großer Medienunternehmen bergen einen unschätzbar wertvollen Fundus zeitgeschichtlicher Informationen, die sich erst durch automatische semantische Verknüpfungen leicht zu verständlichen Themenclustern zusammenfassen lassen. Eigens für das Archiv der Süddeutschen Zeitung und weiterer Partner der Dokumentations- und Informationszentrum München GmbH (DIZ) hat Retresco eine Software entwickelt, die das themengenaue Archivieren und die Recherche im Archiv dank semantischer Technologien zielgerichteter und ergebnisorientierter gestaltet.

Machine-Learning: Training des Klassifikators auf Basis semantischer Textanalyse
Basis der Lösung ist das Training eines Klassifikators anhand der Retresco-eigenen Semantik. Dazu werden im ersten Schritt sämtliche verfügbaren Archivinhalte hinsichtlich ihrer semantischen Grundstruktur analysiert. Anhand identifizierter Texteigenschaften wird im zweiten Schritt der Klassifikator trainiert, um zu bestimmen, welche Art von Texten in welche der mehr als 20.000 Themencluster und Personen- und Institutionen-Dossiers des Archivs gehört. Zur finalen Zuordnung erhält jeder Inhalt Scorewerte, die die Wahrscheinlichkeit ausdrücken, mit der diese zu den verschiedenen Themenclustern passen.

Jeder Artikel inklusive Scorewerte wird im Workflow-Tool von DIZ/SZ Archiv dargestellt. Der Lektor bestätigt die Vorschläge, die zum Artikel passen oder sucht ggf. nach anderen Clustern. Die Cluster-Vorschläge des Klassifikators beschleunigen und vereinfachen die Arbeitsprozesse für die Lektoren des DIZ Die Recherchearbeiten werden mit einem Visualisierungstool unterstützt. Es werden die Themencluster als Wissensnetz dargestellt, das die Parent-Child-Beziehungen der Inhalte zeigt (siehe Grafik)„Diese Prozessoptimierung erleichtert die tägliche Arbeit unserer Mitarbeiter in entscheidendem Maße. Vor allem die automatische Vorauswahl und dadurch die schnelle Zuordnung vereinfachen den Archivierungsprozess. Die Software bietet eine hohe Qualität und Zuverlässigkeit bei der Zuordnung von Texten zu Themenclustern“, bestätigt Hella Schmitt, Geschäftsführerin des DIZ München GmbH, die Vorteile der neuen Lösung.

„Wir standen hier vor der spannenden Herausforderung, eine Vielzahl an täglich neuen Artikeln einer noch größeren Anzahl an Clustern zuzuordnen. Besonders wichtig war für uns, via Mashine-Learning den hohen Ansprüchen des DIZ und dessen Premiumkunden an die Qualität der automatisch erzeugten Ergebnisse zu entsprechen. Zudem musste der Prozess für die Mitarbeiter des DIZ einfach nachvollziehbar sein und die Ergebnisse übersichtlich visualisiert werden“, erklärt Alexander Siebert, Computerlinguist und CEO von Retresco. „Das gemeinsame Projekt mit dem DIZ nutzt Retrescos Semantik für die

Klassifizierung von Texten in einem herausfordernden Big Data Umfeld und hebt die Arbeit mit dem Archiv damit auf ein höheres Level. Dieses Projekt ist ein wegweisendes Beispiel für die gelungene Automatisierung im Verlagswesen“, so Siebert weiter.



Beispielhafte Darstellung des DIZ-Wissensnetzes

Über Retresco

Retresco ist Experte und Partner für Semantik und die Automatisierung contentgetriebener Geschäftsmodelle. Auf Basis moderner Open-Source-Suchtechnologien und semantischer Verfahren entwickeln wir hochleistungsfähige Lösungen und automatisieren die effektive Verwertung von Inhalten entlang der gesamten Wertschöpfungskette. Unsere Lösungen erhöhen das User Engagement und die Relevanz in Suchmaschinen, optimieren Produktionsprozesse und steigern die Umsätze unserer Kunden.
Zu diesen zählen unter anderem United Internet, N24, FAZ.net, Rheinische Post Digital, Augsburger Allgemeine, Axel Springer SE, BÜNDNIS 90/DIE GRÜNEN und das Bundesministerium für Gesundheit.

Unternehmenskontakt
Johannes Sommer
Retresco GmbH
Fon +49 30 609 839 605
E-Mail sommer@retresco.de
www.retresco.de

Weiterführende Informationen unter: www.retresco.de