

Geschwindigkeit und Relevanz entscheiden

Mit Machine Learning und automatisierter Spracherkennung erweitert APA-DeFacto ihre Auswertungstools in der professionellen Medienbeobachtung.

••• Von Dinko Fejzuli
und Herbert Hirner

In Zeiten von Social Media, Fake News, der Expansion visueller Elemente und nie da gewesener Datenmengen stehen nicht nur Medienunternehmen selbst, sondern auch Anbieter darauf aufbauender Dienstleistungen vor zahlreichen Herausforderungen. Klemens Ganner ist Geschäftsführer des marktführenden Medienbeobachtungsunternehmens APA-DeFacto, das für seine Kunden Tausende redaktionelle Publikationen aus Print, Web und Radio/TV und diverse Social Media-Kanäle im Blick hat und über ein Archiv verfügt, das mehr als 130 Mio. Beiträge umfasst. Die APA-Tochter setzt in dieser Phase besonders stark auf den Einsatz neuer Technologien.

Zentrale Themen sind laut Ganner das Sammeln der Daten und die Interpretation derselben. „Wir investieren große Res-

ourcen in unsere Mediendatenbank, damit diese so vollständig und komplett ist, wie sie ist“, erklärt Ganner.

Sammeln und interpretieren

Die größte Herausforderung beim Sammeln seien die Sozialen Medien, so Ganner. Denn dort sei man einerseits abhängig von den gewährten Zugängen

„Technologie ist ein Kernthema von APA-DeFacto, wie auch der APA-Gruppe als Ganzes“, so Ganner. „Grundsätzlich sind wir sehr, sehr gut dabei, Dinge zu antizipieren. Wir haben gut aufgebaute Systeme für Innovations-Screening und -Management, sind top bei Forschung und Entwicklung. Und wir kooperieren mit Universitäten, um

dadurch innerhalb weniger Minuten einen Link zum Transkript des gesamten Beitrags, wenn ein definierter Suchbegriff in einer Nachrichtensendungen genannt wird. Derzeit noch auf die Bundesländer-Radios beschränkt, wertet die Software eigenständig aus, wie und in welchem Zusammenhang Unternehmen, politische Parteien und Personen erwähnt werden. „Das funktioniert mittlerweile sehr zeitnah“, ist Ganner zufrieden. „Wir können damit ein wesentlich größeres Sendungsspektrum beobachten und wollen dieses Service bald auch auf andere Nachrichtenformate ausdehnen.“



Mit Innovationsmanagement und Kooperationen mit Universitäten sichert APA-DeFacto die Relevanz und Aktualität der eigenen Services.

Bild- & Videoerkennung

Technologische Fortschritte hat man auch bei einem anderen Thema gemacht. Wie im letzten medianet-Gespräch angekündigt, kann Ganner mittlerweile auch Erfolge bei der maschinellen Erkennung von Artikeln innerhalb von Zeitungen und Zeitschriften aufweisen, die als PDF-Dokumente vorliegen. Auch bei dieser Anwendung lernt das System selbstständig dazu und wird bereits testweise eingesetzt. Am nächsten großen Thema – der Bild- und Videoerkennung mittels Machine Learning – arbeite man bereits intensiv, versichert Ganner.

Selbst bei einem derart hohen Automatisierungsgrad lässt sich nicht auf den Menschen verzichten. Nicht nur bei der Qualitätskontrolle, einer Verdichtung durch Relevanzauswahl oder im Hinblick auf die Glaubwürdigkeit sind manuelle Überprüfungen durch Menschen unabdingbar, berichtet der Me-

”

Wir investieren große Ressourcen in unsere Mediendatenbank, damit diese so vollständig und komplett ist, wie sie ist.

Klemens Ganner
GF APA-DeFacto

“

und Schnittstellen, und zum anderen verfüge Social Media nicht über eine definierte Grundgesamtheit wie die klassischen Medien. Bei der Interpretation schließlich sind heute die *Meta-Daten* das entscheidende Kriterium, also jene Informationen, die Auskunft über Reichweite, Likes, Engagement, Platzierung und die Anzahl von Nennungen geben, die also über den Inhalt hinausgehen. Um in beiden Bereichen den Wünschen und Anforderungen der Kunden gerecht zu werden, heißt es, am Puls der Zeit zu sein.

genau diesen Anforderungen Rechnung zu tragen, denn entscheidend sind *Relevanz* und *Geschwindigkeit*“, erklärt der APA-DeFacto-Geschäftsführer.

Speech-to-Text & Algorithmen

Für das Monitoring von TV- und Radiobeiträgen setzt APA-DeFacto seit April beispielsweise auf automatisierte Spracherkennung (Speech-to-Text). Ganner und sein Team haben dazu einen Sprachalgorithmus auf Nachrichtenformate angesetzt und mit Tausenden Stunden an Material trainiert. Kunden erhalten