



© APA (2)

Konvolut

Die APA-DeFacto ist mit mehr als 140 Millionen Einzeldokumenten der größte Medien- und Fachdatenbank-Anbieter Österreichs.

140
Mio.

Mitarbeiterinnen und Mitarbeitern Hilfestellungen bei der Erfüllung ihrer Arbeit in die Hand zu geben“, erklärt Ganner.

Zusammengefasst

Ein weiterer Punkt, an dem APA-DeFacto derzeit verstärkt arbeitet, ist die automatisierte Inhaltsanalyse von Texten und die Erstellung maschineller Zusammenfassungen. „Wir sind in diesem Bereich bereits relativ weit und haben jetzt noch einmal einen großen Sprung in der Qualität der Zusammenfassung gemacht, die die Einschätzung der Relevanz verbessert“, führt Ganner aus. Ein weiterer Schritt in Verbindung mit automatisierten inhaltlichen Zusammenfassungen ist die Bildung thematischer „Cluster“. Damit können beispielsweise alle relevanten Texte zu einem bestimmten

Thema oder Schlagwort gefiltert werden. Auch Resonanzmessungen – etwa der Niederschlag einer OTS-Aussendung in den Medien – sind so möglich.

Eine Technologie, die in den vergangenen Jahren nicht nur bei der APA, sondern weltweit große Fortschritte gemacht hat, ist die Speech-to-Text-Funktionalität. Bei APA-DeFacto soll diese vor allem zum Radio-, TV- und Web-Monitoring eingesetzt werden. Das Qualitätslevel eines manuellen Transkripts erreicht die Spracherkennung nicht, der Nutzen ist aber ein anderer: „Die automatisierte Transkription passiert sehr zeitnah – zwar mit Ungenauigkeiten, aber gut genug, um relevante Treffer, z.B. für Echtzeit-Alarmierungen, zu liefern“, beschreibt Ganner, der davon ausgeht, damit auch neue Kundensegmente erschließen zu können. Durch laufend trainierte Sprachmodelle können in diesem Bereich auch Informationsmengen verarbeitet werden, die bisher zu umfangreich waren.

Metadaten und Analyse

Wichtig für fast alle Anwendungsbereiche sind jedenfalls entsprechende Metadaten, die eine relevante Zuordnung ermöglichen – etwa, ob ein Artikel auf der Titelseite oder in kleiner

Form weiter hinten erschienen ist. Im Online-Bereich, so Ganner, sei man derzeit dabei, zu filtern, ob Inhalte prominent auf der Startseite oder nur auf einer Subseite aufscheinen. Ziel ist es also, ein Informationsangebot zu schaffen, das die Relevanz gefilterter Inhalte für den Kunden beurteilen kann. „Es geht um Relevanzfaktoren für Reichweite oder Position, also Kennzahlen, die schnelle Rückschlüsse erlauben“, führt Ganner aus.

Laufende Entwicklung

Viele der im Moment entwickelten Technologien werden in den nächsten Jahren weitere Fortschritte machen und so neue Einsatzfelder erschließen. Ob im Bereich der Bilderkennung – wo künftig das automatisierte Erkennen von Gesichtern und Logos möglich sein wird – oder in den erwähnten Speech-to-Text-Lösungen, die in Echtzeit Transkripte auf einem Niveau erstellen, das zumindest eine deutlich schnellere manuelle Weiterverarbeitung ermöglicht. Künstliche Intelligenz sei dementsprechend sicher eines der Themen, das künftig wesentliche Investitionsmittel erhalten werde. „Maschinelles Lernen bietet beinahe unendlich viele Anwendungsmöglichkeiten und Nutzungsfälle“, zeigt sich Ganner überzeugt. (fej/shr)

Multi-Medien-Dienstleister

Im APA-Konzern arbeiten derzeit rund 600 Mitarbeiter.

Forschungsfelder

Gebot der Stunde

Als Machine Learning wird das maschinelle Lernen von intellektuellen Arbeitsabläufen bezeichnet. Bei APA-DeFacto kommt diese Technologie beispielsweise bei der PDF-Segmentation, der Kategorisierung von Texten oder der Relevanz-Bewertung zur Selektion zum Einsatz. Bei der PDF-Segmentation geht es etwa darum, die Systematik hinter unterschiedlichen Publikationen zu erkennen und so eine Zuordnung von Bildern, Grafiken und Texten zu einem zusammenhängenden Artikel zu ermöglichen.

