

Von **DATENGETRIEBENEN PROZESSEN** und **DATENGETRIEBENEN ORGANISATIONEN**



Auszug

Architektur für serienmäßige Prozessautomatisierung
mit Natural Language Processing

Versicherungsforen-**Themendossier**

Eine Publikation der

 **Versicherungs**foren Leipzig

Architektur für serienmäßige Prozessautomatisierung mit Natural Language Processing (NLP)

Richtig und konsequent angewendet ist Natural Language Processing (NLP) eine Technologie, mit der sich wissensbasierte Prozesse automatisieren und langfristige strategische Vorteile generieren lassen.

So können zum Beispiel eingehende Anträge von Maklern und Kunden (teil-)automatisiert verarbeitet werden. Unstrukturiert eingehende Schadenmitteilungen in Kfz- und Haftpflichtversicherung werden für die Sachbearbeitung automatisiert zusammengefasst, oder direkt dunkelverarbeitet, aktuelle Verträge und Altverträge automatisch klassifiziert und relevante Informationen, wie Kündigungsfristen, direkt extrahiert.

Die initialen Mehrwerte solcher einfach gelagerter Lösungen schlagen sich oft in einer Kostenreduktion und kürzeren Bearbeitungszeiten nieder. Letzteres ist vor allem im Vertrieb besonders wertvoll, da die Reaktionszeit gegenüber dem Kunden einer der maßgeblichen Treiber für Abschlussquote und Kundenzufriedenheit ist.

Im Weiteren können auch aufwendige, wissensbasierte Prozesse automatisiert werden. So können zum Beispiel die Ermittlung des Gesamtexposure einzelner Deckungsbausteine wie Pandemie oder Silent Cyber in der Industrieversicherung oder die webbasierte Ergänzung der Unternehmensbranche anhand von Website-Informationen für Automatisierungen im Underwriting herangezogen werden.

So entsteht ein strategischer Wissensvorsprung, der im Vertrieb, der Produktentwicklung, der Bildung von Rückstellungen und im Risikomanagement gewinnbringend eingesetzt werden kann.

Um allerdings die Vorteile von Natural Language Processing über die ersten Pilotprojekte hinaus zu nutzen und auch komplexe Modelle für lange Dokumente trainieren zu können, braucht es einen skalierbaren Entwicklungsprozess und spezialisierte technologische Werkzeuge.

Die Situation

Es geht dabei um nicht weniger als das maschinelle Ver-

stehen aller zur Verfügung stehenden relevanten Informationen und die Umsetzung dieses Verständnisses in geschäftsrelevante Resultate. Immer mit Blick auf die vollständige Automatisierung des Geschäftsbetriebes und aller dazugehörigen Prozesse. Dieser Anspruch mag manch einem überspitzt erscheinen.

Allerdings ist es mit Natural Language Processing der jüngsten Generation möglich, selbst menschliches Leseverständnis abzubilden.

Ein Beispiel für die semantische Kompetenz solcher Modelle bietet folgender einfacher Satz mit zwei verschiedenen Fortsetzungen:

*„Ich gehe heute zu meiner Lieblings-Bank, ...
1. ... denn das Wetter ist schön und ich sitze gern im Park.“
2. ... denn ich habe etwas Geld anzulegen.“*

Handelt es sich hierbei um die Bank zum Sitzen, oder das Finanzinstitut? Diese semantische Unterscheidung war bis vor wenigen Jahren Menschen vorbehalten.

Natural-Language-Processing-Modelle der jüngsten Generation (sog. Transformermodelle), sind heute in der Lage, genau diese Unterscheidung aus dem Kontext korrekt zu treffen.

So ist es möglich, fast alle Informationen, die bislang unstrukturiert in Texten vorlagen, in strukturierte Daten zu überführen. Dadurch kann ein immenser Vorsprung an Wissen und Automatisierung realisiert werden – welcher sich wiederum in geringeren Kosten, sinkenden Reaktionszeiten, transparenteren Prozessen und letztlich einer höheren Kundenzufriedenheit bemerkbar macht.

Die Herausforderung

Damit ist ein Wettlauf eingeleitet, den viele andere Industrien bereits hinter sich haben und der vielen etablierten Unternehmen erheblichen Schaden zugefügt hat. Denn der Wettbewerb um Marktanteile wird im Umfeld von auf Automatisierung konzentrierten Marktbegleitern schnell zum gefährlichen Kostentreiber, wenn Marketingausgaben ineffizienten internen Prozessen gegenüberstehen.

Diese Situation wird sich verschärfen und bedarf einer adäquaten Antwort.

Demgegenüber stehen eine oft überalterte IT, begrenzte Mittel und zu wenige qualifizierte Mitarbeiter. Von mangelnder Agilität und Bereitschaft zum Wandel nicht zu sprechen.

Doch gerade die daraus erwachsende notwendige Konzentration auf das Wesentliche bietet demjenigen, der sie ergreift, eine erhebliche Chance.

Die Lösung

Es braucht keine gewaltige Infrastruktur, Unmengen an Daten, Heerscharen an Mitarbeitern oder Kooperationen mit universitären Forschungsinstituten, um wissensbasierte Prozesse mit Natural Language Processing zu digitalisieren.

Die praktische Anwendung von Natural Language Processing ist vielmehr eine Frage der effizienten und skalierbaren Anpassung existierender Modelle auf spezifische Domänen oder Dokumentenbestände.

Der Unterschied zwischen einer schnellen, erfolgreichen Umsetzung und einem nie enden wollenden Proof of Concept liegt dabei zu 90 Prozent nicht im Modelltraining, sondern in der Auswahl, Konvertierung, Bereinigung, Erstellung und Annotation von Trainingsdaten.

Die größten Hebel für eine kürzere Time to Market sind hierbei: ein robuster Trainingsprozess, die Erfahrung der beteiligten Data Scientists und das optimale Werkzeug.

Ein Best-Practice-Trainingsprozess besteht dabei in der Regel aus fünf klar definierten Schritten.

1. Vorbereitung und Synthetisieren von Trainingsdaten

Vorliegende Dokumente werden in ein passendes Format konvertiert – hier spielt auch die passende OCR-Engine mit Layout-Erkennung eine große Rolle. Die konvertierten Dokumente werden bereinigt, dedupliziert und eine passende Auswahl für ein ausgewogenes Training getroffen.

Gerade wenn nur wenige Beispieldokumente vorliegen oder die zu erkennenden Strukturen semi-regulär sind, bietet sich das Synthetisieren von Trainingsdaten mit spezialisierten generativen Grammatiken an. Durch dieses Verfahren können innerhalb kurzer Zeit tausende Trainingsdaten mit einer angemessenen Varianz erzeugt werden – und der folgende Schritt entfällt ganz.

2. Manuelle Annotation einiger Beispiele

Durch den Einsatz von Transfer Learning lässt sich ein erstes Modell heutzutage schon mit 50 bis 150 annotierten Beispielen trainieren. Die Annotation erfolgt am besten in Kooperation zwischen Data Team und Fachabteilung. So wird sichergestellt, dass innerhalb weniger Tage die optimale Trainingsgrundlage geschaffen wird.

3. Active Learning

Das so gewonnene Modell wird anschließend auf nicht annotierte Rohdaten angewendet. Aus den Vorhersagen des Modells wird eine Menge von 30 bis 50 Dokumenten ausgewählt und von Data Scientists oder Fachexperten überprüft und ggf. korrigiert.

Die Auswahl erfolgt dabei nach zwei Kriterien: Wo ist sich das Modell unsicher („uncertainty sampling“) und welche Dokumente unterscheiden sich stark von den vorherigen Trainingsdaten („diversity sampling“). Dadurch wird sichergestellt, dass die ausgewählten Dokumente beim erneuten Training einen maximalen Lerneffekt bieten – sogenanntes „Active Learning“.

In wenigen, kurzen Iterationsschleifen wird das Modell dadurch auf eine optimale Vorhersagegüte gebracht – ein Prozess, der sonst oft ein Vielfaches der Zeit benötigt.

4. Deployment

Das so gewonnene Modell ist in aller Regel nach wenigen Wochen für den Einsatz in Produktion gut geeignet. Moderne Trainingslösungen bieten ein automatisches Deployment im standardisierten Format und oft bereits Konnektoren für handelsübliche IT-Systeme. Das Modell kann via standardisierter API in Backend-Systeme, Workflow-Tools und sogar Frontends eingebunden werden.

5. Kontinuierliche Verbesserung

Im laufenden Betrieb wird das Modell abgesichert mit einem sogenannten „Human in the Loop“: Vorhersagen, bei denen die Konfidenz einen vorher festgelegten Schwellwert unterschreitet, werden zur Überprüfung an einen menschlichen Bearbeiter gegeben, um Fehler zu vermeiden.

Die Korrekturen des „Human in the Loop“ werden anschließend für ein Fine-Tuning des Modells verwendet – so entsteht ein kontinuierlicher Ver-

besserungsprozess.

Fazit

Wichtig für eine effiziente und skalierbare Entwicklung von Natural Language Processing Modellen ist, dass dieser Prozess möglichst ohne Medienbrüche abläuft.

Generische KI-Plattformen sind regelmäßig nicht auf die spezifischen Anforderungen von Natural Language Processing für komplexe oder lange Dokumente ausgerichtet – mit dem Effekt, dass das Data Team eine Menge Code schreiben muss, um das nötige Tooling um die Plattform herum sicherzustellen. Spezialisierte Lösungen bieten hier eine gute Alternative oder Ergänzung, um Natural Language Processing Modelle erfolgreich und in Serie zu entwickeln.

Gepaart mit der unproblematischen Integration in vorhandene Software über standardisierte Schnittstellen ist der erfolgreiche Einsatz von Natural-Language-Processing weniger eine Frage der technischen Voraussetzungen, sondern eher eine organisatorische Herausforderung.

Wem es gelingt, sich dem Thema schnell und aus dem Fachbereich heraus getrieben zu

nähern, dem bietet sich die Chance, schnelle und effektive Prozessverbesserungen in Serie umzusetzen.

Autor



Johannes Humbert

Geschäftsführer
tetrel GmbH



Impressum

Autor(en) des vorliegenden Themendossiers

Vincent Wolff-Marting et al.

T +49 341 98988-284

E vincent.wolff-marting@versicherungsforen.net

Feedback zum vorliegenden Themendossier

Wenn Sie uns Ihre Meinung mitteilen möchten, würde uns das sehr freuen. Vielleicht gibt es ja ein spezielles Thema, über das Sie im Themendossier einmal lesen möchten? Haben Sie weitere Fragen und Anregungen oder Anlass zur Kritik? In jedem Fall freuen wir uns über eine Nachricht von Ihnen.

Bitte senden Sie Ihre Kommentare an

Elisa Strey

T +49 341 98988-235

E elisa.strey@versicherungsforen.net

Abonnement des Versicherungsforen-Themendossiers

Aufgrund der Partnerschaft Ihres Unternehmens mit den Versicherungsforen Leipzig steht Ihnen das Abonnement des Versicherungsforen-Themendossiers unternehmensweit zur Verfügung! Gern können Sie deshalb weitere Empfänger aus Ihrem Haus registrieren lassen. Nutzen Sie dazu einfach unser Anmeldeformular unter www.versicherungsforen.net/abo_themendossier. Eine Übersicht über alle Partnerunternehmen finden Sie unter www.versicherungsforen.net/partner.

Abbestellen des Versicherungsforen-Themendossiers

Sie wollen das Themendossier in Zukunft nicht mehr empfangen? Senden Sie einfach eine E-Mail mit dem Betreff „unsubscribe Themendossier“ an kontakt@versicherungsforen.net.

Downloadbereich mit aktuellen Zahlen und Fakten zur Versicherungswirtschaft

Im Downloadbereich unter www.versicherungsforen.net/daten-fakten finden unsere Partner aktuelle Zahlen, Daten und Fakten zu verschiedenen versicherungswirtschaftlichen Themen. Diese werden fortwährend erweitert und regelmäßig aktualisiert. Aussagekräftige Charts sorgen für einen schnellen und detaillierten Überblick über relevante Branchenthemen. Sie möchten diese Informationen in Ihre Präsentationen einbinden? Dann stellen wir Ihnen die Zahlen, Daten und Fakten gern im neutralen Power-Point-Format zur Verfügung, das Sie mit einem Klick in das eigene Corporate Design überführen können.

Versicherungsforen-Newsletter

Wenn Sie regelmäßig per E-Mail über Aktualisierungen im Bereich „Wissen“ auf www.versicherungsforen.net, Veröffentlichungen, gegenwärtige Veranstaltungen sowie Nachrichten aus unserem Partnernetzwerk informiert werden möchten, können Sie sich auf www.versicherungsforen.net/newsletter anmelden. Diesen Service bieten wir auch für Nicht-Partnerunternehmen kostenfrei an.

