

snapADDY GmbH verbessert Adresserkennung mittels Deep Learning

01. Februar 2019

Die snapADDY GmbH aus Würzburg verwendet mit dem aktuellen Update ihrer Produkte einen neuen Parser, der durch den Einsatz von Deep-Learning-Techniken die Erkennung der Adress- und Kontaktdaten weiter verbessert. Dieses System stellt in allen Produkten (snapADDY Grabber, VisitReport, Visitenkartenscanner und Assistant) die zentrale Instanz zur Identifizierung der Kontaktdaten aus unstrukturiertem Text dar und ist so das technische Herzstück der snapADDY Produkte.

Funktionsweise des snapADDY Parsers

Der Adressparser von snapADDY sorgt für die Identifizierung von Adress- und Kontaktdaten aus unterschiedlichen Textquellen. Dies kann beispielsweise der reine Text eines Impressums sein, oder auch eine E-Mail, in deren Signatur die für snapADDY relevanten Daten enthalten sind. Auch das Foto einer Visitenkarte, das zunächst mittels Texterkennung „gelesen“ wird, ist ein Anwendungsfall für den snapADDY Parser. Für Maschinen ist dabei die Erkennung von Kontaktdaten im Gegensatz zum Menschen keine triviale Aufgabe und erfordert eine Mischung von ausgefeilten Regeln und einer intelligenten Bewertung der identifizierten Inhalte. Dafür werden die Bestandteile eines Kontaktes einzeln betrachtet. Manche Kontaktinformationen, wie zum Beispiel E-Mail-Adressen oder URLs, sind einheitlich formatiert und daher für den Parser durch Regeln leicht zu identifizieren.

Erweiterung des snapADDY Parsers mit Deep-Learning-Techniken

Zusätzlich zum regelbasierten Vorgehen geht die snapADDY GmbH mit dem zum neuen Jahr veröffentlichten Parser einen neuen Weg: Bei Kontaktwerten, die nicht standardisiert sind, ist die Erkennung durchweg komplexer und nicht mehr zufriedenstellend durch Regeln abbildbar. Das betrifft zum Beispiel Namen und Berufsbezeichnungen: Der Begriff „Schneider“ kann sowohl einen Beruf bezeichnen, als auch der Nachnamen einer Person sein. Der Mensch folgert den richtigen Sinn einfach aus dem Kontext. Der Einsatz von Deep Learning ermöglicht der Maschine nun ein ähnliches Vorgehen, bei der einzelne Kontaktinformationen anhand der benachbarten Begriffe korrekt zugeordnet werden können. Die so gewonnenen Erkenntnisse werden im snapADDY Parser dann mit den aus Regeln gewonnenen Informationen kombiniert, um eine möglichst gute Kontakterkennung zu erreichen.



Gestiegene Qualität bei der Kontakt- und Adresserkennung

Mit dem Anfang Januar veröffentlichten Update des Parsers konnte die snapADDY GmbH eine signifikante Verbesserung der Erkennungsqualität in ihren Produkten erreichen. Besonders bei nicht standardisierten Kontaktfeldern helfen die Kontextinformationen, die mittels Deep Learning gewonnen werden können. So hat sich die Erkennung der korrekten Berufsbezeichnung um ca. 19% im Vergleich zur vorherigen Version des Parsers verbessert. Auch Vor- und Nachnamen sowie Telefonnummern werden besser erkannt: die Verbesserung beträgt hier 7% bzw. 16%. Abbildung 1 zeigt die erreichten Verbesserungen in einer Grafik.