

TXTWerk - Kurzüberblick

Die Neofonie entwickelt für ihre Kunden individuelle Lösungen und bietet Beratung, Konzeption, Design, Entwicklung und Betrieb aus einer Hand. Gemeinsam mit dem Tochterunternehmen Neofonie Mobile zählt der Auftragnehmer laut Internetagentur-Rankings zu den bundesweit führenden Fullservice-Agenturen. Das Unternehmen wurde 1998 im Zuge der Entwicklung der ersten deutschen Suchmaschine "Fireball" als Spin-off der TU Berlin gegründet und arbeitet seitdem an Themen der Künstlichen Intelligenz (KI) mit dem Fokus auf Sprachverarbeitung und Textanalyse (NLP).

Dadurch ist die Neofonie einer der wenigen Anbieter im deutschsprachigen Raum, der mit TXTWerk sowohl eine produktreife NLP Lösung als auch ein KI Team zur Umsetzung von kundenspezifischen Anforderungen hat. Das Framework TXTWerk der Neofonie ist ein modulbasiertes System, das auf die individuellen Anforderungen der Kunden zugeschnitten werden kann. Zudem sind TXTWerk und die eingesetzten Verfahren speziell auf die deutsche Sprache optimiert, und unterscheiden sich so von allgemeinen mehrsprachigen Lösungen anderer Anbieter.

TXTWerk ermöglicht es, aus Texten Metadaten zu extrahieren und so aus unstrukturierten Texten strukturierte Daten zu liefern. Durch die Anwendung maschineller Lernverfahren in Kombination mit regelbasierten Ansätzen kann TXTWerk Texte lesen und verstehen, und das mit hoher Geschwindigkeit. Ob Sie 1000 oder 10 Milliarden Dokumente verarbeiten wollen, TXTWerk erkennt die wichtigsten Schlüsselwörter, Personen, Orte, Organisationen, Events und andere Konzepte und verlinkt diese zu weiteren Quellen wie zum Beispiel Wikidata oder unternehmensinternen Daten. Mit diesen strukturierten Metadaten erschließen sich völlig neue Möglichkeiten der Weiterverarbeitung. Ebenfalls Teil des Frameworks sind KI-Verfahren zur Klassifikation in vom Kunden definierte Klassen, eine Sentimentanalyse von Texten, Phrasen- und Rollenerkennung sowie die automatische Verlinkung von Entitäten nach eigens definierten Relationen. Neben den KI-Verfahren bringt TXTWerk einen Wissensgraphen mit über 7 Millionen Einträgen mit.

TXTWerk wird aktuell in verschiedenen Bereichen eingesetzt. So wird es bei Kunden wie der DPA, der PMG, dem Weserkurier oder dem ZDF zur semantischen Textanalyse und Metadatengenerierung genutzt. TXTWerk arbeitet aber auch als Komponente in Suchinfrastrukturen zum Beispiel bei der HUK und Barmer GEK, um hier dem Nutzer besser mit den gewünschten Informationen zu versorgen. Außerdem ist TXTWerk Teil der Neofonie Chatbotlösung, wo es sowohl Nutzeranfragen als auch vorhandene Informationen aufbereitet, um so eine möglichst natürliche Kommunikation zu ermöglichen.

TXTWerk Module

Services	Core Service	Extension	Beschreibung
Language Detection	x		Spracherkennung
Subjectivity Detection	x		Erkennung der Subjektivität eines Textes
Classification	x	x	Klassifikation des Textes in vorgegebene Kategorien
Keyphrase Extraction	x		Extraktion von Schlüsselwörtern
Named Entity Recognition	x	x	Erkennung von Entitäten (z.B. Personen, Orte, Organisationen)
Named Entity Linking	x		Verlinkung von Entitäten auf eine Wissensbasis (Wikidata)
Date Recognition	x		Datumserkennung
Measure Detection	x		Erkennung von Maßzahlen (z.B. 1Kg, 2m, 3l)
Near Duplicate Detection	x		Berechnung von Fingerprints zur Erkennung von Duplikaten
Author Extraction	x		Extraktion des Autors
Quote Recognition		x	Zitaterkennung
Relation Extraction		x	Extraktion von Beziehungen zwischen Entitäten
Sentiment Analysis		x	Erkennung des Sentiments eines Textes
Quality Detection		x	Bestimmung der Text-Qualität
Function Extraction		x	Erkennung der Funktion einer Person im Text