



COGNITIVE SERVICES

KANN EINE MASCHINE
DEN MENSCHEN
VERSTEHEN?



KOGNITIVE SERVICES

Kann eine Maschine seine Umwelt wahrnehmen? Was lässt sich aus der Umwelt erfahren oder sogar lernen? Wie kann eine Maschine den Menschen verstehen?

Mit diesen Themen beschäftigen sich kognitive Software-Systeme: Ausgehend von der Aufnahme von Daten z.B. über Sensoren wie Kameras und Mikrophone oder Dokumenten und Daten aus den eigenen Systemen werden systematisch Informationen erzeugt. Dabei können Wissensdatenbanken oder bereits in der Vergangenheit erlerntes Wissen herangezogen werden um neues Wissen zu erzeugen. Auf Basis dieses Wissens können dann Entscheidungen getroffen werden. In der Interaktion mit dem Menschen, fängt die Maschine an, das Gegenüber Mensch besser zu verstehen: Die Intention oder die Semantik des gesprochenen Wortes soll ebenso wie die Emotion erlernt werden. Der Einsatz solcher Systeme ist dabei sehr breit: von Support-Hotlines und Chatbots die autonom auf Fragen antworten können bis hin zu Entscheidungssystemen, die auf Basis ihrer Daten Prozessentscheidungen treffen können. Im Kern dieser Systeme arbeiten in der Regel KI-Algorithmen, die ihr Wissen über Trainingsphasen verfeinern und aufbauen.

DIE BAUSTEINE VON KOGNITIVEN SYSTEMEN – KOGNITIVE SERVICES

Steht man vor der Herausforderung, ein kognitives System zu bauen, so stellt sich unmittelbar die Frage, wie man ein solches System baut. Ein neues System von Grund auf zu bauen, welches komplexe Daten nicht nur erfassen, sondern auch interpretieren kann, ist aufwendig und komplex. Soll dabei noch eine breite Wissensbasis zum Lernen zur Verfügung stehen, so stößt man schnell an die Grenzen der Machbarkeit. Wem stehen zum Training z.B. viele Millionen Bilder zur Verfügung? Cloud-Services vereinfachen die Entwicklung kognitiver Systeme deutlich, da neben den KI Algorithmen oft auch schon die erforderliche Datenbasis und trainierte KIs vorliegen.

Betrachtet man den Markt der kognitiven Services, so lassen sich bei den Anbietern die folgenden Schwerpunkte ausmachen:

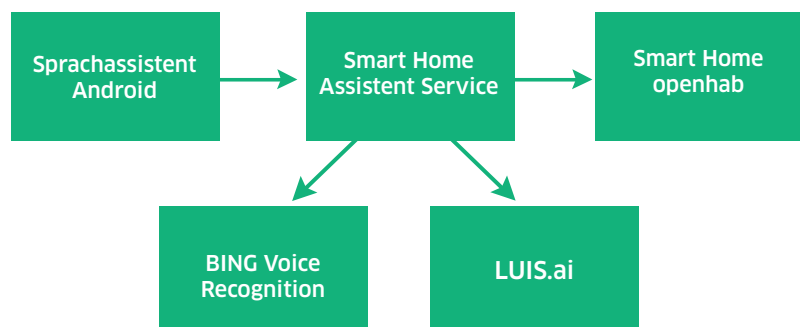
- Services zur Erfassung von gesprochener Sprache (Speech2Text), Bild und Film (Video Recognition)
- Auswerten von natürlicher Sprache (Native Language Understanding und Processing)
- Services zu Charakteristiken von Sprache und Bildern (Geschlecht, Alter, Emotionen)
- Spezifische Bilddienste wie Gesichtserkennung und Identifikation

Wir wissen nun was kognitive Systeme sind und was sie leisten können. Doch welchen Nutzen bieten diese Systeme konkret und wie setzen wir diese bei Accso für unsere Kunden ein? Die Einsatzgebiete bei unseren Kunden sind vielfältig: Von der Gesichtserkennung in Xbox-Anwendungen auf Basis von UWP-Anwendungen, Steuerung von Mediatheken oder dem Smart Home mit Sprachassistent oder die Analyse von Bild- und Videobeiträgen im Bereich Broadcast/Medien. Zwei dieser Systeme stellen wir euch im Detail vor.

SCOTTY: „COMPUTER?“

Kann man mit Sprache programmieren? In unserem ersten Szenario stellten wir uns diese Frage im Rahmen einer bei Accso durchgeführten Master-Arbeit in Zusammenarbeit mit der Hochschule Darmstadt. Betrachtet man heutige Sprachassistenten im Smart Home und deren zugrunde liegende Systeme, so sind die Interaktionsmuster oft eingeschränkt auf einfache Befehle wie „Mache das Licht an!“ oder einfache Konversationen „Mache das Licht an!“ - „Welches?“ - „Im Bad“. Eine Konfiguration oder weiter gefasst eine Programmierung in der Form „Wenn ich abends einen Film schaue, soll das Licht gedimmt und das Telefon stumm geschaltet werden“ sind nicht möglich. Regelsysteme im Smart Home müssen über Webfrontends oder Apps konfiguriert/programmiert werden, entsprechendes Know-how vorausgesetzt.

Dass es auch anders geht, zeigt das von uns implementierte System: Über eine Android-App kann der Benutzer dem System Regeln für das Smart Home vorgeben. Diese werden über die Services von Microsoft BING Voice Recognition maschinenlesbar zu Text übersetzt. Auf Basis dieses Textes wurde nun mit Hilfe von LUIS.ai, einem Microsoft Dienst zur Analyse von Sprache, die Semantik des Satzes extrahiert. Logische Wenn-Dann Regeln und Operatoren (Und, Oder) ließen sich gut mit dem Vokabular des Smart Homes (Ak-



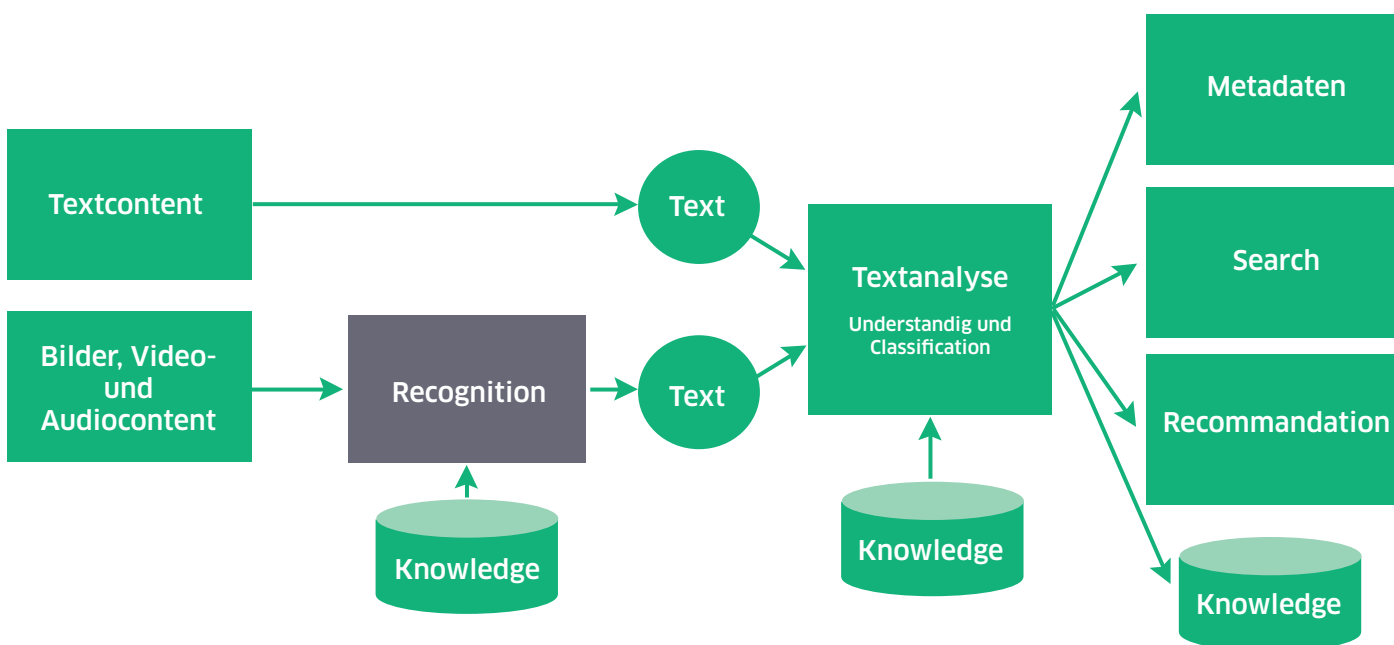
toren wie Licht) und den Aktionen bzw. Zuständen (an, aus, zu, auf) kombinieren. So gut, dass sich die Regeln für das Smart Home per Sprache generieren lassen.

Weiterführen kann man das Konzept für Systeme, die ebenfalls mit regelartigen Konfigurationen arbeiten und deren Szenario sich für die Sprachsteuerung eignen. Hierzu zählt der Einsatz von gesprochenen Regeln für die Rule Engine Drools oder von Geschäftsprozessen mit visualisierender BPM Unterstützung.

KOGNITIVE FILMANALYSE

Kognitive Systeme zur Analyse von Video- und Tonmaterial haben den Zweck, Filme und Videos zu verstehen. Ideal wäre es, durch die Analyse des Materials das Drehbuch weitgehend rekonstruieren zu können. Wozu sollte dies nützlich sein? Viele Fernsehsender verfügen über Archive, die viele Jahrzehnte Beiträge gesammelt haben. Zu diesen Beiträgen und Filmen sind nur wenige Informationen vorhanden. Nachträglich Metadaten, z.B. eine Liste der Schauspieler, zu erstellen ist ein langwieriger und oft auch manueller Prozess. Bei großen Archiven würde diese Arbeit Jahre in Anspruch nehmen. Auch die Art der erfassten Daten ähnelt in ihrer Form oft eher Schlagworten, so dass selten die Abfolge oder Dialoge festgehalten werden. Diese Metadaten sind in Archiven und Mediatheken jedoch wichtig, um Inhalte überhaupt auffindbar zu machen: Nutzer wollen in den Beiträgen suchen können oder auf Basis der bislang gesehenen Sendungen Empfehlungen ausgesprochen bekommen. Abfragen sollten dabei möglichst natürlich-sprachlich erfolgen können: „Wer hat >>Dumm ist der, der Dummes tut<< in welchem Film gesagt?“

Zu diesem Zweck haben wir ein Softwaresystem konzipiert, welches Bildmaterial über eine Recognition Engine erfasst. Erprobt wurden dabei die gängigen kognitiven Services auf dem Markt, unter anderen von Microsoft, Google und Amazon, um einen Eindruck über die Leistungsfähigkeit der kognitiven Services zu erhalten. Die auf diesem Weg entstehenden Textphrasen werden weiter über eine Textanalyse verarbeitet um dann in Suchmaschinen, Empfehlungssystemen oder (graphen-basierten) Metadaten-Systemen Eingang zu finden.



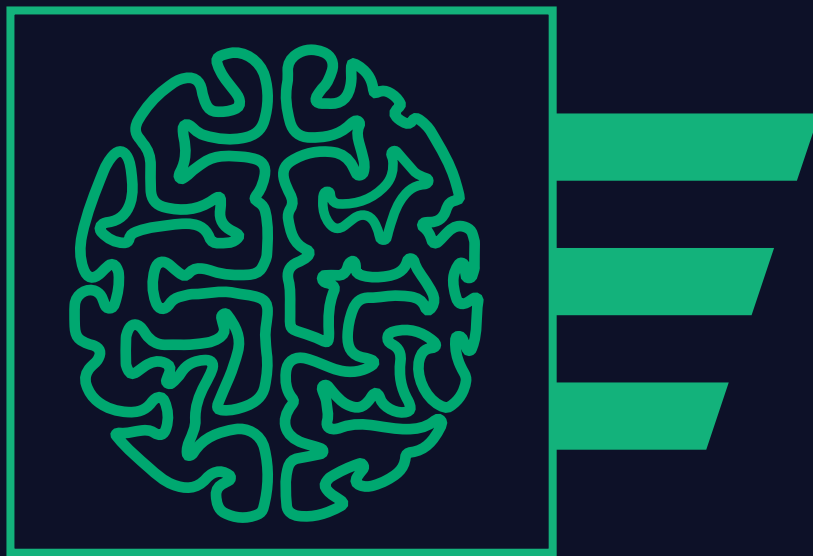
WIE GEHT ES WEITER?

Kognitive Systeme können die Interaktion mit der IT unterstützen, bis hin zu Systemen, die autonom mit dem Menschen und der Umwelt agieren. Die Spracherkennung (Speech2Text) ist dabei bereits auf einem Niveau, welches den menschlichen Fähigkeiten nahekommt, wenn nicht sogar teilweise überlegen ist. Die Intension des gesprochenen Wortes jedoch richtig zu verstehen, ist eine Herausforderung. Wer Sprachassistenten verwendet, verzweifelte schon mal an einem „das habe ich leider nicht verstanden“. Wechselt man das Sprachvokabular jedoch auf spezifische Bereiche mit einem reduzierten Wortschatz oder auf weniger komplexe Kontexte wie das Smart Home, so lassen sich bereits erstaunliche Ergebnisse erzielen.

Bei der Bild- und Videoanalyse steht man erst am Anfang. Viele, für den Menschen einfach zu erkennende Situationen, überfordern noch die Algorithmen der vorhandenen Services. Während Personen schon recht gut erkannt werden können und Schlagwörter zu Szenen erzeugt werden können, ist es für die Systeme noch sehr schwer Tätigkeiten zu erkennen.



Thomas Jäger, Cheftecnologe
Accso - Accelerated Solutions GmbH



**Sie haben Fragen rund um Cognitive Services?
Auf diesem Gebiet ist Accso Spezialist !
Kontaktieren sie uns, und wir helfen ihnen weiter.
www.accso.de/cognitive-services**