

# Virtuelle Berater, die aus Dokumenten lernen

Paul Caspers, CEO, Theum AG

*Sie sollten Dokumente und Bücher nicht nur speichern, damit Sie von Menschen gelesen werden. Was drin steckt ist nämlich nichts weniger als der Rohstoff für virtuelle Berater und andere Anwendungen der künstlichen Intelligenz.*

## Rohstoff für künstliche Intelligenz

In Dokumenten aller Art ist Wissen eines Unternehmens, einer Organisation und der ganzen Menschheit niedergelegt. Als unstrukturierte Daten unterschiedlichster Couleur lagern sie in riesigen Mengen in Trägersystemen aller Art – vom lokalen Laufwerk oder Dateiserver über Dokumentverwaltungssysteme bis hin zu Wikis, Websites und Cloud Lösungen. Das in ihnen festgehaltene und fortgeschriebene Wissen schlummert oft im Verborgenen, da es schwer zugänglich ist und sich den üblichen Auswertungen entzieht. Und für die heute nicht mehr wegzudenkende mobile Welt sind Dokumente gar völlig untauglich. Trotz oder gerade deswegen wird die Erschließung unstrukturierter Daten für Wissensprozesse ein wesentliches Element künftiger KI Anwendungen sein.

## 1.000 Dokumente aber nichts gefunden

Mit welchem Werkzeug immer Sie heute suchen (Betriebssystem, DMS, Wiki, Suchmaschine ...), sobald dabei Dokumente (Web, Office, PDF ...) durchforstet werden – erhalten Sie als kleinste Einheit immer nur ganze Dokumente. Diese werden in meist langen und uns allen hinlänglich bekannten Trefferlisten präsentiert.

Das gilt übrigens auch für modernste kognitive Software und KI-Systeme: Solange diese auf strukturierte Daten (Big Data) zurückgreifen können, liefern sie präzise Antworten: Was haben Leute noch gekauft, die denselben Artikel bestellt haben? Wie alt ist Angela Merkel? Wie wird das Wetter morgen? Welche Route nehme ich heute zum Büro? Wenn diese Systeme aber wie bei einer Recherche zum Thema *Künstliche Intelligenz und Digitale Transformation* auf Dokumenten operieren, finden wir wieder nur unsere altbekannten Sammelsurien von Dateien aufgelistet in Trefferlisten – anstatt brauchbarer Ergebnisse (<http://bit.ly/2j9NpO1>). Und je mehr „Antwort-Skills“ man den Systemen versucht anzutrainieren, desto exorbitant teurer wird deren Einsatz und ist deswegen nur für hochprofitable ausgewählte Anwendungen überhaupt eine Option.

Bestehende Systeme können also scheint es, benötigtes Wissen gar nicht, nur sehr unzureichend oder mit hohem Aufwand aus unterschiedlichen Dokumenten filtern und feingranular in einem neuen Dokument quasi wie eine Antwort zusammenstellen.

## Anwender wollen Antworten

Anwender suchen aber zumeist keine Dokumente – sondern feingranular gefilterte und „in einem Stück“ angebotene Informationen zu anstehenden Aufgaben oder Themen. Also sozusagen ein aus Teilen verschiedener Quellen erstelltes Exzerpt mit dem Ergebnis zu einer Suchanfrage. Anstatt endlos Zeit mit der

Recherche nach Informationen beim Suchen, Öffnen und Durchblättern von Dokumenten zu verschwenden, wünschen sich Anwender schlicht direkte Antworten. Sie wollen das gerade benötigte Wissen mit einem Klick – extrahiert aus allen relevanten Dokumenten, kontext-orientiert zusammengestellt und sofort nutzbar erhalten.

Also defacto eine Art virtueller Berater irgendwo auf dem Wege von der Gegenwart zur KI, der - von den Inhalten zugeordneter Dokumentpools gespeist – auf Suchanfragen seiner Nutzer komplette Ergebnisse in „gebundener Form“ liefert.

## Die Matrix, die weiß was Ihr Unternehmen weiß

Wir haben bereits gesehen: In unterschiedlichen Dokumenten zu einer definierten Thematik verpacktes Wissen kann nicht allein durch bloßen Zugang zu diesen Dokumenten effizient erschlossen werden. Hier fehlt bildhaft gesprochen das Zoom für die erforderliche Tiefenschärfe.

Es gibt jedoch einen anderen Weg: Man kann Dokumentinhalte normalisieren. Der Begriff kommt aus der Datenbankwelt und bedeutet so etwas wie „in strukturierte maschinell auswertbare Daten umwandeln“. Man zerlegt Dokumente dabei in ihre atomaren Bestandteile (hier: Kapitel, Absätze, Listen, Tabellen, Bilder, Wörter ...), ermittelt und addiert thematische Bezüge („Verwandtschaften“) zwischen den Elementen, reichert das Ganze mit vorhandenen oder anhand von Mustern erkannten Metadaten an und verpasst den Wissensatomen last not least noch ein einheitliches Erscheinungsbild. Wenn man solchermaßen in Form gebrachte Dokumentinhalte in einem clever konstruierten Index ablegt, enthält dieser das dokumentierte Wissen einer Organisation zu jedem entsprechenden Themenkomplex in vernetzter Form. Und dieser Index kann bei Änderung von - und Hinzufügen neuer Quellen organisch wachsen. Ein bisschen wie die Matrix aus dem gleichnamigen Science-Fiction Film.

Unsere hier diskutierte Matrix des Wissens erlaubt nicht nur die „dokument-befreite“ geräte-unabhängige Bereitstellung von dokumentierter Information in Form durchgängiger, präzise zu navigierender Wissensbanken, digitaler Bibliotheken und Bücher – sie liefert vor allem komplette Antworten auf Rechercheanfragen aller Art.

Und das Beste: Die Lösung erfordert weder die Abschaffung von Dokumenten (wir werden sie noch lange Zeit benötigen), noch die Anpassungen bestehender Systeme und Abläufe rund um das Management von Dokumenten – und schon gar nicht verlangt sie langwierige Konzeptionsarbeiten zur ihrer Einführung.

Es reicht - dem System einfach mitzuteilen, wo die Dokumente liegen, aus denen sich der Inhalt der Matrix speisen soll – etwa in Verzeichnissen auf einem Fileserver, in einem DMS, in einer SharePoint Dokumentbibliothek, in einem Wiki, einer Website oder in der Cloud. Es müssen keine Unternehmenstaxonomien, Nomenklaturen oder andere aufwändige Konstrukte konzipiert werden.

Damit stellen Sie Dokumentinhalte zu beliebigen Themen - basierend auf einer mächtigen aus Dokumenten lernenden Wissensmatrix - als komfortabel zugreifbares Wissen für direkte Nutzung am PC, Tablet und Smartphone intern oder wenn gewünscht weltweit in wenigen Tagen bereit.

## Virtuelle Berater die aus Dokumenten lernen

Der langfristig aber viel durchschlagendere Nutzen ist, dass die Matrix die einfache Implementation virtueller Berater ermöglicht, die Antworten zu Anfragen auf Basis der Inhalte von Dokumenten und Büchern geben können.



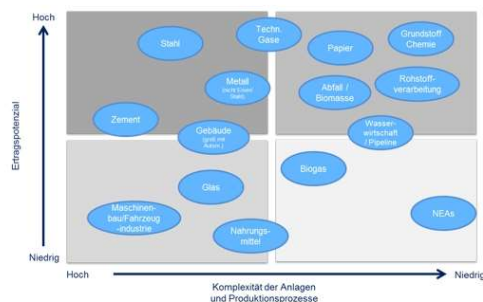
Wie ist das Lastverschiebungspotential in der Industrie (<http://bit.ly/2hSvgAD>) ... um ein weiteres Beispiel zu geben:

Um das rauszufinden, müssen Sie sofern Sie keinen menschlichen Experten als Berater befragen können – im Zweifelsfalle viele umfängliche Dokumente zum Thema Energieversorgung und Netze lesen. Sobald ein mit denselben Dokumenten gespeister virtueller Berater befragt wird – gibt er Ihnen sofort diese Antwort:

## Was ist das Lastverschiebungspotential der Industrie

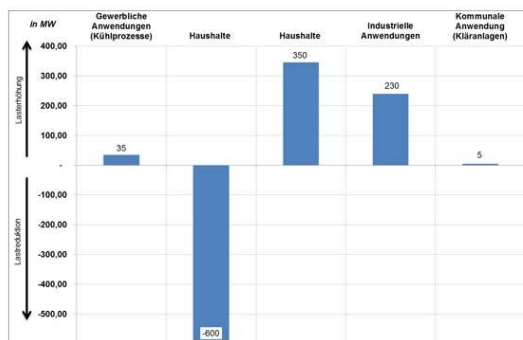
PUBLIC AUTHORITY PUBLICATIONS / SMART GRIDS 2.0 WHITE PAPER / 5 TECHNISCHE ANSATZPUNKTE

Aktuelle Studien zeigen ein Lastverschiebungspotential in der österreichischen Industrie von rund 400 MW [3]. Aus dem Blickwinkel von Aggregatoren können die Branchen hinsichtlich ihrer wirtschaftlichen Attraktivität folgendermaßen in ein Portfolio einsortiert werden:



PUBLIC AUTHORITY PUBLICATIONS / SMART GRIDS 2.0 WHITE PAPER / 6 KRITISCHE EIGENSCHAFTEN VON FLEXIBILITÄT

Die unterschiedlichen Flexibilitätsoptionen und deren Potentiale wurden in unterschiedlichen Forschungsprojekten untersucht ([2],[3],[4]). Die Ergebnisse einer Meta-Studie bezüglich der Lastverschiebungspotentiale in Industrie, Gewerbe und Haushalten sind in Abbildung 5 dargestellt.



Virtuelle Berater bieten also **antwort-orientierten Zugriff auf Wissen** – und beziehen dabei ihre Fähigkeiten aus der mit dem Inhalt zugeordneter Dokumente gespeisten Matrix. Dabei gibt es 3 verschiedene Szenarien – eine Antwort zu liefern:

1. Es wird **ein Antwortdokument** mit Extrakten aus den Inhalten beliebig vieler relevanter Dokumente erzeugt.
2. Sobald die Antwort eine definierte max. Länge überschreitet (was bei recht allgemein gehaltenen Anfragen der Fall sein kann) bildet das System selbsttätig Kategorien aus dem Ergebnis und bietet diese als Kontext-Optionen für die Präzisierung der weiteren Suche an und führt damit den Anwender gewissermaßen zum richtigen Ziel. Wenn man etwa den Begriff Cholera in einem Berater zu Seuchen recherchiert, könnte man **Kontext-Optionen** wie Behandlung, Verbreitung, Inkubationszeit etc. erhalten.
3. Beides kann kombiniert sein mit „Topics“ – das sind Abschnitte eines Dokuments – bei denen bereits die Kapitel- oder Seitenüberschrift (im Falle von Folien z.B.) Übereinstimmungen mit der Anfrage zeigen.

Es gibt noch einen weiteren gewichtigen Vorteil dieser Technik: Ein virtueller Berater liefert immer einen umfassenden aus allen verfügbaren Quellen gespeisten Gesamtüberblick zu einem Thema. Das im Gegensatz zu Trefferlisten so vermittelte ganzheitliche Bild eines Sachverhaltes beugt Fehlentscheidungen mit vielleicht gravierenden Folgen infolge übersehener Information vor.

Mit virtuellen Beratern kann also die in Bergen von Dokumenten versteckte Information antwort-orientiert auf jedem Endgerät passgenau als Wissen erschlossen werden. Da sich die Inhalte der virtuellen Berater aus Quelldokumenten unterschiedlichster Herkunft aufbauen – ist eine weitere Synergie das Verschwinden oft vorhandener Informationssilos samt ihren Zugangsbarrieren.

Stellen Sie sich nur vor – und das ist schon fast greifbar nahe: Alexa, Cortana oder Siri würden Ihnen auch auf komplexe gesprochene Fragen etwa zur Rente wie **Welche Beitragszeiten sind auf Wartezeiten anrechenbar** (Link erhalten Sie auf Anfrage) die Antwort vorlesen!

## Der Nutzen von virtuellen Beratern

Virtuelle Berater die aus Dokumenten lernen, erschließen das oftmals brachliegende Knowhow in zahllosen Dokumenten und Systemen. Das führt zu folgenden Vorteilen:

- Der Aufwand für Aufbereitung und Verteilung von dokumentiertem Wissen wird nahezu vollständig eliminiert.
- Der Aufwand für Informationsrecherchen reduziert sich signifikant und gleichzeitig werden Geschäftsabläufe beschleunigt.
- Die Redundanz von Wissen lässt sich durch dessen simultane Wiederverwendbarkeit reduzieren.

Gewiss - virtuelle aus Dokumenten lernende Berater können noch nicht immer eine perfekte Antwort geben. Aber ihre Antworten sind heute schon – wenn's um unstrukturierte Information geht - immer besser als die Ergebnisse, die herkömmliche Technologien liefern. Zu Letzteren sind virtuelle Berater im hier definierten Sinne im Übrigen komplementär – und können einfach integriert werden.

Virtuelle Berater können überall dort sinnvoll eingesetzt werden, wo umfangreiches Wissen in Dokumenten verpackt und regelmäßig von vielen Menschen benötigt wird. Man kann sich virtuelle Berater für Steuern, Förderprogramme, Rentenfragen, Wettbewerbertracking, Normen, Richtlinien, Verfahrensanweisungen, Projekte, F&E u.v.m. vorstellen. Sie arbeiten alle nach denselben Prinzipien und können sich sogar untereinander verständigen und damit ihre Auskunftsfähigkeit noch steigern.

Die Technik für die hier geschilderten kognitiven Lösungen ist verfügbar und binnen Tagen einsatzfähig. Im Wege steht eigentlich nur ein der Vergangenheit entlehntes Paradigma: Wir dokumentieren Wissen in

Dokumenten (Office, Wiki, Web ...) – wir verteilen Wissen in Dokumenten – und wir suchen Wissen (mit viel Aufwand) in Dokumenten. Mit der hier geschilderten Wissenstechnologie kann man alle Vorteile von Dokumenten weiter nutzen – befreit die Inhalte aber von ihrer limitierenden sperrigen Hülse – und verteilt und nutzt diese dann wirksam und mühelos als vernetztes strukturiertes Wissen – mit dem man Antworten auf Suchanfragen erhält anstatt lediglich Listen von Dokumenten.

Noch eine Anmerkung zu den Beispielen: Wir können Ihnen die Wirkungsweise der beschriebenen Technik anhand Ihrer eigenen Website oder eigener Dokumente aufzeigen – unsere in diesem Artikel verlinkten Demoinhalte sind wegen deren Öffentlichkeit limitiert.

**Mehr**

<https://www.theum.com>

[pcaspers@theum.com](mailto:pcaspers@theum.com)