

Masterarbeit im Studiengang
Informationswissenschaften
Modul: Masterprüfung MAP
4. Semester

Die Rolle des Dokuments in Zeiten der digitalen Transformation

Chancen und Herausforderungen bei der Nutzung unstrukturierter Daten mithilfe KI-basierter Wissenstechnologien

vorgelegt von
Isanne Liana Baumann
an der Fachhochschule Potsdam
am 24.10.2020

zur Erlangung des akademischen Grades eines
Master of Arts

Erstprüfer: Prof. Dr. Wolfgang Schulz
Zweitprüferin: Prof. Dr. Angela Schreyer

Inhaltsverzeichnis

Abbildungsverzeichnis	5
Tabellenverzeichnis	6
Abkürzungsverzeichnis	7
Einleitung.....	9
1. Bedeutung des Begriffs „Dokument“	12
1.1 Etymologie und frühe Entwicklung des Begriffs	13
1.2 Theorien zur Bedeutung	14
1.2.1 Die Anfänge: Paul Otlet und Suzanne Briet	14
1.2.2 „Embodied thought“: S.R. Ranganathan	17
1.2.3 Moderne Dokumenttheorie.....	18
1.3 Theorien zum digitalen Dokument	25
1.4 Unterscheidung der Begriffe „Daten“ und „Dokument“	29
2. Technologische Aspekte des digitalen Dokuments	32
2.1 Datenstruktur.....	32
2.2 Auszeichnungssprachen	33
2.3 Semantic Web und Linked Data	38
2.3.1 Resource Description Framework	38
2.3.2 Ontologien	39
2.3.3 Wissensgraphen	40
2.4 Enterprise Content Management.....	41
2.4.1 Aufgaben des ECM.....	42
2.4.2 Funktionen moderner ECM-Systeme	44
2.4.3 Zukunft des ECM	45
3. Künstliche Intelligenz	48
3.1 Geschichte der Künstlichen Intelligenz	49
3.2 Bedeutung des Begriffs „Künstliche Intelligenz“	51
3.3 Maschinelles Lernen, künstliche neuronale Netze und Deep Learning	54
3.4 KI-basierte Wissenstechnologien	57

3.4.1 Text Mining	58
3.4.2 Natural Language Processing	58
4. Untersuchung dokumentenbasierter KI-Software	66
4.1 Produktvorstellungen	66
4.1.1 Theum	66
4.1.2 PhlexDistiller	71
4.1.3 Serviceware Knowledge	73
4.2 Produkttest Theum	74
4.2.1 Methodik	75
4.2.2 Ergebnisse	77
4.2.3 Diskussion	82
4.3 Experteninterviews	83
4.3.1 Methodik	84
4.3.2 Ergebnisse	86
4.3.3 Diskussion	92
5. Fazit	94
Eidesstattliche Erklärung	97
Literaturverzeichnis	98
Anhang	107
Anhang 1: Interviewleitfaden für Isanne Baumann	107
Anhang 2: Fragenkatalog für Sachkundige	110
Anhang 3: Interview-Transkriptionen	112
Interview-Transkription Daniel Koppers, Phlexglobal	112
Transkription Interview mit Paul Caspers, Theum	125
Transkription Interview mit Carolin Hischer, Serviceware	144
Anhang 4: Quellen von im Softwaretest genutzten Dokumenten	165
Gesetze	165
Bedienungsanleitungen	170
Prosa, Dramatik, Epik	174
Wissenschaftliche Artikel	178

Shirley Jackson	182
Anhang 5: Ergebnisse des Softwaretests in elektronischer Form	183
Anhang 6: Zusammengefasste Aussagen aus den Interviews in elektronischer Form	184

Abbildungsverzeichnis

Abb. 1: Lektüre-Vereinbarung nach Salaün	24
Abb. 2: Vergleich der Struktur von Daten und Dokumenten	30
Abb. 3: HTML-Quelltext und dessen Darstellung im Browser.....	35
Abb. 4: Prozess der Generierung eines Ergebnisdokuments	37
Abb. 5: Einfache Darstellung von RDF-Tripeln.....	39
Abb. 6: Beispiel für die Infobox zu Elon Musk	41
Abb. 7: ECM-Modell.....	43
Abb. 8: Vereinfachte Darstellung eines KNN.....	56
Abb. 9: Links Katze, rechts Guacamole	57
Abb. 10: Vereinfachte Darstellung einer Tokenisierung.....	60
Abb. 11: POS-Tagging nach dem Stuttgart-Tübingen-Tagset	60
Abb. 12: Beispiel für Chunking.....	61
Abb. 13: Beispiel für Entity Linking mittels Wikidata-Datensätzen	62
Abb. 14: Google-Antwort zum Alter von Brad Pitt.....	63
Abb. 15: Vereinfachte Darstellung von Relation Extraction	64
Abb. 16: Theum-Wissensbank mit verschiedenen Bibliotheken und geöffnetem Knowledge Navigator.....	68
Abb. 17: Auszug zu einer Antwort auf eine Suchanfrage	69
Abb. 18: Unscharfe Suchanfrage mit Kontext-Vorschlägen.....	70
Abb. 19: Extraktion von Nebenwirkungen aus einem Dokument im PhlexDistiller	72
Abb. 20: Antwortrelevanz der unterschiedlichen Textsorten.....	77
Abb. 21: Relevanz nach Anfragenart beim Textsortenvergleich	79
Abb. 22: Relevante Trefferanzahl in den unterschiedlichen Formaten	80
Abb. 23: Relevante Trefferanzahl in den unterschiedlichen Sprachen	80

Tabellenverzeichnis

Tabelle 1: Auszug aus Interview Block 1.....	86
Tabelle 2: Auszug aus Interview Block 2.....	88
Tabelle 3: Auszug aus Interview Block 3.....	89
Tabelle 4: Auszug aus Interview Block 4.....	90

Abkürzungsverzeichnis

BPM	Business Process Management
CIA	Central Intelligence Agency
CSS	Cascading Style Sheets
DDR	Deutsche Demokratische Republik
DM	Dokumentenmanagement
DTD	Document Type Definition
ECM	Enterprise Content Management
EDV	Elektronische Datenverarbeitung
FID	Federation Internationale de Documentation
GInAS	Global Ingredient Archival System
GML	Generalized Markup Language
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IDMP	Identification of Medicinal Products
ISO	International Organization for Standardization
IT	Informationstechnik
KI	Künstliche Intelligenz
KNN	Künstliche neuronale Netze
LIS	Library and Information Sciences
NLG	Natural Language Generation
NLP	Natural Language Processing
OCR	Optical Character Recognition
OWL	Web Ontology Language
POS	Part of Speech
RDF	Resource Description Framework
RM	Records Management
SaaS	Software as a Service
SGML	Structured General Markup Language
SPARQL	SPARQL Protocol And RDF Query Language
STTS	Stuttgart-Tübingen-Tag-Set
TEI	Text Encoding Initiative
Tf/idf	Term frequency/inverse document frequency
UFOD	Union Française des Organismes de Documentation
URI	Uniform Resource Identifier
WCM	Web Content Management

W3C

World Wide Web Consortium

WWW

World Wide Web

XML

Extensible Markup Language

Einleitung

Dokumente begleiten die Menschheit bereits seit Tausenden von Jahren und nahmen stets eine wichtige Rolle in der Weitergabe von Informationen ein. Über 40.000 Jahre zählen die ältesten bisher entdeckten Höhlenmalereien, die von Steinzeitmenschen angefertigt wurden und denen Informationen über deren damalige Lebensweise entnommen werden können. Die ersten Schriftzeugnisse sind ca. 6000 Jahre alt, wobei es sich um sumerische Symbole auf Tontafeln handelt, die genutzt wurden, um geschäftliche Transaktionen festzuhalten.¹ Seitdem änderten sich die Schriftarten und die Medien, auf denen das Geschriebene festgehalten wurde. Über Keilschriften und Hieroglyphen bis zur Buchstabenschrift², von Tontafeln, Leder, Papyrus, Wachstafeln, Pergament, Papier und vielen weiteren Materialien³ führte der Weg der Dokumente schließlich durch das Aufkommen der Computer erstmals zu einer immateriellen, digitalen Form, bestehend aus Bits und Bytes.

Die Entwicklung vom Analogen zum Digitalen stellt den bisher grundlegendsten Einschnitt in der Geschichte der Dokumente dar. Viele der Eigenschaften, die bis zu diesem Zeitpunkt für sie galten, veränderten sich durch die Digitalisierung maßgeblich. Die heutigen digitalen Dokumente können ohne jegliche Abweichungen vom Original reproduziert und innerhalb von Millisekunden um die Welt geschickt werden. Sie können, ohne Spuren auf dem Original zu hinterlassen, beliebig oft verändert werden, sind flexibel und weniger statisch. Durch Cloud-Techniken sind sie nicht mehr an einen bestimmten Ort oder ein Gerät gebunden, sondern können praktisch von überall jederzeit abgerufen werden. Aufgrund der Einbindung von Hyperlinks sind Dokumente stark untereinander vernetzt, was ihre Grenzen mehr und mehr verschwimmen lässt.

Die Informationsdichte der Dokumente nahm im Lauf des Digitalisierungsprozesses exponentiell zu. Konnte man auf Tontafeln noch durchschnittlich 5,3 Zeichen pro Quadratzentimeter abbilden, waren es bereits ca. 16.465 auf einer Diskette und mehr als 7,7 Millionen auf einer optischen Bildplatte.⁴ Es können immer mehr Informationen auf immer kleinerem Raum festgehalten werden, was jedoch nicht nur Vorteile mit sich bringt. Die unverhältnismäßig umfangreichen Datensammlungen, die in Dokumenten heute im gesamten Internet, am Arbeitsplatz und sogar auf privaten Rechnern vorliegen, sind auf

¹ Vgl. Liu, Z. (2004): The evolution of documents and its impacts, S. 280 [elektronische Quelle].

² Vgl. Funke, F. (1999): Buchkunde, S. 14.

³ Vgl. ebd., S. 54 f.

⁴ Conway, P. (1996): Preservation in the Digital World, S. 4.

manuelle Art und Weise kaum mehr überschaubar. Strukturierte Dateiformate, Metadaten und Auszeichnungssprachen, wie sie z.B. das Semantic Web dominieren, versuchen, dieser Entwicklung entgegenzuwirken, reichen jedoch nicht aus, um das Wissen, welches in Dokumenten enthalten ist, effektiv zu extrahieren.

Auf dem Vormarsch sind daher Technologien, die eine automatische Verarbeitung von digitalen Dokumenten ermöglichen. Hier, wie auch in vielen anderen Bereichen, kommt künstliche Intelligenz (KI) immer häufiger zum Einsatz. Laut dem Unternehmen Gartner Inc., welches Marktanalysen im Bereich der Informationstechnik (IT) publiziert, werden die Vorteile von KI momentan zur Realität in Unternehmen und die Technologien beginnen, ihr Potenzial voll auszuschöpfen.⁵ Mittels maschinellem Lernen, Text Mining und Natural Language Processing (NLP) wird das Auffinden von Wissen in Dokumenten unterstützt. Durch syntaktische und semantische Analyse wird es möglich, exakt die gesuchten Informationen aus den vorliegenden Dokumentquellen zu extrahieren und für die Suchenden in einer benutzungsfreundlichen Form aufzubereiten. Die sich stetig optimierenden KI-Technologien könnten für die weitere Entwicklung der Dokumente die nächste Zäsur bedeuten, welche ihre Funktionen und ihr Erscheinungsbild stark verändern könnten.

Ziel der vorliegenden Arbeit ist es, die gegenwärtige Rolle des Dokuments und deren mögliche Veränderungen in der Zukunft zu erforschen. Dies wird durch die Auswertung einschlägiger Forschungsliteratur sowie der Durchführung qualitativer Interviews mit Sachkundigen aus dem Bereich der KI-basierten Wissenstechnologien und einem Test eines der in diesem Zusammenhang vorgestellten Produkte ermöglicht. So wird auf die Fragen eingegangen, was analoge und digitale Dokumente voneinander unterscheidet, welche Funktionen das digitale Dokument heute besitzt und wie diese durch KI-Technologien stetig erweitert und verbessert werden.

Das erste Kapitel befasst sich mit der Bedeutung des Begriffs „Dokument“. Zu diesem Zweck wird zunächst die Etymologie des Wortes beschrieben, gefolgt von mehreren unterschiedlichen Theorien und Definitionsversuchen renommierter Wissenschaftler und Wissenschaftlerinnen aus diesem Bereich. Dabei wird chronologisch vorgegangen. So werden zunächst Theorieansätze zum analogen und anschließend zum digitalen Dokument besprochen. Zum Abschluss des Kapitels folgt ein Vergleich der Begriffe

⁵ Vgl. Goasduff, L. (2020): 2 Megatrends Dominate the Gartner Hype Cycle for Artificial Intelligence, 2020 [elektronische Quelle].

„Dokument“ und „Daten“, da diese sehr häufig miteinander in Verbindung gebracht werden und Daten für den weiteren Verlauf der Arbeit von hoher Relevanz sind.

Kapitel zwei behandelt die wichtigsten technologischen Aspekte des Dokuments, welche die digitale Transformation begünstigt haben, sowie deren Auswirkungen im Bereich des Internets und auf das Management von Dokumenten. Unterschiedliche Grade von Strukturierung in Dokumenten, welche einen zentralen Punkt in der Verarbeitung derselben ausmachen, werden besprochen, gefolgt von Auszeichnungssprachen, Semantic Web und Linked Data. Das letzte Teilkapitel behandelt das Enterprise Content Management (ECM) und zeigt die Umbrüche auf, die in diesem Bereich in den letzten Jahren stattfanden und möglicherweise in Zukunft folgen werden.

Mit einem der wichtigsten Themen der Zukunft befasst sich auch Kapitel drei, welches sich auf künstliche Intelligenz konzentriert. Der Begriff und seine Geschichte werden umrissen, Lernmethoden der KI werden erklärt. Anschließend folgt eine detaillierte Beschreibung KI-basierter Wissenstechnologien, die vor allem das Thema Natural Language Processing ausführlich behandelt und die dazugehörigen Techniken anhand von Beispielen erklärt.

Auf diesen Theorieteil folgen praktische Beispiele aus dem Bereich der KI-basierten Wissenstechnologien. Drei Produkte verschiedener Firmen, die mit KI-Techniken in Verbindung mit Dokumenten arbeiten, werden vorgestellt. Eines dieser Produkte konnte durch die Zurverfügungstellung der Demoversion einem Softwaretest unterzogen werden. Methodik und Ergebnisse dieses Tests werden im zweiten Unterkapitel ausführlich erläutert und diskutiert. Im Anschluss folgt die Auswertung dreier qualitativer Interviews, die mit Fachkundigen aus den Firmen geführt wurden, deren Produkte in der Arbeit vorgestellt werden. Das Fazit, in welchem die Erkenntnisse dieser Masterarbeit zusammenfassend aufgearbeitet werden, bildet das fünfte und letzte Kapitel der Thesis.

1. Bedeutung des Begriffs „Dokument“

Wenn man Menschen nach Zufallsprinzip auf der Straße fragen würde, was ein Dokument ist, fühlte sich höchstwahrscheinlich die Mehrheit der Befragten dazu in der Lage, den Begriff zu umschreiben oder zumindest einige Beispiele dafür zu nennen. Denn es ist ein Begriff, dem man tagtäglich begegnet. Er taucht nicht so selten auf wie z.B. „Axiom“, ist nicht wie „Gravitationskonstante“ auf das Vorkommen in einem speziellen wissenschaftlichen Bereich reduziert und scheint leichter zu veranschaulichen zu sein als „Kreativität“. Folglich wären diese zum Begriff des Dokuments befragten Personen vermutlich überrascht, wüssten sie, dass die Definition desselben schon einigen Forschenden Kopfzerbrechen bereitet hat und sich hinter ihm mehr verbirgt als man vermutet.

Ein Blick in ein Wörterbuch bringt die folgenden vier Bedeutungen von „Dokument“ hervor:

1. Urkunde, amtliches Schriftstück
2. Beweisstück, Zeugnis
3. Kurzform für Parteidokument in der Deutschen Demokratischen Republik (DDR)
4. strukturierte, als Einheit erstellte und gespeicherte Menge von Daten; [Text]datei in der Elektronischen Datenverarbeitung (EDV)⁶

Schon hier tauchen nach kurzem Überlegen Fragen auf. Ist ein Roman kein Dokument? Was ist mit einer Werbebroschüre, was mit einem Tagebuch? Es stellt sich heraus, dass die zunächst vergleichsweise naheliegende Definition wohl doch nicht so trivial ist und sich der Sinn des Begriffs allein anhand eines Wörterbuchs nicht ermitteln lässt. Im Folgenden werden daher die Herkunft des Begriffs „Dokument“ sowie mehrere Theorien renommierter Wissenschaftler und Wissenschaftlerinnen über dessen Bedeutung aufgezeigt. Die Betrachtung folgt dabei einem chronologischen Verlauf vom analogen bis hin zum digitalen Dokument und beschäftigt sich zuletzt mit der Frage nach dem Unterschied zwischen den Begriffen „Dokument“ und „Daten“.

⁶ Duden (2020): Dokument [elektronische Quelle].

1.1 Etymologie und frühe Entwicklung des Begriffs

Zur Nachverfolgung der Entwicklung eines Begriffs lohnt zu Beginn stets ein Blick in ein etymologisches Wörterbuch. Das deutsche Wort „Dokument“ wurde im 16. Jahrhundert aus dem lateinischen Substantiv „documentum“ entlehnt, was so viel bedeutet wie Lehre, warnendes Beispiel, Beweis, Probe. Das entsprechende lateinische Verb „docēre“ bedeutet lehren, unterrichten.⁷ In der ursprünglichen Bedeutung des Begriffs erscheint das Dokument also eher als etwas Ideelles denn etwas Materielles, allein die Übersetzung „Beweis“ deutet auf eine mögliche Materialität hin.

Auch das französische Pendant „document“ wurde primär als “ce qui sert à instruire, enseignement, leçon”⁸ [das, was zum Unterrichten dient, Unterweisung, Lektion] verstanden. Diese Bedeutung wandelte sich erst durch Aufkommen der europäischen Bürokratie im 17. Jahrhundert. In diesem Zeitraum wurde es als “écrit servant de preuve ou de renseignement”⁹ [Schrift, die zum Beweis oder zur Auskunft dient] beschrieben.¹⁰ Erstmals wird hier also klar ausgedrückt, dass es sich beim Dokument um etwas Materielles, nämlich etwas Schriftliches handelt, das einem bestimmten Zweck dient. Nach und nach löste das Dokument damals, zu Beginn der europäischen Moderne, orale oder anderweitig vereinbarte Abmachungen ab. Stattdessen wurden immer häufiger schriftliche Verträge aufgesetzt. Zudem wurden Dokumente nun auch als Beweise verwendet, beispielsweise in Form von Besitzurkunden o.Ä. und auch als Art der Auskunft, der Information, z.B. in Form eines politischen Traktats. Diese damaligen Eigenschaften bzw. Gebrauchsformen des Dokuments vereint der norwegische emeritierte Professor Niels W. Lund in seiner „Document Theory“ unter dem Schlagwort „written, true knowledge“¹¹.

Im 19. Jahrhundert wurden Dokumente dann auch in der Verwaltung und vor allem der Wissenschaft zu einem wichtigen Bestandteil der gängigen Praxis. Von Forschenden wurden nun empirische Beweise in Form von Schriftstücken gefordert, um ihre Thesen zu untermauern.¹² Die gesellschaftliche Bedeutung des Dokuments nahm immer mehr zu. Es war nötig zur Verbreitung von Wissen und Technologien und auch, um die

⁷ Vgl. Pfeifer, W. et al. (1993): Etymologisches Wörterbuch des Deutschen, Stichwort „Dokument“ [elektronische Quelle].

⁸ Rey, Alain (1992): Dictionnaire historique de la langue française contenant les mots français en usage et quelques autres délaissés, S. 620.

⁹ Ebd.

¹⁰ Vgl. Lund, N. W. (2011): Document Theory, S. 2 [elektronische Quelle].

¹¹ Vgl. ebd., S. 3 [elektronische Quelle].

¹² Vgl. ebd. [elektronische Quelle].

Stabilität von Organisationen und die Entwicklung des Handels zu gewährleisten.¹³ Der französische Professor für Informationswissenschaften Jean-Michel Salaün wagt daher sogar die These, dass der Status, den das Dokument heutzutage besitzt, daher rührt, dass es in Zeiten der wissenschaftlichen und industriellen Revolution als wirkungsmächtiges Instrument zum Zweck des Transfers, Überzeugens und Beweisens unerlässlich war.¹⁴

Nachdem die Etymologie des Wortes „Dokument“ und seine frühe Entwicklung erläutert wurden, werden in den folgenden Abschnitten verschiedene Theorien aus dem 20. und 21. Jahrhundert zu dessen Bedeutung vorgestellt und erörtert.

1.2 Theorien zur Bedeutung

Seit den 1930er Jahren wurden einige Theorien zur Bedeutung des Begriffs „Dokument“ aufgestellt, auch wenn ihre Anzahl beispielsweise weit unter jenen zur Bedeutung von „Information“ liegt, was u.a. darauf zurückzuführen ist, dass der Begriff der Information im Vergleich zum Begriff des Dokuments institutionellen Charakter hat. Nichtsdestotrotz gab es vor allem in der jüngeren Vergangenheit doch einige Forschende im Bereich der Library and Information Sciences (LIS), die sich stark mit dem Dokument auseinandergesetzt haben.

1.2.1 Die Anfänge: Paul Otlet und Suzanne Briet

Den Anfang der Überlegungen zur Bedeutung des Dokuments machte Paul Otlet. Der 1868 geborene Belgier und gelehrte Jurist wandte kurze Zeit nach seinem Einstieg in das Berufsleben seine Aufmerksamkeit dem damaligen Bibliothekssystem zu, welches er als verbesserungswürdig empfand. Daher gründete er zusammen mit dem Anwalt und zukünftigen Nobelpreisträger Henri La Fontaine 1895 das „Office International de Bibliographie“ mit dem Vorhaben, eine universelle Bibliothek, genannt „Mundaneum“, zu erschaffen. Erreicht wurde dieses ambitionierte Ziel zwar nie komplett, doch 1934 war immerhin die beachtliche Summe von 15,6 Millionen Titeln im bibliografischen Katalog des Mundaneums vorhanden. Otlets Ideen hinsichtlich der Organisation von Wissen waren seiner Zeit weit voraus und noch nicht realisierbar, traten aber Jahre später, wenn auch

¹³ Vgl. Salaün, J. (2014): *Why the Document Is Important... and How it Is Becoming Transformed*, S. 188 [elektronische Quelle].

¹⁴ Vgl. ebd. [elektronische Quelle].

technisch auf andere Art und Weise als von ihm gedacht, in Form des Internets in Erscheinung.¹⁵

Otlet schrieb sehr viel über seine Theorien zur Organisation von Information. Eines seiner bekanntesten Werke ist das Buch „Traité de documentation“ von 1934. In dieser Arbeit beschäftigt er sich unter anderem mit der Frage, was unter einem Dokument zu verstehen ist. Er beschreibt das geschriebene oder grafische Dokument als eine Repräsentation von materiellen Objekten oder abstrakten Bildern. Die Objekte selbst können seiner Meinung nach ebenfalls Dokumente sein, wenn sie dem Betrachtenden zu Erkenntnissen, also einem Gewinn an Information, verhelfen. Als Beispiele für solche Objekte nennt er natürliche Objekte, künstliche, vom Menschen geschaffene Objekte (auch solche, die Konzepte darstellen sollen, wie z.B. Modelle), archäologische Funde und Kunst.¹⁶ Otlets Verständnis zufolge wären also auch antike Werkzeuge Dokumente, ebenso ein Gemälde von Leonardo Da Vinci. Er fasst den Dokumentbegriff folglich sehr weit. Walter Schürmeyer, ein Zeitgenosse Otlets und deutscher Bibliothekar, schrieb in einem Aufsatz über Dokumentation, das Wort „Dokument“ beschränke sich nicht auf das geschriebene oder gedruckte Wort, sondern sei „jede materielle Unterlage zur Erweiterung unserer Kenntnisse, die einem Studium oder Vergleich zugänglich ist.“¹⁷ Mit dieser Definition liegt er sehr nahe bei Otlet, auch wenn er nicht deutlich macht, was er unter einer „materiellen Unterlage“ versteht.

Ebenfalls recht weit fasste den Dokumentbegriff eine Bibliothekarin und Historikerin, die in etwa zur selben Zeit wie Otlet und Schürmeyer in Frankreich lebte: Suzanne Briet. Ihr Status als Vorreiterin auf dem Gebiet der Dokumentation trug ihr den Spitznamen „Madame Documentation“ ein. Sie war beteiligt an der Gründung und späteren Führung der „Union Française des Organismes de Documentation“ (UFOD), wurde 1951 Studiendirektorin des „Institut National de Techniques de la Documentation“ und später Vizepräsidentin der „Federation Internationale de Documentation“ (FID).¹⁸ Ihr bekanntestes Werk, „Qu'est-ce que la documentation?“, in dem sie das Dokumentationswesen behandelt, veröffentlichte sie 1951. In diesem beschreibt sie ein Dokument als „tout indice concret ou symbolique, conservé ou enregistré, aux fins de représenter, de reconstituer ou de prouver un phénomène ou physique ou intellectuel“¹⁹ [jedes konkrete oder symbolische Indiz, das erhalten oder aufgezeichnet wird, um ein physisches oder intellektuelles

¹⁵ Vgl. Wright, A. (o.J.): Paul Otlet; Hartmann, F. (2006): Von Karteikarten zum vernetzten Hypertext-System [elektronische Quelle].

¹⁶ Vgl. Otlet, P. (1934): Traité de documentation, S. 217.

¹⁷ Schürmeyer, W. (1935): Aufgaben und Methoden der Dokumentation, S. 537.

¹⁸ Vgl. Buckland, M. (o.J.): Michael Buckland's Suzanne Briet Page [elektronische Quelle].

¹⁹ Briet, S. (1951): Qu'est-ce que la documentation?, S. 7.

Phänomen darzustellen, zu rekonstruieren oder zu beweisen]. Wie sie selbst anmerkt, ist diese Beschreibung sehr abstrakt und schwer zu begreifen.²⁰ Daher versucht sie anhand von Beispielen, näher auf die Bedeutung einzugehen:

„Une étoile est-elle un document? Un galet roulé par un torrent est-il un document? Un animal vivant est-il un document? Non. Mais sont des documents les photographies et les catalogues d'étoiles, les pierres d'un musée de minéralogie, les animaux catalogués et exposés dans un Zoo.“²¹

[Ist ein Stern ein Dokument? Ist ein Kieselstein, der durch einen Strom getrieben wird, ein Dokument? Ist ein lebendes Tier ein Dokument? Nein. Aber die Fotografien und Kataloge der Sterne, die Steine in einem mineralogischen Museum, die Tiere, die in einem Zoo katalogisiert und ausgestellt sind, sind Dokumente.]

Briet verdeutlicht ihre Theorie anhand eines in der Welt der Informationswissenschaften berühmt gewordenen Beispiels, jenem der Antilope. Ein Forscher trifft auf einer Expedition in Afrika auf eine neue Antilopenart. Es gelingt ihm, eine Antilope zu fangen und zurück nach Europa zu bringen, um sie dort „auszustellen“. Zu diesem Zweck werden Pressemitteilungen, Radio- und Filmnachrichten verfasst, um der Gesellschaft den Fund mitzuteilen. Das Tier im Zoo wird katalogisiert und nach seinem Tod in ausgestopfter Form im Museum ausgestellt. Audiovisuelle Aufnahmen des Tiers werden erstellt, es erscheinen Einträge über seine Art in Enzyklopädiën. Diese Werke tauchen in Bibliotheken und Buchhandlungen auf, es werden Zeichnungen, Statuen, Fotos und Filme der Kreatur geschaffen. Die katalogisierte Antilope selbst ist laut Briet ein „document initial“ [Ursprungsdokument], alle anderen folgenden Dokumente, die sich mit ihr beschäftigen, sind „documents seconds“ [sekundäre Dokumente].²²

Briet bezeichnet also nicht nur das Katalogisat der Antilope als Dokument, sondern die Antilope im Zoo selbst. Wie Otlet legt auch sie sich nicht auf Geschriebenes fest, geht jedoch noch weiter als er und bindet sogar Lebewesen in ihre Vorstellung von einem Dokument mit ein. Michael Buckland, der sich ausführlich mit der Person Suzanne Briet und ihrem Werk beschäftigt hat, geht darauf ein, dass aus ihrem Text nicht klar hervorgeht, ab wann genau etwas zum Dokument wird. Er leitet daraus aber vier Regeln für Dokumente aus Briets Sicht ab: (1) Materialität, das Objekt ist auf jeden Fall physisch vorhanden, (2) Absicht, das Objekt wird vorsätzlich als Beweisstück behandelt, (3) das

²⁰ Vgl. Briet, S. (1951): *Qu'est-ce que la documentation?*, S. 7.

²¹ Ebd.

²² Vgl. ebd., S. 7 f.

Objekt muss zu einem Dokument verarbeitet werden und (4) das Objekt muss von der Außenwelt als Dokument wahrgenommen werden. Er wirft die Frage auf, ob Briet wohl meinte, dass ein Objekt erst dann zum Dokument wird, wenn es als solches behandelt wird.²³ Leider wird dieser Punkt in ihren Ausführungen nicht ganz klar. Fest steht jedoch, dass Briet der Meinung war, dass jedes beliebige Objekt ein Dokument sein kann, sofern es in ein Dokumentationssystem eingebunden ist. Der bereits erwähnte Professor Jean-Michel Salaün sieht den Unterschied zwischen Otlet und Briet darin, dass Otlet den Fokus auf die Übermittlung von Erkenntnissen legt, während es Briet mehr um den Beweischarakter von Dokumenten geht.²⁴ Daher wären aus der Sicht Otlets rein quantitativ gesehen wohl mehr Objekte Dokumente als aus Briets Sicht. Denn für Briet muss ein Objekt mehr Bedingungen erfüllen, um als Dokument zu gelten als für Otlet. Ein archäologischer Fund, z.B. ein antiker Meißel, ist für Otlet bereits ein Dokument, seine Betrachtung verschafft uns einen Gewinn an Information. Für Briet wäre der Meißel erst dann ein Dokument, wenn er in einem Dokumentationssystem angelegt wäre.

1.2.2 „Embodied thought“: S.R. Ranganathan

Eine weitaus reduziertere Betrachtung des Dokuments, widersprüchlich zu der Otlets oder Briets, verfolgte dagegen der indische Mathematiker und Bibliothekar Shiyali Ramamrita Ranganathan. In seinem Werk „Documentation and its facets“ berichtet er von einem Brüsseler Kongress in den 1960er Jahren zum Thema audiovisuelle Materialien, auf dem mehrere Teilnehmende darüber diskutierten, was unter einem Dokument zu verstehen sei. Dabei wurden unter anderem Radio, Fernsehen, Statuen und Ausstellungsstücke in Museen genannt, da alle diese Dinge in irgendeiner Weise Gedanken übermitteln, eine Betrachtung ganz im Sinne Otlets also. Ranganathan sieht jedoch keines dieser Objekte als Dokument an, da seiner Meinung nach nur als Dokument bezeichnet werden kann, was ein „record on a more or less flat surface“²⁵ ist. An anderer Stelle bezeichnet er konventionelle Dokumente als „embodied thought“²⁶, Aufzeichnungen auf Papier oder einem anderen Material, das sich leicht transportieren und erhalten lässt. Als Beispiele für konventionelle Dokumente nennt er unter anderem Bücher, Zeitschriften, Landkarten und Atlanten. Als unkonventionelle Dokumente benennt er z.B. Mikrofilme oder Hörbücher, denn auch diese sind Aufzeichnungen auf mehr oder weniger flachem Material, auch wenn es sich nicht um Geschriebenes handelt.²⁷ Bei

²³ Vgl. Buckland, M. (1997): What Is a „Document“?, S. 806 [elektronische Quelle].

²⁴ Vgl. Salaün, J. (2014): Why the Document Is Important... and How it Is Becoming Transformed, S. 189 [elektronische Quelle].

²⁵ Ranganathan, S. R. (1963): Documentation and its facets, S. 41.

²⁶ Ebd., S. 29.

²⁷ Vgl. ebd., S. 30-37.

Ranganathan wird der Dokumentbegriff folglich stark auf die Form, das Äußere reduziert. Es geht ihm nicht wie Otlet um die Übermittlung von Gedankengut und auch nicht wie Briet um den Beweisaspekt. Die Theorie Ranganathans hat ihren Reiz, da sie gegenüber anderen Theorien sehr viel einfacher klingt und sich die Unterscheidung zwischen Dokument und Nicht-Dokument leichter treffen lässt. Aber sie ist vielleicht auch zu vereinfachend, da sie andere Aspekte als jenen der Form, wie z.B. den Inhalt des Dokuments, außer Acht lässt bzw. nur sehr schwammig beschreibt.

1.2.3 Moderne Dokumenttheorie

Michael Buckland, bereits im Zusammenhang mit seinen Studien zu Suzanne Briet erwähnt, beschäftigt sich ebenfalls seit Jahrzehnten mit dem Dokumentbegriff. Der ehemalige Bibliothekar und mittlerweile emeritierte Professor für Bibliotheks- und Informationswissenschaften hat mit seiner Forschung vieles zum Thema beigetragen. Anstatt von einer „Informationsgesellschaft“ zu sprechen, spricht Buckland von einer „Dokumentengesellschaft“, in der wir heute leben. Diese Äußerung begründet er dadurch, dass Menschen schon seit Anbeginn der Zeit dadurch überleben, dass sie Informationen untereinander austauschen. Information spielte schon immer eine zentrale Rolle. Verändert hat sich dagegen die Rolle der Dokumente. Ihre Anzahl nahm im Lauf der Zeit stetig zu und ebenso ihre Stellung im Leben der Menschen, da diese immer mehr von Aufzeichnungen abhängig sind. Mit dem Zuwachs an Dokumenten entstanden Berufsfelder, die damit zusammenhängen, wie die Bibliotheks- und Dokumentationswissenschaft oder das Informationsmanagement. Durch diese Entwicklung wurde es nötig, den Begriff des Dokuments genauer zu definieren und die moderne Dokumententheorie setzte um 1990 ein.²⁸

Einer der bekanntesten Aufsätze Bucklands spricht über die dreifache Bedeutung von Information: Information als Prozess, als Wissen und als Ding. Auf die beiden ersteren soll an diesem Punkt nicht eingegangen werden, jedoch auf „Information-as-thing“, wie Buckland es nennt. Unter diesem Stichwort versteht er „representations of knowledge (and of events)“²⁹, also beispielsweise Texte und Daten, aber auch Objekte wie Fossilien und Artefakte. Seiner Meinung nach deckt sich diese Bedeutung von Information mit einer weit gefassten Definition des Dokuments. Wie es zu dieser allgemeinen Beschreibung des Begriffs kommt, zeigt sich gut in folgendem Zitat von Buckland:

²⁸ Vgl. Buckland, M. (2018): Document Theory, S. 426 f [elektronische Quelle].

²⁹ Buckland, M. (1991): Information as Thing, S. 352 [elektronische Quelle].

„A printed book is a document. A page of hand-writing is a document. A diagram is a document. A map is a document. If a map is a document, why should not a three-dimensional contour map also be a document. Why should not a globe also be considered a document since it is, after all, a physical description of something.[...] If a globe, a model of the earth, is a document, why should one not also consider a model of a locomotive or of a ship to be a document?“³⁰

Es zeigt sich, dass es also schwerfällt, den Dokumentbegriff klar zu definieren, wenn man keine ausdrücklichen Grenzen setzt, wie es z.B. bei Ranganathan der Fall ist. Dieser beschränkt sich, wie bereits erwähnt, auf eine mehr oder weniger ebene Fläche, ein Globus würde also nicht in diese Kategorie fallen. Aber gerade bei diesem Beispiel erscheint Ranganathans Theorie auch ein wenig willkürlich gewählt. Wieso sollte eine papierne, also flache Karte der Erde ein Dokument sein, eine Karte in Form eines Globus aber nicht? Und das, obwohl die papierne Karte durch die Mercator-Projektion und die einhergehenden Verzerrungen genau genommen sogar falsch ist, die des Globus jedoch korrekt? Buckland beurteilt eine solche Definition, die sich sehr stark auf die äußere Form fokussiert, als zu eingeschränkt.³¹

Er kritisiert jedoch genauso Briet, nämlich in dem Punkt, dass sie sich zu stark auf den Beweisaspekt von Dokumenten konzentriert. Seiner Meinung nach werden dadurch z.B. viele belletristische Bücher nicht als Dokumente berücksichtigt, da sie nicht als Beweis dienen, sondern primär zur Unterhaltung geschaffen wurden. Bei Briet muss ein Objekt nicht nur von der Form her ein Dokument sein (wie es z.B. bei belletristischen Büchern der Fall ist), sondern auch den „Beweisaspekt“ erfüllen, also in gewisser Weise einen Anspruch auf Wahrheit, Rechtsgültigkeit oder moralische Korrektheit haben, um als Dokument zu gelten.

Buckland zeigt jedoch eine zweite Option auf, nämlich diejenige, sich von dem Pochen auf sachliche Beweise zu befreien und auch unterhaltende Werke als Dokumente zu verstehen, denn auch diese können die Lesenden etwas lehren, wenn auch auf andere Weise als wissenschaftliche Werke³², womit man wieder beim ursprünglichen lateinischen „docēre“ wäre.

Auch einige norwegische Wissenschaftler und Wissenschaftlerinnen haben sich mit dem Dokumentbegriff beschäftigt. 1989 wurde die norwegische Nationalbibliothek gegründet.

³⁰ Buckland, M. (1991): Information as Thing, S. 354 [elektronische Quelle].

³¹ Vgl. Buckland, M. (1997): What Is a „Document“?, S. 806 f. [elektronische Quelle].

³² Vgl. Buckland, M. (2018): Document Theory, S. 430 [elektronische Quelle].

Damit ging ein Gesetz einher, in dem beschrieben ist, welche Medien die Bibliothek zu sammeln hat. In diesem Gesetz wurden nicht nur sowohl analoge als auch digitale, schriftliche Dokumente erwähnt, sondern auch Filme sowie Radio- und Fernsehbeiträge, die ebenfalls als Dokumente betitelt wurden. In Anlehnung an dieses Gesetz wurde 1996 ein neues Programm an der Universität von Tromsø mit dem Namen „Documentation Studies“ ins Leben gerufen.³³ Zudem wurde 2001 die „Document Academy“ gegründet, ein interdisziplinäres, globales Forschungskollektiv, das im Bereich Dokumente und Dokumentation forscht, publiziert und Workshops abhält.³⁴ Einer der Gründer der Academy und ebenso ehemaliger Leiter des Programms an der Universität Tromsø ist Niels W. Lund. Dieser stellt sich die Frage, ob der eigentliche Kern der Meinungsverschiedenheiten rund um die Bedeutung des Begriffs „Dokument“ nicht dessen physische Form ist, sondern vielmehr seine sozialen und kulturellen Dimensionen.³⁵

Auf diesen Punkt geht auch eine über 100-köpfige französische Forschergruppe ein, die unter dem Namen Roger T. Pédauque veröffentlicht. Das Besondere an Pédauque ist, dass zu der Gruppe nicht nur Forschende aus dem informationswissenschaftlichen Bereich gehören, sondern auch aus anderen Bereichen, in denen Dokumente zwar ebenso eine wichtige Rolle spielen, jedoch nicht den Kern der Wissenschaft ausmachen, wie es beispielsweise in den Geschichtswissenschaften der Fall ist. Dadurch soll ein multidisziplinärer Ansatz geschaffen werden.³⁶ Aus diesem Grund wird bei Pédauque das Dokument in drei Dimensionen unterteilt, die für jeweils einen Teil der Forschenden die größte Rolle spielen:

1. „The document as form“: die materielle bzw. immaterielle Struktur von Dokumenten
2. „The document as sign“: der Inhalt des Dokuments, der ihm seine Bedeutung verleiht
3. „The document as medium“: das Dokument als eine Spur von Kommunikation, die außerhalb von Raum und Zeit existiert³⁷

Aufbauend auf der ersten der drei Dimensionen kommt Pédauque zu einer ersten Definition, bzw. Gleichung: *Traditional document = medium + inscription*³⁸. Mit „traditionell“ ist dabei ein analoges, also nicht-elektronisches Dokument gemeint. Unter dem Medium wird beispielsweise Papier oder Pergament verstanden, das darauf Geschriebene ist die

³³ Vgl. Lund, N.W. (2007): Building a Discipline, Creating a Profession: An Essay on the Childhood of „dokvit“, S. 12 [elektronische Quelle].

³⁴ Vgl. Document Academy (o.J.): About [elektronische Quelle].

³⁵ Vgl. Lund, N. W. (2011): Document Theory, S. 21 [elektronische Quelle].

³⁶ Vgl. Pédauque, R. T. (2003): Document: Form, Sign and Medium, As Reformulated for Electronic Documents, S. 1 f. [elektronische Quelle].

³⁷ Vgl. ebd., S. 3. [elektronische Quelle]

³⁸ Ebd., S. 5. [elektronische Quelle]

inscription. Im Falle von Audiodateien oder Videos reicht allein das Medium allerdings nicht mehr aus, es wird außerdem ein Gerät benötigt, welches das „Geschriebene“ sichtbar bzw. für den Menschen verständlich macht.

Die zweite Dimension führt zur zweiten Gleichung: *Document = inscription + meaning*.³⁹ Der Inhalt ist in Form des Geschriebenen materialisiert und vermittelt die Bedeutung, welche davon abhängt, in welchem Kontext das Dokument produziert und verteilt und somit interpretiert wird. Diese Dimension ist von drei Ideen geprägt: der Erstellung von Dokumenten, ihrer Interpretation und den Zeichen, aus denen sie bestehen. Die Erstellung eines Dokuments geht automatisch mit einer Klassifikation einher, das Dokument wird einer bestimmten Art zugeordnet. Diese Klassifikationen können unterschiedlichster Art sein, konkret oder abstrakt, spezifisch oder unspezifisch, aber eine Zuordnung zu einer Gruppe erfolgt. Die Interpretation findet beim Lesen des Dokuments statt und kann je nach Lesendem sehr unterschiedlich ausfallen. Viele Faktoren spielen dabei eine Rolle, beispielsweise die Zeit, in der die interpretierende Person lebt oder ihr sozialer Status. Die dritte Idee betrifft die Zeichen, also die Dokumente selbst. Pédauque geht davon aus, dass die meisten Dokumente aus geschriebener oder gesprochener Sprache bestehen und bezeichnet sie als „discrete pieces, more or less separable and recombable, analyzable, subjected to syntax, discourse structuring and style rules.“⁴⁰

Die dritte und letzte Gleichung lautet: *Document = inscription + legitimacy*.⁴¹ Sie soll den sozialen Prozess der Dokumenterstellung beschreiben. Legitimität erreicht ein Dokument durch Erfüllen zweier Bedingungen: Das Geschriebene muss über private Kommunikation hinausgehen und es darf nicht kurzlebig sein, es darf nicht nur in seiner Gegenwart auftauchen, sondern muss auch in der Zukunft zugänglich sein. Daher sagt Pédauque, dass z.B. ein Tagebuch kein Dokument ist, da es die erste Bedingung nicht erfüllt (es sei denn, es wird publiziert) und ein Livestream ebenfalls nicht, da er die zweite Bedingung nicht erfüllt (es sei denn, er wird auch aufgezeichnet und nicht nur live gesendet). Pédauque merkt jedoch an, dass nicht alle Mitglieder der Forschungsgruppe mit diesem Punkt einverstanden sind. Zudem enthält dieser Punkt auch Ausnahmen. Beispielsweise werden laut Pédauque medizinische Akten ebenfalls nicht publik gemacht, trotzdem sind sie legitim, da sie einen sozialen Charakter haben, dadurch, dass sie nach bestimmten Regeln erstellt werden.⁴²

³⁹ Pédauque, R. T. (2003): Document: Form, Sign and Medium, As Reformulated for Electronic Documents, S. 12. [elektronische Quelle]

⁴⁰ Vgl. ebd., S. 12 f. [elektronische Quelle]

⁴¹ Ebd., S. 17. [elektronische Quelle]

⁴² Vgl. ebd., S. 17 f. [elektronische Quelle]

Dass sich die Anzahl der Dokumente seit langer Zeit kontinuierlich erhöht, hängt mit der gesellschaftlichen Entwicklung zusammen, und zwar mit der sozialen Nutzung der Dokumente und ihrer internen Ökonomie. Beispielsweise explodierte die Dokumentenanzahl in Zeiten der Industrialisierung durch die damals stattfindenden Umwälzungen. Neue Technik, neue Organisationsformen, neues Know-how in vielen Bereichen führt zu vielen neuen Dokumenten. Die soziale Nutzung nimmt also extrem zu. Die Ökonomie der Dokumente verändert sich ebenfalls, nämlich aufgrund sich weiter entwickelnder Technologien, die bei der Erstellung genutzt werden und wegen anderen Veränderungen im Erstellungsprozess. „The creation of a document can be analyzed as an ordinary act of communication with one (or more) senders and one (or more) receivers.“⁴³ Kommunikation wird dabei im organisationalen Zusammenhang (das Dokument als Geschäftsprozess) und im Zusammenhang mit den Medien betrachtet (das Dokument als Publikationsprozess). Ersterer umfasst dabei Vorgänge wie z.B. die Nutzung von Dokumenten als Organisations- oder Strategieelement, letzterer beinhaltet Prozesse wie Peer Reviews oder Zitationen. Den Punkt der Dokumentökonomie fasst Pédaque folgendermaßen zusammen: „The more the existence of a document is known, the more it is read, and the more it is read, the more its existence becomes known“.⁴⁴

Pédaque zeigt, dass das Dokument für viele Forschungsdisziplinen eine wichtige Rolle spielt, das Dokumentenkonzept aber stark vom Kontext abhängt.⁴⁵ Allerdings sind die Gleichungen nicht eben einfach nachzuvollziehen und die Theorie allgemein erscheint recht abstrakt.

Besser verständlich präsentiert Jean-Michel Salaün, der selbst Teil von Pédaque ist, seine Vorstellung von den verschiedenen Dimensionen des Dokuments. Er ist der Meinung, dass das Dokument nach den meisten heutigen Standards nur zwei der drei Dimensionen besitzt, die Pédaque anspricht, und zwar die Dimension der Form („le vu, das Gesehene“⁴⁶) und die Dimension des Inhalts („le lu, das Gelesene“⁴⁷). Die erste Dimension ist anthropologisch und bezeichnet unsere Beziehung zum Dokument als Objekt. Ein Dokument muss für uns sofort als solches verstanden werden. Damit ist gemeint, dass man beispielsweise weiß, wie ein Pass oder eine Rechnung aussieht und

⁴³ Pédaque, R. T. (2003): Document: Form, Sign and Medium, As Reformulated for Electronic Documents, S. 19. [elektronische Quelle].

⁴⁴ Ebd., S. 20 [elektronische Quelle].

⁴⁵ Vgl. Francke, H. (2005): What's in a Name? Contextualizing the Document Concept, S. 61 [elektronische Quelle].

⁴⁶ Salaün, J. (2018): Die drei Dimensionen des Dokuments und ihre Auswirkungen auf die Bibliotheks- und Informationswissenschaft, S. 352 [elektronische Quelle].

⁴⁷ Ebd. [elektronische Quelle].

dass es sich dabei um ein Dokument handelt, auch wenn man den Inhalt noch nicht kennt.⁴⁸

Die zweite Dimension, jene des Inhalts, ist eine intellektuelle oder auch neuroanthropologische Dimension, die nur durch gewisse Eigenschaften erfasst werden kann. Um den Inhalt eines Buches zu verstehen, muss man in der Lage sein, es zu lesen. Ein Buch, das in einer dem Lesenden unbekanntem Sprache verfasst ist, kann dieser zwar in der ersten Dimension als Dokument erkennen, jedoch nicht in der zweiten Dimension. Der Inhalt bleibt verborgen, da die Sprache nicht verstanden wird.⁴⁹

Was in vielen Theorien über das Dokument nach der Meinung Salaüns jedoch nicht zur Sprache kommt, ist die dritte Dimension, die soziale oder auch sozio-anthropologische Funktion des Dokuments, die Vermittlung („le su, das Gewusste“⁵⁰). Gemeint ist hierbei die Beziehung unseres Menschseins zur Funktion des Dokuments in seiner Vermittlungsfähigkeit, unabhängig von seiner Form oder seinem Inhalt. Das Dokument muss etwas übermitteln und setzt sich hier über die Grenzen von Raum und Zeit hinweg. Die Schriften altgriechischer Gelehrten wären längst in Vergessenheit geraten, besäße das Dokument diese Eigenschaft nicht.⁵¹

Ein Dokument muss laut Salaün alle drei Dimensionen durch eine sogenannte Lektürevereinbarung in sich zusammenführen, welche die Produzierenden oder Dokumentverantwortlichen mit den Lesenden verbindet. So gelangt er zu seiner Definition: „A document is a trace used to interpret a past event through a reading contract“⁵².

⁴⁸ Vgl. Salaün, J. (2014): *Why the Document Is Important... and How it Is Becoming Transformed*, S. 189 ff [elektronische Quelle].

⁴⁹ Vgl. Salaün, J. (2018): *Die drei Dimensionen des Dokuments und ihre Auswirkungen auf die Bibliotheks- und Informationswissenschaft*, S. 352 [elektronische Quelle]; vgl. Salaün, J. (2014): *Why the Document Is Important... and How it Is Becoming Transformed*, S. 189 ff. [elektronische Quelle].

⁵⁰ Salaün, J. (2018): *Die drei Dimensionen des Dokuments und ihre Auswirkungen auf die Bibliotheks- und Informationswissenschaft*, S. 352 [elektronische Quelle].

⁵¹ Vgl. Salaün, J. (2014): *Why the Document Is Important... and How it Is Becoming Transformed*, S. 192 [elektronische Quelle]; vgl. Salaün, J. (2018): *Die drei Dimensionen des Dokuments und ihre Auswirkungen auf die Bibliotheks- und Informationswissenschaft*, S. 352 [elektronische Quelle].

⁵² Salaün, J. (2014): *Why the Document Is Important... and How it Is Becoming Transformed*, S. 194 [elektronische Quelle].

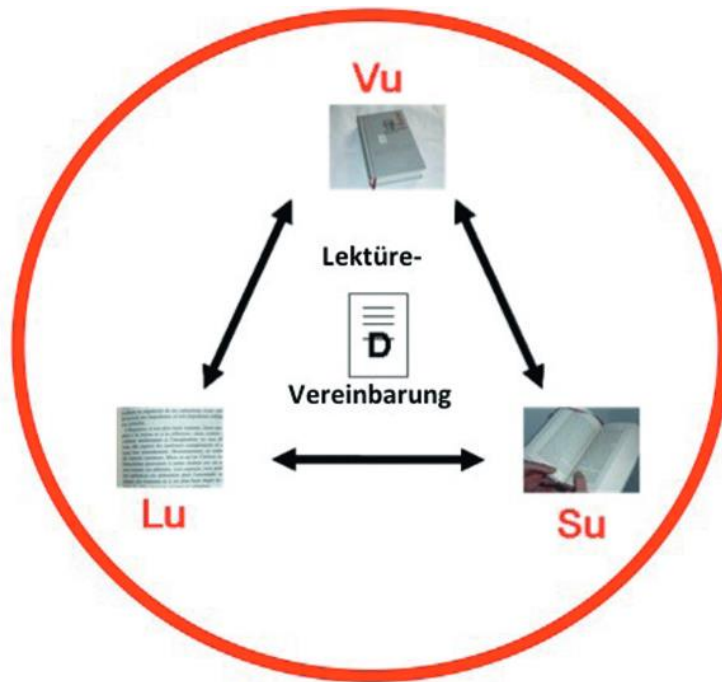


Abb. 1: Lektüre-Vereinbarung nach Salaün⁵³

In dieser Definition vereint er die drei Dimensionen: das Materielle (Vu), das Intellektuelle (Lu) und die Vermittlungsfunktion (Su). Anhand eines Beispiels verdeutlicht er diese nochmals: Ein Reisepass wird von uns als solcher aufgrund seiner äußeren Form erkannt (Vu), schlagen wir ihn auf und lesen wir ihn, so erkennen wir seinen Inhalt (Lu) und wir wissen, dass der Pass benötigt wird, um sich auszuweisen und zu reisen (Su). Nur durch die von uns stark verinnerlichte, reflexartige Lektüre-Vereinbarung wird dies möglich. Mit dem „past event“ ist nicht gemeint, dass das Dokument von einem vergangenen Ereignis spricht, sondern dass das Dokument selbst, seine Erstellung, eine Spur ist, die Vergangenheit und Gegenwart miteinander verknüpft.⁵⁴

Salaün stimmt mit der Theorie Briets in dem Punkt überein, dass auch für ihn z.B. Tiere Dokumente sein können, sofern sie in ein Informationssystem mit Metadaten eingebettet sind und von diesem System korrekt interpretiert werden können. Er weist auch auf die Definition Ranganathans hin, die er zwar anerkennt, jedoch für unvollständig hält, da sie nicht darauf eingeht, dass, wie bei Briet, Objekte jeglicher Art Dokumente sein können und auch die soziale Komponente, die Vermittlung, nicht beachtet wird.⁵⁵ Da er

⁵³ Abb. entnommen aus: Salaün, J. (2018): Die drei Dimensionen des Dokuments und ihre Auswirkungen auf die Bibliotheks- und Informationswissenschaft, S. 352 [elektronische Quelle].

⁵⁴ Vgl. Salaün, J. (2018): Die drei Dimensionen des Dokuments und ihre Auswirkungen auf die Bibliotheks- und Informationswissenschaft, S. 352 f. [elektronische Quelle]; vgl. Salaün, J. (2014): Why the Document Is Important... and How it Is Becoming Tranformed, S. 194 [elektronische Quelle].

⁵⁵ Vgl. Salaün, J. (2014): Why the Document Is Important... and How it Is Becoming Tranformed, S. 194 f. [elektronische Quelle].

Ranganathans Definition jedoch trotzdem sehr hilfreich findet, verbindet er diese mit seiner eigenen, um zu einer endgültigen Definition zu gelangen: „A document is a trace for the interpretation of a past event in accordance with a reading contract. In most cases, this trace has been recorded on an appropriate medium for easy physical handling, transport across space and preservation over time.“⁵⁶ Genau wie bei Briet kann bei Salaün jedes Objekt ein Dokument sein, sofern es die von ihm beschriebenen Bedingungen erfüllt, die meisten Dokumente sind jedoch auf Medien festgehalten, die sich leicht transportieren und erhalten lassen, wie eben Papier o.Ä.

Man kann annehmen, dass die meisten Menschen mit dem Begriff des Dokuments wohl zuallererst etwas Textuelles auf Papier assoziieren. Wie gezeigt wurde, können den meisten Meinungen von Forschenden zufolge aber auch viele andere Objekte Dokumente sein. Es gibt keine klare Definition und keine Einigkeit darüber, aber es ist wichtig zu verstehen, dass es nicht um das reine Äußere geht, das, was man als erstes wahrnimmt, ohne sich weiter mit dem Objekt auseinanderzusetzen. Es geht auch um den Inhalt und die Funktion, die das Dokument erfüllt. Wie entscheidend diese beiden Dimensionen sind, zeigt sich vor allem dann, wenn man von dem Typ Dokument spricht, den es erst seit einigen Jahrzehnten gibt: dem digitalen Dokument.

1.3 Theorien zum digitalen Dokument

Wie sich gezeigt hat, ist es schwierig bis unmöglich, den Begriff „Dokument“ einheitlich zu definieren. Noch komplizierter wird die Sache, wenn man von digitalen Dokumenten spricht. Viele Eigenschaften des analogen Dokuments haben sich im digitalen Dokument gewandelt. Beispielsweise sind sie viel leichter zu verändern bzw. manipulieren als analoge Dokumente. Innerhalb eines digitalen Dokuments können verschiedene Medientypen vernetzt sein, beispielsweise Videos oder Audios innerhalb von ansonsten textuellen Dokumenten. Speicherorte, Formate und Ansichten können schnell ausgetauscht werden. Digitale Dokumente sind weitaus besser durchsuchbar als analoge Dokumente. Sie können sofort um den gesamten Globus geschickt und ohne Abweichungen vom Original reproduziert werden. Die Quantität spielt dabei keine Rolle, da die Kopien keinerlei Qualitätsverlust unterliegen. Viele Eigenschaften des Dokuments haben sich jedoch auch verkompliziert, so zum Beispiel ihre Dynamik. Ein digitales Dokument kann sich sekundlich verändern, es kann sich sogar für verschiedene Nutzende verschieden darstellen, je nachdem, was der Nutzende benötigt. Das Beschreiben, Indexieren,

⁵⁶ Ebd., S. 195 [elektronische Quelle].

Speichern und Abrufen von digitalen Dokumenten wird dadurch erheblich erschwert. Auch Machtverhältnisse verschieben sich, da es heute ein Leichtes ist, Dokumente im Web zu veröffentlichen und man keinen kritischen Publikationsprozess mehr durchlaufen muss. Zwar gibt es v.a. im wissenschaftlichen Bereich sogenannte Peer-Reviews, bei denen die Inhalte einer Publikation von Fachleuten vor Veröffentlichung auf ihre Qualität hin überprüft werden, das Meiste, was jedoch im Internet veröffentlicht wird, durchläuft keine derartigen Prozesse. So kommt es auch zu der häufigen Verbreitung von Fake News. Die Hilfe von Experten und Expertinnen wird durch Suchmaschinen weniger benötigt, um geeignete Dokumente aufzuspüren. Geistiges Eigentum ist nicht mehr so gut geschützt, da das Kopieren derart einfach und schnell vonstattengeht. Häufig wird das Urheberrecht im digitalen Umfeld mit Füßen getreten.⁵⁷ Die technischen Hintergründe vieler dieser sich verändernden Eigenschaften werden in Kapitel 2 noch näher besprochen werden.

Zurückkommend auf die Theorien zum „traditionellen“, analogen Dokument wird sehr deutlich, dass sich eine gewisse Form, oder ein gewisses Medium, nicht zum alleinigen Unterscheidungsmerkmal von Dokumenten und Nicht-Dokumenten eignet, denn in der digitalen Welt ist diese Form aufgelöst.⁵⁸ Digitale Dokumente lassen sich nicht in die Hand nehmen, nicht anfassen und sie alle bestehen aus demselben „Material“, nämlich Bits. Ein wissenschaftlicher Artikel besteht aus Bits ebenso wie ein Blog oder eine E-Mail und auch nicht-textuelle Dateien wie Videos, Fotos und Audios bestehen aus nichts anderem als einer enormen Anzahl von Bits. Bits sind nur für Maschinen lesbar, daher unterscheiden sich digitale Dokumente von analogen in dem Punkt, dass sie nur durch Maschinen für den Menschen verständlich werden. Ohne das entsprechende Gerät gibt es keine Möglichkeit für den Menschen, das Dokument zu lesen.⁵⁹

Es gibt auch andere Interpretationsweisen der Form eines digitalen Dokuments. Auch in diesem Punkt ist das Digitale schwerer zu definieren als das Analoge. Die Forschergruppe Pédaque hat sich neben dem analogen Dokument auch stark mit dem digitalen Dokument beschäftigt. Rückblickend auf die bereits im vorigen Kapitel erwähnte erste Gleichung *traditional document = medium + inscription*, die das Dokument als Form beschreibt, überprüft Pédaque diese für das digitale Dokument. Bei einem Dokument auf Papier kann ebendieses klar als das *medium* benannt werden. Beim digitalen Dokument

⁵⁷ Vgl. Schamber, L. (1996): What Is a Document? Rethinking the Concept in Uneasy Times, S. 669 f. [elektronische Quelle].

⁵⁸ Vgl. Buckland; M. (1998): What is a „digital document“?, S. 215 [elektronische Quelle].

⁵⁹ Vgl. Pédaque, R. T. (2003): Document: Form, Sign and Medium, As Reformulated for Electronic Documents, S. 6 [elektronische Quelle].

ist das *medium* viel mehrdeutiger. Es kann von der aus Bits bestehenden Datei dargestellt werden, ebenso aber von der Hardware, auf der sie gespeichert ist, beispielsweise einer externen Festplatte. Drittens könnte auch das, was auf dem Bildschirm im Endeffekt angezeigt wird und man sehen kann, als das *medium* des Dokuments betrachtet werden. Und auch dieser Punkt ist kritisch, denn ein und dasselbe Dokument kann beispielsweise auf einem Smartphone anders angezeigt werden als auf einem PC. Auch verschiedene Browser können die Anzeige eines Dokuments verändern.

Laut Pédauque hat das *medium* seine Klarheit im digitalen Bereich aufgrund all dieser Argumente verloren. Noch stärker betroffen von dem Ausmaß der Veränderungen ist jedoch der zweite Teil der Gleichung, die *inscription*. Die erste Gleichung *traditional document = medium + inscription* wird daher umgewandelt in *electronic document = structures + data*⁶⁰. Die Struktur variiert je nach Dokumenttyp stark. Ein digitales Dokument kann unstrukturiert, semi-strukturiert oder strukturiert sein. Es gibt außerdem zwei verschiedene Ebenen der Dokumentstruktur: die logische, die das Dokument in verschiedene Teile gliedert (z.B. Kapitel und Unterkapitel) und die formale Struktur, das Layout (z.B. verschiedene Schriftarten).⁶¹ Das *medium* spielt keine zentrale Rolle mehr, sondern die Struktur, wodurch der „Lesevertrag“, der zwischen Produzierendem und Lesendem besteht, zu einer Art „Lesbarkeitsvertrag“ wird. Auf Basis dieser Überlegungen gelangt Pédauque zu der ersten Definition: „An electronic document is a data set organized in a stable structure associated with formatting rules to allow it to be read both by its designers and its readers.“⁶² Diese Definition bezieht sich jedoch lediglich auf den Aspekt des Dokuments als Form. Die Beziehung zwischen Dokumenten und Daten wird im anschließenden Kapitel noch näher erörtert.

Im Zusammenhang mit Pédauques im vorangegangenen Kapitel besprochener zweiten Gleichung *Document = inscription + meaning*⁶³ spricht er im digitalen Kontext „Dokumentationssprachen“, wie z.B. Thesauri oder Indexe an. Digitale Dokumente werden anhand dieser Dokumentationssprachen klassifiziert und mit Metadaten versehen, was ihre zukünftige Nutzung beeinflusst, da durch die Klassifizierung ein gewisser Kontext entsteht. Personen, die im Feld von Informations-Retrieval tätig sind, greifen in die Interpretation eines Dokuments ein, indem sie gesonderte Teile daraus extrahieren und dem Lesenden zur Verfügung stellen. Für den Lesenden ist dies zwar eine Erleichterung, da er oder sie

⁶⁰ Pédauque, R. T. (2003): Document: Form, Sign and Medium, As Reformulated for Electronic Documents, S. 9 [elektronische Quelle].

⁶¹ Vgl. ebd., S. 8 [elektronische Quelle].

⁶² Ebd., S. 9 [elektronische Quelle].

⁶³ Ebd., S. 12 [elektronische Quelle].

nun nicht mehr das gesamte Dokument lesen muss, aber die Interpretation des Dokuments kann dadurch verändert werden. Die initiale Gleichung wird verändert zu *Electronic document = informed text + knowledge*⁶⁴. Grund dafür ist, dass das Dokument individuell strukturiert werden kann und dass Wissen speziell für einen Nutzenden in Form von Metadaten hinzugefügt werden kann. Die zweite Definition Pédaupes, die sich auf das Dokument als Zeichen bezieht, lautet daher: „An electronic document is a text whose elements can potentially be analyzed by a knowledge system in view of its exploitation by a competent reader.“⁶⁵

Schließlich behandelt Pédaupue den dritten und letzten Aspekt, das Dokument als Medium, wobei Medium in sehr weitem Sinne verstanden wird als ein Gegenstand der Kommunikation zwischen Menschen. Die Anzahl an analogen Dokumenten ist in den vergangenen Jahren relativ stabil geblieben, die Vorstellung vom „paperless office“ hat sich noch längst nicht erfüllt.⁶⁶ Die Anzahl an digitalen Dokumenten jedoch nimmt rapide zu. Die initiale Gleichung *Document = inscription + legitimacy* wird verändert zu *Electronic document = text + procedure*, allerdings gilt diese Gleichung nur für die Kommunikation in Organisationen. Im Web jedoch transformiert sich die Gleichung zu *Web document = publication + cued access*. Damit ist gemeint, dass allein die Publikation eines Textes im Web diesen noch nicht legitimiert. Dies geschieht erst dadurch, dass das Dokument einen höherrangigen Status durch zahlreiche Zugriffe darauf erhält. Die dritte und abschließende Definition lautet: „An electronic document is a trace of social relations reconstructed by computer systems“.⁶⁷ Allerdings merkt Pédaupue an, dass die Definition recht allgemein ist, da es schwer ist, die teilweise chaotischen Entwicklungen in diesem Bereich umfassend zu beschreiben.

Niels W. Lund merkt zu Pédaupes Theorien treffend an, dass die Arbeit als eine Art systematisches, interdisziplinäres Brainstorming angesehen werden kann, es jedoch schwer ist, daraus eine allgemeine Dokumenttheorie herzuleiten.⁶⁸ Die Betrachtungen von Pédaupue sind nicht leicht nachzuvollziehen, zeigen jedoch gerade dadurch, dass es ein hochkompliziertes Unterfangen darstellt, das digitale Dokument zu definieren. Fakt ist, dass das digitale Dokument ebenso wie das analoge Dokument niemals nur auf einen Faktor wie z.B. seine äußere Form reduziert werden darf, da noch viel mehr

⁶⁴ Pédaupue, R. T. (2003): Document: Form, Sign and Medium, As Reformulated for Electronic Documents, S. 14 [elektronische Quelle].

⁶⁵ Ebd., S. 15 [elektronische Quelle].

⁶⁶ Vgl. Sellen, A.J.; Harper, R.H.R. (2003): The myth of the paperless office, S. 1 f.

⁶⁷ Pédaupue, R. T. (2003): Document: Form, Sign and Medium, As Reformulated for Electronic Documents, S. 23 [elektronische Quelle].

⁶⁸ Vgl. Lund, N. W. (2011): Document Theory, S. 38 [elektronische Quelle].

Komponenten hinter diesem Konzept stecken. Die im vorigen Kapitel besprochene Theorie von Jean-Michel Salaün zeigt dies besonders gut. Dokumenttheorien, die sich lediglich auf einen Aspekt des Dokuments konzentrieren, schließen zwangsläufig einen Teil von Dokumenten aus ihrer Definition aus, da ihre Vielfalt durch nur ein Kriterium, wie z.B. den Dokumentinhalt, nicht abgebildet werden kann. Das Vu-Lu-Su-Konzept von Salaün stellt daher einen gelungenen Versuch dar, sowohl analoge als auch digitale Dokumente anhand derselben Kriterien zu beurteilen, auch wenn selbst hier nicht zweifelsfrei bestätigt werden kann, dass die Definition tatsächlich alle Formen von Dokumenten erfasst.

Die heutige Gesellschaft erlebt eine Art „Redokumentalisierung“, vergleichbar mit dem Dokumenttrend zu Otlets Zeiten, jedoch in viel gewaltigerem Ausmaß. Vor etwa 40 Jahren sah es noch so aus, als sei das Dokumentkonzept veraltet, jetzt jedoch hat man den Eindruck, dass es relevanter ist als je zuvor. Otlets Traum von besserer Ordnung im Dokumentbereich durch das Aufkommen neuer Informationstechnologien hat sich leider nicht erfüllt, stattdessen erscheint das mit digitalen Dokumenten gefüllte Web oft als ein „Ort der Anarchie“.⁶⁹ Viele Hoffnungen, die mit der Digitalisierung des Dokuments verbunden waren, sind gescheitert. Die Umstände haben sich dadurch nicht in jeder, aber in vielerlei Hinsicht nicht wie erwartet verbessert, sondern verkompliziert. Neue Technologien sind jedoch auf dem Vormarsch und verändern Dokumente auf bemerkenswerte Art und Weise. In Kapitel 2 und 3 werden diese Veränderungen und die zugrundeliegenden Technologien aufgezeigt und erläutert. Zuvor wird jedoch noch kurz auf den Unterschied zwischen den Begriffen „Daten“ und „Dokument“ eingegangen, da dies für das Verständnis der anschließenden Kapitel relevant ist.

1.4 Unterscheidung der Begriffe „Daten“ und „Dokument“

Häufig wird das Dokument mit anderen Begriffen verglichen und man versucht, es von diesen abzugrenzen. Einer dieser Begriffe ist „Daten“. Grundsätzlich kann der Begriff „Daten“ sehr verschieden aufgefasst werden, je nachdem, in welchem Kontext er auftaucht. Tabellen mit Messwerten können genauso als Daten bezeichnet werden wie textuelle Notizen, die sich Forschende während eines Experiments machen, später ausarbeiten und als „Forschungsdaten“ publizieren.⁷⁰ Wie Buckland anmerkt, werden längere Textdatensätze meistens als Dokumente bezeichnet, kurze numerische Datensätze

⁶⁹ Vgl. Lund, N. W. (2011): Document Theory, S. 33 f. [elektronische Quelle].

⁷⁰ Vgl. Voß, J. (2013): Was sind eigentlich Daten?, S. 6 [elektronische Quelle].

dagegen als Daten.⁷¹ Diese Meinung wird auch von Götzer et al. geteilt, die versuchen, Daten und Dokumente anhand der zugrundeliegenden Struktur zu beschreiben:

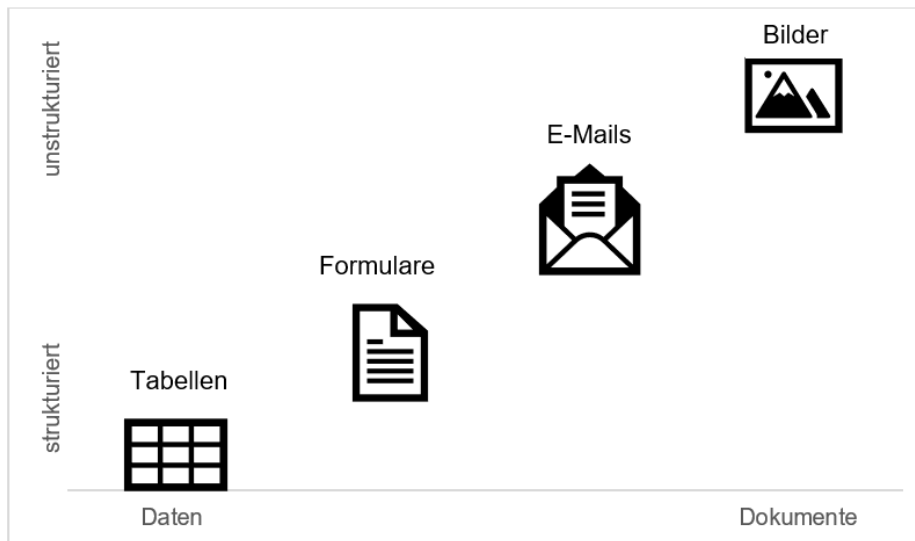


Abb. 2: Vergleich der Struktur von Daten und Dokumenten⁷²

Sie merken an, dass die meisten Dokumente eher schwach- bis unstrukturiert sind, wohingegen reine Daten sehr strukturiert sind.⁷³ Auf die Datenstruktur wird in Kapitel 2.1 noch näher eingegangen.

Von einer klaren Definition sind diese Aussagen weit entfernt, ob es so eine Definition jedoch überhaupt geben kann, ist fragwürdig. Der Bibliothekar und Informatiker Jakob Voß merkt an, dass die Grenzen des Dokuments im Digitalen verwischen, habe ich z.B. ein 20-seitiges Textdokument und teile dieses in zwei Dateien mit jeweils zehn Seiten auf, so habe ich nun zwei Dokumente. Auch diese könnte ich weiter aufteilen in noch kürzere Dokumente. Jedes aus Daten bestehende Objekt, beispielsweise eine PDF-Datei oder eine Webseite, bildet laut Voß ein Dokument, „sobald es als publizier- oder kommunizierbares Objekt aufgefasst wird“.⁷⁴ Was jedoch ein publizier- oder kommunizierbares Objekt ist, ist wiederum eine Frage, die sich subjektiv beantworten lässt. Eine grundsätzliche Unterscheidung zwischen Daten und digitalen Dokumenten sei daher nicht möglich.

Laut dem australischen Philosophen Brian Ballsun-Stanton können Daten unter anderem als Folge von Bits verstanden werden, als computerbasierte Zeichen, die zur Kommunikation dienen. Die Betonung liegt auf dem semiotischen Charakter. Dabei macht es

⁷¹ Vgl. Buckland, M. (2018): Document Theory, S. 428 [elektronische Quelle].

⁷² Abb. verändert entnommen aus: Götzer, K. et al. (2013): Dokumenten-Management, S. 85.

⁷³ Vgl. ebd., S. 85.

⁷⁴ Voß, J. (2013): Was sind eigentlich Daten?, S. 7 [elektronische Quelle].

keinen Unterschied, in welcher Form (textuell, numerisch) die Daten vorliegen und welche Inhalte sie haben.⁷⁵ Voß merkt dazu an, dass Daten, wenn sie als binäre Nachrichten aufgefasst werden, mit dem digitalen Dokumentenbegriff zusammenfallen und somit alle Daten auch Dokumente seien.⁷⁶ Eine klare Abgrenzung der beiden Begriffe ist nicht möglich, eher gehen sie ineinander über. Was als Dokument und was als Daten bezeichnet wird, hängt häufig vom Kontext ab und wird je nachdem unterschiedlich beurteilt. Vereinfachend könnte man sagen, dass Dokumente aus Daten bestehen. Hat man z.B. eine Datei, in welche u.a. numerische Tabellen, aber auch lange Textpassagen eingebunden sind, spricht man von einem Dokument. Würde man die numerischen Tabellen jedoch außerhalb des Gesamtdokuments betrachten, ohne jeglichen textuellen Inhalt, der die Zahlen erläutert, so würde man das Ganze eher als Datensatz denn als Dokument bezeichnen. Es gibt jedoch keine Vorschrift und keine klare Definition dessen, was als Dokument und was als Daten zu bezeichnen ist. In der vorliegenden Arbeit bezeichnen Daten kleinere Teile, aus denen ein digitales Dokument besteht. Es wird kein klarer Schnitt zwischen Daten und Dokumenten gemacht, vielmehr werden die Begriffe je nach Kontext passend eingesetzt.

⁷⁵ Vgl. Ballsun-Stanton, B. (2010): Asking about Data: Experimental Philosophy of Information Technology, S. 122 [elektronische Quelle].

⁷⁶ Vgl. Voß, J. (2013): Was sind eigentlich Daten?, S. 8 [elektronische Quelle].

2. Technologische Aspekte des digitalen Dokuments

Im vorangegangenen Kapitel wurden Theorien zu Eigenschaften der analogen und digitalen Dokumente ausführlich behandelt. Im folgenden Abschnitt soll es um die zugrundeliegenden technologischen Aspekte gehen, welche das digitale Dokument verändert haben und auch in Zukunft noch weiter verändern werden und zur digitalen Transformation beitragen. Die „digitale Transformation“ wird häufig mit dem Begriff „Digitalisierung“ synonym verwendet. Tatsächlich handelt es sich bei der „Digitalisierung“ lediglich um die Umwandlung von analogen Formen in digitale Formate, z.B. um das Scannen eines papiernen Dokuments. Sie ist eine Voraussetzung für die digitale Transformation, welche den Umbruch in der heutigen Gesellschaft in Bezug auf neue Geschäftsmodelle, Organisations- und Arbeitsformen, Produkte, Technologien und Verfahren beschreibt.⁷⁷

Man begegnet digitalen Dokumenten nicht nur tagtäglich im Berufs- und Privatleben in Form von Word-, Excel-, PDF-Dateien etc., sondern auch auf allen Webseiten, die man aufruft, denn diese sind nichts anderes als Dokumente. Sie stellen den Kern des Internets dar, auch wenn sie hier oft auf ihre kleinere Einheit, die Daten, reduziert werden. Mit neuen Erfindungen im IT-Bereich gehen stets auch Änderungen im Dokumentbereich einher, denn da Dokumente ein wesentlicher Bestandteil des Informationszeitalters sind, hat jede bahnbrechende Innovation, die sich auf die IT und das Internet auswirkt, ebenso Auswirkungen auf die Dokumente.⁷⁸ Es werden daher einige grundlegende Technologien vorgestellt, die zu großen Änderungen und Herausforderungen im Dokumentbereich führten, die teilweise noch nicht bewältigt sind. Der Fokus liegt dabei auf textuellen Dokumenten. Es soll zwar nicht der Eindruck entstehen, Videos, Audios, Abbildungen und andere digitale Dateien seien keine Dokumente, jedoch sind diese in vielerlei Hinsicht in ihrer Komplexität noch höher angesiedelt. Zudem bestehen die meisten Dokumente, mit denen man tagtäglich zu tun hat, hauptsächlich aus Text.

2.1 Datenstruktur

Wie bereits in Kapitel 1.3 und 1.4 kurz erwähnt, können in Bezug auf die Struktur grundsätzlich drei Arten von Daten unterschieden werden: strukturierte, semistrukturierte und unstrukturierte Daten. Strukturierte Daten enthalten Metadaten, also „Daten über Daten“.

⁷⁷ Vgl. Franzetti, C. (2019): Essenz der Informatik, S. 223.

⁷⁸ Vgl. Ragnet, F. (2009): Document 3.0 How the next generation of dynamic documents will improve the way we work, S. 5 [elektronische Quelle].

Sie sind nach einem festgelegten Datenschema strukturiert und können dadurch leicht gespeichert, verarbeitet und ausgewertet werden. Strukturierte Daten sind beispielsweise in hochformatierten Datenbanken abgelegt, in Tabellenkalkulationen oder Buchhaltungssystemen.

Semistrukturierte Daten können in einzelnen Bestandteilen strukturiert sein, weisen jedoch in der Gesamtheit keine eindeutige Struktur auf, da sie häufig zu komplex sind, um sie mit den flachen Tupeln (Attributen) eines relationalen Datenmodells wiederzugeben. Sie sind flexibler als strukturierte Daten, gleichzeitig jedoch schwerer zu verarbeiten. Ein Beispiel dafür sind Extensible Markup Language (XML)-Daten, auf die später noch genauer eingegangen wird.

Unstrukturierte Daten folgen keiner formalen Struktur. Sie können zwar ebenfalls strukturierte Informationen enthalten, jedoch sind diese in der Regel nicht eindeutig erkennbar. Beispiele dafür sind einfache PDF-Textdokumente sowie Bild- und Videodokumente.⁷⁹

Der größte Anteil aller Daten ist Schätzungen zufolge unstrukturiert, die Rede ist von rund 85%.⁸⁰ Dies hat zur Folge, dass eine exorbitante Menge an Daten zwar nützliche Informationen enthält, diese jedoch aufgrund der fehlenden Struktur nicht weiterverarbeitet werden können. Vergleichbar ist dieses heute herrschende Szenario mit einem Schatz auf dem Meeresgrund, von dessen Existenz man weiß, der jedoch aufgrund seiner zu tiefen Lage nicht geborgen werden kann. Allerdings gibt es mittlerweile Techniken, die sich mit ebendiesem Problem befassen und durch die es ermöglicht wird, Informationen auch aus semi- und unstrukturierten Daten (und somit Dokumenten) zu gewinnen. Näher wird darauf in Kapitel 3 eingegangen.

2.2 Auszeichnungssprachen

Eine Auszeichnungssprache (markup language) ist eine formale, nicht-natürliche Sprache, durch die unterschiedliche Bestandteile eines Textes mittels Markups, auch Tags genannt, gekennzeichnet werden können. Beliebigen Textelementen können Eigenschaften zugewiesen werden, um deren Bedeutung auszudrücken.⁸¹

⁷⁹ Vgl. Piro, A.; Gebauer, M. (2011): Definition von Datenarten zur konsistenten Kommunikation im Unternehmen, S. 146; Bry, F. et al. (2001): Semistrukturierte Daten, S. 230 ff.

⁸⁰ Vgl. Bitcom (2012): Big Data im Praxiseinsatz–Szenarien, Beispiele, Effekte, S. 12 [elektronische Quelle].

⁸¹ Vgl. Neumann, G. (2014a): Auszeichnungssprache [elektronische Quelle].

Der Begriff der Auszeichnungssprache geht auf William Tunnicliffe zurück und wurde von ihm erstmals auf einer Konferenz im Jahre 1967 erwähnt. Tunnicliffe legte großen Wert darauf, dass in Texten nicht nur die Informationsinhalte dargestellt würden, sondern auch die typographische Formatierung. Eine Trennung von Inhalt und Format war ihm dabei sehr wichtig, da gleiche Inhalte auf verschiedene Art und Weise dargestellt werden können. Die erste Auszeichnungssprache wurde bereits zwei Jahre nach Tunnicliffes Vortrag entwickelt, es handelte sich um die Generalized Markup Language (GML). 1986 wurde dann eine Weiterentwicklung von GML, die Structured Generalized Markup Language (SGML), von der International Organisation for Standardization (ISO) als Standard verabschiedet (ISO 8879).⁸²

Einer der Nachfahren von SGML ist die Hypertext Markup Language (HTML). Sie wurde von Tim Berners-Lee, dem Begründer des World Wide Web (WWW), entwickelt. HTML bildet zusammen mit dem Uniform Resource Identifier (URI) und dem Hypertext Transfer Protocol (HTTP) die Basis des WWW.⁸³ Ein HTML-Quelltext besteht aus dem darzustellenden Text, Kommentaren, die nur im Quelltext sichtbar sind und Tags, welche die Struktur des Dokuments abbilden. Wird ein HTML-Dokument mittels eines Browsers ins Netz hochgeladen, so werden die HTML-Elemente durch den Browser erkannt und auf dem Bildschirm formatiert dargestellt, wobei verschiedene Browser das gleiche HTML-Dokument verschieden darstellen können, was auf unterschiedlichen Formatkonvertierungen beruht.⁸⁴

In der nachfolgenden Abbildung ist ein HTML-Quelltext und seine tatsächliche Repräsentation im Browser dargestellt. Auffällig sind hier die Hyperlinks, welche mittels des <a>-Tags dargestellt werden. Sie verweisen, wie in diesem Fall, auf ein anderes HTML-Dokument (also eine andere Webseite) oder auf eine bestimmte Stelle in demselben Dokument.

⁸² Vgl. Neumann, G. (2014a): Auszeichnungssprache [elektronische Quelle].

⁸³ Vgl. Fink, A. (2014): HTML [elektronische Quelle].

⁸⁴ Vgl. Elsner, F. (2004): Einführung in HTML, S. 8-12 [elektronische Quelle].

```

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html>
<head>
<title>HTML</title>
<link href="mystyle.css" rel="stylesheet" type="text/css">
</head>
<body>
<h1>Hinweise zu HTML</h1>
<p>Grundlegende Informationen und Spezifikationen sind beim
<a href="http://www.w3.org/html/">World Wide Web
Consortium (W3C) im Bereich zu HTML</a> zu finden.</p>
<p>In diesem zweiten Absatz verweisen wir auf eine
beliebte Dokumentation zu HTML:
<a href="http://de.selfhtml.org/">SELFHTML</a>.</p>
</body>
</html>

```



Abb. 3: HTML-Quelltext und dessen Darstellung im Browser⁸⁵

HTML zeichnet sich vor allem durch seine Einfachheit aus, was jedoch auch große Nachteile mit sich bringt. Die Sprache enthält ein festgelegtes Vokabular, spezifische Strukturelemente können also nicht kreiert werden. Dadurch geht die reiche Datenstruktur verloren, komplexe Strukturen, die zur Darstellung von Datenbankschemata oder objektorientierten Hierarchien erforderlich sind, können nicht abgebildet werden. Des Weiteren können Daten nicht semantisch ausgezeichnet werden und so findet bei der Umsetzung in Web-Dokumente stets ein Informationsverlust statt.⁸⁶ Abhilfe für dieses Problem schafft XML.

XML wurde 1996 vorgestellt und hat seit 1998 den Status einer Empfehlung des World Wide Web Consortium (W3C). Die Sprache wurde entwickelt, weil der Vorgänger SGML zu umfangreich und komplex war und HTML die bereits angesprochenen Nachteile mit sich brachte. Auch wenn der Name *Extensible Markup Language* darauf hinweist, dass

⁸⁵ Abb. verändert entnommen aus: Fink, A. (2014): HTML [elektronische Quelle].

⁸⁶ Vgl. Bosak, J. (1997): XML, Java, and the future of the Web [elektronische Quelle].

es sich bei XML um eine Auszeichnungssprache handelt, ist sie genau genommen eher eine Metasprache, mittels der andere Auszeichnungssprachen wie beispielsweise HTML definiert werden können. Bei XML werden Dokumente in einzelne Teile zerlegt (modularisiert), diese Teile werden durch Metainformationen beschrieben und können so einzeln angesteuert, kopiert, gefiltert, aktualisiert oder referenziert werden. Im Gegensatz zu HTML muss sich XML nicht an ein vorgegebenes Vokabular halten, um Dokumentinhalte auszuzeichnen. Es können individuelle Tags erstellt werden, um anwendungsspezifische Datenstrukturen zu beschreiben. XML bietet die Möglichkeit zum Austausch, zur Recherche und Verwaltung semantisch qualifizierter Daten.⁸⁷

Ein digitales Dokument besteht grundsätzlich aus einer Einheit aus Inhalt, Struktur und Layout. Der Inhalt besteht aus den zu übermittelnden Informationen, die Struktur spiegelt Aufbau und Abfolge derselben wider und das Layout visualisiert die strukturierten Informationen.⁸⁸ In XML sind Inhalt und Struktur strikt vom Layout getrennt. Die Struktur wird in einer Document Type Definition (DTD) festgelegt. Ein Dokument wird als valide bezeichnet, wenn es diese DTD einhält. Allerdings hat die DTD gerade in Bezug auf die Validierung einige Schwächen, daher wurde sie bereits 1999 durch XML-Schema abgelöst, einer Sprache, um Strukturdefinitionen zu verfassen.⁸⁹ Die strikte Trennung von Inhalt und Struktur vom Layout unterscheidet XML u.a. von einfachen Textverarbeitungsprogrammen wie Microsoft Word.⁹⁰ In diesem Programm kennzeichnet man beispielsweise eine Überschrift, indem man sie entsprechend anders als den Rest des Inhalts gestaltet, z.B. durch eine größere Schrift oder Fettdruck. Die Struktur wird also durch das Layout gekennzeichnet.

In XML aber wird eine Überschrift als solche ausgezeichnet, sie wird also mit dieser Strukturinformation versehen und alle Formatierungsinformationen, die das Layout der Überschrift betreffen, werden getrennt davon in einem sogenannten Stylesheet gespeichert. Dies hat die Vorteile, dass verschiedene Dokumente mit demselben Stylesheet oder auch ein Dokument mittels mehrerer kombinierter Stylesheets formatiert werden können und durch die Trennung auch die Arbeit unter mehreren Personen aufgeteilt werden kann. Das Dokument mit seinen beiden getrennten Elementen wird dann einem Programm (Prozessor) übergeben, der daraus ein Ergebnisdokument generiert.⁹¹

⁸⁷ Vgl. Neumann, G. (2014b): XML [elektronische Quelle]; Spitta, T.; Bick, M. (2008): Informationswirtschaft. Eine Einführung, S. 161 ff.

⁸⁸ Vgl. Bodendorf, F. (2006): Daten- und Wissensmanagement, S. 69.

⁸⁹ Vgl. Becher, M. (2009): XML, S. 35; S. 85.

⁹⁰ Vgl. ebd., S. 4 ff.

⁹¹ Vgl. ebd., S. 6 ff.

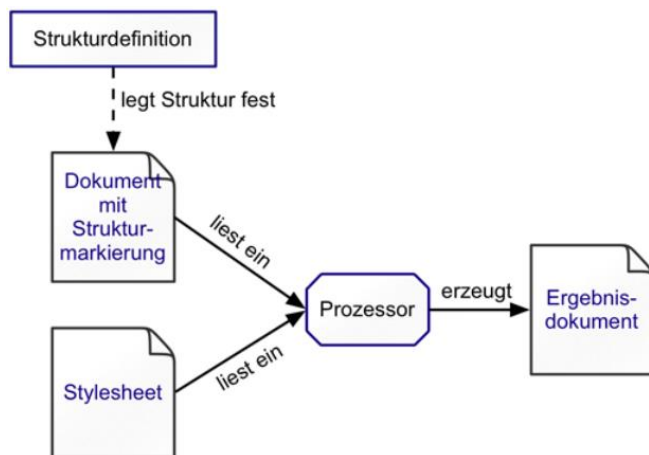


Abb. 4: Prozess der Generierung eines Ergebnisdokuments⁹²

Ein bekanntes Beispiel für ein Schema, das als XML-Anwendung codiert ist, wurde durch die Text Encoding Initiative (TEI) eingeführt. Das TEI-Format existierte bereits vor XML als SGML-Codierung, wechselte dann aber durch die Einführung von XML zu dieser Auszeichnungssprache.⁹³ TEI ist ein Standard für die Darstellung von digitalen Texten und hat sich vor allem in den Geistes-, Sozial- und Sprachwissenschaften etabliert.⁹⁴ Durch TEI können Texte mit wichtigen Strukturen wie Paragraphen, Überschriften, Versen etc. und allgemeinen Metadaten wie dem Erscheinungsjahr, dem Titel des Dokuments etc. ausgezeichnet werden. Auch inhaltlich wichtige Elemente wie z.B. Namen von Personen und Orten, die in einem Dokument auftauchen, können ausgezeichnet werden.⁹⁵

Was sind nun die konkreten Vorteile von XML in einem digitalen Dokument? XML-Dokumente sind leicht auszutauschen, da sie in jedem beliebigen Textprogramm geöffnet werden können, leicht zu durchsuchen und ermöglichen automatisierte Auswertung. Durch die Anreicherung mit Metadaten lässt sich beispielsweise ein umfangreiches Handbuch für ein beliebiges Produkt in Bausteine unterteilen (z.B. Herstellerangaben, Reinigung, Nutzung, etc.). Da Handbücher nicht für jedes Produkt individuell verfasst werden, sondern zum Teil gleiche Abschnitte enthalten, können diese Bausteine aufgrund ihrer Auszeichnung ganz einfach ausgetauscht werden. In einfachen Textverarbeitungsprogrammen müsste dies durch aufwendiges Kopieren und Einfügen bei jedem Dokument einzeln durchgeführt werden. Die Metadaten können auch erst nachträglich,

⁹² Abb. entnommen aus: Becher, M. (2009): XML, S. 8.

⁹³ Vgl. The TEI Consortium (2020): TEI P5. Guidelines for Electronic Text Encoding and Interchange, S XXIII [elektronische Quelle].

⁹⁴ Vgl. The TEI Consortium (o.J.): Home [elektronische Quelle].

⁹⁵ Vgl. The TEI Consortium (2020): TEI P5. Guidelines for Electronic Text Encoding and Interchange, S. 72 [elektronische Quelle].

also nach der Erstellung eines Dokuments hinzugefügt werden.⁹⁶ Auszeichnungssprachen wie XML machen es also möglich, dass unstrukturierte Dokumente durch die hinzugefügten Metadaten maschinenlesbar und somit verarbeitbar und auswertbar werden.

2.3 Semantic Web und Linked Data

Wie bereits im vorherigen Kapitel erwähnt wurde, besteht das Web aus einer Vielzahl an Dokumenten, die in diesem Zusammenhang auch Webseiten genannt werden. Diese Dokumente sind größtenteils nicht maschinenlesbar, eben weil sie unstrukturierte Daten enthalten. Zwar gibt es einige Elemente, die ein Computer nachvollziehen kann – beispielsweise kann er Hyperlinks zu anderen Dokumenten erkennen und diesen folgen – aber den Großteil der Webseiten „versteht“ die Maschine schlichtweg nicht, da sie nicht in der Lage ist, die Semantik zu begreifen. Aufgrund dieses Umstands entstand die Idee zum Semantic Web. Dieses ist kein separates Web, sondern eine Erweiterung des gegenwärtig vorhandenen. Während im herkömmlichen WWW die Betonung auf Dokumenten liegt, die miteinander vernetzt sind, liegt sie im Semantic Web auf den Daten, sozusagen also auf der kleineren Einheit, aus der Dokumente bestehen. Diesen Daten wird eine Bedeutung zugeschrieben, damit Computer in der Lage sind, sie zu „verstehen“ und zu verarbeiten, anstatt sie lediglich abzubilden.⁹⁷ Die Vernetzung dieser Daten wird als Linked Data bezeichnet. Das Semantic Web wird daher auch Web of Data genannt.⁹⁸

2.3.1 Resource Description Framework

Ein wichtiger Bestandteil zur Realisierung des Semantic Web ist XML, ein weiterer das Resource Description Framework (RDF). Wie bereits erwähnt, macht XML es möglich, Dokumenten eine spezifische Struktur zu geben. Über die Bedeutung der Strukturelemente sagt XML jedoch nichts aus. Dies wird durch RDF ermöglicht. Die Bedeutungen der Elemente werden hier in Form von Tripeln, die aus Subjekt, Prädikat und Objekt bestehen, ausgedrückt. Alle Bestandteile eines Tripels werden dabei jeweils durch eine URI identifiziert, wie sie auch für Webseiten verwendet wird.⁹⁹

⁹⁶ Vgl. Spitta, T.; Bick, M. (2008): Informationswirtschaft. Eine Einführung, S. 178.

⁹⁷ Vgl. Berners-Lee, T.; Hendler, T.; Lassila, O. (2001): The Semantic Web, S. 36 f. [elektronische Quelle].

⁹⁸ Vgl. Bizer, C.; Heath, T.; Berners-Lee, T. (2011): Linked Data – The Story so Far, S. 205 f. [elektronische Quelle].

⁹⁹ Vgl. Berners-Lee, T.; Hendler, T.; Lassila, O. (2001): The Semantic Web, S. 38 ff. [elektronische Quelle].

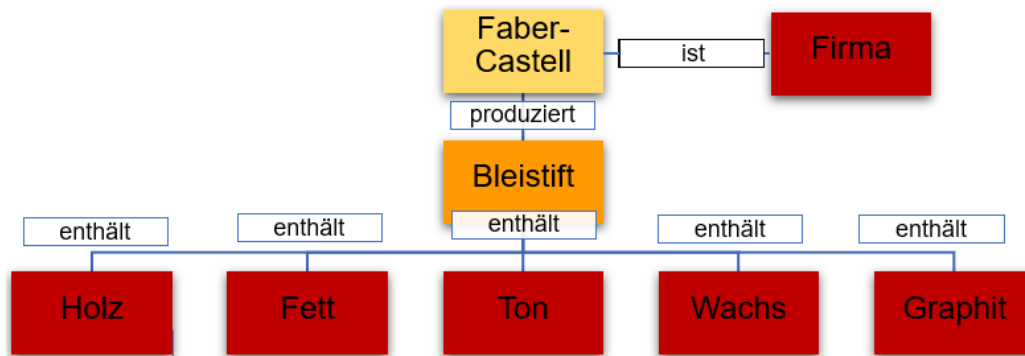


Abb. 5: Einfache Darstellung von RDF-Tripeln

Die Abbildung zeigt das Prinzip von RDF-Tripeln, wobei die gelbe Farbe ein Subjekt, die rote Farbe ein Objekt und ein weißes Kästchen ein Prädikat darstellt. Die Farbe Orange ist eine Mischung aus Gelb und Rot, da der Bleistift in diesem Beispiel sowohl Subjekt als auch Objekt ist. Durch Abfragesprachen wie beispielsweise die SPARQL Protocol And RDF Query Language (SPARQL) können Daten in RDF abgefragt werden.

2.3.2 Ontologien

Trotzdem gibt es auch beim RDF-Konzept noch ein Problem, nämlich jenes, dass natürliche Sprachen reich an Synonymen sind und nicht jeder für eine Sache denselben Begriff verwendet. Beispielsweise kann man „Kochsalz“ sagen, aber auch „Natriumchlorid“, wenn man es gerne wissenschaftlicher ausdrückt. Ein Computer „versteh“ jedoch nicht automatisch, dass Kochsalz und Natriumchlorid dasselbe bedeuten, wenn man diese Verbindung zwischen den beiden Wörtern nicht explizit auszeichnet. Eine solche Sammlung von Informationen, in der die Beziehungen zwischen verschiedenen Begriffen definiert werden, wird im Bereich der Informatik Ontologie genannt. Ontologien sind ein weiterer wichtiger Bestandteil des Semantic Web.

Eine einfache Ontologie enthält Inferenzregeln und eine Taxonomie. Die Taxonomie definiert Klassen von Entitäten und ihre Beziehungen untereinander. Die Inferenzregeln erlauben es dem System, logische Rückschlüsse zu ziehen.¹⁰⁰ Ein einfaches Beispiel hierfür ist die Transitivitätsregel:¹⁰¹ Der Neckar fließt durch Stuttgart und Stuttgart liegt in Baden-Württemberg. Daraus ergibt sich die logische Schlussfolgerung, dass der Neckar durch Baden-Württemberg fließt. Es sind aber auch noch viel komplexere Schlussfolgerungen möglich, z.B. durch domänenspezifische Regeln. Das System kann dann z.B. aus den Angaben, dass ein Hotel einen Spielplatz und Kinderbetten besitzt sowie

¹⁰⁰ Vgl. Berners-Lee, T.; Hendler, T.; Lassila, O. (2001): The Semantic Web, S. 40 [elektronische Quelle].

¹⁰¹ Vgl. Busse, J. et al. (2014): Was bedeutet eigentlich Ontologie?, S. 289 [elektronische Quelle].

Kindermenüs anbietet, schlussfolgern, dass dieses Hotel als familienfreundlich gilt. Eine Sprache, die zum Erstellen von Ontologien sehr verbreitet ist, ist die Web Ontology Language (OWL).

2.3.3 Wissensgraphen

Wissensgraphen sind sehr große semantische Netze, die heterogene Informationsquellen integrieren, um Wissen über bestimmte Diskursbereiche darzustellen.¹⁰² Dabei bedienen sie sich der Werkzeuge, die in den vorangegangenen Kapiteln besprochen wurden. Mindestens ein Wissensgraph ist wohl jedem, wenn teilweise auch unbewusst, bekannt: der sogenannte Knowledge Graph von Google. Dieser wurde 2012 eingeführt. Es gibt keine öffentliche Dokumentation bezüglich der zugrundeliegenden Technologie und des Formalismus dieses Wissensgraphen, bekannt ist aber, dass Google Daten aus verschiedenen Quellen wie z.B. Wikipedia, Weltbank oder Eurostat in seinen Wissensgraphen einbindet.¹⁰³ Enthält eine Google-Anfrage bestimmte Stichwörter (z.B. Namen von Personen, Organisationen, Produkten oder Orten), die im System als semantische Entitäten ausgezeichnet sind, werden diese erkannt und angereichert mit weiteren Informationen rechts neben den Treffern in einer Infobox präsentiert.

¹⁰² Vgl. Fensel, D. et al. (2020): Knowledge Graphs, S. 6.

¹⁰³ Vgl. ebd., S. 9 f.



Abb. 6: Beispiel für die Infobox zu Elon Musk¹⁰⁴

Führt man beispielsweise eine Google-Anfrage zu dem Unternehmer Elon Musk durch, erhält man neben den üblichen Treffern die in Abb. 6 gezeigte Infobox, die auf einen Blick die wichtigsten Fakten zu Musk zusammenfasst. Außerdem werden Fotos von ihm sowie Personen, die mit ihm in irgendeiner Weise in Verbindung stehen, angezeigt. Neben dem wohl bekanntesten Knowledge Graph von Google gibt es noch viele weitere, z.B. DBpedia, YAGO, Wikidata, KBpedia oder auch den Entitätengraph von Facebook.¹⁰⁵

2.4 Enterprise Content Management

Dokumente veränderten sich im Lauf der Zeit über verschiedene Materialien bis hin zum Digitalen. Gleichsam veränderten sich auch die Methoden und Technologien, die dazu

¹⁰⁴ Google (2020a): Suchanfrage zu „Elon Musk“ [Screenshot].

¹⁰⁵ Vgl. Fensel, D. et al. (2020): Knowledge Graphs, S. 6-10.

dienen, Dokumente zu registrieren, zu ordnen, bereitzustellen, aufzubewahren und auszusondern. Im 20. Jahrhundert wurden diese Tätigkeiten, die bei der Anlage und Verwaltung von Dokumenten anfallen, unter den Begriffen Schriftgutverwaltung, Aktenhaltung oder Aktenführung zusammengefasst.¹⁰⁶ Ab Ende der 1980er-Jahre wurden diese dann durch eine Vielfalt an neuen Begriffen wie Dokumentenmanagement (DM), Records Management (RM), Enterprise Content Management (ECM), Web Content Management (WCM) und einige mehr ersetzt.¹⁰⁷ Der Funktionsumfang und auch das Verständnis der Begriffe variieren stark je nach Marktsegment, Branche und Zielgruppe des konkreten Produkts.¹⁰⁸ ECM kann als eine Art Oberbegriff betrachtet werden, denn es umfasst alle Strategien, Methoden und Werkzeuge zur Erfassung, Verwaltung, Speicherung, Bewahrung und Bereitstellung von Inhalten (Content) und Dokumenten im Zusammenhang mit Organisationsprozessen in Unternehmen.¹⁰⁹ Durch die Verwaltung des Content sollen gesetzliche Vorschriften eingehalten, Geschäftswissen beibehalten und verbreitet, Prozesse optimiert, neue Arbeitsweisen entdeckt und im besten Falle auch noch Geld gespart werden.

2.4.1 Aufgaben des ECM

ECM kann jegliche Arten von elektronischen Dokumenten enthalten, digitalisierte Objekte, genauso wie „born-digital“-Dateien, z.B. Excel-Tabellen, reine Textdokumente, Digitalfotografien oder Dateien gemischten Inhalts, wie beispielsweise eine PDF-Datei, die Text und Bilder enthält.¹¹⁰ Der Begriff „Content“ ist nicht klar definierbar, u.a. aufgrund des Problems der Definition der Begriffe Daten, Information, Wissen und Dokument. Man kann Content jedoch als „alles, was an inhaltlicher Information in Systemen vorgehalten wird“¹¹¹ bezeichnen.

¹⁰⁶ Archivschule Marburg (o.J.): Terminologie. Schriftgutverwaltung [elektronische Quelle].

¹⁰⁷ Vgl. Digitales Wörterbuch der Deutschen Sprache (o.J.): Dokumentenmanagement [elektronische Quelle].

¹⁰⁸ Vgl. Schwalm, S.; Ulrich, R.; Dippel, A. (2008): Lexikon Dokumentenmanagement und Archivierung, S. 53.

¹⁰⁹ Vgl. AAIM (2020): What is Enterprise Content Management (ECM)? [elektronische Quelle].

¹¹⁰ Vgl. Kampffmeyer, U. (2006): Dokumentenmanagement, S. 6 ff [elektronische Quelle].

¹¹¹ Project Consult (o.J.a): Enterprise Content Management (ECM) [elektronische Quelle].

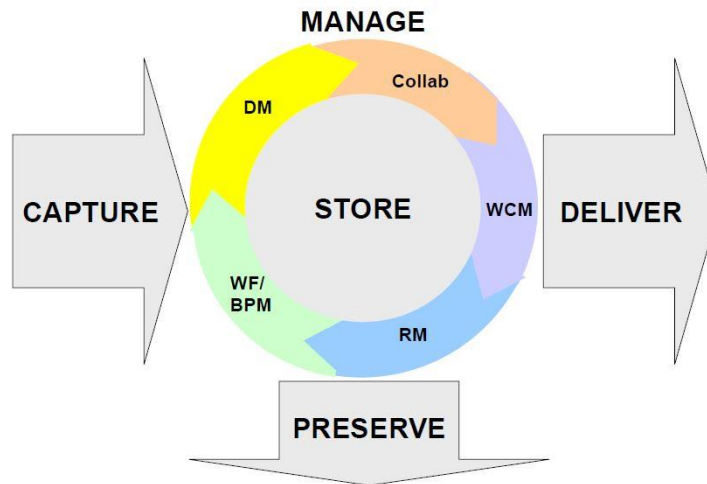


Abb. 7: ECM-Modell¹¹²

Das ECM-Modell beinhaltet die wichtigsten Komponenten eines ECM-Systems: das Erfassen (Capture), Verwalten (Manage), Speichern (Store), Ausgeben (Deliver) und langfristige Sichern (Preserve) von Inhalten. Zuständig für diese Funktionen sind DM, kollaborative Systeme, WCM, RM und Workflow (WF) bzw. Business Process Management (BPM).

Das DM umfasst die Suche und Navigation im System zum Auffinden von Informationen (beispielsweise durch Volltextsuche), Check-In/Check-Out-Mechanismen, Versionsmanagement, Revisionen und Renditionen (gleiche Information in einem unterschiedlichen Format) sowie die Visualisierung des Ganzen. „Collab“ steht für „Collaboration“ und meint dementsprechend Funktionen, welche die Zusammenarbeit mehrerer Personen erleichtern bzw. ermöglichen, z.B. Videokonferenzen, die gleichzeitige und kollaborative Arbeit an Content, gemeinsam nutzbare Wissensbasen, Terminpläne usw. Beim WCM geht es, wie der Name schon sagt, um Inhalte, die im Internet oder auf Portalen bereitgestellt werden. Diese werden kontrolliert erstellt und veröffentlicht, automatisch für unterschiedliche Anzeigeformate konvertiert und für die Darstellung in Browsern visualisiert.¹¹³

RM wird im deutschen Sprachgebrauch in der DIN ISO 15489-1:2001 als Schriftgutverwaltung übersetzt und beschreibt die „als Führungsaufgabe wahrzunehmende, effiziente, und systematische Kontrolle und Durchführung der Erstellung, Entgegennahme, Aufbewahrung, Nutzung und Aussonderung von Schriftgut einschließlich der Vorgänge zur Erfassung und Aufbewahrung von Nachweisen und Informationen über

¹¹² Abb. entnommen aus: Project Consult (o.J.a): Enterprise Content Management (ECM) [elektronische Quelle].

¹¹³ Vgl. ebd. [elektronische Quelle]

Geschäftsabläufe und Transaktionen in Form von Akten.“¹¹⁴ Records meinen hier aufbewahrungspflichtige oder aufbewahrungswürdige Informationen.¹¹⁵ Folglich werden beim RM in einem ECM-System Aufbewahrungs- und Vernichtungsfristen verwaltet, Informationen indiziert und mittels Metadaten beschrieben und strukturierte Verzeichnisse zur geordneten Ablage von Informationen abgebildet. WF und BPM schließlich sind u.a. zuständig für das Abbilden von Ablauf- und Organisationsstrukturen, die Einbindung von Bearbeitungswerkzeugen (z.B. Office-Produkte) sowie Kontrolle und Dokumentation von Bearbeitungsständen und Laufwegen.¹¹⁶

2.4.2 Funktionen moderner ECM-Systeme

Der Bereich der Schriftgutverwaltung, der in jedem Unternehmen unerlässlich ist, veränderte sich durch die digitale Transformation grundlegend. Ohne auf die technischen Hintergründe einzugehen, seien im Folgenden einige dieser Veränderungen durch neue Funktionen genannt.

Einen der großen Einschnitte bildet die räumliche Unabhängigkeit. Durch die Digitalisierung fallen weniger Papierdokumente an, die physischen Raum einnehmen. Dokumente können zeitgleich von mehreren Personen standortübergreifend bearbeitet und/oder gelesen werden. Immer häufiger werden serverbasierte On-Premises-Installationen durch Cloud-Services erweitert, wodurch die problemlose Nutzung an jedem Ort und auf verschiedenen mobilen Endgeräten möglich wird.¹¹⁷ Gerade in diesem bemerkenswerten Jahr 2020, in dem Millionen von Menschen aufgrund der Corona-Krise im Home-Office arbeiten müssen und ihre Büroräume nicht betreten können, bewährte sich dieser Aspekt des ECM.

Weitere wichtige Funktionen moderner ECM-Systeme sind das ausgefeilte Versionsmanagement, die Check-In/Check-Out-Mechanismen und kontrollierte Änderungsdienste, welche sicherstellen, dass alle Änderungsschritte, die ein Dokument durchläuft, erhalten bleiben. Automatisierungen machen es möglich, dass Dokumente nach Ablauf eines festgelegten Zeitraums vernichtet werden. So werden keine Fristen mehr verpasst. Ebenso werden Dokumente, die für einen längeren Zeitraum gesichert werden sollen, entsprechend archiviert.¹¹⁸ Berechtigungssysteme sichern den Zugriff auf Content vor

¹¹⁴ DIN ISO (2002): DIN ISO 15489-1:2001: Information und Dokumentation. Schriftgutverwaltung. Teil 1: Allgemeines (DIN ISO 1548-1:2002)

¹¹⁵ Vgl. Götzer et al. (2013): Dokumenten-Management, S. 162.

¹¹⁶ Vgl. Project Consult (o.J.a): Enterprise Content Management (ECM) [elektronische Quelle].

¹¹⁷ Vgl. PROJECT CONSULT Unternehmensberatung GmbH (2019): ECM, EIM, Content Services, IIM – what's next?, 31.46-33.06 [YouTube].

¹¹⁸ Vgl. Kampffmeyer, U. (2014): Was ist ECM Enterprise Content Management?, S. 10-14 [elektronische Quelle].

Unbefugten. Dokumente können in Module unterteilt werden, die dann in mehreren Dokumenten problemlos eingesetzt werden können. Bei der Änderung von Modulen werden diese gleichsam in allen Dokumenten, in denen sie vorkommen, geändert.¹¹⁹ Sehr hilfreich ist diese Funktion z.B. bei der Erstellung von Packungsbeilagen oder Gebrauchsanweisungen, die oft in gewissen Abschnitten für verschiedene Produkte identisch sind. Durch die Volltextsuche und automatische Klassifikation und Indexierung können gesuchte Inhalte leichter gefunden werden. Optical Character Recognition (OCR), zu Deutsch Texterkennung, macht dies z.B. auch bei gescannten Dateien möglich.¹²⁰ Die OCR-Technik analysiert Dokumentinhalte und zerlegt sie in Zeilen, Wörter und schließlich Buchstaben, welche dann mit Musterbildern von Buchstaben verglichen werden. So wird der gesamte Text abgearbeitet, bis das System eine Lösung präsentiert.¹²¹

2.4.3 Zukunft des ECM

Michael Woodbridge, Research Director bei Gartner, versetzte die ECM-Fachwelt 2017 mit seinem Artikel „The Death of ECM and Birth of Content Services“ in Aufruhr. In diesem erklärt er die Zeit des ECM für vorüber, jedenfalls in der bis dahin gekannten Form. Als Grund dafür gibt er an, dass ECM als einzelne, unternehmensweite Plattform nicht mehr in der Lage sei, die erforderlichen Leistungen zu erbringen und die Anforderungen von Kunden und Kundinnen ausreichend zu befriedigen. Dies liege u.a. an zu hohen Kosten und Zeitaufwänden, die durch die zentralisierte Plattform entstehen. Stattdessen spricht Woodbridge von „Content Services“.¹²² Content Services sind separate oder integrierte Anwendungen, die ein ECM sozusagen um zusätzliche Funktionen erweitern können.

Allerdings kritisiert der deutsche Unternehmensberater Ulrich Kampffmeyer, der sich seit Jahrzehnten mit den Themen rund um ECM befasst, dass Content Services lediglich eine neue Bezeichnung für ein und dasselbe Konzept sei.¹²³ Gleiches prangert er bei der Association for Intelligent Information Management (AIIM) an, welche zunächst die Bezeichnung von Gartner übernahm und dann innerhalb weniger Monate von „Systems of Understanding“, über „Digital Workplace Land“ und „Information Management“

¹¹⁹ Vgl. Styrz (o.J.): Content Management [elektronische Quelle].

¹²⁰ Vgl. Götzer et al. (2013): Dokumenten-Management, S. 147 ff.

¹²¹ Vgl. Abbyy (2020): Was ist OCR? [elektronische Quelle].

¹²² Vgl. Woodbridge, M. (2017): The Death of ECM and Birth of Content Services [elektronische Quelle].

¹²³ Vgl. PROJECT CONSULT Unternehmensberatung GmbH (2019): ECM, EIM, Content Services, IIM – what's next?, 22.45-23.22 [YouTube].

schließlich bei dem Term „Intelligent Information Management“ landete.¹²⁴ Die vielen verschiedenen Begriffe für die meist mehr oder weniger gleiche Sache sind für Endnutzende sehr verwirrend. Kampffmeyer kritisiert dies, benutzt inzwischen selbst aber auch häufig den Term „Information Management“, da er die Auffassung vertritt, dass es im ECM um jegliche Informationen geht. Das Wort „Enterprise“ könne man inzwischen getrost weglassen, da es nicht mehr nur um Informationen des eigenen Unternehmens gehe, sondern genauso um Informationen, die bei den Partner-Unternehmen oder Kunden und Kundinnen liegen.¹²⁵

Welchen Begriff man nun auch verwendet, sicher ist, dass ECM-Systeme heute nicht mehr wegzudenken sind und benötigt werden, um der ansteigenden Informationsflut in Unternehmen gewachsen zu sein. Diese Systeme sind nicht veraltet oder gar tot, sondern entwickeln sich durch zusätzliche, neue Services weiter. Einer der größten Trends im Information Management ist, wie auch in fast allen anderen Bereichen der Informations- und Kommunikationstechnik, momentan die künstliche Intelligenz und das damit verbundene maschinelle Lernen, was sich bereits in zahlreichen Funktionen widerspiegelt. Sie kommen vor allem bei der Informationserfassung und dem Retrieval zum Einsatz. Durch KI-Technologien werden z.B. Prozesse wie die Indexierung und Klassifizierung von Dokumenten automatisiert, die Weiterleitung von Dokumenten an die richtigen Personen, die Prüfung auf Zuständigkeit, Lesbarkeit und Vollständigkeit. Informationen werden zielgerichtet aufbereitet, für die passenden Personen zur Verfügung gestellt und reduziert, um Mitarbeitende von der Durchsuchung von zu viel Content zu befreien und so Zeit (und Geld) zu sparen. Außerdem können Informationen durch diese Techniken gezielt herausgefiltert werden, um sie beispielsweise speziell zu schützen oder zu löschen. Dadurch, dass sich bisherige Dokumentformate immer mehr auflösen und zu strukturierten Komponenten werden, können sie beliebig kombiniert werden und für die Anwender und Anwenderinnen auf praktische Art und Weise in neuen Darstellungsformen präsentiert werden.¹²⁶

Durch die Digitalisierung haben sich viele neue Möglichkeiten für die Dokumentverwaltung- und Verarbeitung erschlossen, die maßgeblich zur digitalen Transformation beigetragen haben. Einige Funktionen existierten bereits in ähnlicher Form im Analogen, beispielsweise Indexe oder Schlagwortsuche. Im digitalen Format sind diese ebenfalls

¹²⁴ Vgl. PROJECT CONSULT Unternehmensberatung GmbH (2019): ECM, EIM, Content Services, IIM – what's next?, 12.09-13.36 [YouTube].

¹²⁵ Vgl. ebd., 18:34-18:54 [YouTube].

¹²⁶ Vgl. PROJECT CONSULT Unternehmensberatung GmbH (2020): Die 10 PROJECT CONSULT Trends für das Information Management 2020, 14.32-19.34; 27.51-32.46; 38.12-41.41 [YouTube]

vorhanden, funktionieren für die Nutzenden jedoch einfacher und schneller. Zudem gibt es sehr viele neue Funktionen, die in analoger Form schlicht unmöglich zu realisieren sind, wie z.B. die Volltextsuche, die Möglichkeit des Zugriffs unabhängig vom Standort und die Dynamik der sich verändernden Dokumente. Zusammenfassend lässt sich sagen, dass das Dokument vor allem in den letzten zwanzig Jahren erhebliche Veränderungen durchlaufen hat und sich nicht als veraltet und überholt, sondern im Gegenteil als wichtiger denn je in unserem Alltag und Berufsleben im modernen Gewand erweist. Zwar imitieren Dokumente immer noch das physische Dokument, indem sie versuchen, es digital möglichst ähnlich aussehend nachzubilden (sehr gut zu beobachten im PDF-Format).¹²⁷ Doch auch hier bahnen sich Veränderungen an, und zwar durch die Möglichkeiten, die sich durch das alles beherrschende Thema KI in diesem Bereich auftun. Das folgende Kapitel befasst sich näher mit dem Thema KI allgemein und erklärt, wie KI-Techniken unser Verständnis von der Funktion und Gestalt eines Dokuments möglicherweise beeinflussen werden.

¹²⁷ Vgl. Dudek, S. (2012): Die Zukunft der Buchstaben in der alphanumerischen Gesellschaft. Text und Dokument unter digitalen Bedingungen, S. 190 [elektronische Quelle].

3. Künstliche Intelligenz

Künstliche Intelligenz ist eines der absoluten Trendthemen der letzten Jahre, und zwar nicht nur in einem bestimmten Bereich, sondern in vielen verschiedenen Branchen. Ob in der Industrie, in verschiedensten Wissenschaftsbereichen von den Natur- über die Geistes- und Sozialwissenschaften bis zur Medizin, in Kriminologie und Psychologie, sogar in der Kunst, dem Umweltschutz, im Marketing, Finanz-, Sicherheits-, Mobilitäts- und auch Militärbereich, überall begegnet man dem Thema heutzutage. Die Investitionen in KI-Startups steigen jährlich rasant an, zwischen 2011 und 2017 wurde weltweit ein Wachstum um den Faktor 50 verzeichnet. China und die USA sind dabei die Hauptinvestoren.¹²⁸

Die Meinungen zum Thema KI gehen jedoch weit auseinander. Die einen sehen darin *den* Lösungsansatz für die Zukunft, die anderen kritisieren sie stark und warnen vor den Folgen, wieder andere stehen dem Thema neutral gegenüber oder sind in ihrer Meinung zwiegespalten.¹²⁹ Kritisch zu sehen ist, dass die zunehmende Verbreitung von KI-Anwendungen nicht Schritt hält mit der Aufklärung in der Bevölkerung. Viele Menschen überschätzen ihr Wissen zu dem Thema erheblich, wie eine Studie der Firma Arithnea aus dem Jahr 2020 zeigt. 81% der Befragten glauben, dass sie nicht täglich mit KI in Berührung kommen, obwohl 98% mindestens einen Service nutzen, der auf KI zurückgreift (z.B. die Google-Suche).¹³⁰ Der Term „KI“ selbst führt zu Missverständnissen und kann Misstrauen auslösen, was vor allem am „Treibstoff“ der KI liegt, nämlich den Daten und der damit verbundenen Möglichkeit der Überwachung und Manipulation. 33% der Studienteilnehmenden halten KI für gefährlich, 30% haben sogar Angst davor.¹³¹ Teilweise stützen sich derartige Ängste auf reale Gefahren, manches entspringt aber auch dem offensichtlich fehlenden Hintergrundwissen, was auch daran liegt, dass das Thema technisch sehr komplex ist. Festzuhalten ist, dass KI-Anwendungen sich rasant in der Arbeitswelt, aber auch dem Privatleben, wenn auch oft unbewusst, ausbreiten. Sie bringen viele neue positive Effekte mit sich, genauso jedoch auch viele Risiken und teilweise unvorhersehbare Folgen für jeden Einzelnen.

Die Einsatzfelder der KI lassen sich grob unterteilen in Sprachverarbeitung, Bildverarbeitung, Expertensysteme und Roboter, wobei diese Bereiche alle Hand in Hand gehen

¹²⁸ Vgl. Buxmann, P; Schmidt, H. (2019a): Ökonomische Effekte der Künstlichen Intelligenz, S. 21.

¹²⁹ Vgl. Lenzen, M. (2020): Künstliche Intelligenz. Fakten, Chancen, Risiken, S. 7 f.

¹³⁰ Vgl. Arithnea (2020): Wie künstliche Intelligenz Kunden und Unternehmen dient, S. 6 ff.

¹³¹ Vgl. ebd., S. 13 [elektronische Quelle].

und nicht klar voneinander abgetrennt sind.¹³² Ein Expertensystem wie Amazon Echo mit der persönlichen Assistentin „Alexa“ muss beispielsweise auch Sprache verarbeiten können.

Das umfassende Ziel von KI ist es, auf *mindestens* menschlichem Niveau bestimmte Aufgaben auszuführen. Dafür muss das System in der Lage sein zu lernen, planen, entscheiden, wahrzunehmen und zu handeln.¹³³ Dies wird durch unterschiedliche Techniken bewerkstelligt, durch die große Datenmengen verarbeitet, deren Zusammenhänge und Muster erkannt und die Ergebnisse bewertet werden können. Dies wird auch als „Knowledge Discovery“ bezeichnet.¹³⁴

In den folgenden Kapiteln wird die Geschichte der KI kurz zusammengefasst, gefolgt von einer Erörterung des Begriffs und Erklärungen zu den wichtigsten Technologien in diesem Bereich, wobei der Fokus auf der Sprachverarbeitung, auch Natural Language Processing (NLP) genannt, liegt, da diese unerlässlich für KI-basierte Wissenstechnologien ist.

3.1 Geschichte der Künstlichen Intelligenz

Künstliche Intelligenz wird von vielen Menschen als eine sehr moderne Wissenschaft angesehen. Dabei reichen einige ihrer Grundgedanken sogar bis zur antiken Logik zurück, schon im Mittelalter wurde, natürlich auf „primitive“ Art und Weise, an automatisierten Beweisverfahren gefeilt, Logiker und Mathematiker wie George Boole und Wilhelm Leibniz zählen ebenfalls zu Vorvätern dessen, was wir heute unter KI verstehen. Bereits in den 1930er Jahren, also noch vor Aufkommen der Computer, entwarf Alan Turing seine heute weltbekannte Turing-Maschine, durch die er beweisen konnte, dass es prinzipiell möglich ist, eine Maschine zu bauen, die jedes Problem lösen kann. Er war es auch, der 1947 in einem Vortrag vor der London Mathematical Society von intelligenten Maschinen sprach, die von Menschen lernen und beim Schachspiel getestet werden könnten. Schon drei Jahre später beschrieb Claude Shannon, Mathematiker, Elektrotechniker und Begründer der Informationstheorie, ein erstes solches Schachprogramm. Ebenfalls in den 1940er-Jahren wurden von dem Logiker Walter Pitts und dem

¹³² Vgl. Kreuzer, R.T.; Sirrenberg, M. (2019): Künstliche Intelligenz verstehen. Grundlagen – Use-Cases – unternehmenseigene KI-Journey, S. 27.

¹³³ Vgl. ebd., S. 19.

¹³⁴ Vgl. ebd., S. 9.

Neurophysiologen und Kybernetiker Warren McCulloch erste sogenannte künstliche neuronale Netze (KNN) entworfen, die dem menschlichen Gehirn nachempfunden sind.¹³⁵

Als Beginn der modernen KI gilt jedoch eine Tagung, die im Jahre 1956 am Dartmouth College in New Hampshire stattfand und an der viele renommierte Forschende teilnahmen, um über die Frage zu diskutieren, wie man es ermöglichen könnte, Maschinen Probleme lösen zu lassen, Sprache zu verwenden und Begriffe zu bilden. Die Forschenden waren zunächst überaus optimistisch und versprachen sich große Erfolge innerhalb kurzer Zeit. Jedoch wurde schon bald klar, dass sie die Komplexität der Probleme und vor allem das menschliche Gehirn in hohem Maße unterschätzt hatten. Ein weiterer Faktor, der die Arbeit behinderte, waren die damals noch sehr geringen Rechenleistungen der Computer. Nach einigen Fehlschlägen kam es Ende der 1950er-Jahre zum ersten sogenannten „KI-Winter“, die Fördergelder für die KI-Forschung wurden massiv gekürzt. In den 1970er- und 1980er-Jahren etablierte sich KI dann trotzdem auch langsam in Europa, die ersten Forschungsgruppen und Förderprogramme entstanden.¹³⁶ Es wurden Expertensysteme entwickelt, indem explizites Wissen, oft in Form von Wenn-dann-Regeln, für den Computer kodiert wurde. Solche Systeme werden „wissensbasiert“ genannt, auch heute noch spielen sie eine wichtige Rolle in der KI. Allerdings können sie lediglich mit deklarativem Wissen umgehen, Wissen, das beschrieben werden kann, größtenteils also Fakten. Nachfolger dieser frühen Expertensysteme sind die bereits in Kapitel 2.3.3 erwähnten Wissensgraphen. Doch auch diese Expertensysteme stießen bald an die Grenzen ihres Leistungsvermögens, Grund dafür waren Aufgaben von zu hoher Komplexität, die sich nicht programmieren ließen. So kam es in den 1980ern also zum zweiten KI-Winter, wieder wurden Forschungsgelder gekürzt, viele Ziele konnten nicht umgesetzt werden.¹³⁷

Die schon von Turing vorgeschlagene Idee, Maschinen von Menschen lernen zu lassen, wurde zu diesem Zeitpunkt wieder aufgegriffen, es wurden sogenannte lernende Verfahren entwickelt, die sich an der Arbeit des menschlichen Gehirns orientierten. Seit dem Eintritt ins 21. Jahrhundert gibt es immer schnellere Fortschritte im Bereich des maschinellen Lernens, vor allem beim „Deep Learning“, von dem in Abschnitt 3.3 noch die Rede sein wird.¹³⁸ Wie bereits erwähnt, ist KI heute populärer denn je, durch die immer stärkere Rechenleistung der Computer sind heute Verfahren möglich, die vor einigen Jahren

¹³⁵ Vgl. Lenzen, M. (2020): Künstliche Intelligenz. Fakten, Chancen, Risiken, S. 18 ff.

¹³⁶ Vgl. Lenzen, M. (2020): Künstliche Intelligenz. Fakten, Chancen, Risiken, S. 19-23.

¹³⁷ Vgl. ebd., S. 23 f.

¹³⁸ Vgl. ebd., S. 24.

noch nicht realisierbar waren. Der Begriff KI ist zwar in aller Munde, jedoch ist die Bezeichnung recht problematisch, was im folgenden Kapitel näher erläutert wird.

3.2 Bedeutung des Begriffs „Künstliche Intelligenz“

KI ist ein Teilgebiet der Informatik, in dem versucht wird, „kognitive Berechnungsprozesse nachzubilden, die ein Computer ausführen kann.“¹³⁹ John McCarthy, einer der „Väter“ der KI, der zur damaligen Konferenz in New Hampshire einlud und diesen Begriff prägte, bezeichnet KI als „the science and engineering of making intelligent machines, especially intelligent computer programs.“¹⁴⁰ Nun ist diese Definition an sich leicht zu verstehen, beinhaltet jedoch ein Problem: was bedeutet „intelligent“? Intelligenz ist ein Begriff, den wir mit dem Menschen und teilweise auch Tieren verbinden. Man kann sagen, „dieser oder jener Mensch ist intelligent“, oder „dieses Tier zeigt Anzeichen von Intelligenz“. Was genau ist aber mit Aussagen wie diesen gemeint? Intelligenztests lassen vermuten, dass es sich bei Intelligenz um eine Eigenschaft handelt, die sich wie das Körpergewicht oder die Größe eines Menschen messen lässt. Dem ist aber nicht so. Zwar haben derartige Tests eine gewisse Aussagekraft, sie behandeln jedoch nur einen Teil der Intelligenz, die aber viele weitere Gebiete umfasst. Der Psychologe Howard Gardner entwarf in den 1980er-Jahren seine „Theorie der multiplen Intelligenzen“, in der von mindestens acht verschiedenen Intelligenzen die Rede ist: sprachlicher, musikalischer, logisch-mathematischer, räumlicher, körperlich-kinästhetischer, naturalistischer, intrapersonaler und interpersonaler Intelligenz. Laut Gardners Analysen werden in Schulen lediglich zwei davon, die sprachliche und die logisch-mathematische Intelligenz, getestet. Seiner Meinung nach ist es möglich, dass ein Mensch in den verschiedenen Feldern unterschiedlich hohe Intelligenzen aufweist, beispielsweise eine hohe logisch-mathematische, jedoch niedrigere interpersonale Intelligenz.¹⁴¹

Gardners Theorie wird von vielen anderen Forschenden aus dem Bereich der Psychologie bezweifelt, da sie Intelligenz als eine einheitliche, nicht multiple Eigenschaft verstehen. Genauso gibt es jedoch auch Wissenschaftler und Wissenschaftlerinnen, die den multiplen Ansatz vertreten, wenn auch mit verschiedenen Anzahlen und Benennungen von Intelligenzen.¹⁴² Es herrscht zwar Konsens darüber, dass Intelligenz verschiedene Prozesse wie Erkenntnisvermögen, abstraktes Denken, Repräsentation, Urteilsfähigkeit,

¹³⁹ Lenzen, M. (2020): Künstliche Intelligenz. Fakten, Chancen, Risiken, S. 11.

¹⁴⁰ McCarthy, J. (2007): What is Artificial Intelligence?, S. 2 [elektronische Quelle].

¹⁴¹ Vgl. Gardner, H. et al. (2011): The Theory of Multiple Intelligences, S. 485 f.

¹⁴² Vgl. ebd., S. 486.

Problemlösen und Entscheidungsfindung beinhaltet,¹⁴³ Dennoch gibt es bisher keine allgemeingültige Definition von Intelligenz. Aufgrund dieser Unklarheit ist es nicht möglich zu bestimmen, ab wann ein System als „intelligent“ bezeichnet werden kann.¹⁴⁴ Zwar gab es auch hierfür mehrere Ansätze, z.B. den Turing-Test, bei dem ein Mensch unwissentlich mit einer Maschine kommuniziert. Fällt dem Menschen nicht auf, dass es sich bei seinem Gegenüber um eine Maschine handelt, so kann diese als intelligent bezeichnet werden. Doch auch der Turing-Test ist, wie viele andere Verfahren, umstritten. Denn ein großes Problem von künstlichen Intelligenzen ist, dass sie nur exakt das können, wozu sie programmiert sind. Der Philosoph Richard David Precht beginnt sein Buch „Künstliche Intelligenz und der Sinn des Lebens“ mit folgendem Zitat von Paul Valéry: „A cyclone can raze a city, yet not even open a letter or untie the knot in this piece of string.“¹⁴⁵ In diesem Satz steckt genau dieses Problem der KI. Das Programm „Alpha-Zero“ beispielsweise lernte innerhalb von wenigen Stunden, Schach auf weltmeisterlichem Niveau zu spielen. Dasselbe Programm wäre jedoch nicht einmal in der Lage, eine Partie Memory gegen ein Kind zu gewinnen, einzig und allein aus dem Grund, weil es dafür nicht programmiert ist.¹⁴⁶ Jeder menschliche Schachweltmeister ist hingegen sicherlich in der Lage, dies zu bewerkstelligen. An diesem Beispiel wird klar, warum es (bislang) nicht möglich ist, eine menschenähnliche Intelligenz zu programmieren. Die Wissenschaft weiß bisher schlichtweg zu wenig darüber, wie sie funktioniert, um in der Lage zu sein, dies in einem Programm umzusetzen.

Bemerkenswert ist, dass KIs in der Lage sind, Dinge zu tun, die Menschen besonders schwerfallen, beispielsweise hochkomplizierte Berechnungen anzustellen. Haben sie erst einmal die klaren mathematischen Regeln erlernt, die dafür benötigt werden, ist ihnen keine Rechenaufgabe zu schwer. Auf der anderen Seite stellen viele Dinge, die für Menschen buchstäblich kinderleicht sind, teilweise unlösbare Aufgaben für Anwendungen der KI dar. Ein gutes Beispiel hierfür ist die Objekterkennung. Schon ein Kind im Grundschulalter wird keinerlei Probleme damit haben, einen Wolf von einem Schäferhund zu unterscheiden. Dafür reicht es aus, dass das Kind einige Male Bilder von beiden Rassen gesehen hat. Eine KI kann ebenfalls durch Bilder erlernen, was typische Merkmale eines Wolfs und eines Schäferhunds sind, um diese zu unterscheiden. Allerdings lässt sich das System leicht verwirren, indem z.B. ein Wolf mit Halsband gezeigt wird.¹⁴⁷ Schon allein aufgrund dieser minimalen Änderung wird die KI Probleme haben zu

¹⁴³ Vgl. Görz, G.; Nebel, B. (2003): Künstliche Intelligenz, S. 10.

¹⁴⁴ Vgl. Lenzen, M. (2020): Künstliche Intelligenz. Fakten, Chancen, Risiken, S. 12.

¹⁴⁵ Valéry, P. (1970): Collected Works of Paul Valéry, Volume 14.

¹⁴⁶ Vgl. ebd., S. 13.

¹⁴⁷ Vgl. Kreutzer, R.T.; Sirrenberg, M. (2019): Künstliche Intelligenz verstehen. Grundlagen – Use-Cases – unternehmenseigene KI-Journey, S. 4.

entscheiden, ob es sich nun um Wolf oder Schäferhund handelt. Das Kind würde jedoch sofort erkennen, dass es sich um einen Wolf handelt, der eben seltsamerweise ein Halsband trägt. Die Prozesse, die im menschlichen Gehirn ablaufen, um dies zu erkennen, sind noch nicht weit genug verstanden, um sie in einer Maschine nachzuahmen.¹⁴⁸

Viele KI-Forschende plädieren gerade aufgrund dessen, dass die sogenannte künstliche Intelligenz keinesfalls mit der menschlichen Intelligenz gleichgesetzt werden kann, dafür, einen anderen Begriff zu verwenden, beispielsweise „erweiterte Intelligenz“. Dadurch, dass der Begriff wörtlich aus dem Englischen ins Deutsche übersetzt wurde, werden vollkommen andere Assoziationen erzeugt, denn „artificial“ und ebenso die umstrittene „intelligence“ können auf Englisch auch andere Bedeutungen haben. Schließlich ist die „Central Intelligence Agency“ (CIA) auch keine Intelligenzbehörde, sondern ein Geheimdienst. „Intelligence“ kann auch Einsicht, Verstand, Information oder Nachrichten bedeuten, „artificial“ kann ebenso als „unecht“ oder „imitiert“ übersetzt werden. Der Begriff KI ist irreführend und bewirkt bei fachkundigen Personen die Entstehung von Ängsten und überhöhten Erwartungen.¹⁴⁹ Von einer künstlichen Intelligenz zu sprechen, kann den Eindruck erwecken, dass es sich um eine Art Ersatz für die menschliche Intelligenz handelt, so wie ein künstliches Hüftgelenk ein menschliches Hüftgelenk ersetzen kann. Dies ist jedoch nicht der Fall. KI funktioniert anders als der Mensch und hat deshalb auch andere Fähigkeiten. Sie ist momentan nicht in der Lage, den Menschen zu ersetzen, schon allein deshalb, weil hinter jeder KI Menschen stecken, die sie programmieren müssen, damit sie funktioniert.

Oft wird von starker und schwacher KI gesprochen. Eine schwache KI bezeichnet eine Maschine, die in der Lage ist, dem Menschen vergleichbare oder sogar bessere Ergebnisse hervorzubringen, allerdings nur auf einem speziellen, eingeschränkten Gebiet. Momentan sind also alle existierenden KIs schwache KIs. Eine starke KI hingegen wäre dem Menschen in vielen oder allen Bereichen ebenbürtig oder gar überlegen. Auch wenn von wenig wissenschaftlichen Medien häufig suggeriert wird, dass man bereits kurz vor der Entwicklung einer solchen starken KI steht, sieht die Realität doch anders aus. Von einer derartigen KI ist man u.a. aufgrund der bereits genannten Punkte noch sehr weit entfernt. Viele Wissenschaftler und Wissenschaftlerinnen schließen sogar aus, dass es jemals möglich sein wird, eine starke KI zu kreieren. Genauso gibt es aber Verfechter der starken KI, wie z.B. den US-amerikanischen Erfinder und Futurist Ray Kurzweil,

¹⁴⁸ McCarthy, J. (2007): What is Artificial Intelligence?, S. 4 [elektronische Quelle].

¹⁴⁹ Vgl. Bonin, G. (2020): „Künstliche Intelligenz gibt es eigentlich nicht“. Gespräch mit Ladan Pooyan-Weihs [elektronische Quelle].

welcher deren Fertigstellung bereits auf das Jahr 2045 prognostiziert.¹⁵⁰ Letztendlich werden die zukünftigen Jahre zeigen, welche Behauptung zutrifft.

Der Begriff „KI“ ist also durchaus umstritten und problematisch aufgrund der fehlenden allgemeingültigen Definition von Intelligenz, allerdings ist er bereits in einem Maße verbreitet, dass es für einen Umschwung zu einem anderen Begriff wohl zu spät ist. Wie bereits erwähnt, ist die Kenntnis der meisten Menschen über KI und ihre Funktionen trotz der Allgegenwärtigkeit des Themas eher gering. Das folgende Kapitel soll zur Aufklärung beitragen und behandelt das Feld des Lernens, welches eine der wichtigsten Komponenten von KI ist.

3.3 Maschinelles Lernen, künstliche neuronale Netze und Deep Learning

Das Konzept des maschinellen Lernens ist ebenso wie die Idee der KI schon älter, jedoch haben sich die Rahmenbedingungen, die zu ihrem Funktionieren erforderlich sind, in den letzten Jahren stark verändert, sodass das maschinelle Lernen erst jetzt seinen großen Erfolgzug antreten kann. Zu diesen Rahmenbedingungen zählen u.a. die stark verbesserte Rechenleistung von modernen Computern, der immer günstiger werdende Speicherplatz und das Vorhandensein von nutzbaren Daten in horrendem Ausmaß.¹⁵¹

Letzteres macht automatisierte Methoden der Datenanalyse immer wichtiger, welche als maschinelles Lernen bezeichnet werden. Dabei werden Algorithmen eingesetzt, die gewisse Lernprozesse durchlaufen und so Muster in Daten automatisch erkennen können, wodurch sie Zukunftsprognosen anstellen oder Entscheidungen treffen können.¹⁵² Der große Vorteil des maschinellen Lernens ist, dass Wissen nicht explizit codiert werden muss, sondern die Systeme auf Basis von Trainingsdaten selbst „lernen“. Zur Veranschaulichung soll nochmals das Wolf/Schäferhund-Problem betrachtet werden. Ohne maschinelles Lernen müsste in einem Softwarecode genaustens formuliert sein, worin sich Schäferhund und Wolf unterscheiden. Dadurch könnte das System dann klare Bilder der richtigen Rasse zuordnen. Der Nachteil bei der Codierung von bestimmten Merkmalen der beiden Tiere liegt darin, dass sich vieles, was man weiß, schwer in Worte

¹⁵⁰ Vgl. Kurzweil, R. (2005): *The Singularity Is Near: When Humans Transcend Biology*, S. 136.

¹⁵¹ Vgl. Buxmann, P.; Schmidt, H. (2019b): *Grundlagen der Künstlichen Intelligenz und des Maschinellen Lernens*, S. 7.

¹⁵² Vgl. Murphy, K.P. (2012): *Machine Learning. A Probabilistic Perspective*, S. 2.

fassen und daher schwer codieren lässt. Dieser Umstand wird nach dem Philosophen Michael Polanyi als Polanyi-Paradox bezeichnet.¹⁵³

Beim maschinellen Lernen hingegen würde das System mit etlichen Fotos von Wölfen und Schäferhunden trainiert, bis es gelernt hätte, worin die Unterschiede zwischen den beiden Tieren bestehen (bei beiden Methoden würde jedoch das Halsband am Wolf die Maschine verunsichern).

Das maschinelle Lernen lässt sich in drei Kategorien gliedern: überwachtes Lernen (supervised learning), unüberwachtes Lernen (unsupervised learning) und verstärkendes Lernen (reinforcement learning). Beim überwachten Lernen, welches am häufigsten angewendet wird, wird der Algorithmus mit beschrifteten Wolfs- und Schäferhundbildern trainiert, d.h. der Mensch gibt ihm vor, welches Bild einen Wolf und welches einen Schäferhund darstellt. Beim unüberwachten Lernen hingegen sind die Bilder nicht beschriftet, dem Algorithmus werden keinerlei Vorgaben gemacht und er bildet selbst Kategorien aus, in die er die Bilder gliedert. Das verstärkende Lernen basiert auf einem klassischen Belohnungsprinzip: gibt der Algorithmus eine richtige Antwort, erhält er eine positive Rückmeldung, bei einer falschen Antwort eine negative. Ziel des Algorithmus ist es, die positiven Rückmeldungen zu maximieren, dadurch wird er in eine bestimmte Richtung gelenkt.¹⁵⁴

Künstliche neuronale Netze sind derzeit im Bereich der KI in aller Munde. Wie bereits erwähnt, wurde das erste KNN schon in den 1950er-Jahren entwickelt, doch erst durch den heutigen Stand der Technik sind diese Netze in der Lage, herausragende Leistungen zu produzieren. Ein KNN ist dem Aufbau des menschlichen Gehirns nachempfunden. Es besteht aus Knoten, die Neuronen verkörpern, und Kanten, welche Synapsen darstellen. Die Hirnrinde besteht aus mehreren übereinanderliegenden Schichten, welche in einem KNN mathematisch simuliert werden. Es werden drei verschiedene Arten von Knoten, auch Units genannt, unterschieden: Input-Units, welche die Eingangsdaten erhalten, Output-Units, die die Ausgangsdaten enthalten und Hidden-Units, die sich zwischen Input- und Output-Units befinden und die inneren Schichten des KNN repräsentieren.¹⁵⁵

¹⁵³ Vgl. Buxmann, P.; Schmidt, H. (2019b): Grundlagen der Künstlichen Intelligenz und des Maschinellen Lernens, S. 8.

¹⁵⁴ Vgl. Buxmann, P.; Schmidt, H. (2019b): Grundlagen der Künstlichen Intelligenz und des Maschinellen Lernens, S. 9 ff.

¹⁵⁵ Vgl. ebd., S. 13 f.; Lenzen, M. (2020): Künstliche Intelligenz. Fakten, Chancen, Risiken, S. 37.

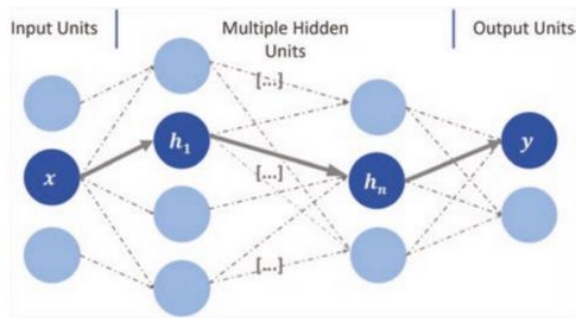


Abb. 8: Vereinfachte Darstellung eines KNN¹⁵⁶

Die Kanten, welche die Units miteinander verbinden, haben verschiedene Gewichtungen, welche zu Beginn zufällig verteilt sind und sich dann während des Umgangs mit den Trainingsdaten verändern, bis die richtige Gewichtungverteilung gefunden ist. Während das erste KNN nur eine Schicht enthielt, können moderne KNN Hunderte solcher Schichten mit Millionen Units umfassen. Je mehr Schichten und Units, desto komplexere Klassifikationen kann ein KNN erlernen und desto mehr Rechenleistung und Trainingsdaten benötigt es, um zu funktionieren.¹⁵⁷

Das sogenannte Deep Learning ist eine besondere Art des maschinellen Lernens. Es kann größere Datenmengen verarbeiten, benötigt weniger Datenvorbereitung durch den Menschen und liefert genauere Ergebnisse als herkömmliches maschinelles Lernen. Zusätzlich werden besondere Optimierungsmethoden eingesetzt. Ermöglicht wird dies durch eine hohe Anzahl von Schichten im KNN, woher das „deep“ im Namen rührt. Deep Learning wird beispielsweise bei Handschriftenerkennung eingesetzt, eine Aufgabe, die sich durch klassische Programmierung nicht lösen lässt.¹⁵⁸

Ein Nachteil der maschinellen Lernverfahren mit KNN ist, dass sich nicht genau nachvollziehen lässt, wie sie zu ihren Schlussfolgerungen kommen. Sie werden daher auch als Black-Box-Systeme bezeichnet. Ein Beispiel dafür ist eine Bilderkennungs-KI von Google aus dem Jahr 2017, die ein Katzenfoto, an dem minimale Änderungen vorgenommen wurden, plötzlich nicht mehr als Katze erkannte, sondern zu dem Schluss kam, dass es sich um Guacamole handelt. Wie die KI zu diesem Fazit kam, bleibt im Dunkeln.

¹⁵⁶ Abb. entnommen aus: Buxmann, P.; Schmidt, H. (2019b): Grundlagen der Künstlichen Intelligenz und des Maschinellen Lernens, S. 14.

¹⁵⁷ Vgl. Lenzen, M. (2020): Künstliche Intelligenz. Fakten, Chancen, Risiken, S. 37 f.

¹⁵⁸ Vgl. Kreutzer, R.T.; Sirrenberg, M. (2019): Künstliche Intelligenz verstehen. Grundlagen – Use-Cases – unternehmenseigene KI-Journey, S. 8.

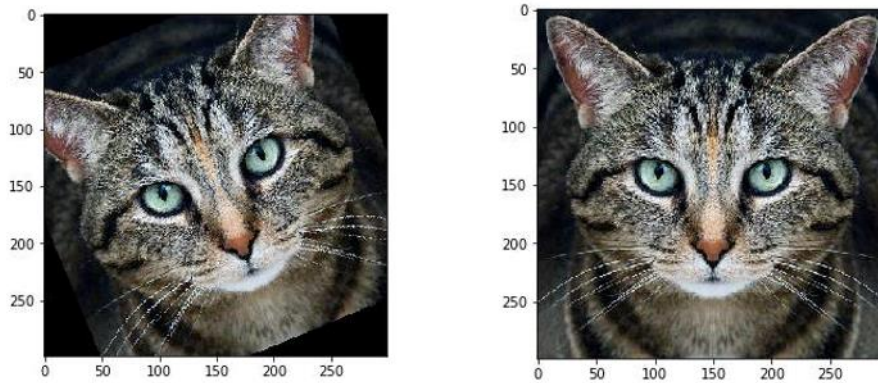


Abb. 9: Links Katze, rechts Guacamole¹⁵⁹

In diesem Fall schadete die KI niemandem mit ihrem Fehltrail. Vorsicht ist jedoch in anderen Bereichen geboten, beispielsweise der Medizin. Ob eine KI hier einen Tumor richtig oder falsch erkennt, kann weitreichende Folgen haben. Die schwer bis gar nicht nachvollziehbaren Black-Box-Systeme können also durchaus gefährlich werden, wenn auch auf eine andere Art und Weise, als es in manchen Science-Fiction-Fantasien dargestellt ist.

3.4 KI-basierte Wissenstechnologien

In Kapitel 2 standen die Techniken und Entwicklungen aus dem informationstechnischen Bereich im Fokus. Alle diese Ansätze haben Auswirkungen auf die zukünftige Entwicklung des Dokuments.

Nachdem bereits u.a. Technologien wie Auszeichnungssprachen, Linked Data etc. dazu beigetragen haben, dass sich das digitale Dokument durch seine Eigenschaften mittlerweile stark vom analogen Dokument unterscheidet, beginnt durch die KI-Technologien nun voraussichtlich eine weitere Ära des Dokuments. Mittlerweile gibt es sehr viele Techniken, die in der einen oder anderen Weise mit Dokumenten arbeiten. Die vorliegende Arbeit beschränkt sich aufgrund deren Vielfalt auf solche Verfahren, die bei der Wissensextraktion aus hauptsächlich textuellen Dokumenten angewandt werden. Themen wie z.B. Sentiment Detection oder Chatbots werden nicht näher erläutert. Zunächst wird die Technik hinter diesen Tools erklärt, in Kapitel 4 folgt dann die Vorstellung von Produkten, welche diese konkret umsetzen.

¹⁵⁹ Abb. verändert entnommen aus: Quach, K. (2017): How we fooled Google's AI into thinking a 3D-printed turtle was a gun: MIT bods talk to El Reg [elektronische Quelle].

3.4.1 Text Mining

Text Mining bezeichnet den Prozess der Prüfung großer Dokumentensammlungen, um neue Informationen zu entdecken oder um spezifische Forschungsfragen zu beantworten.¹⁶⁰ Diese KI-Technik filtert aus der riesigen Menge von unstrukturierten Daten in Dokumenten mittels Text Extraction Fakten, Beziehungen und Behauptungen heraus, die ansonsten für den Menschen schwer auffindbar und größtenteils ungenutzt im Verborgenen bleiben würden. Text Mining verwendet mehrere Methoden, um dieses Ziel zu erreichen. Die wichtigste davon nennt sich „Natural Language Processing“ (NLP).¹⁶¹ Dabei ist es schwer auszumachen, wo Text Mining aufhört und NLP beginnt. Text Mining kann eher als der übergeordnete Begriff verstanden werden, wohingegen NLP eine wichtige Technik des Text Mining darstellt. Verbreiteter als Text Mining ist der Begriff „Data Mining“, auch hier ist die Abgrenzung nicht ganz klar, da bereits darüber gesprochen wurde, dass Daten und Dokumente bzw. Text nicht klar voneinander getrennt werden können. Meist wird unter Data Mining jedoch die Analyse strukturierter Daten verstanden, während das Text Mining eher semi- bis unstrukturierte Daten analysiert. Beim Text Mining werden also Methoden des Data Mining auf Textdokumente angewendet. Den semi- bis unstrukturierten Textdaten kann mittels NLP-Techniken Struktur hinzugefügt werden, wodurch dann Fragen beantwortet, Texte zusammengefasst und Analysen oder Algorithmen des maschinellen Lernens betrieben werden können.¹⁶² Im folgenden Kapitel werden derartige Techniken vorgestellt.

3.4.2 Natural Language Processing

Natural Language Processing (NLP) zählt zu den großen Gewinnern des KI-Bereichs der letzten Jahre. Durch NLP werden Computer dazu befähigt, nicht mehr nur auf formalisierte Programmiersprachen zu reagieren, sondern auch auf natürliche Sprache. Mit dieser können sie in gesprochener oder geschriebener Form arbeiten, sie können selbst Text erzeugen oder Bedeutung aus vorgegebenen Texten extrahieren.¹⁶³ Unterschieden werden die vier Formen Speech-to-Text, Speech-to-Speech, Text-to-Speech und Text-to-Text. Speech-to-Text überführt das gesprochene Wort in einen digitalen Text, dies passiert z.B. bei Transkriptionssoftware. Der umgekehrte Fall, Text-to-Speech, ist ein hilfreiches Feature z.B. für Menschen mit eingeschränkter Sehfähigkeit, die sich dadurch Geschriebenes vorlesen lassen können. Ein Beispiel für Speech-to-Speech ist die

¹⁶⁰ Linguamatics (o.J.): What is Text Mining, Text Analytics and Natural Language Processing? [elektronische Quelle].

¹⁶¹ Vgl. ebd. [elektronische Quelle].

¹⁶² Vgl. Rentzmann, R. (2006): Text Mining [elektronische Quelle].

¹⁶³ Vgl. Gentsch, P. (2018): Künstliche Intelligenz für Sales, Marketing und Service, S. 31.

Kommunikation mit einem Assistenzsystem wie z.B. „Alexa“ von Amazon, Text-to-Text kommt u.a. bei Übersetzungsprogrammen wie dem Google Translator zum Einsatz.¹⁶⁴

Es stellt kein Problem dar, ein komplettes Wörterbuch in eine Maschine einzuspeisen. Schwierig wird es, wenn es um Faktoren wie Dialekt, Akzent, Phonologie, Morphologie, Syntax, Semantik und Pragmatik geht. Diese sind in verschiedenen Sprachen unterschiedlich und hängen von verschiedenen Faktoren wie z.B. der Intention des Gesagten oder Geschriebenen ab. Im Bereich des NLP konnten jedoch in den letzten Jahren aufgrund der großen vorhandenen Datenmengen und der schnelleren Hardware große Fortschritte erzielt werden.¹⁶⁵ Hat man in den letzten Jahren den Google Translator benutzt, konnte man diese Fortschritte mitverfolgen. Kam es vor einigen Jahren noch durchaus vor, dass Sätze wie „Ich ging zur Botschaft“ als „I went to the message“ anstatt als „I went to the embassy“ übersetzt wurden, bleiben derartige Übersetzungsfehler heute aus. Sogar wissenschaftliche Texte können im Google Translator zumindest in bestimmten Sprachen auf gut verständlichem Niveau übersetzt werden, auch wenn sich noch sprachliche Fehler einschleichen und die Texte stilistisch gesehen keine hohe Qualität aufweisen. Sprachen, die eine geringere Anwendung finden, weil sie von weniger Menschen gesprochen werden, wie z.B. Polnisch oder Thai enthalten aufgrund der geringeren Datenmengen oft mehr Fehler.¹⁶⁶ Trotzdem arbeitet der Google Translator insgesamt mittlerweile auf einem guten Niveau.

NLP-Techniken stützen sich entweder auf regelbasierte Ansätze, die von spezialisierten Personen codiert werden müssen und daher sehr zeit- und geldintensiv sind, oder nutzen das bereits vorgestellte maschinelle Lernen, wobei hier große Datenmengen für das Trainieren des System benötigt werden und Fehler durch das Black-Box-Phänomen nicht gut nachvollzogen werden können. Grundsätzlich gibt es NLP-Techniken, die sich um die sprachliche Verarbeitung kümmern und solche, die für die Wissensextraktion zuständig sind. Einige dieser Techniken sollen im Folgenden vorgestellt werden, wobei mit der sprachlichen Verarbeitung begonnen wird.¹⁶⁷

Die *Segmentierung* zerlegt Texte in Sätze oder andere Einheiten. Dies klingt zwar einfach, kann aber erschwert werden durch Strukturen wie „z.B.“, wo ein Punkt kein

¹⁶⁴ Vgl. Kreutzer, R.T.; Sirrenberg, M. (2019): Künstliche Intelligenz verstehen. Grundlagen – Use-Cases – unternehmenseigene KI-Journey, S. 28 f.

¹⁶⁵ Vgl. ebd., S. 30; Gentsch, P. (2018): Künstliche Intelligenz für Sales, Marketing und Service, S. 31.

¹⁶⁶ Vgl. Meinhold, A. (2020): Google Translate vs. DeepL [elektronische Quelle].

¹⁶⁷ Vgl. Abdelaal, H.S. (2019): Knowledge Extraction from Simplified Natural Language Text, S. 20 f. [elektronische Quelle].

Satzende markiert, sondern ein Element der Abkürzung darstellt. Schwieriger sind auch Texte, die anders aufgebaut sind als Fließtexte, z.B. Strichlisten. Bei der *Tokenisierung* werden Sätze in kleinere Einheiten, sogenannte Tokens, aufgeteilt. Das System registriert die Leerzeichen und Interpunktionen zwischen den Einheiten und kann so erkennen, wo Tokens beginnen und enden. Tokens können z.B. Wörter, Zahlen oder Symbole sein.¹⁶⁸ Für westliche Sprachen ist dies einfacher als für nicht-segmentierende Sprachen wie z.B. Thai, in denen es keine Leerzeichen gibt. Hier müssen alternative Verfahren angewandt werden.¹⁶⁹

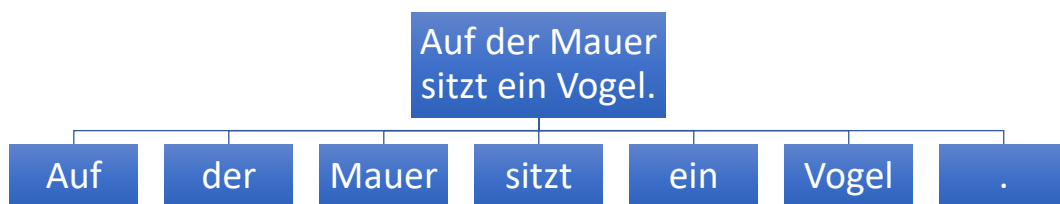


Abb. 10: Vereinfachte Darstellung einer Tokenisierung

Beim *Part-of-Speech (POS)-Tagging* werden den einzelnen Tokens Annotationen wie Präposition, Substantiv etc. zugewiesen. Die Präzision der Ergebnisse hängt davon ab, mit welcher Art von Texten das System trainiert wird. Ähneln die Trainingstexte den Texten, die dem trainierten System vorgelegt werden, können sehr gute Ergebnisse erzielt werden, wohingegen andere Sorten von Texten (z.B. Social-Media-Posts) schlechter verarbeitet werden.¹⁷⁰ Im Folgenden sieht man ein Beispiel für ein POS-Tagging, bei dem das für die deutsche Sprache häufig eingesetzte Stuttgart-Tübingen-Tagset (STTS) verwendet wurde.¹⁷¹



Abb. 11: POS-Tagging nach dem Stuttgart-Tübingen-Tagset

(APPR = Präposition, ART = Artikel, NN = normales Nomen, VFIN = finites Verb ADJA = attributives Adjektiv, \$.= satzbeendende Interpunktion)¹⁷²

¹⁶⁸ Vgl. Abdelaal, H.S. (2019): Knowledge Extraction from Simplified Natural Language Text, S. 22 f.

¹⁶⁹ Vgl. Carstensen et al. (2010): Computerlinguistik und Sprachtechnologie. Eine Einführung, S. 225.

¹⁷⁰ Vgl. Abdelaal, H.S. (2019): Knowledge Extraction from Simplified Natural Language Text, S. 24 f. [elektronische Quelle].

¹⁷¹ Vgl. Universität Stuttgart (1999): STTS Tag Table [elektronische Quelle].

¹⁷² Vgl. ebd.

Das *syntaktische Parsen* erkennt die syntaktischen Beziehungen der Wörter innerhalb der Sätze, basierend auf der Grammatik der zugrunde liegenden Sprache, also die Verbindungen der Wörter zu Wortgruppen und Sätzen. Es gibt dabei verschiedene Ansätze, z.B. *Dependency Parsing*, wobei die Abhängigkeiten der Wörter voneinander analysiert werden, oder *statistisches Parsen*, bei dem Wahrscheinlichkeiten zum Einsatz kommen. Ein Satz kann mehrere verschiedene grammatikalische Analysen liefern, daher muss das System schätzen, welche Analyse eher infrage kommt, um die relative Häufigkeit jeder Regel zu definieren.¹⁷³

Beim sogenannten *Chunking* können mehrere Tokens zu einem Chunk zusammengefasst werden. Chunks sind Textsegmente, die syntaktisch korreliert sind¹⁷⁴, wie an folgendem Beispiel gezeigt wird:

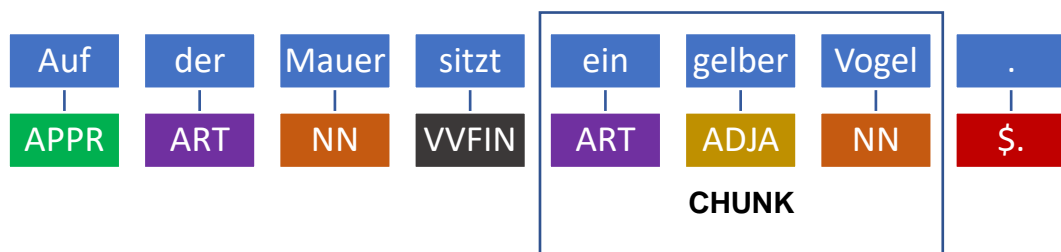


Abb. 12: Beispiel für Chunking

Die *unscharfe Suche*, auch *Fuzzy-Suche* genannt, ist ein String-Matching-Algorithmus, der bei der Eingabe von Suchwörtern nach Wörtern sucht, die Ähnlichkeiten in der Schreibweise oder Phonetik aufweisen.¹⁷⁵ Auch Google arbeitet mit Fuzzy-Suche, so dass das System bei einer Suche mit Schreibfehlern in der Lage ist, dem Suchenden die korrekte Schreibweise anzubieten. Fuzzy-Suche kann auf verschiedene Arten programmiert werden, beispielsweise mit der Levenshtein-Distanz, welche die minimale Anzahl von Änderungen (Hinzufügen, Entfernen oder Austauschen von Zeichen) beschreibt, die benötigt wird, um aus einer ersten Zeichenkette eine zweite Zeichenkette zu generieren.¹⁷⁶

Dies sind die wichtigsten NLP-Techniken, die natürliche Sprache verarbeiten. Es folgen die schwierigeren Verfahren, deren Aufgabe die Wissensextraktion aus den Dokumenten ist.

¹⁷³ Vgl. Abdelaal, H.S. (2019): Knowledge Extraction from Simplified Natural Language Text, S. 25 f. [elektronische Quelle]; NLP-progress (o.J.a): Dependency parsing [elektronische Quelle].

¹⁷⁴ Vgl. Abdelaal, H.S. (2019): Knowledge Extraction from Simplified Natural Language Text, S. 27 [elektronische Quelle].

¹⁷⁵ Vgl. ITWissen.info (2020): Fuzzy-Suche [elektronische Quelle].

¹⁷⁶ Vgl. Franz, S. (2018): Wie Suchalgorithmen funktionieren: Die Levenshtein-Distanz [elektronische Quelle].

Die sogenannte *Koreferenz-Auflösung* stellt verschiedene sprachliche Ausdrücke für dieselbe Entität fest.¹⁷⁷ Enthält ein Dokument beispielsweise folgende Sätze: „Auf der Mauer sitzt ein gelber Vogel. Er singt ein Lied.“, besteht die Aufgabe der Koreferenz-Auflösung darin zu erkennen, dass es sich bei dem Pronomen „er“ um den gelben Vogel aus dem vorigen Satz handelt, also um dieselbe Entität. Die *Eigennamenerkennung* identifiziert und klassifiziert Entitäten mit Eigennamen in einem Text. Die Identifikation ist aufgrund von Homonymen nicht immer einfach zu lösen und auch die Klassifikation ist manchmal nicht eindeutig ausführbar.¹⁷⁸ Klassifiziert wird in Gruppen wie z.B. Personen, Organisationen, geografische oder geopolitische Entität etc.¹⁷⁹ Ein Beispiel für eine irreführende Identifikation ist der Lebensmitteldiscounter „Netto“, dessen Eigenname das System von dem Adverb „netto“ unterscheiden muss.

Beim *Entity Linking* werden die herausgefilterten Entitäten einem Eintrag in einer Ontologie oder Wissensdatenbank, wie z.B. die in Kapitel 2.3.3 erwähnte DBpedia oder Wikidata, zugeordnet.¹⁸⁰ Auch hier gibt es Fehlerrisiken dadurch, dass es oft mehrere Einträge zu einer Entität gibt oder dadurch, dass verschiedene Entitäten denselben Namen tragen, z.B. ist die Stadt Stuttgart sowohl in Baden-Württemberg als auch in Arkansas zu finden.

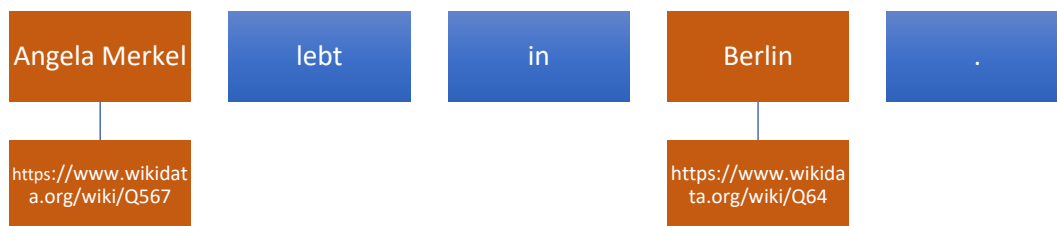


Abb. 13: Beispiel für Entity Linking mittels Wikidata-Datensätzen

Entity-Linking ist ein wichtiger Bestandteil des *semantischen Parsens*. Nachdem beim syntaktischen Parsen die Syntax der Texte erlernt wird, kommt beim semantischen Parsen die Semantik und somit der Sinn hinzu. Erst an dieser Stelle „versteht“ das System die natürlichsprachlichen Sätze, indem es sie in eine formale Bedeutungsrepräsentation umsetzt.¹⁸¹ Semantisches Parsen wird z.B. bei Übersetzungsprogrammen benötigt, da es dem System ermöglicht, nicht nur Wort für Wort zu übersetzen, sondern den Text zu „verstehen“ und somit richtig zu übersetzen.¹⁸² Auch bei einfachen Google-Anfragen

¹⁷⁷ Vgl. ebd., S. 28 [elektronische Quelle].

¹⁷⁸ Vgl. ebd., S. 29 [elektronische Quelle].

¹⁷⁹ Vgl. Li, S. (2018): Named Entity Recognition and Classification with Scikit-Learn [elektronische Quelle].

¹⁸⁰ Vgl. NLP-progress (o.J.b): Entity Linking [elektronische Quelle].

¹⁸¹ Vgl. Mooney, R.J. (2007): Learning for Semantic Parsing, S. 311 [elektronische Quelle].

¹⁸² Vgl. Kreuzer, R.T.; Sirrenberg, M. (2019): Künstliche Intelligenz verstehen. Grundlagen – Use-Cases – unternehmenseigene KI-Journey, S. 31 f.

findet u.a. semantisches Parsen Verwendung. Stellt man z.B. die Frage: „Wie alt ist Brad Pitt?“ erhält man nicht nur die übliche Trefferliste an Webseiten, sondern auch eine Infobox mit seinem momentanen Alter.

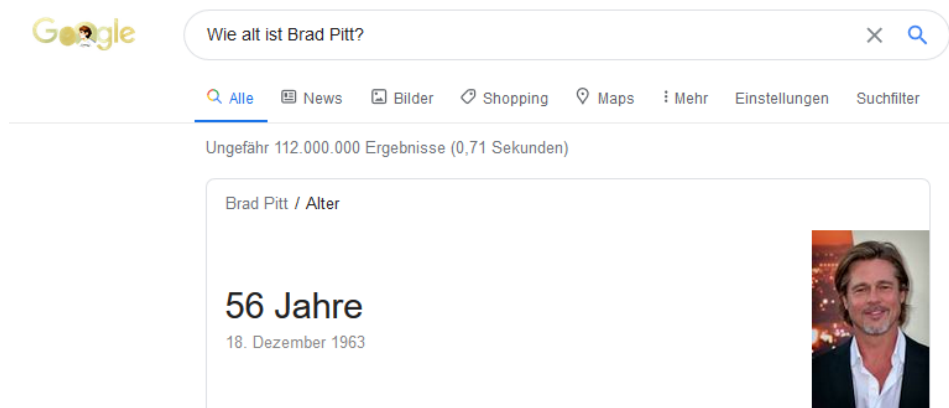


Abb. 14: Google-Antwort zum Alter von Brad Pitt¹⁸³

Eine solche Beantwortung einer natürlichsprachlichen Frage wird nur durch semantisches Parsen ermöglicht. Bisher funktioniert dies allerdings nur mit einfachen Fragen. Gut feststellen lässt sich diese Technik auch bei virtuellen Assistenzsystemen wie Alexa oder Chatbots, an die man auf vielen Websites beim Kundensupport gerät. Diese kombinieren NLP-Techniken zusätzlich mit Natural Language Generation (NLG), um dem Gegenüber „antworten“ zu können.¹⁸⁴

Keyword-Extraktion ist eine weitere NLP-Technik, die aus einem Text die relevantesten Schlüsselbegriffe herausfiltert. Auch hier gibt es mehrere verschiedene Ansätze. Grundsätzlich werden alle Wörter eines Texts gezählt und die Interpunktionen und Stoppwörter entfernt (z.B. Präpositionen, Pronomen, Konjunktionen). Die übrigen Wörter werden normalisiert.¹⁸⁵ Diese Technik wird auch als *Stammformreduktion* oder *Stemming* bezeichnet. Dabei werden verschiedene grammatikalische Formen eines Wortes, wie sein Substantiv, Adjektiv, Verb, Adverb usw. auf die Wurzelform des Wortes reduziert.¹⁸⁶ Anschließend werden die normalisierten Wörter gezählt und z.B. mit der term frequency/inverse document frequency (Tf/idf)-Methode im Verhältnis zu ihrer Seltenheit in der gesamten Dokumentensammlung gewichtet. So können die wichtigsten Schlüsselbegriffe

¹⁸³ Google (2020): Brad Pitt [Screenshot].

¹⁸⁴ Vgl. Kreuzer, R.T.; Sirrenberg, M. (2019): Künstliche Intelligenz verstehen. Grundlagen – Use-Cases – unternehmenseigene KI-Journey, S. 30 ff.

¹⁸⁵ Vgl. Vivek, S. (2018): Automated Keyword Extraction from Articles using NLP [elektronische Quelle].

¹⁸⁶ Vgl. Jivani, A. (2011): A Comparative Study of Stemming Algorithms, S. 1930 [elektronische Quelle].

herausgefiltert werden. Diese Methode wird auch im Informations-Retrieval verwendet.¹⁸⁷

Die *Relation Extraction* kann erst stattfinden, wenn im Vorfeld Entity-Linking durchgeführt wurde. Nachdem alle relevanten Entitäten festgestellt wurden, können ihre Beziehungen untereinander durch Relation Extraction festgestellt werden. Solche Beziehungen können beispielsweise „verheiratet mit“, „Kind von“, „wohnhaft in“, „angestellt bei“ etc. sein.¹⁸⁸

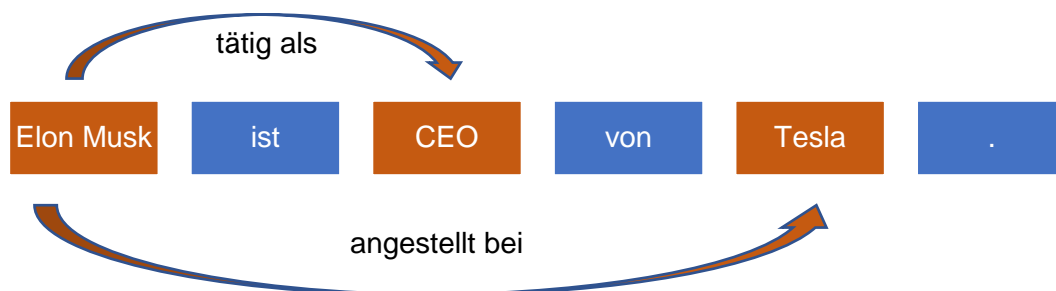


Abb. 15: Vereinfachte Darstellung von Relation Extraction

Auch Relation Extraction kann auf verschiedene Arten stattfinden, manuell codiert, was sehr aufwendig ist, oder durch verschiedene Abstufungen von überwachtem oder unüberwachtem maschinellem Lernen.¹⁸⁹

Wie dieses Kapitel gezeigt hat, können in der Sprachverarbeitung durch komplizierte Algorithmen mittlerweile einige Funktionen des menschlichen Gehirns im Ansatz nachgeahmt werden. Jedoch wird deutlich, wie anders Computer im Vergleich zum Menschen funktionieren. Prozesse, die im menschlichen Gehirn wie von selbst ablaufen, müssen mühevoll und in jahrelanger Arbeit modelliert und trainiert werden. Zurückkommend auf das Vu-Lu-Su-Konzept von Salaün aus Kapitel 1.2 wird der Unterschied nochmals deutlicher: Zwar erkennt ein Computer z.B. eine Textdatei mühelos (Vu), doch bereits das Verstehen des Inhalts (Lu) ist nur in beschränktem Maße gegeben, da von einem wirklichen Verständnis im menschlichen Sinn nicht gesprochen werden kann, sondern lediglich von einer Art nachgeahmten Verständnis. Auf dasselbe Hindernis stößt ein Computer auch bei der sozio-anthropologischen Funktion eines Dokuments (Su). Eine Art „Wissen“ um diese kann zwar modelliert werden, ist aber nur künstlich und nicht wie beim

¹⁸⁷ Vgl. Abdelaal, H.S. (2019): Knowledge Extraction from Simplified Natural Language Text, S. 32 [elektronische Quelle].

¹⁸⁸ Vgl. Abdelaal, H.S. (2019): Knowledge Extraction from Simplified Natural Language Text, S. 33 [elektronische Quelle].

¹⁸⁹ Vgl. Herman, A. (2019): Different ways of doing Relation Extraction from text [elektronische Quelle].

Menschen reflexhaft veranlagt. Trotzdem wurde aufgezeigt, dass künstlich modellierte Intelligenzen auch viele Vorteile bieten. Im folgenden Kapitel werden einige dieser Vor-, aber auch Nachteile anhand von praktischen Beispielen dargestellt.

4. Untersuchung dokumentenbasierter KI-Software

Wie KI-Systeme grundsätzlich und konkret im Bereich von NLP arbeiten, wurde im vorangegangenen Kapitel erläutert. Im Folgenden geht es nun um konkrete Anwendungsbeispiele hierzu. Zu diesem Zweck werden verschiedene Produkte vorgestellt, die mit KI-basierten Wissenstechnologien im Dokumentbereich arbeiten. Eines der Produkte konnte durch die Zurverfügungstellung einer Testversion explizit erprobt werden. Zudem wurden Interviews mit drei Fachkundigen aus diesem Bereich zu ihren Produkten, KI im Allgemeinen und KI-basierten Wissenstechnologien geführt.

4.1 Produktvorstellungen

Ein Einblick in die Produkte konnte durch die bereitgestellten Informationen auf den Webseiten und Videos auf den YouTube-Kanälen der Firmen sowie anhand virtueller Einführungen durch die Interviewten gewonnen werden. Zudem konnten durch die Interviews weiterführende Informationen integriert werden, die durch sonstige Quellen nicht in Erfahrung zu bringen waren. Im Folgenden werden die Produkte Theum, PhlexDistiller und Serviceware Knowledge vorgestellt.

4.1.1 Theum

Die Theum AG wurde 1994 von Paul Caspers gegründet und hat ihren Sitz in Ostfildern in der Region Stuttgart. Paul Caspers studierte Wirtschaftswissenschaften und Geografie und arbeitete 15 Jahre lang im IT-Bereich der Industrie und als freiberuflicher IT-Berater.¹⁹⁰ Die Inspiration zum Produkt Theum kam durch die Arbeit mit verschiedenen Organisationen, bei denen analog vorliegende Dokumente in digitale Systeme integriert werden mussten. Zu diesem Zweck wurden viele verschiedene individuelle Anwendungen mit demselben Nutzen entworfen. Die Idee war daher, stattdessen ein Standardprodukt zu entwerfen. Dieses hieß zunächst „Hyper.Net“ und konnte Dokumente bereits zerlegen und teilweise analysieren, jedoch noch nicht auf dem heutigen Niveau. Ab dem Jahr 2015 fing die Firma an, KI-Technologien in ihr Produkt einzubinden. Schätzungsweise sind seit Beginn etwa 150 Personenjahre in die Entwicklung von Theum geflossen.¹⁹¹

¹⁹⁰ Vgl. Theum AG (2019): KI Einsatz zur präzisen Bereitstellung von Wissen, 1:34 [YouTube].

¹⁹¹ Vgl. Caspers, P. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch].

Theum versteht den Menschen als narratives Geschöpf, das sein Wissen schon seit Tausenden von Jahren in Form von Dokumenten weitergibt. Durch die Digitalisierung änderte sich nichts an diesem Fakt, jedoch änderte sich die Art und Anzahl der Dokumente erheblich, wie bereits mehrfach in den vorigen Kapiteln festgestellt. Aus diesem Grund ist Theum der Meinung, dass die Art des Wissenszugriffs durch Dokumente sich ändern muss. Dieser erfolgt selektiv, doch wie soll ein Mensch die Fülle an Dokumenten heute noch zufriedenstellend selektieren? Richard Reed, Vorstand der Theum AG, ist der Meinung, dass es nicht darum geht, KI-Methoden für eine bessere Dokumentensuche einzusetzen, sondern um Folgendes:

„ [...] to enable knowledge to transcend the confines of document containers, and to enrich it with implied or inferred relationships with the goal of making it increasingly valuable, instantly accessible, and able to guide users efficiently and correctly from A to B in any changing business process.“¹⁹²

Daher machte Theum es sich zur Aufgabe, nicht den Dokument-, sondern den Wissenszugriff so einfach wie möglich zu gestalten, denn „der Zugriff auf Wissen sollte so mühelos und intuitiv sein,[sic!] wie der Zugang zu einem Gedanken“.¹⁹³

Das Produkt Theum arbeitet mit NLP-Techniken, um Wissen aus Dokumenten jeder Art (z.B. originäre Word-Dokumente, PDF, Webseiten, Wikis, etc.) in einer Art und Weise zu präsentieren, die das schnelle Auffinden der gewünschten Informationen möglich macht. Die Hauptaufgabe von Theum ist es, Antworten auf Fragen zu liefern, und zwar nicht in Google-Manier, indem Hunderttausende von Dokumenten (Webseiten) als „Antwort“ auf eine Frage präsentiert werden, sondern indem das Wissen aus allen relevanten Dokumenten antwort-orientiert auf einer einzigen Seite zusammengestellt wird.¹⁹⁴ Dadurch bleibt den Nutzenden das mühevoll und zeitraubende Durchsuchen von allen möglichen Dokumenten, um zu einer passenden Antwort zu gelangen, erspart. Nutzende müssen nicht wissen, wo die Dokumente liegen oder wie sie benannt sind, um auf die Inhalte zuzugreifen. Theum stellt die passenden Inhalte aus allen relevanten Dokumenten zusammen, auch wenn sie auf verschiedenen Plattformen wie z.B. in ECM-Systemen, auf SharePoint, im Intranet, auf Dateiservern, etc. gespeichert sind. Änderungen an Dokumenten werden durch Updates sofort in Theum integriert.¹⁹⁵ Da das System nicht auf eine bestimmte Art von Dokument oder bestimmte Inhalte fixiert ist, kommt die Kundschaft der Firma aus unterschiedlichsten Bereichen. Beispiele sind Siemens, die

¹⁹² Reed, R. (2017): Documents, AI, and the Future [elektronische Quelle].

¹⁹³ Theum (o.J.a): Wir über uns [elektronische Quelle].

¹⁹⁴ Vgl. Theum (o.J.b): Was ist Theum? [elektronische Quelle].

¹⁹⁵ Vgl. Theum (o.J.c): Antworten aus SharePoint und anderen DMS [elektronische Quelle].

Deutsche Rentenversicherung, Bertelsmann oder die Hochschule für angewandte Wissenschaften München.¹⁹⁶

Funktional werden die Inhalte aller vorhandenen Dokumente aus diesen „herausgezogen“ und in, wie Theum es beschreibt, „Wissensatome“ zerlegt. Diese Wissensatome werden miteinander zu einer Wissensmatrix verknüpft, welche in HTML5 vorliegt. Die Wissensmatrix oder Wissensbank ist eine Art „Single Point of Access“ für all das Wissen, das aus den Dokumenten extrahiert wurde.¹⁹⁷ Es ist also nicht nötig, alle verschiedenen Punkte, an denen die Dokumente liegen, einzeln nach den relevanten Dokumenten zu durchsuchen. Die Wissensbank kann auf einem beliebigen Server entweder On-Premises oder als private Cloud installiert werden, wobei weder eine spezielle Software noch eine spezielle Wartung nötig ist.¹⁹⁸

Das Aussehen der Wissensbank kann durch Cascading Style Sheets (CSS) angepasst werden, als Sprachen für die Benutzeroberfläche werden momentan Deutsch und Englisch unterstützt.¹⁹⁹ Für die Suche unterstützt Theum alle möglichen Sprachen, sogar logografische Sprachen wie z.B. Chinesisch. Die verschiedenen Themen einer Wissensbank werden in Form von Kacheln dargestellt, die Bibliotheken genannt werden. An der rechten Bildseite befinden sich Icons, die zu verschiedenen Funktionen führen (z.B. „Drucken“), am oberen Bildrand die Suchleiste für die Recherche und links das Menüsymbol, welches den „Knowledge Navigator“ öffnet. Navigiert werden kann entweder über diesen, der aussieht wie ein ganz normales Drill-Down-Menü, oder über die Kachelnsymbole.

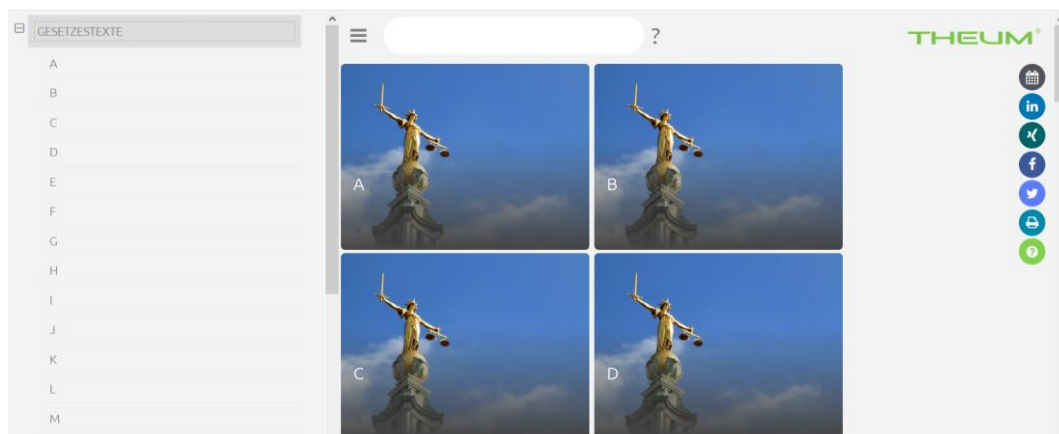


Abb. 16: Theum-Wissensbank mit verschiedenen Bibliotheken und geöffnetem Knowledge Navigator²⁰⁰

¹⁹⁶ Vgl. Theum (o.J.a): Wir über uns. [elektronische Quelle].

¹⁹⁷ Vgl. Theum AG (2019): KI Einsatz zur präzisen Bereitstellung von Wissen, 6:58-10:03 [YouTube]

¹⁹⁸ Vgl. Theum (2019): Theum 3.7 User Guides [elektronische Quelle].

¹⁹⁹ Vgl. Theum (o.J.d): Arbeiten mit Theum Wissensbanken [elektronische Quelle].

²⁰⁰ Theum (o.J.e): Gesetzestexte [Screenshot].

Durch das Klicken auf eine Bibliothek öffnen sich die „Bücher“, welche in dieser enthalten und ebenfalls in Kacheln angeordnet sind. Die Bücher wiederum enthalten dann die Hauptpunkte der jeweiligen Inhalte.²⁰¹ Im Fall der deutschen Gesetzestexte, die Theum aufbereitet hat, bilden alle Gesetzestexte zusammen eine Bibliothek, die verschiedenen Gesetze (z.B. das Grundgesetz) bilden die einzelnen Bücher und die Paragraphen schließlich die weiteren Hauptpunkte.

Theum kann auch auf mobilen Endgeräten online und offline angewendet werden, wobei das Layout dementsprechend angepasst ist. Für die Offline-Verwendung müssen aus den jeweiligen „Büchern“ „E-Books“ im epub-Format erstellt und heruntergeladen werden.²⁰² Das System allgemein ist benutzerfreundlich, da die Navigation recht intuitiv verwendbar ist.²⁰³ Präzise bzw. thematisch enger gefasste Suchanfragen führen direkt zu fertigen „Antworten“, sprich Dokumentexzerpten, welche auf einer einheitlichen Seite dargestellt werden. Durch Links können Nutzende trotzdem erkennen, aus welchen Dokumenten die jeweiligen Wissensatome stammen und diese bei Bedarf per Klick aufrufen.²⁰⁴ Mittels Metadatenfilter können Suchanfragen verfeinert werden, beispielsweise auf einen bestimmten Zeitraum oder Ort eingegrenzt werden.²⁰⁵

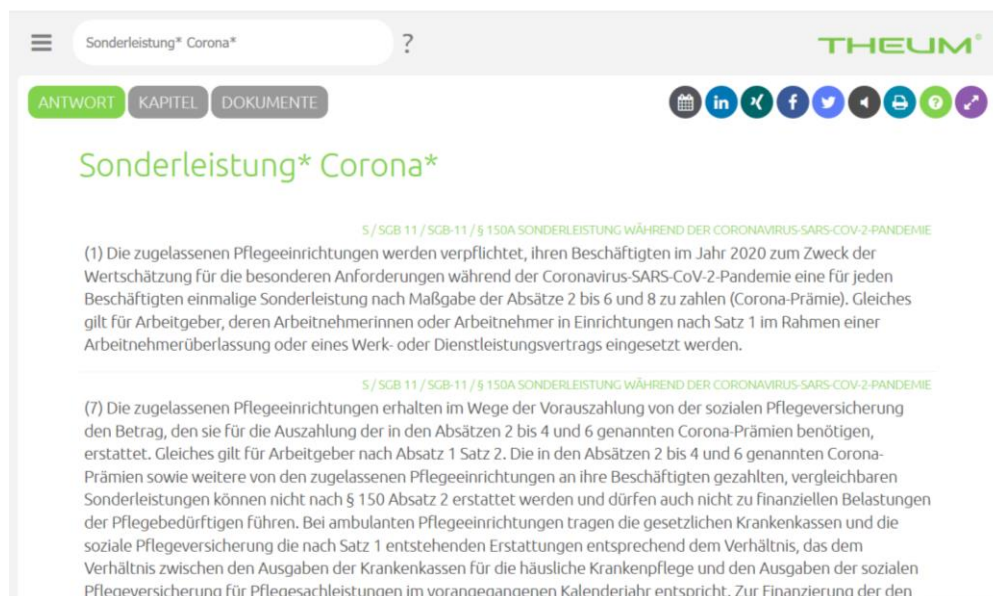


Abb. 17: Auszug zu einer Antwort auf eine Suchanfrage²⁰⁶

Thematisch weiter gefasste Anfragen an das System ergeben zunächst eine Art Menü mit Vorschlägen, aus denen ausgewählt werden kann, in welchem Kontext die Anfrage

²⁰¹ Vgl. Theum (o.J.d): Arbeiten mit Theum Wissensbanken [elektronische Quelle].

²⁰² Vgl. Theum (o.J.d): Arbeiten mit Theum Wissensbanken [elektronische Quelle].

²⁰³ Vgl. Theum AG (2019): KI Einsatz zur präzisen Bereitstellung von Wissen, 6:58-10:03 [YouTube].

²⁰⁴ Vgl. ebd., 13:37-14:27 [YouTube].

²⁰⁵ Vgl. ebd., 16:06-16:31 [YouTube].

²⁰⁶ Theum (o.J.e): Gesetzestexte [Screenshot].

gestellt wird. Durch Klicken auf die jeweilige Kachel führt das System dann auf die Seitenansicht, die das relevante Wissen enthält.²⁰⁷

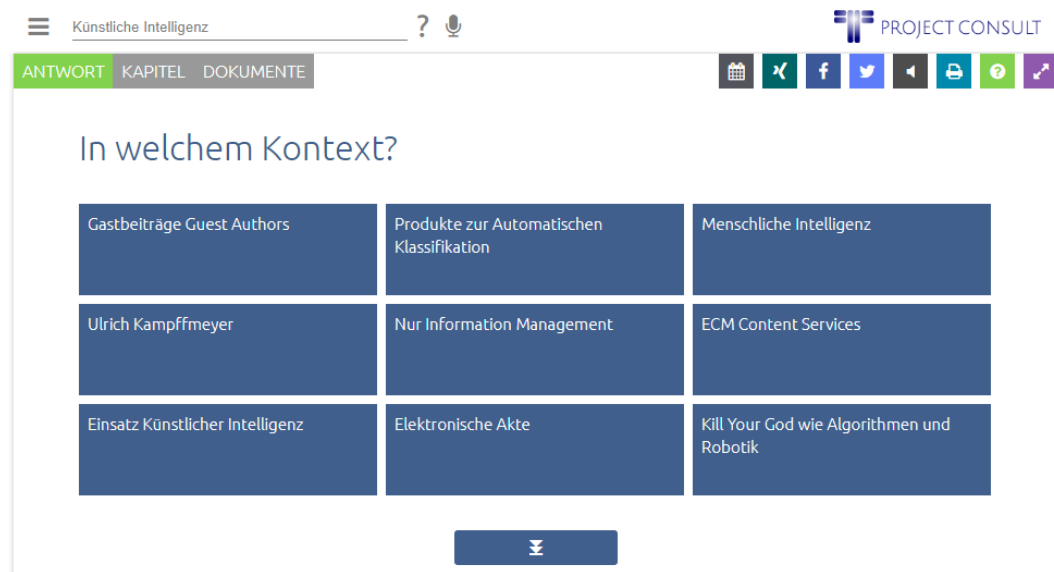


Abb. 18: Unschärfe Suchanfrage mit Kontext-Vorschlägen²⁰⁸

Durch das Einrichten einer Äquivalenzliste können Synonyme festgelegt werden, die fachfremden Personen nicht zugänglich sind oder nur unternehmensintern benutzt werden, damit alles relevante Wissen auch bei der Verwendung dieser Begriffe gefunden wird. Ebenso wird die Stammformreduktion (Stemming) eingesetzt, damit alle verschiedenen grammatikalischen Varianten eines Wortes erkannt werden.²⁰⁹ Stoppwörter werden bei der Suche ignoriert und können bei Bedarf zu einer Liste hinzugefügt werden. Zudem werden verschiedene Suchoperatoren wie z.B. Asterisk-Suche, Platzhaltersuche, Einschränkung auf einen gewissen Zeitraum etc. unterstützt.²¹⁰ Durch die Verknüpfung mit dem jeweiligen Browser, in dem die Anwendung läuft, können Funktionen des Browsers genutzt werden. So kann z.B. Fuzzy-Suche angewandt und per Mausklick korrigiert werden.

Die Dokumente, die Theum verarbeitet, können, wie gesagt, verschiedener Herkunft sein. Offiziell für die Suche unterstützt werden die Formate .doc, .docx, .txt, .log, .ppt, .pptx, .xls, .xlsx, .jpg, .png, .gif, .htm, .html und .pdf. Jedoch begünstigen thematische Strukturen die Verarbeitung. Wie schon in Kapitel 2.1 erklärt wurde, können stark strukturierte Dokumente leichter verarbeitet werden als schwach strukturierte Dokumente. Daher kann Theum z.B. mit einer originären .doc-Datei, bei der Überschriften, Abbildungen etc. als solche gekennzeichnet sind, besser umgehen als mit einem mittels OCR

²⁰⁷ Vgl. Theum AG (2019): KI Einsatz zur präzisen Bereitstellung von Wissen, 16:32-17:53 [YouTube].

²⁰⁸ Abb. entnommen aus: Theum (o.J.f): Recherchieren mit Theum [Screenshot].

²⁰⁹ Vgl. Theum (o.J.f): Recherchieren mit Theum [elektronische Quelle].

²¹⁰ Vgl. Theum (o.J.g): Help [elektronische Quelle].

bearbeitetem Scan in .pdf-Format, der oft keinerlei oder kaum Strukturen enthält. Die Überschriften werden beispielsweise genutzt, um Themen in einem Dokument zu erkennen. Das Prinzip “garbage in, garbage out” trifft hier zu, mit schlecht strukturierten oder inhaltlich unzureichenden Dokumenten lassen sich keine guten Antworten auf Fragen produzieren.²¹¹

4.1.2 PhlexDistiller

Zu Beginn dieser Arbeit hieß das hier vorgestellte Produkt noch CuneDistiller, entworfen von der 2013 in München gegründeten Firma Cunesoft. Cunesoft wurde jedoch zu Beginn des Jahres 2020 von der Firma Phlexglobal mit Sitz in Amersham, England, übernommen. Dementsprechend heißt das Produkt jetzt PhlexDistiller. Phlexglobal ist eine führende Technologie- und Dienstleistungsfirma im Bereich der Pharmabranche, weshalb sich der Kundenkreis auf dieses Gebiet beschränkt. Die Firma bietet Produkte auf KI-Basis für das regulatorische Datenmanagement als Software-as-a-Service (SaaS) an. Die Algorithmen der Software werden mittels Daten aus der Pharmabranche vor ihrem Einsatz trainiert.²¹² Grundsätzlich wird ein Standardprodukt verkauft, das jedoch in den meisten Fällen individuell an die Bedürfnisse der Kundschaft angepasst wird.²¹³

In den Life Sciences sind KI-Technologien bereits Standard. PhlexDistiller ist ein System, das mehrere der in Kapitel 3 vorgestellten Techniken anwendet: Deep Learning, KNN, NLP und Mustererkennung. Dadurch, dass verstärkendes Lernen angewendet wird, entwickelt sich das System permanent weiter und verbessert seine Einstellungen und somit die Ergebnisse. Der PhlexDistiller extrahiert Wissen aus unstrukturierten Datenquellen jeglicher Art, z.B. Webseiten oder PDF-Dokumenten. Auch Scans analoger Dokumente können mittels OCR verwendet werden. Dazu werden die Dokumente z.B. aus einem CM-System hochgeladen, der Dateityp wird automatisch erkannt, anschließend werden Metadaten extrahiert. Das System kann mit Dokumenten in 26 verschiedenen Sprachen arbeiten.²¹⁴

2012 beschloss das EU-Parlament die Identification of Medicinal Products (IDMP), um die Pharmakovigilanz EU-weit zu verbessern. Informationen zu Arzneimitteln sollten in einer Datenbank in einheitlicher Struktur vorliegen. Daher müssen Arzneimittelhersteller ihre Daten zu den Produkten in dieser IDMP-konformen Form abgeben.²¹⁵ Der

²¹¹ Vgl. Theum (2019): Theum 3.7 User Guides [elektronische Quelle].

²¹² Vgl. Phlexglobal (2020a): Startseite [elektronische Quelle].

²¹³ Vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch].

²¹⁴ Vgl. Phlexglobal (2020b): PhlexDistiller [elektronische Quelle].

²¹⁵ Vgl. Arcondis (2017): IDMP (Identification of Medicinal Products) [elektronische Quelle].

PhlexDistiller extrahiert Daten aus Dokumenten wie z.B. Nebenwirkungen, Produktname, Haltbarkeit etc. automatisch und bringt sie in ebendiese Form. Zudem können die Daten durch Zugriff auf externe Datenbanken aufgewertet und bereinigt werden.²¹⁶ Beispielsweise können die Global Ingredient Archival System (GIAS)-Codes für Substanzen automatisch auf die gefundenen entsprechenden Substanzen gemapped werden.²¹⁷ Einer Schätzung zufolge flossen etwa 55 Personenjahre in die Entwicklung des PhlexDistiller, das Produkt wird jedoch immer noch kontinuierlich weiterentwickelt.²¹⁸

In Abb. 19 ist beispielhaft ein Ausschnitt aus dem PhlexDistiller dargestellt, bei dem spezifische Daten aus einem Dokument, nachdem es in das System hochgeladen wurde, extrahiert wurden. Die Abbildung zeigt das betreffende Dokument links, mittig sieht man ein Drill-Down-Menü, in dem die extrahierten Daten gegliedert sind, und rechts befinden sich die hier angeklickten Nebenwirkungen, die der PhlexDistiller in diesem Dokument gefunden hat.

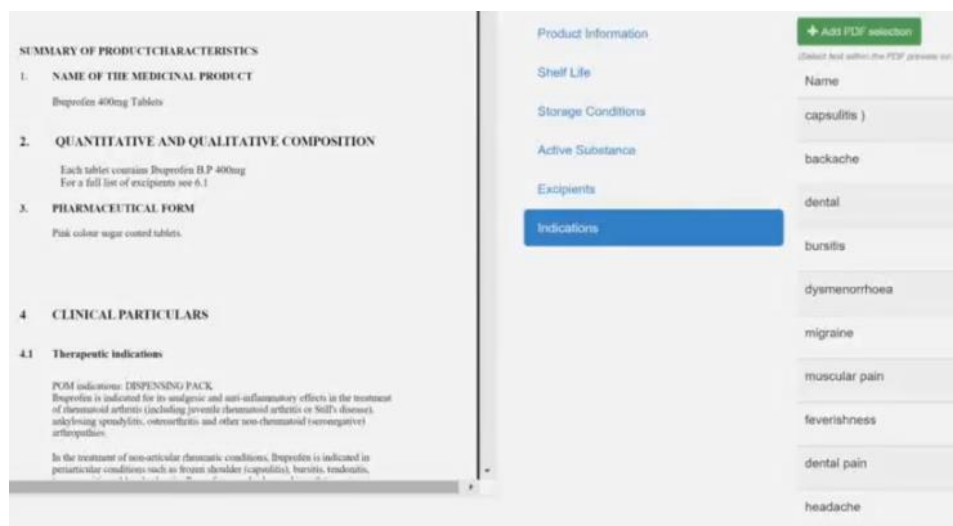


Abb. 19: Extraktion von Nebenwirkungen aus einem Dokument im PhlexDistiller²¹⁹

Werden Daten unsauber extrahiert, wie hier z.B. sichtbar an der unerwünschten geöffneten Klammer hinter dem Wort „capsulitis“ in der ersten Zeile der Nebenwirkungen, kann ein solches Datenfeld markiert und mit den korrekten Daten gefüllt werden. Daraufhin greift das verstärkende Lernen, sodass der Fehler in Zukunft vermieden werden kann.²²⁰ Nachdem die Daten extrahiert und ggf. kontrolliert wurden, kann der PhlexDistiller sich z.B. mit einem IDMP-System verbinden und die Daten per Klick direkt dorthin

²¹⁶ Vgl. Cunesoft Regulatory Compliance – Online (2016): Automatic Regulatory Data Mining Done Right – PhlexDistiller, 0:11-0:49 [YouTube]

²¹⁷ Vgl. Cunesoft Regulatory Compliance – Online (2016): Data mining, extraction, mapping and cleaning software PhlexDistiller, 3:00-3:20 [YouTube]

²¹⁸ Vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch].

²¹⁹ Abb. verändert entnommen aus: Cunesoft Regulatory Compliance – Online (2016): Automatic Regulatory Data Mining Done Right – PhlexDistiller, 3:27 [Screenshot].

²²⁰ Vgl. ebd., 3:48-4:25 [YouTube].

senden oder es kann eine Datei im .xml-, .xls-, .json- oder .csv-Format erstellt und heruntergeladen werden.²²¹ Die European Medicines Agency (EMA) Policy 0070 gibt vor, dass gewisse Informationen, die in klinischen Studien enthalten sind, anonymisiert werden müssen. Solche Daten kann der PhlexDistiller ebenfalls aus den entsprechenden Dokumenten herausfiltern und automatisch anonymisieren.²²²

4.1.3 Serviceware Knowledge

Serviceware ist ein Unternehmen mit Sitz in Bad Camberg, das Strategien und Softwarelösungen für Enterprise Service Management anbietet. Der Kundenkreis ist branchentechnisch und auch von der Unternehmensgröße her sehr gemischt, u.a. zählen dazu Onlinehändler wie z.B. Otto, Banken, Versicherungen und viele mehr.²²³ Die Firma ist der Meinung, dass die reine Digitalisierung von Prozessen heute nicht mehr ausreicht, denn diese sind häufig zu langsam und komplex. Daher hat Serviceware es sich zur Aufgabe gemacht, die Effizienz und Effektivität derartiger Prozesse zu steigern.²²⁴ Eines der Produkte, das von Serviceware vertrieben wird, ist Serviceware Knowledge. Ziel dieser Softwarelösung ist es, das organisatorische Wissen standortübergreifend und zeitunabhängig auf dem aktuellen Stand in einer benutzerfreundlichen Form für die Mitarbeitenden zu präsentieren.²²⁵

Die erste Version von Serviceware Knowledge, damals noch unter dem Namen Sabio, wurde bereits vor 20 Jahren entwickelt, wobei es sich zum damaligen Zeitpunkt um ein Minimal Viable Product handelte. Ziel von Sabio war es, möglichst schnell in den Markt einzusteigen, um das Produkt anhand des Kundenfeedbacks zeitnah in eine gewünschte Richtung zu entwickeln. Die initiale Idee war, ein System zu kreieren, welches Mitarbeitende im Service durch exakt das im jeweiligen Moment benötigte Wissen unterstützt. Aus dem initialen Produkt entwickelte sich nach und nach ein DMS bis hin zu einem Wissensmanagement, wie die Firma es nennt. Zurzeit sind etwa 15 Entwickelnde an dem Produkt beteiligt, welche in Zyklen von je zwei Wochen entwickeln, um dann wieder Feedback einzuholen. Die Idee dahinter ist, sich nicht zu lange zurückzuziehen, da der Markt und die Bedürfnisse der Kunden und Kundinnen sich schnell verändern können

²²¹ Vgl. Cunesoft Regulatory Compliance – Online (2016): Data mining, extraction, mapping and cleaning software PhlexDistiller, 3:21-3:42 [YouTube]

²²² Vgl. ebd., 3:49-4:00 [YouTube].

²²³ Vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch].

²²⁴ Serviceware (2020): Über Serviceware [elektronische Quelle].

²²⁵ Serviceware (2020): Serviceware Knowledge [elektronische Quelle].

und die Firma darauf möglichst zeitnah eingehen möchte. Alle sechs Wochen wird daher ein neues Release herausgegeben.²²⁶

Die Dokumente, welche in allen bekannten Formaten vorliegen können, werden in das System von Serviceware Knowledge hochgeladen und in einer Cloud gespeichert. Dadurch können die Dokumente ortsunabhängig abgerufen werden, auch die Nutzung auf mobilen Endgeräten wie Smartphones ist möglich. Im System werden die Dokumente mittels NLP-Techniken wie z.B. Stemming, Fuzzy-Suche und semantischem Parsen verarbeitet. Die Verarbeitung von Bildern in dem Sinne, dass z.B. das Bild einer Katze bei der Suche nach dem Stichwort „Katze“ angezeigt wird, ist derzeit noch nicht möglich, jedoch in Bearbeitung. Auch OCR in Scans wird momentan vom System noch nicht unterstützt, an einer Lösung wird jedoch ebenfalls gearbeitet.

Die Suche nach einem Wort oder einer Wortkette wird durch Elemente wie Boolesche Operatoren oder Asteriske unterstützt. Auf der linken Seite befindet sich dabei das Drill-Down-Menü, welches alle Dokumente enthält, rechts werden die Treffer angezeigt, welche im entsprechenden Dokument farbig markiert sind. Die KI-gesteuerte Suche ist derzeit für Deutsch und Englisch nutzbar.²²⁷

Das Produkt unterstützt nicht nur die Suche nach Wissen in Dokumenten, sondern auch die Dokumenterstellung. Hierbei werden ebenfalls KI-Techniken angewendet, welche eventuell bereits vorhandene, ähnliche Inhalte erkennen und Vorschläge zur Strukturierung machen. So wird das Produzieren von Dubletten verhindert und die Dokumente werden an einem sinnvollen Ort gespeichert.²²⁸

4.2 Produkttest Theum

Paul Caspers von Theum stellte eine Demoversion von Theum zur Verfügung, damit das Produkt einem Test unterzogen werden konnte. In der Demoversion sind nicht alle, aber die meisten Funktionen freigeschaltet, die in der Vollversion 3.7 vorhanden sind. Auf der Webseite von Theum und dem dazugehörigen YouTube-Kanal gibt es viele nützliche Informationen und Beispiele zur Funktionalität der Software. Auch können drei Wissensbanken, die zu Demonstrationszwecken zur Verfügung gestellt sind, von jedem Besuchenden der Webseite ausprobiert werden, um ein Gefühl für das System zu entwickeln.

²²⁶ Vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch]

²²⁷ Vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch]

²²⁸ Vgl. ebd. [pers. Gespräch].

Im Softwaretest sollten jedoch Funktionalitäten an ausgewählten Dokumenten getestet werden, weshalb es nötig war, eigene Dokumente hochzuladen und eine Wissensbank zu erstellen. Dies wurde durch die Einrichtung eines Zugangs zum System sowie eines Sharepoint-Verzeichnisses, in welchem die betreffenden Dokumente hochgeladen und anschließend mit dem System verbunden wurden, ermöglicht.

4.2.1 Methodik

Grundsätzlich sollte der Softwaretest zeigen, wie gut das System seine Aufgabe ausführt, nämlich die Beantwortung von Fragen bzw. das Auffinden von Wissen. Zu diesem Zweck wurden drei Bereiche getestet: unterschiedliche Textsorten, Formate und Sprachen. Zusätzlich wurden die Äquivalenzfunktion, das Stemming, welches in Theum bei Suchanfragen automatisch durchgeführt wird und die erweiterten Suchoperatoren „*“ als Platzhalter für mehrere Zeichen und „?“ als Platzhalter für genau ein Zeichen getestet.

Prinzipiell gibt es viele verschiedene Arten von Klassifikationen für Texte, einige fokussieren sich auf die Funktion, andere auf die Form oder den Kontext, in dem der Text entstanden ist. Für den Test der Textsorten wurde beschlossen, sich auf die Funktion der Texte zu fokussieren, da die Form für dieses Vorhaben nicht von Belang war. Wichtig war lediglich, dass alle Dokumente im gleichen Format vorlagen, um eine gleichwertige Analyse zu ermöglichen und dass die betreffenden Dokumente nicht zu wenig Text enthielten, damit genügend unterschiedliche Fragen gestellt werden konnten. Es wurden daher je 50 Dokumente im PDF-Format für die folgenden fünf Textsorten getestet:

- Belehrende (kognitive) Texte: wissenschaftliche Artikel
- Regelnde (normative) Texte: Gesetze
- Beschreibende (deskriptive) Texte: Bedienungsanleitungen
- Unterhaltende (trivial-narrative) sowie poetisch-deutende (ästhetisch-kreative) Texte: Prosa, Dramatik und Epik²²⁹

Ein vollständiges Quellenverzeichnis der Dokumente, die zum Test eingesetzt wurden, findet sich in Anhang 4. Es wurden insgesamt 15 Fragen pro Textsorte von unterschiedlicher Granularität an das System gestellt, um zu testen, ob sich diese auf die Relevanz der Antworten auswirkt. Die Granularität reichte dabei von der Eingabe eines einzelnen Wortes bis hin zu vollständig formulierten Fragen. Im Bereich der Textsorten ging es nicht um die Anzahl der gelieferten Antworten, da dies stark von den jeweiligen Inhalten abhängt. Weniger Antworten sind nicht automatisch gleichbedeutend mit einem schlechteren Funktionieren des Systems, sondern zeigen lediglich, dass die durchsuchten

²²⁹ Vgl. Brendel, A.; Brack-von Wins, I.; Schmitz, V. (1984): Textanalysen II, S. 23 ff.

Dokumente weniger Informationen zur jeweiligen Suchanfrage enthalten. Stattdessen war die Relevanz der gefundenen Antworten das zentrale Untersuchungskriterium.

Diese wurde in dem Sinne beurteilt, ob die gefundenen Antworten die Suchanfrage zufriedenstellend „beantwortet“ hatten. Beispielsweise lieferte die Suche nach „Gründe für Asylantrag“ insgesamt sieben Antworten, doch nur zwei davon waren relevant. Die Antwort „(5) Die Absätze 2 bis 4 finden keine Anwendung, wenn der Ausländer einen vor dem 6. August 2016 liegenden Termin zur Stellung des Asylantrags nach §23 Absatz 1 aus Gründen, die er zu vertreten hat, nicht wahrgenommen hat“²³⁰, ist beispielsweise nicht relevant, da sie keine Gründe für einen Asylantrag aufzeigt. Zwar ist durch die auftretenden Wörter „Gründe“ und „Asylantrag“ nachvollziehbar, weshalb das System die Antwort lieferte, trotzdem ist sie nicht relevant. Um die Relevanz für die Textsorten festzulegen, wurden alle Antworten auf jede Frage gezählt und durch die Anzahl der relevanten Antworten geteilt. So konnte nicht nur ermittelt werden, welche Textsorte die relevantesten und demnach die besten Antworten liefert, sondern auch, wie sich der Zusammenhang zwischen der Granularität der Suchanfrage und der Relevanz der Antworten darstellt. Eine detaillierte Auflistung aller Suchanfragen mit Anzahl der Antworten und ihrer Relevanz findet sich aufgrund der Tabellengröße ausschließlich im elektronischen Anhang.

Zum Test der verschiedenen Formate wurde ein identischer Dokumentinhalt jeweils als DOCX-, PDF- und HTML-Datei im System hochgeladen, wobei es sich um das Arbeitszeitgesetz der Bundesrepublik Deutschland handelte. Die OCR-Funktion ist in der Demoversion nicht verfügbar, weshalb es leider nicht möglich war, einen Scan zu testen. Auch bei diesem Test wurden 15 Fragen unterschiedlicher Granularität an das System gestellt, wobei in diesem Fall die Anzahl der Antworten ausschlaggebend war, um das Format zu ermitteln, welches sich am besten zur Durchsuchung eignet. Der Grund dafür ist, dass die Inhalte dieselben sind und daher nicht von einem Unterschied in der Relevanz auszugehen war. Die Anzahl der Antworten jedoch kann aufgrund der unterschiedlich starken Struktur der Dokumente variieren.

Um die Funktion des Systems auf verschiedene Sprachen zu testen, wurden mehrere Dokumente, die in identischer Form sowohl auf Deutsch als auch auf Englisch und Spanisch vorlagen, getestet. Erneut wurden je Sprache 15 identische Fragen gestellt, wobei die Suchoption des Systems zuvor in den Einstellungen für die jeweilige Sprache

²³⁰ §87c Absatz 5 Satz 1 AsylG

optimiert wurde. Hier wurde aufgrund der gleichen Inhalte wieder nur die Anzahl der relevanten Antworten betrachtet.

Zuletzt wurden einige der erweiterten Funktionen, die Theum anbietet, an mehreren zufällig ausgewählten Dokumenten getestet. Es ging hier lediglich darum zu testen, ob die Funktionen wie beschrieben funktionieren, oder ob es Probleme gibt. Getestet wurden die Äquivalenzfunktion, das Stemming, der Platzhalter für ein Zeichen (?) und eine beliebige Zahl von Zeichen (*).²³¹

4.2.2 Ergebnisse

Textsorten

Insgesamt zeigte die Relevanz der erhaltenen Antworten über die verschiedenen Textsorten hinweg ein zufriedenstellendes Ergebnis von 75%. Hinsichtlich der Relevanz der erhaltenen Antworten schnitten die Textsorten Bedienungsanleitung (82,6%) und wissenschaftliche Aufsätze (81,9%) am besten ab, gefolgt von der Gruppe der Prosa, Dramatik und Epik (71,7%) und schließlich den Gesetzen (62,9%).²³²

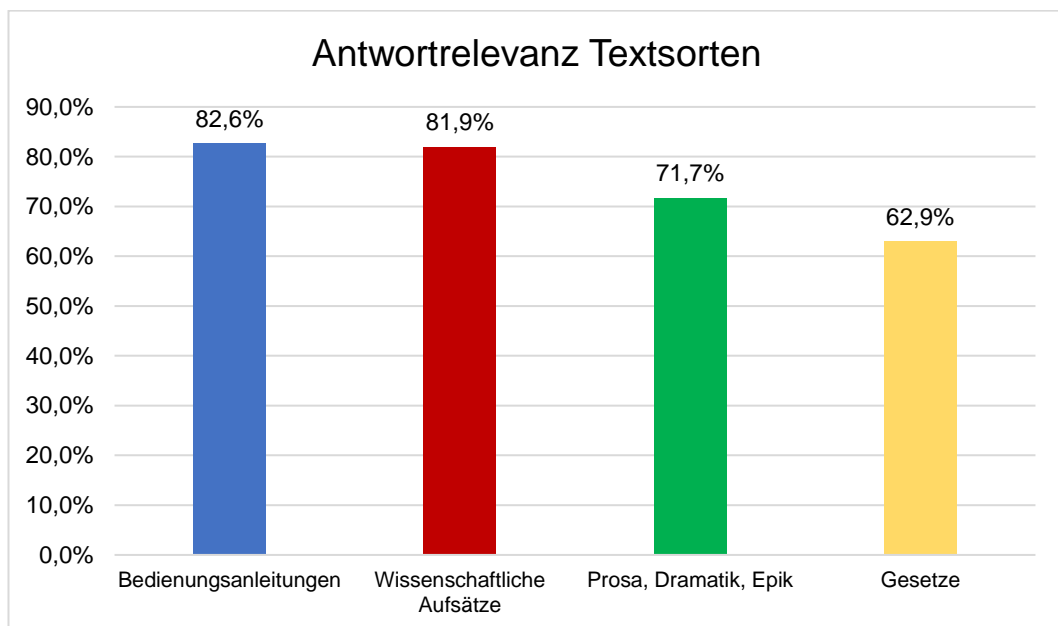


Abb. 20: Antwortrelevanz der unterschiedlichen Textsorten²³³

Die besseren Ergebnisse der Bedienungsanleitungen und der wissenschaftlichen Aufsätze gegenüber den Prosa-, Epik- und Dramatik-Texten sowie den Gesetzen lässt sich durch den Schreibstil und die gewählten Begriffe in den jeweiligen Gattungen erklären. Bedienungsanleitungen unterscheiden sich nicht stark voneinander, egal, um welches

²³¹ Vgl. Theum (o.J.f): Recherchieren mit Theum [elektronische Quelle].

²³² Vgl. Anhang 5.

²³³ Vgl. ebd.

Produkt es sich handelt. Stets geht es um die Funktionsweise des jeweiligen Geräts, es gibt Garantie- und Warnhinweise, Erklärungen zur korrekten Reinigung usw. Dabei tauchen in den unterschiedlichen Dokumenten immer wieder die gleichen oder zumindest sehr ähnlichen Begriffe auf. Suchanfragen wie z.B. „Patrone wechseln Drucker“ führen daher zu klaren Ergebnissen, die das System aus der jeweiligen Bedienungsanleitung herausfiltert. Möchte man wissen, wie die Druckerpatrone gewechselt wird, bieten sich nicht viel mehr Konstruktionen an, mit denen man suchen könnte. Das „wechseln“ könnte durch „austauschen“ ersetzt werden, aber an dieser Stelle erschöpfen sich die Synonyme bereits. Anders stellt sich die Suche bei Gesetzestexten dar. Hier zeigt sich deutlich, wie wichtig es ist, den Jargon der durchsuchten Texte zumindest ein wenig zu kennen. Gesetze erscheinen fachfremden Personen häufig sehr umständlich und teilweise sogar unverständlich formuliert zu sein. Folglich kann sich dieser Umstand in der Suche widerspiegeln. An folgendem Beispiel wird dies deutlich: Sucht man nach „Gründe für Asylantrag“ erwartet man als Antwort die im Asylgesetz aufgezählten Gründe, die einen Antrag auf Asyl rechtfertigen. Diese Antwort befindet sich im Asylgesetz, ist jedoch hier beschrieben als „Zuerkennung der Flüchtlingseigenschaft“²³⁴. Das System hat somit keine Chance, die richtige Antwort mit einer derartigen Anfrage herauszufiltern, zumal in §3 AsylG das Wort „Asylantrag“ nicht auftaucht. Es lässt sich die Vermutung anstellen, dass die Ergebnisse, wäre die Suche von einer Person mit juristischem Hintergrund durchgeführt worden, mehr relevante Antworten enthalten hätten.

Die wissenschaftlichen Aufsätze folgen ebenfalls meist einem gewissen Muster und sind allesamt in einem wissenschaftlichen Fachjargon geschrieben. Kennt man sich mit den Themen aus, die in den Dokumenten behandelt werden, so sind die Suchanfragen einfach zu formulieren. Hier braucht es lediglich etwas Gefühl für die wissenschaftliche Sprache und die Inhalte, wobei die Formulierungen der Suchanfragen auch für fachfremde Personen weniger schwierig als bei den Gesetzestexten sein dürften. In der Gruppe der Prosa, Dramatik und Epik ist die Suche nach Antworten ähnlich den Gesetzestexten teilweise schwierig zu formulieren, da sich die Texte sowohl inhaltlich als auch sprachlich stark voneinander unterscheiden. Sucht man z.B. nach „Teufel“, so werden die Treffer, die den Teufel als „Satan“ oder „Luzifer“ betiteln, nicht angezeigt (es sei denn, man legt Äquivalenzen fest, worauf später noch eingegangen wird). Welches Wort in einem Text verwendet wird, ist bei Dokumenten dieser Art schwer zu sagen. Ein zweiter Aspekt, der die geringere Antwortrelevanz dieser Textsorte erklärt, wird an folgendem Beispiel deutlich: Sucht man nach „Deichgraf Pferd“, sollen Antworten gefunden werden, die sich um das Pferd des Deichgrafen drehen. Neben diesen Antworten gibt es z.B.

²³⁴ §3 AsylG

jedoch auch eine, in welcher der Deichgraf erwähnt wird und es gleich darauf heißt: „Hauke stieg vom Pferde [...]“²³⁵. Dass es hier nicht um das Pferd des Deichgrafen, sondern um Haukes Pferd geht, kann das System nicht nachvollziehen. So lassen sich die verschiedenen Grade der Relevanz in den unterschiedlichen Textsorten erklären.

Hinsichtlich der Art der Anfragen wurden die relevantesten Antworten in den unspezifischen Anfragen, die nur aus einem Wort bestanden, geliefert (100,0%). Bei den spezifischeren Anfragen war kein allzu großer Unterschied in der Relevanz festzustellen. Anfragen, die aus zwei Wörtern bestanden, lieferten eine durchschnittliche Relevanz von 73,0%, Anfragen aus drei Wörtern eine Relevanz von 75,7% und Anfragen mit mehr als drei Wörtern eine Relevanz von 68,4%.²³⁶ Je mehr Wörter in einer Anfrage vorkommen, desto spezifischer muss die Antwort ausfallen, dementsprechend war hier eine etwas niedrigere Relevanz zu erwarten.

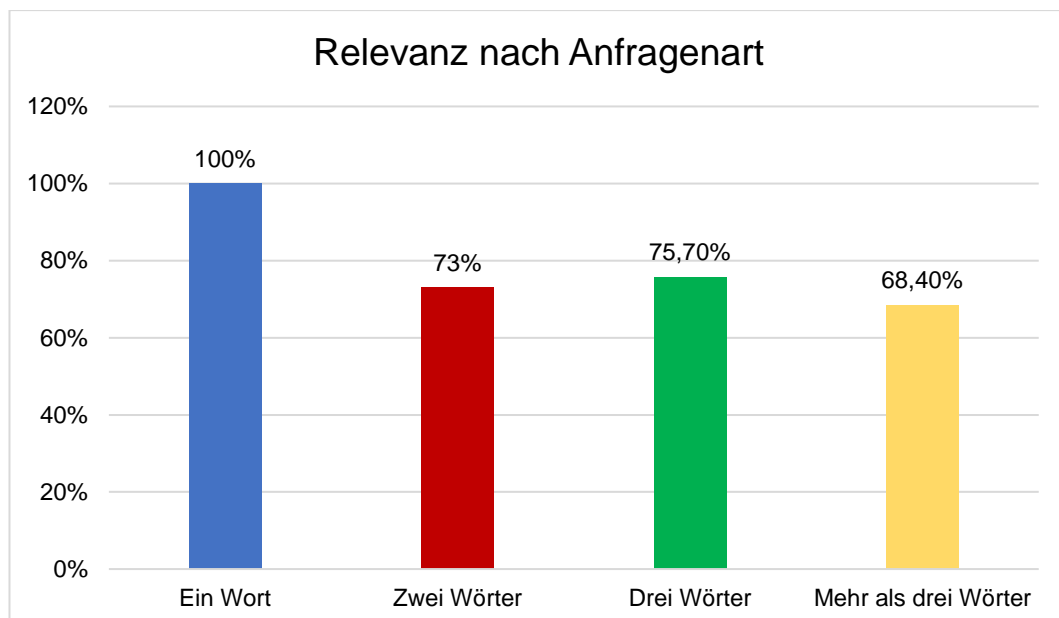


Abb. 21: Relevanz nach Anfragenart beim Textsortenvergleich²³⁷

Formate

Die Anfragen an die verschiedenen Formate zeigten ein deutliches Plus für das HTML-Format mit 22 relevanten Antworten, gefolgt von PDF und DOCX, beide mit jeweils neun relevanten Antworten.²³⁸

²³⁵ Storm, Theodor (2004): Der Schimmelreiter, S. 42.

²³⁶ Vgl. Anhang 5.

²³⁷ Vgl. ebd.

²³⁸ Vgl. ebd.

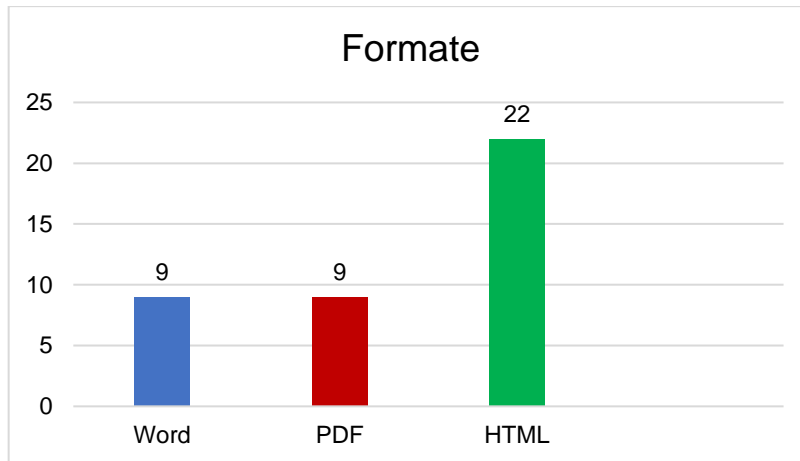


Abb. 22: Relevante Trefferanzahl in den unterschiedlichen Formaten²³⁹

Die deutlich bessere Trefferquote für HTML ist auf die feingliedrigere Struktur dieses Formats zurückzuführen. Wie bereits häufig im Lauf der Arbeit erwähnt wurde, ist eine gute Struktur ein ausschlaggebendes Kriterium für die Verarbeitbarkeit eines Dokuments. Dieser Umstand spiegelt sich im Test wider.

Sprachen

Der sprachliche Test zeigte einen deutlichen Vorteil für Englisch gegenüber Deutsch und Spanisch. Englisch lieferte mit 32 relevanten Antworten mehr als doppelt so viele Treffer wie Spanisch mit 14 und Deutsch mit nur zwölf relevanten Antworten.²⁴⁰

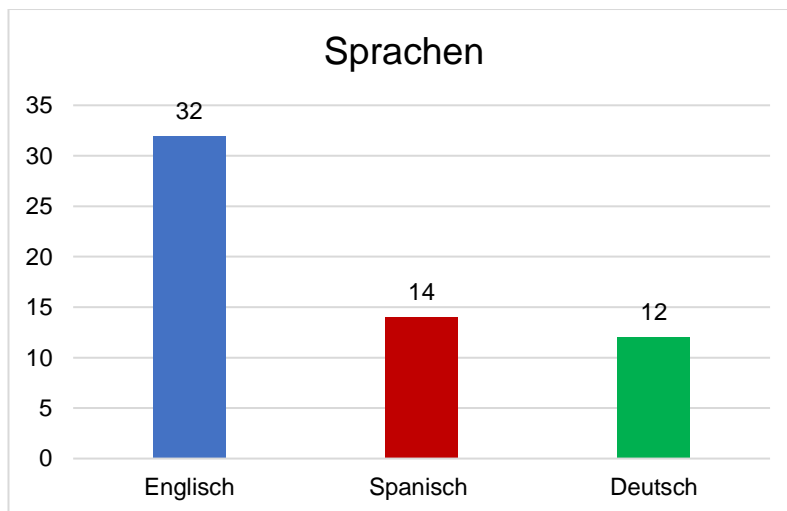


Abb. 23: Relevante Trefferanzahl in den unterschiedlichen Sprachen²⁴¹

Ein Teil der Ergebnisse ist dabei durchaus nachvollziehbar. Beispielsweise wurden bei der Suche nach „Universität“ im Deutschen und „universidad“ im Spanischen keine Treffer erzielt, bei der Suche nach „university“ im Englischen jedoch ein Treffer, was daran

²³⁹ Vgl. Anhang 5.

²⁴⁰ Vgl. ebd.

²⁴¹ Vgl. ebd.

lag, dass in einem Dokument von der „Syracuse University“ die Rede war, die im deutschen und spanischen Text aufgrund des Eigennamens nicht als „Syracuse Universität“ bzw. „Universidad de Syracuse“ übersetzt war.²⁴² Andere Ergebnisse waren jedoch nicht nachvollziehbar. Die deutsche Suche „Wie reinigt man den Grill“ ergab keine Treffer, ebenso wenig wie die spanische Suche nach „Cómo limpiar la parrilla“. Die englische Suche nach „How to clean the grill“ ergab jedoch sieben relevante Treffer.²⁴³ Ein Zusammenhang aufgrund von Stoppwörtern und Stemming kann hier vermutet werden. Es ist nicht ganz deutlich, wie Stoppwörter zur Anwendung kommen. Beispielsweise erzielte die Suche nach „Was war ihr letztes Buch“ bzw. „Cúal fue su ultimo libro“ keine Treffer, obwohl die Wortketten „Ihr letztes Buch“ bzw. „Su ultimo libro“ exakt so in den Texten vorkommen. Im inhaltlich identischen englischen Dokument erzielte „Which was her last book“ jedoch einen relevanten Treffer, beginnend mit „Her last book“.²⁴⁴ Es bleibt unklar, aufgrund welcher Regeln hier nur im Englischen Treffer erzielt wurden.

Allgemein liegt die Vermutung nahe, dass die englische Sprache mehr Treffer aufgrund ihrer Grammatik und Lexikologie erzeugt. So erzielt beispielsweise die Suche nach „Descalc the kettle“ vier Treffer, die Suche nach „Descalcificar la jarra“ jedoch keine Treffer.²⁴⁵ Betrachtet man die Suchergebnisse genauer, fällt auf, dass dies oft am Stemming liegt. Da die Grundform des Verbs im Englischen (descalc) dieselbe ist wie der Imperativ (descalc), werden hier u.a. Aufforderungen zur Entkalkung eines Wasserkochers gefunden. Im Spanischen unterscheiden sich jedoch Grundform (descalcificar) und Imperativ (im vorliegenden Fall decalcifique). Durch das Stemming sollten Probleme dieser Art behoben werden, was aber nicht immer der Fall ist. Ein weiterer vermuteter Grund für eine höhere Trefferanzahl im Englischen sind deutsche Komposita, die in ihrer zusammengesetzten Art und Weise im Englischen nicht existieren. Aus den zwei Wörtern „aspiration speed“ wird beispielsweise im Deutschen ein Wort, „Ansaugstärke“. Das System kann mit dem Stichwort „ansaugen“ die „Ansaugstärke“ nicht finden, mit dem Stichwort „aspire“ jedoch „aspiration speed“, vermutlich, weil „aspiration“ sehr nahe an „aspire“ liegt, „ansaugen“ jedoch nicht ganz so nah an „Ansaugstärke“. Um die Antworten dennoch zu finden, empfiehlt sich daher im Deutschen auf jeden Fall eine Suche mit Platzhaltern (? oder *).

Zudem wurde festgestellt, dass die spanischen Akute zu Problemen bei der Suche führen können. So liefert die Suche nach „Advertencia niños cortacésped“ keine Treffer,

²⁴² Vgl. Anhang 5.

²⁴³ Vgl. ebd.

²⁴⁴ Vgl. ebd.

²⁴⁵ Vgl. ebd.

schreibt man jedoch das Wort „Advertencia“ ohne Akut auf dem „i“, erhält man einen Treffer. Grund dafür ist, dass das Wort im vorliegenden Dokument in Großbuchstaben geschrieben ist und daher keine Akute enthält. Trotzdem müsste das System in der Lage sein, das entsprechende Wort aufgrund des Stemming zu erkennen.

Erweiterte Suchfunktionen

Die Äquivalenzfunktion, die Theum anbietet, funktioniert sehr gut und ist einfach zu handhaben. Die Wörter müssen lediglich in der entsprechenden Liste durch Kommata getrennt angegeben werden. So konnte z.B. durch das Hinzufügen von „Brandgefahr“ und „Stromschlaggefahr“ zum Wort „Gefahr“ erreicht werden, dass auch Dokumente, die Brandgefahr oder Stromschlaggefahr enthielten, bei der Suche nach „Gefahr“ gefunden wurden. Dies ist nicht der Fall, wenn die Wörter nicht als Äquivalenzen eingespeichert sind. Das Stemming wurde bereits im Zusammenhang mit den verschiedenen Sprachen angesprochen. Teilweise funktioniert es gut, manchmal aber auch nicht. Wird z.B. das Verb „durchwühlen“ gesucht, werden andere Treffer gefunden, als wenn man nach „durchwühlt“ sucht, obwohl sich aufgrund des Stemming dieselben Treffer ergeben sollten.

Die Platzhalterfunktionen „*“ und „?“ funktionieren problemlos. So findet z.B. die Suche nach „u?a“ Wörter wie „USA“ oder „una“. Die Suche nach „Entkalkung*“ findet Wörter wie „Entkalkungsmittel“, „Entkalkungsvorgang“, „Entkalkungsprogramm“ etc.

4.2.3 Diskussion

Die für den Softwaretest angewandte Methodik war gut ausführbar. Mit noch mehr Dokumenten hätte sich eventuell noch besser testen lassen, da es auch vorkam, dass manche Anfragen keinerlei Ergebnisse erzielten. Jedoch ist es in Bezug auf den Test der Formate und Sprachen schwierig, Dokumente zu finden, die inhaltlich vollständig identisch sind, sich jedoch in Format oder Sprache unterscheiden. Zwar können einige Formate konvertiert werden, jedoch ist es beispielsweise nicht sinnvoll, ein DOCX-Dokument aus einem PDF-Dokument zu generieren. Von Vorteil im Sprachtest waren die Bedienungsanleitungen, da diese stets in mehreren Sprachen vorliegen. Leider konnten nicht mehr als die vorgestellten Sprachen getestet werden, da es einer gewissen Kenntnis der jeweiligen Sprache bedarf, um die Suchanfragen richtig zu formulieren und die Ergebnisse auf Relevanz zu überprüfen.

Allgemein lässt sich sagen, dass Theum für Unternehmen, die ihre zahlreichen Dokumente mit dem System verknüpfen, sicherlich von Vorteil ist. Dokumente können zwar

mit Volltextsuchen durchsucht werden, dafür muss man jedoch jedes einzelne Dokument öffnen und die Suche darin durchführen. Diese Art des Vorgehens kostet enorm viel Mühe und Zeit und somit auch Geld. Eine andere Möglichkeit ist, eine Suchanfrage direkt an dem Ort, an dem die Dokumente abgelegt sind, zu stellen, beispielsweise in einem ECM. Doch auch diese Methode ist sehr zeitaufwendig und bringt oft keine zufriedenstellenden Ergebnisse. Zudem können in beiden Methoden keine speziellen Suchoperatoren eingesetzt werden, es finden keine syntaktischen oder semantischen Analysen statt und es werden keine Synonyme erkannt. Außerdem kann es vorkommen, dass es Formate gibt, die sich auf dem entsprechenden Gerät nicht öffnen lassen, beispielsweise aufgrund eines anderen Betriebssystems. All das sind die Punkte, an denen Theum ansetzt und die eine Erleichterung sowie eine Zeit- und Geldersparnis für die nutzenden Firmen darstellen.

Kritik lässt sich äußern an der Stammformreduktion, die manchmal nicht ganz eindeutig arbeitet und es daher vorkommen kann, dass manche Antworten nicht gefunden werden. Wünschenswert für die Zukunft wäre eine verbesserte semantische Analyse, um auch durch Suchanfragen, die vom Sinn her identisch, jedoch anders als die Dokumenttexte formuliert sind, mehr Treffer zu erhalten. Der Softwaretest hat jedoch deutlich gemacht, dass eine erfolgreiche Suche nach Wissen in digitalen Dokumenten stets nicht nur von der Software abhängt, welche die Dokumente verarbeitet, sondern auch von den Dokumentenerstellenden. Dokumente mit wenig Struktur sind vom System schwerer zu verarbeiten. Es ist daher äußerst wichtig, dass Firmen nicht nur auf Theum als Wissenssystem vertrauen, sondern sich auch darum kümmern, dass ihre Dokumente eine Struktur aufweisen und in Formaten gespeichert sind, die vom System gut verarbeitet werden können. Nur wenn Dokumente auf diese Weise erstellt und abgespeichert werden, kann die Software sie gut verarbeiten.

4.3 Experteninterviews

Um mehr über die in Kapitel 4.1 vorgestellten Produkte und die Themen Dokument, KI allgemein und KI-basierte Wissenstechnologien zu erfahren, wurden drei Interviews mit fachkundigen Personen durchgeführt, die in den jeweiligen Firmen, welche die Produkte vertreiben, tätig sind: Paul Caspers, Vorsitzender der Firma Theum AG, Daniel Koppers, Gründer der Cunesoft GmbH und seit Kurzem Head Of Research And Development bei Phlexglobal und Carolin Hischer, Product Owner AI bei Serviceware.

Interviews mit Experten und Expertinnen dienen zusammengefasst stets dazu, bestimmtes Wissen zu erheben, da zwischen der Person, welche das Interview führt und den zu Interviewenden ein Wissensunterschied vorliegt. Dieser Wissensunterschied ist sowohl auf privilegierte Zugänge zu bestimmten Informationen als auch auf persönliche Erfahrungen und subjektive Sichtweisen, Bewertungen und Deutungen der fachkundigen Personen zurückzuführen.²⁴⁶ Bei den durchgeführten Interviews handelte es sich um qualitative, explorative Gespräche, welche eine Ergänzung zu den bereits gesammelten Informationen und eine Erweiterung derselben durch Wissen darstellen, welches nur durch direkte Kommunikation mit den entsprechenden Personen erhoben werden kann. Es ging dabei vor allem um die Erhebung von technischem Wissen, sprich Daten, Fakten und Tatsachen, die nur über die Sachverständigen zu erlangen sind sowie Deutungswissen, welches auf die subjektiven Ansichten und Erlebnisse der Befragten fokussiert ist, um herauszufinden, welche verschiedenen Vorstellungen und Interpretationen im Forschungsfeld vorkommen.²⁴⁷

4.3.1 Methodik

Zunächst wurden sachkundige Personen gesucht, die in ihrem Berufsalltag mit KI-basierten Wissenstechnologien arbeiten. Nachdem mehrere Anfragen an unterschiedliche Firmen gestellt wurden, welche in diesem Bereich tätig sind, konnten letztendlich mit drei Personen aus drei verschiedenen Organisationen Termine für Interviews arrangiert werden.

Als Hilfsmittel wurde für die qualitativen Interviews je ein Leitfaden für die Interviewende und ein Fragenkatalog für die sachkundigen Personen entworfen, welchen diese im Voraus erhielten, um sich vorbereiten zu können. Beide Dokumente können in Anhang 1 und 2 eingesehen werden. Der Leitfaden für die Interviewende enthielt zusätzlich zu den zu stellenden Fragen Anmerkungen darüber, welche Informationen zu Beginn und Ende des Interviews außerdem zu nennen waren. Dazu zählten der Dank für die Bereitschaft zum Interview, eine kurze Vorstellung der Person und des Themas, das Einholen des Einverständnisses zum Aufzeichnen des Interviews, der zeitliche Rahmen sowie eine Einstiegsfrage bezüglich der Position des jeweiligen Interviewten in der Firma.

Die Fragen des Leitfadens wurden auf Basis der Forschungsfragen formuliert, welche die Arbeit beantworten soll. Zusätzlich wurden technischen Fragen zu den Produkten

²⁴⁶ Vgl. Bogner, A.; Littig, B.; Menz, W. (2014): Interviews mit Experten, S. 17 f.

²⁴⁷ Vgl. ebd. S. 17 f. und S. 24.

formuliert, die durch andere Arten der Recherche in Ermangelung von vorhandenen Informationen nicht in Erfahrung gebracht werden konnten. Nachdem alle relevant erscheinenden Fragen gesammelt waren, wurden sie zu Themenblöcken zusammengefasst. So entstanden folgende vier Interviewblöcke mit insgesamt 21 Fragen:

- Die Begriffe „Dokument“ und „Daten“ (fünf Fragen)
- KI (drei Fragen)
- Produkt (neun Fragen)
- Zukunft (vier Fragen)

Anschließend wurden die Fragen unterteilt in Schlüsselfragen, die auf jeden Fall gestellt werden mussten und ergänzende Fragen, die je nach Zeit und Gesprächssituation gestellt werden konnten. Die Schlüsselfragen wurden auf dem Leitfaden der Interviewerin entsprechend farblich gekennzeichnet, jedoch nicht auf jenem der Experten und der Expertin, um deren Aufmerksamkeit nicht zu stark auf diese zu fokussieren. Da beim qualitativen Interview nicht alle Fragen exakt gleich gestellt werden müssen, wie es beispielsweise bei einem Fragebogen zur quantitativen Erhebung der Fall ist, konnten die ergänzenden Fragen je nach Gesprächssituation und zeitlichem Rahmen gestellt oder nicht gestellt oder an den Gesprächsverlauf angepasst werden.

Die Interviews wurden allesamt als Online-Gespräche mit verschiedenen Tools wie z.B. Microsoft Teams durchgeführt, je nachdem, was in der jeweiligen Firma verwendet wurde. Die Aufnahme der Audiospuren wurde nach der Bitte um Erlaubnis per Smartphone durchgeführt. Nach Beendigung der Interviews wurden die Aufnahmen mittels „IBM Watson Speech to Text Service“²⁴⁸ transkribiert und anschließend händisch verbessert und korrigiert. Die Transkriptionen der Interviews sind im Anhang zu finden. Im Anschluss an die Transkription wurden die Inhalte der Interviews zu Vergleichszwecken knapp zusammengefasst und in Form einer Excel-Tabelle gespeichert. Diese kann aus Gründen der Formatierung ausschließlich im elektronischen Anhang 6 eingesehen werden. Fragen, die auf bloße Fakten rund um die Produkte abzielten, wie beispielsweise die Anzahl an Entwickelnden für das jeweilige Produkt, sind bereits in die Produktvorstellungen in Kapitel 4.1 eingeflossen. Die Ergebnisse aller übrigen Fragen (insgesamt 13), welche auf die Erfahrungen und subjektiven Meinungen der sachkundigen Personen eingehen, werden im Folgenden vorgestellt.

²⁴⁸ IBM Watson (o.J.): Speech to Text

4.3.2 Ergebnisse

Block 1: Die Begriffe „Dokument“ und „Daten“

Es gibt viele verschiedene Interpretationen des Begriffs "Dokument". Beispielsweise betrachtete der indische Mathematiker und Bibliothekar S.R. Ranganathan sie als Gedanken auf Papier oder einem anderen Material, das physische Handhabung, Transport durch den Raum und Konservierung in der Zeit erlaubt ²⁴⁹ , die französische Bibliothekarin Suzanne Briet hingegen hielt sogar Tiere im Zoo, solange sie katalogisiert seien, für Dokumente ²⁵⁰ Was ist Ihre persönliche Meinung zu diesem Thema?
Machen Sie bei Ihrer Definition einen Unterschied zwischen analogem und digitalem Dokument? Wenn ja, welchen?
Wie differenzieren Sie den Begriff des Dokuments von jenem der Daten? Ist eine klare Differenzierung möglich?
Ab wann ist für Sie ein Dokument ein Dokument? Sind beispielsweise Tabellen mit Messwerten für Sie Dokumente? Wo ziehen Sie die Grenze?
Wenn Sie einige Jahre zurückdenken: hat sich Ihre Vorstellung von dem, was ein Dokument ist, verändert? In welcher Weise?

Tabelle 1: Auszug aus Interview Block 1²⁵¹

Alle drei Interviewten stimmen weder der als Beispiel dienenden Aussage von S.R. Ranganathan noch jener von Suzanne Briet voll zu. Herr Koppers pflichtet Ranganathan jedoch teilweise bei und definiert das Dokument als alles, was Information beinhaltet, die transportiert, weitergegeben und geteilt werden kann.²⁵² Für Frau Hischer ist ein Dokument eine Ansammlung von Daten, die einen Wissensgehalt transferieren. Sie betont als Kriterien vor allem, dass ein Dokument statisch sein und einen gewissen Informationsgehalt enthalten muss.²⁵³ Mit ihrer Definition liegt sie derjenigen von Herrn Koppers also sehr nahe. Die Statik ist für Frau Hischer ein ausschlaggebendes Element, ganz im Sinne von Pédauque, der ein dynamisches Objekt wie z.B. einen Livestream ebenfalls nicht als Dokument ansieht und darauf hinweist, dass Dokumente nicht nur in einem bestimmten Moment in der Zeit auftauchen dürfen, sondern für einen längeren Zeitraum zugänglich sein müssen.²⁵⁴ Neben Textdokumenten können auch andere Formate, wie

²⁴⁹ Vgl. Ranganathan, S. R. (1963): Documentation and its facets, S. 24.

²⁵⁰ Vgl. Briet, Suzanne (1951): Qu'est-ce que la documentation, S. 7.

²⁵¹ Tabelle entnommen aus Anhang 2.

²⁵² Vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁵³ Vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁵⁴ Vgl. Pédauque, R. T. (2003): Document: Form, Sign and Medium, As Reformulated for Electronic Documents, S. 17 f. [elektronische Quelle].

z.B. Bilder, für Frau Hischer als Dokument gelten, solange sie die genannten Kriterien erfüllen. Ein Dokument hat für sie nichts mit einer äußeren Form zu tun.²⁵⁵ Herr Caspers bezeichnet den Menschen als narratives Wesen, welches seine „Erzählungen“ früher auf Materialien wie Stein oder Papyrus, heute v.a. in Form von digitalen Formaten festhält. Er betrachtet Formate wie Bilder oder Videos als „Grenzbereich“, sie seien ebenfalls Dokumente, jedoch könnten ihre Inhalte nicht auf dieselbe Art wie Textdokumente erschlossen werden, weshalb er vornehmlich Textdokumente als Dokumente bezeichnet.²⁵⁶ Damit ähnelt sein Ansatz ebenfalls jenem von Pédaque, der die Meinung vertritt, dass die meisten Dokumente aus geschriebener oder gesprochener Sprache bestehen.²⁵⁷

Frau Hischer sieht in ihrer Definition keinen Unterschied zwischen analogem und digitalem Dokument, was sie auch darauf zurückführt, dass sie bereits mit digitalen Dokumenten aufgewachsen ist und daher nie zwischen den beiden unterschieden hat.²⁵⁸ Herr Koppers hingegen vertritt die Auffassung, dass ein großer Unterschied zwischen den beiden Dokumenttypen besteht. Das digitale Dokument ist für ihn eine Quelle von Informationen, die durch bestimmte Techniken auf verschiedenste Weise genutzt werden können. Das analoge Dokument hingegen bezeichnet er als „dumm“, da es technisch gesehen in dem Sinne wertlos ist, dass es nicht ohne Weiteres verarbeitbar ist.²⁵⁹ Ganz so kritisch äußert sich Paul Caspers nicht gegenüber dem analogen Dokument, er verweist auf Techniken wie OCR, jedoch bemängelt auch er die Möglichkeiten, die analoge Dokumente hinsichtlich technischer Verarbeitung im Vergleich zum digitalen Dokument bieten.²⁶⁰

Die Gegenüberstellung von Dokument und Daten sehen die drei Fachkundigen relativ ähnlich. Paul Caspers sieht einen fließenden Übergang zwischen Daten und Dokumenten.²⁶¹ Diese Aussage ähnelt der von Jakob Voß, der die Grenzen zwischen Daten und Dokumenten als verwischt ansieht.²⁶² Daniel Koppers bezeichnet Dokumente als „Daten

²⁵⁵ Vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁵⁶ Vgl. Caspers, P. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁵⁷ Pédaque, R. T. (2003): Document: Form, Sign and Medium, As Reformulated for Electronic Documents, S. 12 [elektronische Quelle].

²⁵⁸ Vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁵⁹ Vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁶⁰ Vgl. Caspers, P. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁶¹ Vgl. ebd.

²⁶² Voß, J. (2013): Was sind eigentlich Daten?, S. 7 [elektronische Quelle].

mit Klebematerial“, wobei er mit diesem Klebematerial z.B. Stoppwörter meint, all das, was der Mensch braucht, die Maschine jedoch nicht. Auch Carolin Hischer beschreibt Dokumente als aus Daten bestehend. Eine Excel-Tabelle mit numerischen Werten sehen beide beispielsweise als Dokument an, die Werte allein jedoch als Daten.²⁶³ Diese Aussagen korrespondieren mit jener von Buckland, der anmerkt, dass längere Textdatensätze meistens als Dokumente bezeichnet werden, kurze numerische Datensätze hingegen als Daten.²⁶⁴

Ebenfalls einig sind sich die Interviewten darüber, dass sich ihr Grundverständnis vom Dokument über die Jahre hinweg nicht verändert hat. Jedoch verweisen Herr Koppers und Frau Hischer beide darauf, dass sich ihr Verständnis vom Dokument durch den beruflichen Werdegang differenziert hat.²⁶⁵ Herr Caspers spricht davon, dass sich zwar die Rahmenbedingungen und Technologien stark verändert haben, die Grundvorstellung vom Dokument jedoch nicht.²⁶⁶

Block 2: KI

Wie stehen Sie persönlich ganz allgemein zu KI (nicht in Zusammenhang mit Ihrem Produkt)? Sehen Sie darin eine große Chance, ein großes Risiko oder beides?
Wie schätzen Sie die Bemühungen in Deutschland ein, um die Entwicklung von KI voranzutreiben (beispielsweise im Vergleich zu den USA oder China)? Wird aus Ihrer Sicht in Deutschland in diesem Kontext genug getan?

Tabelle 2: Auszug aus Interview Block 2²⁶⁷

Paul Caspers sieht in KI sowohl Chance als auch Risiko. Chancen sieht er vor allem in dem Sinn, dass KI dem Menschen zukünftig immer mehr als lästig geltende Arbeiten abnehmen könnte, z.B. im Fertigungsbereich. Gleichzeitig hält er es jedoch für wichtig, dass die Gesellschaft für sich definiert, was ihr wichtig ist, denn pure Arbeitslosigkeit, selbst wenn es eine Art bedingungsloses Grundeinkommen gäbe, sieht er nicht als gesellschaftliches Ziel an, das sich positiv auf die Bevölkerung auswirkt. Zudem weist er auf Bereiche hin, in denen KI auf negative Weise Einsatz findet, wie beispielsweise in

²⁶³ Vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁶⁴ Vgl. Buckland, M. (2018): Document Theory, S. 428 [elektronische Quelle].

²⁶⁵ Vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch]; vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁶⁶ Vgl. Caspers, P. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁶⁷ Tabelle entnommen aus Anhang 2.

der Waffenbranche.²⁶⁸ Daniel Koppers und Carolin Hischer sind dem Thema gegenüber sehr optimistisch eingestellt. Beide sind der Meinung, dass KI auf dem momentanen Stand kein Bereich ist, der zu Ängsten führen sollte. Frau Hischer macht deutlich, dass Verantwortung zur KI dazugehört, sie jedoch kein Risiko darin sieht.²⁶⁹ Herr Koppers betont, dass die heutigen Datenmengen ein Ausmaß an Komplexität erreicht haben, das manuell nicht mehr verarbeitbar ist und die technische Verarbeitung daher unbedingt gefördert werden muss.²⁷⁰

Einig sind sich Herr Koppers und Frau Hischer auch bei der Einschätzung der Bemühungen in Deutschland im Bereich KI. Herr Koppers beschreibt Deutschland als ein Land, in dem sehr gerne reguliert wird und fürchtet, dass derartige Regelungen Türen in der Entwicklung der KI verschließen werden. Er benennt zwar einige deutsche Einrichtungen wie z.B. das Fraunhofer Institut, die gute Arbeit im KI-Bereich leisten, kritisiert jedoch die zu geringe Nutzung von KI-Innovationen durch deutsche Unternehmen.²⁷¹ Frau Hischer sieht Deutschland im KI-Bereich als „in die Bequemlichkeit des Mittelstands“ abgerutscht und spricht von einem zum Teil verloren gegangenen Innovationshunger.²⁷² Beide sehen also im Bereich der KI Potenziale für Deutschland, die jedoch nicht ausgeschöpft werden.

Block 3: Produkt

Was glauben Sie, wie sich die Nutzungshäufigkeit von Produkten wie Ihrem in der Zukunft entwickeln wird, vor allem im Rahmen von Dokumentenmanagement?
Welche anderen „Einsatzorte“ (außerhalb von Dokumentenmanagement, vornehmlich in Unternehmen) können Sie sich für Produkte dieser Art vorstellen?

Tabelle 3: Auszug aus Interview Block 3²⁷³

Die drei Befragten sind sich einig, dass sich die Nutzung von KI-basierten Wissenstechnologien in der Zukunft erhöhen wird. Herr Caspers und Frau Hischer sehen es als künftig immer wichtiger an, dass das richtige Wissen zur richtigen Zeit am richtigen Ort für

²⁶⁸ Vgl. Caspers, P. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁶⁹ Vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁷⁰ Vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁷¹ Vgl. ebd.

²⁷² Vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁷³ Tabelle entnommen aus Anhang 2.

die Mitarbeitenden bereitsteht, und zwar in einer benutzungsfreundlichen Form.²⁷⁴ Auch Herr Koppers hält Systeme dieser Art zur Verbesserung von Datenqualität und Datenvollständigkeit in Zukunft für unabdingbar.²⁷⁵

Einen Einsatz derartiger Systeme im Privatbereich können sich die Befragten aufgrund der hohen Kosten und Komplexität nicht vorstellen. Paul Caspers könnte sich jedoch heutige Internetsuchmaschinen anstatt mit Dokument-Antworten gut mit Exzerpten, wie sie in Theum ausgegeben werden, vorstellen. Zudem beschreibt er Szenarien, in denen Geräte Fehlercodes melden, die dann z.B. per Smartphone ausgelesen werden und sofort das Wissen bereitstellen, um den Fehler zu beheben.²⁷⁶ Daniel Koppers verweist auf die allgemeine Prozessautomatisierung, in seinem Bereich z.B. die automatische Auswertung von Nebenwirkungsmeldungen oder entscheidungsunterstützende Systeme, die auf Basis von Dokumenten arbeiten.²⁷⁷

Block 4: Zukunft

Sehen Sie Möglichkeiten, Ihr Produkt noch effizienter zu machen, beispielsweise anhand von Technologien, die bisher noch gar nicht verfügbar sind? Wenn ja, in welcher Weise?
Was glauben Sie, wie wird das "Dokument der Zukunft" aussehen?
Werden in Zukunft Ihrer Meinung nach gezielte KI-Gesetze nötig sein? Beispielsweise eine Art "hippokratischer Eid" für Programmierer? Warum?
Halten Sie es für möglich, dass der Mensch jemals fähig sein wird, eine Intelligenz zu erschaffen, die der seinen ebenbürtig ist? ("A cyclone can raze a city, yet not even open a letter or untie the knot in this piece of string" ²⁷⁸ , Zitat als Beispiel für das "Wesen" einer KI). Wie begründen Sie Ihre Meinung dazu?

Tabelle 4: Auszug aus Interview Block 4²⁷⁹

Als verbesserungswürdig empfindet Paul Caspers an Theum die Robustheit des Systems, er würde Softwarefehler gerne noch stärker reduzieren.²⁸⁰ Daniel Koppers sieht in

²⁷⁴ Vgl. Caspers, P. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch]; vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3. siehe Anhang 3.

²⁷⁵ Vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁷⁶ Vgl. Caspers, P. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁷⁷ Vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁷⁸ Valéry, P. (1970): Collected Works of Paul Valéry, Volume 14. Analects, S. 132.

²⁷⁹ Tabelle entnommen aus Anhang 2.

²⁸⁰ Vgl. Caspers, P. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

mehr Rechenleistung einen erfolgsbringenden Faktor, um bessere Extraktions- und Klassifizierungsqualität zu erzeugen und um Experimente schneller auszuführen und größere Modelle bauen zu können.²⁸¹ Carolin Hischer hofft in Zukunft auf noch effektivere Automatismen, die Wissen in strukturierter Form auch aus sehr unstrukturierten Quellen extrahieren können.²⁸²

Herr Caspers sieht keine großen Veränderungen im Dokument der Zukunft auf Seiten derjenigen, die ein Dokument erzeugen, jedoch beim Suchenden, der Informationen recherchiert. Er ist der Meinung, dass die benötigten Informationen aus Dokumenten in Zukunft schneller und auf effektivere Art und Weise bereitgestellt werden können.²⁸³ Frau Hischer vertritt die Auffassung, dass das Dokument der Zukunft aus „kleinen Häppchen“ bestehen wird, sie vergleicht es mit einem Puzzle. Statt den gesamten Dokumenten werden den Suchenden lediglich die Ausschnitte präsentiert, die relevant sind. In der Zukunft sieht sie das Dokument als modularer und flexibler, als weniger starr an.²⁸⁴ Herr Koppers sieht die Zeit von Tools wie Microsoft Word oder Adobe Acrobat Reader als abgelaufen an. Er hält es für sehr wichtig, dass Daten in Zukunft in strukturierterer Form gespeichert werden, um diese im Hintergrund zu transformieren und den maschinengestützten Konsum durch Menschen zu ermöglichen.²⁸⁵

Eventuellen KI-Gesetzen zum jetzigen Zeitpunkt stehen die drei Befragten kritisch gegenüber. Herr Koppers bezeichnet diese als schädlich und hindernd, Frau Hischer als innovationshemmend, verlangsamend und zu Ineffizienz führend. Herr Koppers kann sich in der fernen Zukunft eine Art Aufsichtsbehörde für Programmierende vorstellen, merkt jedoch an, dass es in vielen Bereichen, die ebenfalls eventuelle Risiken bergen, keine Gesetze oder Ähnliches gibt. Frau Hischer vertritt die Meinung, dass es in vielen Bereichen sogenannte Work Ethics gibt, diese jedoch nicht immer von jedem eingehalten werden.²⁸⁶ Sie glaubt, dass im KI-Bereich wie auch in allen anderen Bereichen derartige Beschränkungen in den Fällen versagen würden, in denen Menschen sich schlicht

²⁸¹ Vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁸² Vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁸³ Vgl. Caspers, P. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁸⁴ Vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁸⁵ Vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁸⁶ Vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch]; vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

nicht an die bestehenden Regeln halten möchten. Herr Caspers verweist darauf, dass es für Gesetze zu früh ist und merkt erneut an, dass die Gesellschaft zunächst definieren sollte, welche Ziele durch KI erreicht werden sollen, bevor über die Umsetzung von Gesetzen nachgedacht wird.²⁸⁷

Einig sind sich die Befragten darin, dass es zumindest in den nächsten Jahren keine starke KI geben wird, die der Intelligenz des Menschen ebenbürtig ist. Frau Hischer und Herr Caspers möchten dies jedoch nicht komplett ausschließen, auch wenn sie momentan noch keine Anzeichen dafür sehen.²⁸⁸ Carolin Hischer merkt an, dass der Mensch sich gerne als Krone der Schöpfung betrachtet und daher abgeneigt gegenüber Szenarien ist, die ihm diese Position streitig machen könnten. Herr Koppers hingegen ist sicher, dass die Erschaffung einer starken KI unmöglich ist. Er ist der Meinung, dass das Bewusstsein die Summe aller Erfahrungen ist, die ein Mensch im Lauf seines Lebens sammelt und daher nicht künstlich erzeugt werden kann.²⁸⁹

4.3.3 Diskussion

Fast alle Fragen des Leitfadens konnten in jedem Interview gestellt werden, nur wenige optionale Fragen mussten aus Zeitgründen ausgelassen werden. Die Interviews waren gekennzeichnet durch das Engagement der Befragten, die alle sehr gut vorbereitet und zuvorkommend waren und sich Zeit nahmen, um auch schwierige Sachverhalte zu erklären.

Durch die Interviews konnten sehr viele zusätzliche Informationen zu den Produkten gewonnen werden, die auf den Webseiten oder in anderen Quellen der Firmen, wie z.B. YouTube-Kanälen, nicht einsehbar waren. So konnte wertvolles Insiderwissen in Erfahrung gebracht werden, das zum besseren Verständnis der Technik, die hinter den Produkten steckt, beitrug. Außerdem konnten Aussagen aus der Fachliteratur anhand der Meinungen und Erfahrungen der Sachkundigen aus der Praxis überprüft werden. Die Interviewten zeigten häufig gleiche oder zumindest ähnliche Meinungen in vielen Bereichen, starke Gegensätze waren nicht erkennbar. Auf Basis der Interviews lässt sich spekulieren, dass die Ära der Dokumente noch lange nicht vorbei ist, auch wenn sie in Zukunft eventuell eine andere Form annehmen werden. Strukturierte Daten gewinnen mehr

²⁸⁷ Vgl. Caspers, P. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁸⁸ Vgl. ebd.; vgl. Hischer, C. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

²⁸⁹ Vgl. Koppers, D. (2020): Persönliches Online-Gespräch mit der Verfasserin [pers. Gespräch], siehe Anhang 3.

und mehr an Bedeutung und können durch sich ständig weiterentwickelnde KI-Techniken immer besser verarbeitet werden. Textdokumente, aber auch Bilder, Audios und Videos können in Zukunft besser indiziert und genutzt werden. Nach Meinung der Interviewten sollte KI kein Thema darstellen, das Ängste auslöst. Stattdessen sollten die vielfältigen Chancen, die durch derlei Techniken geboten sind, ergriffen und genutzt und nicht durch zu früh erlassene Gesetze eingeschränkt werden.

5. Fazit

Ziel der vorliegenden Arbeit war, die Gegenwart und mögliche Zukunft des Dokuments zu erforschen. Die in Kapitel 1 analysierte Literatur ließ erkennen, dass die Relevanz des Dokuments als Forschungsthema in den letzten Jahren zugenommen hat. Nachdem die meisten Texte aus dem 20. Jahrhundert dazu bereits 70 bis 80 Jahre alt sind, konnte vor allem innerhalb der letzten 20 Jahre ein wiederaufkommendes Interesse am Dokument festgestellt werden. Dass sich die Forschenden bisher auf keine einheitliche Definition zum Begriff des Dokuments einigen konnten, wozu es bei Betrachtung der gegenwärtigen Lage auch in nächster Zukunft vermutlich nicht kommen wird, zeigt die Komplexität des Themas. Unterschiedliche inhaltliche, formale und sogar soziale Kriterien wurden von den Forschenden für das Dokument festgelegt. Was durch die Vorstellung und den Vergleich derselben klar wurde, ist die Vielfalt, die im Dokumentbereich herrscht. Dokumente können aus unterschiedlichsten Materialien bestehen, sogar immateriell sein und in unterschiedlichen Formaten abgespeichert werden. Sie können statisch sein, aber auch flexibel, lang oder kurz, geheim oder öffentlich. Viele Forschende sind der Meinung, dass sogar das gesprochene Wort in Form von Audiodateien sowie Bilder und Videos als Dokumente gelten können, wobei dieser Punkt umstritten ist.

In Kapitel 2 wurden diejenigen technologischen Aspekte vorgestellt, welche die Entwicklung des digitalen Dokuments stark beeinflusst haben und eine Basis für die Anwendung von KI bilden. Außerdem wurde die Veränderungen im ECM beschrieben, welche mit den Veränderungen im Dokumentbereich einhergehen. In Kapitel 3 konnte auf dieser Wissensbasis mit der Vorstellung von KI allgemein und speziellen KI-Techniken, die im Dokumentbereich eingesetzt werden, aufgebaut werden. Das allgegenwärtige Thema KI ist den meisten Menschen, wie bereits erwähnt, nicht im Detail bekannt und führt daher häufig zu Missverständnissen, überzogenen Erwartungen und teilweise sogar Angst. Die folgenden Beschreibungen trugen daher zur Aufklärung bei und zeigten, wie Programme so modelliert werden können, dass sie fähig sind, menschliche Sprache in Dokumenten zu verarbeiten und dass dies ein sehr kompliziertes Unterfangen ist, welches nichts mit übermenschlichen Fähigkeiten zu tun hat, sondern auf komplexen mathematischen Gleichungen beruht.

Durch die empirischen Untersuchungen in Kapitel 4 konnte die Beziehung zwischen Theorie und Praxis hergestellt werden. Die Vorstellung der Produkte und ihrer Funktionen zeigte Anwendungsmöglichkeiten der zuvor beschriebenen Technologien auf. Die Untersuchung der Software Theum zeigte, wie gut Algorithmen im vorgestellten Bereich

bisher arbeiten und wo noch Verbesserungsbedarf besteht. Die Auswertung der Interviews fasste verschiedene Meinungen von Fachkundigen im Bereich der KI-basierten Wissenstechnologien zusammen und gab Ausblicke in die Zukunft dieses Bereichs.

Die vorliegende Arbeit macht deutlich, dass die heutige Gesellschaft eine Dokumentengesellschaft ist. Statt von „Big Data“ könnte man ebenso von „Big Document“ sprechen. Dokumente existieren im Überfluss im WWW, liegen in Form von Rechnungen im Briefkasten, stecken als Pässe in Geldbeuteln, stapeln sich auf Schreibtischen, stehen im Bücherregal, lagern in allen möglichen Formaten auf Rechnern, Festplatten und in Clouds. Der moderne Mensch ist ständig und an den meisten Orten von ihnen umgeben. Ihre stetig wachsende Zahl macht eine rein menschliche Aufarbeitung all der Informationen, die in ihnen enthalten sind, zunehmend unmöglich. Die Weiterentwicklung von Techniken, wie sie in dieser Arbeit vorgestellt wurden, ist daher unabdingbar, wenn das Wissen, das in Dokumenten enthalten ist, zielführend genutzt werden soll. Die Zeit- und Ressourcenersparnis, die sich ergibt, wenn man mit gut funktionierender KI-Software arbeitet, ist ein Wettbewerbsvorteil, der ausgenutzt werden sollte.

Auf Basis der Erkenntnisse dieser Arbeit lässt sich prognostizieren, dass Dokumente auch in Zukunft von großer Wichtigkeit sein werden, voraussichtlich jedoch in noch flexiblerer und weniger statischen Form. Bereits während des Erstellungsprozesses werden Algorithmen greifen, die den Erstellenden inhaltliche Unterstützung bieten. Die Wissenssuche wird schneller und effektiver durch Technologien, welche ein zunehmendes „Verständnis“ für menschliche Sprache entwickeln. Antworten auf Suchen werden aus Teilen verschiedener Dokumente bestehen und in einer einheitlichen Form präsentiert. Auch wenn analoge Dokumente vermutlich nicht aussterben, ist das digitale Dokument aufgrund seiner Verarbeitbarkeit in jedem Fall das Dokument, welches in Zukunft die wichtigere Rolle einnehmen wird.

Weiterführende Forschung könnte den Markt KI-basierter Wissenstechnologien in Verbindung mit Dokumenten in Deutschland oder international im Hinblick auf die Kundenschaft untersuchen. Welche Firmen setzen Produkte dieser Art bereits ein, wie hoch ist die Zufriedenheit, welche Verbesserungen wünschen sich Nutzende der Software? Interessant wäre in diesem Zusammenhang ein Vergleich zwischen Ländern, die auf unterschiedlichen Ebenen im Bereich der KI-Entwicklung stehen. Auch eine vollständige Untersuchung des aktuellen Markts mit sämtlichen Anbietenden von KI-Produkten im Dokumentbereich und deren Vergleich wäre ein interessantes Forschungsthema. Die gegenwärtige Rolle von analogen und digitalen Dokumenten und deren Vergleich,

beispielsweise im Berufsalltag verschiedener Berufsgruppen, wäre ebenfalls ein Punkt, der weitere Rückschlüsse auf die aktuelle Lage und daher eine Prognose für die Zukunft zulassen würde.

Die Zeit des Dokuments ist eindeutig noch nicht vorbei. Es bleibt daher zu hoffen, dass immer mehr Firmen den Wert von KI-Technologien im Dokumentbereich erkennen und durch den Erwerb sowohl die Branche an sich als auch ihre Mitarbeitenden bei der Suche nach Wissen in Dokumenten unterstützen und entlasten. Nur durch die weiterführende Forschung und Arbeit an KI-basierten Wissenstechnologien können die Bereiche, die algorithmisch bisher noch schwer zu verarbeiten sind, in Zukunft besser erschlossen werden und so den Wert der Dokumente voll ausschöpfen.

Eidesstattliche Erklärung

Ich versichere an Eides statt, dass

- ich die schriftliche Abschlussarbeit selbstständig verfasst habe,
- ich keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe,
- Teile der Arbeit oder die Arbeit an sich nicht an anderer Stelle als Prüfungsleistung vorgelegt wurde und
- die Passagen der Arbeit, die fremden Werken wörtlich oder sinngemäß entnommen sind, unter Angabe der Quellen und unter Beachtung der im Wissenschaftsbereich geltenden allgemeinen verwendeten Zitierregelungen gekennzeichnet sind.

Potsdam, den 23.10.2020,

(Isanne Baumann)

Literaturverzeichnis

AAIM (2020): What is Enterprise Content Management (ECM)? URL: <https://www.aiim.org/what-is-ecm-enterprise-content-management#> (02.03.2020).

Abbyy (2020): Was ist OCR? URL: <https://www.abbyy.com/de/finereader/what-is-ocr/> (29.09.2020)

Abdelaal, Hazem Safwat (2019): Knowledge Extraction from Simplified Natural Language Text. Galway, College of Engineering and Informatics, National University of Ireland, Dissertation. URL: <http://hdl.handle.net/10379/15492> (11.09.2020)

Archivschule Marburg (o.J.): Terminologie. Schriftgutverwaltung. URL: <http://m.archivschule.de/DE/forschung/schriftgut/terminologie/terminologieschriftgutverwaltung.html> (06.03.2020)

Arcondis (2017): IDMP (Identification of Medicinal Products). URL: <https://www.arcondis.com/idmp/> (27.09.2020)

Arithnea (2020): Wie künstliche Intelligenz Kunden und Unternehmen dient. URL: <https://lp.arithnea.de/studie-ki-cx-2020?hsCtaTracking=a7c45c63-8d68-4593-828e-b973171e0815%7C9709ac07-eac6-4927-937d-0bda49ef4abf> (23.09.2020)

Ballsun-Stanton, Brian (2010): Asking about Data: Experimental Philosophy of Information Technology, in: 5th International Conference on Computer Sciences and Convergence Information Technology, S. 119–124. DOI: <https://doi.org/10.1109/IC-CIT.2010.5711041>

Becher, Margit (2009): XML. DTD, XML-Schema, XPath, XQuery, XSLT, XSL-FO, SAX, DOM. Herdecke; Witten, W3L-Verlag

Berners-Lee, Tim; Hendler, Tim; Lassila, Ora (2001): The Semantic Web, in: Scientific American 284, H. 5, S. 34-43. URL: <https://www.jstor.org/stable/10.2307/26059207> (10.08.2020)

Bitcom, Hrsg. (2012): Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte. Berlin, Bitcom.

Bizer, Christian; Heath, Tom; Berners-Lee, Tim (2011): Linked Data – The Story so Far, in: Sheth, Amit (Hrsg.): Semantic Services, Interoperability and Web Applications: Emerging Concepts. Hershey, PA, Information Science Reference, S. 205- 227

Bodendorf, Freimut (2006): Daten- und Wissensmanagement. 2. Aufl., Berlin; Heidelberg, Springer-Verlag

Bogner, Alexander; Littig, Beate; Menz, Wolfgang (2014): Interviews mit Experten. Eine praxisorientierte Einführung. Wiesbaden, Springer VS.

Bonin, Gabriela (2020): „Künstliche Intelligenz gibt es eigentlich nicht“. Gespräch mit Ladan Pooyan-Weihs. URL: <https://hub.hslu.ch/informatik/kunstliche-intelligenz-gibt-es-nicht-wichtig-ist-digitale-ethik/> (24.08.2020)

Bosak, Jon (1997): XML, Java, and the future of the Web. URL: <http://www.ibiblio.org/bosak/xml/why/xmlapps.htm> (07.08.2020)

Brendel, Alfons; Brack-von Wins, Ingard; Schmitz, Victoria (1984): Textanalysen II. Untersuchung von Texten. 10.-13. Jahrgangsstufe. 7 Aufl. München, Manz.

Briet, Suzanne (1951): Qu'est-ce que la documentation? Paris, EDIT.

Bry, François; Kraus, Michael; Olteanu, Dan; Schaffert, Sebastian (2001): Semi-strukturierte Daten, in: Information Spektrum 24, H. 4, S. 230-233

Buckland, Michael (o.J.): Michael Buckland's Suzanne Briet Page. URL: <http://people.ischool.berkeley.edu/~buckland/briet.html> (05.07.2020)

Buckland, Michael (1991): Information as Thing, in: Journal of the American Society for Information Science 42, H. 5, S. 351-360.

Buckland, Michael (1997): What Is a „Document“?, in: Journal of the American Society for Information Science 48, H. 9, S. 804-809

Buckland, Michael (1998): What is a „digital document“?, in: Hahn, T.B.; Buckland, M. (Hrsg.): Historical Studies in Information Science. Medford, NJ, Information Today, S. 215-220.

Buckland, Michael (2018): Document Theory, in: Knowledge Organization 45, H. 5, S. 425-436. DOI: <https://doi.org/10.5771/0943-7444-2018-5-425> (06.07.2020)

Busse, Johannes et al. (2014): Was bedeutet eigentlich Ontologie?, in: Informatik Spektrum 37, H. 4, S. 286–297. DOI: <https://doi.org/10.1007/s00287-012-0619-2> (10.08.2020)

Buxmann, Peter; Schmidt, Holger, (2019a): Ökonomische Effekte der Künstlichen Intelligenz, in: Buxmann, Peter; Schmidt, Holger (Hrsg.): Künstliche Intelligenz. Mit Algorithmen zum wirtschaftlichen Erfolg. Berlin, Springer Gabler, S. 21-37

Buxmann, Peter; Schmidt, Holger, (2019b): Grundlagen der Künstlichen Intelligenz und des Maschinellen Lernens, in: Buxmann, Peter; Schmidt, Holger (Hrsg.): Künstliche Intelligenz. Mit Algorithmen zum wirtschaftlichen Erfolg. Berlin, Springer Gabler, S. 3-19

Carstensen, Kai-Uwe; Ebert, Christian; Ebert, Cornelia; Jekat, Susanne; Klabunde, Ralf; Langer, Hagen (2010): Computerlinguistik und Sprachtechnologie. Eine Einführung. 3. Auflage. Heidelberg, Spektrum Akademischer Verlag

Caspers, Paul (2020): Persönliches Online-Gespräch mit der Verfasserin. Berlin, 15.07.2020

Conway, Paul (1996): Preservation in the Digital World, Washington, D.C., The Commission on Preservation and Access

Cunesoft Regulatory Compliance – Online (2016): Automatic Regulatory Data Mining Done Right – PhlexDistiller. URL: <https://www.youtube.com/watch?v=2NU5XVV6jD0> (27.09.2020)

Digitales Wörterbuch der Deutschen Sprache (o.J.): Dokumentenmanagement. URL: <https://www.dwds.de/?q=Dokumentenmanagement> (19.08.2020)

DIN ISO (2002): DIN ISO 15489-1:2001: Information und Dokumentation. Schriftgutverwaltung. Teil 1: Allgemeines (DIN ISO 1548-1:2002)

Document Academy (o.J.): About. URL: <http://documentacademy.org/?about> (20.07.2020)

Dudek, Sarah (2012): Die Zukunft der Buchstaben in der alphanumerischen Gesellschaft. Text und Dokument unter digitalen Bedingungen, in: Bibliothek Forschung und Praxis 36, H. 2, S. 189-199. DOI: <https://doi.org/10.1515/bfp-2012-0023>

Duden (2020): Dokument. URL: <https://www.duden.de/rechtschreibung/Dokument> (30.09.2020)

Elsner, Frank (2004): Einführung in HTML. URL: <https://pdfs.semanticscholar.org/794f/22284789166d7c2fae385e88f0ce4240a228.pdf> (07.08.2020)

Fensel, Dieter et al. (2020): Knowledge Graphs. Methodology, Tools and Selected Use Cases. Cham, Springer.

Fink, Andreas (2014): HTML, in: Enzyklopädie der Wirtschaftsinformatik. Online-Lexikon. URL: <https://www.encyklopaedie-der-wirtschaftsinformatik.de/wi-encyklopaedie/lexikon/technologien-methoden/Sprache/Auszeichnungssprache/HTML> (07.08.2020)

Francke, Helena (2005): What's in a Name? Contextualizing the Document Concept, in: Literary and Linguistic Computing 20, H.1, S. 61-69. DOI: <https://doi.org/10.1093/lc/fqh043> (19.07.2020)

Franz, Sonja (2018): Wie Suchalgorithmen funktionieren: Die Levenshtein-Distanz. URL: <https://www.integer-net.de/wie-suchalgorithmen-funktionieren-levenshtein-distanz/> (10.10.2020)

Franzetti, Claudio (2019): Essenz der Informatik. Berlin; Heidelberg, Springer.

Funke, Fritz (1999): Buchkunde. 6. überarb. und erg. Aufl. Berlin; Boston, De Gruyter Saur

Gardner, Howard; Davis, Katie; Christodoulou, Joanna; Seider, Scott (2011): The Theory of Multiple Intelligences, in: Sternberg, Robert; Scott Barry Kaufman (Hrsg.): The Cambridge Handbook of Intelligence. Cambridge, Cambridge University Press, S.485–503

Gentsch, Peter (2018): Künstliche Intelligenz für Sales, Marketing und Service. Mit AI und Bots zu einem Algorithmic Business – Konzepte, Technologien und Best Practices. Wiesbaden, Springer Gabler

Goasduff, Laurence (2020): 2 Megatrends Dominate the Gartner Hype Cycle for Artificial Intelligence, 2020. URL: <https://www.gartner.com/smarterwithgartner/2-megatrends-dominate-the-gartner-hype-cycle-for-artificial-intelligence-2020/> (05.10.2020)

Google (2020a): Suchanfrage zu „Elon Musk“. URL: https://www.google.de/search?sxsrf=ALeKk03arS9u3hILR7qjtiQNT6R4SEO-Gjg%3A1603387011042&source=hp&ei=gr6RX4rqO_qdjLsP1rih-As&q=elon+musk&oq=elon+musk&gs_lcp=CgZwc3ktYWIQAzIHCC4QJxCTAjIECAAQQzIECAAQQzIECAAQQzIECAAQQzIFCAAQsQMyBQgAELEDMgIIA-DIECAAQQzIHCAAQFBCHAjoECCMQJzoICAAQsQMqgwE6CggAELEDEBQQhwJQngVY2q1gkQ9oAHAAeACAAe8BiAGcCZIBBTauNy4xmAEAoAEBqgEHZ3dzLXdpeg&sclient=psy-ab&ved=0ahUKEwiKs-iJ2sjsAhX6DmMBHVZcCL8Q4dUDCAg&uact=5 (15.07.2020)

Google (2020b): Suchanfrage zu „Brad Pitt“. URL: <https://www.google.com/search?client=firefox-b-d&q=brad+pitt> (20.07.2020)

Görz, Günther; Nebel, Bernhard (2003): Künstliche Intelligenz. Frankfurt am Main, Fischer Verlag

Götzer, Klaus; Maier, Berthold; Schmale, Ralf; Rehbock, Klaus; Komke, Torsten (2013): Dokumenten-Management. Informationen im Unternehmen effizient nutzen. 5. Aufl. Heidelberg, dpunkt.verlag

Hartmann, Frank (2006): Von Karteikarten zum vernetzten Hypertext-System, in: Telepolis. URL: <https://www.heise.de/tp/features/Von-Karteikarten-zum-vernetzten-Hypertext-System-3408411.html> (05.07.2020)

Herman, Andreas (2019): Different ways of doing Relation Extraction from text. URL: <https://medium.com/@andreasherman/different-ways-of-doing-relation-extraction-from-text-7362b4c3169e> (14.09.2020)

Hischer, Carolin (2020): Persönliches Online-Gespräch mit der Verfasserin. Berlin, 13.10.2020

IBM Watson (o.J.): Speech to Text. URL: <https://speech-to-text-demo.ng.bluemix.net/> (01.10.2020)

ITWissen.info (2020): Fuzzy-Suche. URL: <https://www.itwissen.info/Fuzzy-Suche-fuzzy-search.html> (10.10.2020)

Jivani, Anjali Ganesh (2011): A Comparative Study of Stemming Algorithms, in: International Journal of Circuit Theory and Applications 2, H. 6, S. 1930-1938. URL: https://www.researchgate.net/publication/284038938_A_Comparative_Study_of_Stemming_Algorithms#fullTextFileContent (10.09.2020)

Kampffmeyer, Ulrich (2006): Dokumentenmanagement. URL: https://www.project-consult.de/wp-content/uploads/2017/01/Dokumentenmanagement_W2005.pdf (01.03.2020).

Kampffmeyer, Ulrich (2014): Was ist ECM Enterprise Content Management? URL: <http://www.softselect.de/uploads/pdf/ECM-Enterprise-Content-Management.pdf> (20.09.2020)

Koppers, Daniel (2020): Persönliches Online-Gespräch mit der Verfasserin. Berlin, 13.07.2020

Kreutzer, Ralf T.; Sirrenberg, Marie (2019): Künstliche Intelligenz verstehen. Grundlagen – Use-Cases – unternehmenseigene KI-Journey. Wiesbaden, Springer Gabler

Lenzen, Manuela (2020): Künstliche Intelligenz. Fakten, Chancen, Risiken. München, C.H. Beck.

Li, Susan (2018): Named Entity Recognition and Classification with Scikit-Learn, in: Towards data science. URL: <https://towardsdatascience.com/named-entity-recognition-and-classification-with-scikit-learn-f05372f07ba2> (13.09.2020)

Linguamatics (o.J.): What is Text Mining, Text Analytics and Natural Language Processing? URL: <https://www.linguamatics.com/what-text-mining-text-analytics-and-natural-language-processing> (11.09.2020)

Liu, Ziming (2004): The evolution of documents and its impacts, in: Journal of Documentation 60, H. 3, S. 279-288. DOI: <https://doi.org/10.1108/00220410410534185>

Lund, Niels Windfeld (2007): Building a Discipline, Creating a Profession: an Essay on the Childhood of “dokvit”, in: Skare, Roswitha; Lund, Niels Windfeld; Vårheim, Andreas (Hrsg.): A Document (Re)turn. Frankfurt am Main, Peter Lang.

Lund, Niels Windfeld (2011): Document Theory, in: Annual Review of Information Science and Technology 43, H. 1, S. 1-55- DOI: <https://doi.org/10.1002/aris.2009.1440430116>

McCarthy, John (2007): What is Artificial Intelligence? URL: <http://jmc.stanford.edu/articles/whatisai/whatisai.pdf> (21.08.2020)

- Meinhold, Annabelle (2020):** Google Translate vs. DeepL, in: CIO. URL: <https://www.cio.de/a/google-translate-vs-deepl,3575364> (01.10.2020)
- Mooney, Raymond J. (2007):** Learning for Semantic Parsing, in: Gelbukh, Alexander (Hrsg.): Computational Linguistics and Intelligent Text Processing. 8th International Conference, CICLing 2007, 18.02-24.02 in Mexico City, Mexico. Berlin; Heidelberg, Springer, S.311-324. DOI: https://doi.org/10.1007/978-3-540-70939-8_28 (13.09.2020)
- Murphy, K.P. (2012):** Machine Learning. A Probabilistic Perspective. Cambridge, The MIT Press
- Neumann, Gustaf (2014a):** Auszeichnungssprache, in: Enzyklopädie der Wirtschaftsinformatik. Online-Lexikon. URL: <https://www.encyklopaedie-der-wirtschaftsinformatik.de/lexikon/technologien-methoden/Sprache/Auszeichnungssprache> (07.08.2020)
- Neumann, Gustaf (2014b):** XML. URL: <https://www.encyklopaedie-der-wirtschaftsinformatik.de/wi-encyklopaedie/lexikon/technologien-methoden/Sprache/Auszeichnungssprache/XML> (07.08.2020)
- NLP-progress (o.J.a):** Dependency parsing. URL: http://nlpprogress.com/english/dependency_parsing.html (12.09.2020)
- NLP-progress (o.J.b):** Entity Linking. URL: http://nlpprogress.com/english/entity_linking.html (13.09.2020)
- Otlet, Paul (1934):** Traité de documentation. Brüssel, Editiones Mundaneum
- Pédauque, Roger T. (2003):** Document: Form, Sign and Medium, As Reformulated for Electronic Documents. URL: https://archivesic.ccsd.cnrs.fr/sic_00000594 (11.07.2020)
- Pfeifer, Wolfgang et al. (1993):** Etymologisches Wörterbuch des Deutschen. Digitalisierte und überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache. Stichwort „Dokument“. URL: <https://www.dwds.de/wb/Dokument> (04.07.2020)
- Phlexglobal (2020a):** Startseite. URL: <https://www.phlexglobal.com/> (26.09.2020)
- Phlexglobal (2020b):** PhlexDistiller. URL: <https://www.phlexglobal.com/phlexdistiller> (26.09.2020)
- Piro, Andrea; Gebauer, Marcus (2011):** Definition von Datenarten zur konsistenten Kommunikation im Unternehmen, in: Hildebrand, Knut et al. (Hrsg.): Daten- und Informationsqualität. Auf dem Weg zur Information Excellence. 2. Aufl. Wiesbaden, Vieweg + Teubner, S. 143-156
- Project Consult (o.J.a):** Enterprise Content Management (ECM). URL: <https://www.project-consult.de/themen/ecm-enterprise-content-management/> (13.09.2020).

PROJECT CONSULT Unternehmensberatung GmbH (2019): ECM, EIM, Content Services, IIM – what's next? URL: https://www.youtube.com/watch?v=RgwkdBr_eEo (20.09.2020)

PROJECT CONSULT Unternehmensberatung GmbH (2020): Die 10 PROJECT CONSULT Trends für das Information Management 2020. URL: <https://www.youtube.com/watch?v=oC6Kyk2Pon8> (20.09.2020)

Quach, Katyanna (2017): How we fooled Google's AI into thinking a 3D-printed turtle was a gun: MIT bods talk to EI Reg. URL: https://www.theregister.com/2017/11/06/mit_fooling_ai/ (29.08.2020)

Ragnet, François (2009): Document 3.0 How the next generation of dynamic documents will improve the way we work. URL: https://www.xerox.com/downloads/usa/en/xgs/whitepapers/xgs_whitepaper_FrancoisRagnet.pdf

Ranganathan, Shiyali Ramamrita (1963): Documentation and its facets. Mumbai, Asia Publishing House.

Reed, Richard (2017): Documents, AI, and the Future. URL: <https://project-consult.theum.com/newsletter/index.htm?t=Kunden-Newsletters-ab-2011,2017,201703-Juni,Gastbeitr%C3%A4ge~7c-Guest-Authors,Documents,-AI,-and-the-Future> (26.09.2020)

Rentzmann, Rene (2006): Text Mining, in: Informatiklexikon der Gesellschaft für Informatik e.V. URL: <https://gi.de/informatiklexikon/text-mining> (30.09.2020)

Rey, Alain (1992): Dictionnaire historique de la langue française contenant les mots français en usage et quelques autres délaissés. Paris, Dictionnaires Le Robert

Salaün, Jean-Michel (2014): Why the Document Is Important... and How it Is Becoming Transformed, in: The Monist 97, H. 2, S. 187-199. URL: <https://www.jstor.org/stable/44012678> (04.07.2020)

Salaün, Jean-Michel (2018): Die drei Dimensionen des Dokuments und ihre Auswirkungen auf die Bibliotheks- und Informationswissenschaft, in: Bibliothek Forschung und Praxis 42, H. 2, S. 351-356. DOI: <https://doi.org/10.1515/bfp-2018-0039> (20.07.2020)

Schamber, Linda (1996): What Is a Document? Rethinking the Concept in Uneasy Times, in: Journal of the American Society for Information Science 47, H. 9, S. 669-671. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(199609\)47:9%3C669::AID-ASI3%3E3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-4571(199609)47:9%3C669::AID-ASI3%3E3.0.CO;2-Q) (31.07.2020)

Schürmeyer, Walter (1935): Aufgaben und Methoden der Dokumentation, in: Zentralblatt für Bibliothekswesen 52, H.9, S. 533-543.

Schwalm, Steffen; Ulrich, Rainer; Dippel, Andrea, Hrsg. (2008): Lexikon Dokumentenmanagement und Archivierung. Berlin, Infora

Sellen, Abigail J.; Harper, Richard H.R. (2003): The myth of the paperless office. Cambridge, MA; London, The MIT Press.

Serviceware (2020a): Über Serviceware. URL: <https://serviceware-se.com/de/unternehmen/ueber-serviceware> (30.09.2020)

Serviceware (2020b): Serviceware Knowledge. URL: <https://serviceware-se.com/de/loesungen/serviceware-plattform/serviceware-knowledge> (01.20.2020)

Spitta, Thorsten; Bick, Markus (2008): Informationswirtschaft. Eine Einführung. 2. Aufl. Berlin; Heidelberg, Springer.

Styrz (o.J.): Content Management. URL: <https://www.styrz.de/content-management/> (25.09.2020)

The TEI Consortium (o.J.): Home. URL: <https://tei-c.org/> (03.10.2020)

The TEI Consortium (2020): TEI P5. Guidelines for Electronic Text Encoding and Interchange. URL: <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> (03.10.2020)

Theum (o.J.a): Wir über uns. URL: <https://www.theum.com/kognitive-suche-und-wissensfindung/wirueberuns> (25.09.2020)

Theum (o.J.b): Was ist Theum? URL: <https://www.theum.com/kognitive-suche-und-wissensfindung/virtuelle-experten-die-aus-dokumenten-lernen> (25.09.2020)

Theum (o.J.c): Antworten aus SharePoint und anderen DMS. URL: <https://www.theum.com/kognitive-suche-und-wissensfindung/use-cases/137-antworten-aus-sharepoint-und-anderen-dms> (25.09.2020)

Theum (o.J.d): Arbeiten mit Theum Wissensbanken. URL: <https://support.theum.com/Documentation/Theum3.6.1/index.htm?t=Help,Kurzanleitungen,Leser,Schnellstart,Arbeiten-mit-Theum-Wissensbanken> (25.09.2020)

Theum (o.J.e): Gesetzestexte. URL: <https://gesetze.theum.com/> (25.09.2020)

Theum (o.J.f): Recherchieren mit Theum. URL: <https://support.theum.com/Documentation/Theum3.6.1/index.htm?t=Help,Kurzanleitungen,Leser,Schnellstart,Recherchieren-mit-Theum> (25.09.2020)

Theum (o.J.g): Help. URL: <https://support.theum.com/Documentation/Theum3.7/index.htm?t=Help,DE> (02.10.2020)

Theum (2019): Theum 3.7 User Guides. URL: <https://support.theum.com/Documentation/current/index.htm?t=User-Guides> (25.09.2020)

Theum AG (2019): KI Einsatz zur präzisen Bereitstellung von Wissen. URL: <https://www.youtube.com/watch?v=CdA6w8XTs6M> (25.09.2020)

Storm, Theodor (2004): Der Schimmelreiter, in: DigBib.Org. URL: <http://www.dig-bib.org/Theodor Storm 1817/Der Schimmelreiter .pdf> (30.08.2020)

Universität Stuttgart (1999): STTS Tag Table. URL: <https://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/germantagsets/#id-cfcbf0a7-0> (11.09.2020)

Valéry, Paul (1970): Collected Works of Paul Valéry, Volume 14. Analects. Princeton, N.J., Princeton University Press

Vivek, Sowmya (2018): Automated Keyword Extraction from Articles using NLP. URL: <https://medium.com/analytics-vidhya/automated-keyword-extraction-from-articles-using-nlp-bfd864f41b34> (13.09.2020)

Voß, Jakob (2013): Was sind eigentlich Daten?, in: LIBREAS Library Ideas 23, S. 4-11. URL: <https://libreas.eu/ausgabe23/02voss/>

Woodbridge, Michael (2017): The Death of ECM and Birth of Content Services. URL: <https://blogs.gartner.com/michael-woodbridge/the-death-of-ecm-and-birth-of-content-services/> (18.09.2020)

Wright, Alex (o.J): Paul Otlet, in: Encyclopaedia Britannica. URL: <https://www.britanica.com/biography/Paul-Otlet> (05.07.2020)

Anhang

Anhang 1: Interviewleitfaden für Isanne Baumann

Einstieg

Hallo Herr/Frau X. Vielen Dank, dass Sie sich bereit erklärt haben, dieses Interview mit mir zu führen. Nochmal ganz kurz zu meiner Person: ich bin Isanne Baumann, habe meinen Bachelor in Bibliotheks- und Informationsmanagement in Stuttgart gemacht und bin jetzt in den letzten Zügen meines Masterstudiums in Informationswissenschaft an der FH Potsdam. Meine Masterarbeit, die der Grund für dieses Interview ist, befasst sich mit dem Dokumentenbegriff im Wandel der Zeit und dem Einsatz von KI-basierten Wissenstechnologien im Zusammenhang mit Dokumenten. In diesem Kontext möchte ich Ihnen gerne einige allgemeine Fragen zum Thema und auch spezielle Fragen zu Ihrem Produkt stellen. Das Interview wird ca. eine Stunde in Anspruch nehmen, ist das für Sie in Ordnung? Ich werde das Gespräch aufzeichnen, jedoch lediglich die Audiospur, keine Videoaufzeichnung. Sind Sie damit einverstanden? Wenn es von Ihrer Seite aus keine Fragen mehr gibt, würde ich jetzt mit dem Interview beginnen.

Interview

Schlüsselfragen: rot , ergänzende Fragen: weiß

Block 1: Die Begriffe “Dokument” und “Daten”

Es gibt viele verschiedene Interpretationen des Begriffs “Dokument”. Beispielsweise betrachtete der indische Mathematiker und Bibliothekar S.R. Ranganathan sie als Gedanken auf Papier oder einem anderen Material, das physische Handhabung, Transport durch den Raum und Konservierung in der Zeit erlaubt²⁹⁰, die französische Bibliothekarin Suzanne Briet hingegen hielt sogar Tiere im Zoo, solange sie katalogisiert seien, für Dokumente²⁹¹ Was ist Ihre persönliche Meinung zu diesem Thema?

Machen Sie bei Ihrer Definition einen Unterschied zwischen analogem und digitalem Dokument? Wenn ja, welchen?

Wie differenzieren Sie den Begriff des Dokuments von jenem der Daten? Ist eine klare Differenzierung möglich?

Ab wann ist für Sie ein Dokument ein Dokument? Sind beispielsweise Tabellen mit Messwerten für Sie Dokumente? Wo ziehen Sie die Grenze?

²⁹⁰ Vgl. Ranganathan, S. R. (1963): Documentation and its facets, S. 24.

²⁹¹ Vgl. Briet, Suzanne (1951): Qu’est-ce que la documentation, S. 7.

Wenn Sie einige Jahre zurückdenken: hat sich Ihre Vorstellung von dem, was ein Dokument ist, verändert? In welcher Weise?

Block 2: KI

Mit welchen KI-Aspekten arbeitet Ihr Produkt (künstliche neuronale Netze, Deep Learning, Text Mining, ...)?

Wie stehen Sie persönlich ganz allgemein zu KI (nicht in Zusammenhang mit Ihrem Produkt)? Sehen Sie darin eine große Chance, ein großes Risiko oder beides?

Wie schätzen Sie die Bemühungen in Deutschland ein, um die Entwicklung von KI voranzutreiben (beispielsweise im Vergleich zu den USA oder China)? Wird aus Ihrer Sicht in Deutschland in diesem Kontext genug getan?

Block 3: Produkt

Wie kam es dazu, dass Ihr Produkt entwickelt wurde (Idee, Initiative, ...)?

Wie lang war die Entwicklungszeit Ihres Produkts?

Wie viele Personen waren an der Entwicklung beteiligt?

Können Sie Näheres zur Funktionsweise des Produkts erläutern? Beispielsweise, auf welche Art das Produkt lernt (z.B. Deep Learning, KNN)?

Was glauben Sie, wie sich die Nutzungshäufigkeit von Produkten wie Ihrem in der Zukunft entwickeln wird, vor allem im Rahmen von Dokumentenmanagement?

Wie hat sich das Produkt seit seiner ersten Version verändert?

Wer sind Ihre Kunden bzw. wer ist die Zielgruppe Ihres Produkts?

Welche anderen "Einsatzorte" (außerhalb von Dokumentenmanagement vornehmlich in Unternehmen) können Sie sich für Produkte dieser Art vorstellen?

Spezifische Produktfragen (evtl. bei Tests, Videos aufgetaucht)

Block 4: Zukunft

Sehen Sie Möglichkeiten, Ihr Produkt noch effizienter zu machen, beispielsweise anhand von Technologien, die bisher noch gar nicht verfügbar sind? Wenn ja, in welcher Weise?

Was glauben Sie, wie wird das "Dokument der Zukunft" aussehen?

Werden in Zukunft Ihrer Meinung nach gezielte KI-Gesetze nötig sein? Beispielsweise eine Art "hippokratischer Eid" für Programmierer? Warum?

Halten Sie, es für möglich, dass der Mensch jemals fähig sein wird, eine Intelligenz zu erschaffen, die der seinen ebenbürtig ist? ("A cyclone can raze a city, yet not even

open a letter or untie the knot in this piece of string"²⁹², Zitat als Beispiel für das "Wesen" einer KI). Wie begründen Sie Ihre Meinung dazu?

Quellen:

Briet, Suzanne (1951): Qu'est-ce que la documentation. Paris, EDIT

Ranganathan, S. R., Hrsg. (1963): Documentation and its facets. Mumbai, Asia Publishing House

Valéry, Paul (1970): Collected Works of Paul Valéry, Volume 14. Analects. Princeton, N.J., Princeton University Press

Abschluss:

Ganz herzlichen Dank nochmal für Ihre Bereitschaft, dieses Interview mit mir zu führen und für die Zeit, die Sie sich genommen haben. Falls Sie noch irgendetwas zu ergänzen haben, das Sie für relevant halten und durch die Fragen nicht angesprochen wurde, können Sie mir das gerne noch mitteilen.

²⁹² Valéry, P. (1970): Collected Works of Paul Valéry, Volume 14. Analects, S. 132.

Anhang 2: Fragenkatalog für Sachkundige

Block 1: Die Begriffe “Dokument” und “Daten”

Es gibt viele verschiedene Interpretationen des Begriffs “Dokument”. Beispielsweise betrachtete der indische Mathematiker und Bibliothekar S.R. Ranganathan sie als Gedanken auf Papier oder einem anderen Material, das physische Handhabung, Transport durch den Raum und Konservierung in der Zeit erlaubt ²⁹³ , die französische Bibliothekarin Suzanne Briet hingegen hielt sogar Tiere im Zoo, solange sie katalogisiert seien, für Dokumente ²⁹⁴ Was ist Ihre persönliche Meinung zu diesem Thema?
Machen Sie bei Ihrer Definition einen Unterschied zwischen analogem und digitalem Dokument? Wenn ja, welchen?
Wie differenzieren Sie den Begriff des Dokuments von jenem der Daten? Ist eine klare Differenzierung möglich?
Ab wann ist für Sie ein Dokument ein Dokument? Sind beispielsweise Tabellen mit Messwerten für Sie Dokumente? Wo ziehen Sie die Grenze?
Wenn Sie einige Jahre zurückdenken: hat sich Ihre Vorstellung von dem, was ein Dokument ist, verändert? In welcher Weise?

Block 2: KI

Wie stehen Sie persönlich ganz allgemein zu KI (nicht in Zusammenhang mit Ihrem Produkt)? Sehen Sie darin eine große Chance, ein großes Risiko oder beides?
Wie schätzen Sie die Bemühungen in Deutschland ein, um die Entwicklung von KI voranzutreiben (beispielsweise im Vergleich zu den USA oder China)? Wird aus Ihrer Sicht in Deutschland in diesem Kontext genug getan?
Mit welchen KI-Aspekten arbeitet Ihr Produkt (künstliche neuronale Netze, Deep Learning, Text Mining, ...)?

Block 3: Produkt

Wie kam es dazu, dass Ihr Produkt entwickelt wurde (Idee, Initiative, ...)
Wie lang war die Entwicklungszeit Ihres Produkts?
Wie viele Personen waren an der Entwicklung beteiligt?
Können Sie Näheres zur Funktionsweise des Produkts erläutern? Beispielsweise, auf welche Art das Produkt lernt (z.B. Deep Learning, KNN)?
Was glauben Sie, wie sich die Nutzungshäufigkeit von Produkten wie Ihrem in der Zukunft entwickeln wird, vor allem im Rahmen von Dokumentenmanagement?

²⁹³ Vgl. Ranganathan, S. R. (1963): Documentation and its facets, S. 24.

²⁹⁴ Vgl. Briet, Suzanne (1951): Qu'est-ce que la documentation, S. 7.

Wie hat sich das Produkt seit seiner ersten Version verändert?
Wer sind Ihre Kunden bzw. wer ist die Zielgruppe Ihres Produkts?
Welche anderen "Einsatzorte" (außerhalb von Dokumentenmanagement vornehmlich in Unternehmen) können Sie sich für Produkte dieser Art vorstellen?

Block 4: Zukunft

Sehen Sie Möglichkeiten, Ihr Produkt noch effizienter zu machen, beispielsweise anhand von Technologien, die bisher noch gar nicht verfügbar sind? Wenn ja, in welcher Weise?
Was glauben Sie, wie wird das "Dokument der Zukunft" aussehen?
Werden in Zukunft Ihrer Meinung nach gezielte KI-Gesetze nötig sein? Beispielsweise eine Art "hippokratischer Eid" für Programmierer? Warum?
Halten Sie, es für möglich, dass der Mensch jemals fähig sein wird, eine Intelligenz zu erschaffen, die der seinen ebenbürtig ist? ("A cyclone can raze a city, yet not even open a letter or untie the knot in this piece of string" ²⁹⁵ , Zitat als Beispiel für das "Wesen" einer KI). Wie begründen Sie Ihre Meinung dazu?

Quellen:

Briet, Suzanne (1951): Qu'est-ce que la documentation. Paris, EDIT

Ranganathan, S. R., Hrsg. (1963): Documentation and its facets. Mumbai, Asia Publishing House

Valéry, Paul (1970): Collected Works of Paul Valéry, Volume 14. Analects. Princeton, N.J., Princeton University Press

²⁹⁵ Valéry, P. (1970): Collected Works of Paul Valéry, Volume 14. Analects, S. 132.

Anhang 3: Interview-Transkriptionen

Interview-Transkription Daniel Koppers, Phlexglobal

Berlin, 13.07.2020

Interviewerin: Isanne Baumann (Fettdruck)

Interviewpartner: Daniel Koppers (Normaldruck)

Gut...okay, bevor ich mit dem ersten Block von Fragen anfang, ahm, würd ich gern nochmal wissen, was genau Ihre Rolle im Unternehmen ist und, ahm, vielleicht ganz kurz so bisschen Ihren beruflichen Hintergrund.

Also aktuell, ähm, bin ich ähm Head of R&D, das heißt, ich mach ähm eigentlich alles im Unternehmen, was mit technologischer Weiterentwicklung, ähm mit neuen Technologien, äh zu tun hat. Und ähm helf so 'n bisschen, ähm die Produkte, die ja teilweise, ähm teilweise schon bisschen älter sind bei der „Phlexglobal“ ahm etwas zu modernisieren und auf den neuesten Stand zu bringen. So neulich zum Beispiel die KI bei uns, bei uns im Team ähm. Äh und solche Sachen, wie „Big Data“ Datenbanken, ähm größere Datenbanken, ähm Data...„Data Lakes“, solche Geschichten ähm liegen im Moment bei uns, um zu schauen, was wir mit der Technologie tun können, um die, ähm nützlich für unsere Kunden in die Produktlinien zu integrieren. **Mhm.** Zu mein ähm, ähm, zu mein ähm beruflichen Hintergrund, pfff...wie weit soll ich denn zurückgehen?

(Lacht) Ja halt, was haben Sie gelernt...studiert?

Studiert hab ich nicht, ähm ich hab gelernt, äh ich bin ausgebildeter Software-Entwickler für äh Wirtschaftsinformatik. Das war früher noch ne Ausbildung und heute ist es glaube ich ein Studiengang. Ahm genau und ähm hab mich eigentlich von Anbeginn meiner beruflichen Karriere auf äh die Pharma-Branche konzentriert und grade da dann auch in die Nische Dokumentenmanagement, Dokumentenorganisation und Information...ähm Information-Management, Records-Management, solche Geschichten.

Okay, also Sie kamen jetzt nicht von ganz wo anders her (lacht).

Nee, meine Ausbildung hab ich tatsächlich, interessanterweise äh äh bei 'ner Bank gemacht, **mhm** wie man das ja früher so gemacht hat: „Junge, geh zu 'ner Bank. Dann lernst du was Gescheites“. **(Lacht).** Ahm bei der Landesbank Baden-Württemberg. **Ah!** Und äh ja ja, und hab da, hab da während der Ausbildung auch für mich entschieden, dass äh das nichts für mich ist. **(Lacht) Okay. (Lacht).** Äh genau. Und bin dann relativ schnell eben in die Pharmabranche gewechselt und da bin ich seit 13 Jahren ungefähr, 14 Jahren, und bin auch ganz glücklich.

Okay, das ist natürlich schön. Gut, dann würde ich starten mit dem ersten Block, wo es um die Begriffe „Dokument“ und „Daten“ geht. Hmm... Das ist nicht so einfach, ich weiß (lacht). Ähm ja. Also es gibt ja viele verschiedene Interpretationen von dem Begriff „Dokument“ hab ich jetzt gemerkt, während meiner Recherche, ähm, beispielsweise hat ja eben der indische Mathematiker und Bibliothekar Ranganathan - ja der war sehr auf die Form bedacht, also hat da gesagt eben alles, was auf Papier oder ‘nem anderen Material ähm...steht sozusagen und Transport durch Raum und Konservierung durch die Zeit erlaubt, ist ein Dokument. Und die Suzanne Briet, die wird auch ganz oft zitiert, ähm die war dann eher so auf den Inhalt bedacht, als auf die Form, weil sie hat ja sogar gesagt, dass Tiere im Zoo Dokumente sein können, solange sie eben katalogisiert sind. Und ja, da würd’ ich gern mal wissen, was ist Ihre Meinung zu dem Thema? Was ist für Sie ein Dokument?

Ja da wird es schon sehr philosophisch. Ich hab den Fragenkatalog ja schon vorher bekommen, hab dann auch lange drüber nachge – es war die einzige Frage, wo ich lange drüber nachdenken musste (lacht) (**lacht**) – und damit hab ich mich echt noch nicht beschäftigt. Ähm ja, was ist ein Dokument, also ich würde mit dem Ranganathan da schon schon irgendwo ‘n Stückweit äh äh mitgehen. Ähm Tiere im Zoo das find ich jetzt dann schon etwas, etwas weit hergeholt **ja** das als Dokument zu bezeichnen. Ja der, der Datenpunkt des Katagoli...Kate...Katagoli..KA-TALOGISIERENS ja, das ist äh das ist, das kann dann zu ‘nem Dokument werden, aber, aber nicht das Tier selbst. Also ähm, ich, ich denke auch, ähm, dass egal, ob jetzt gedruckt oder, oder digital ähm alles, was, was, was Information beinhaltet und wenn’s ‘ne Einkaufsliste auf Papier geschrieben ist, die, die eben transportiert werden kann, weitergegeben werden kann, geteilt werden kann - und ich glaube das ist der wichtige Punkt: geteilt werden kann - ahm qualifiziert sich, ahm für mich als Dokument. **Mhm**. Die Frage ist dann, wie nützlich ist dieses Dokument? **Mhm**. Der Einkaufszettel auf Papier ist eben meiner Meinung nach lange nicht so nützlich, wie...äh ‘n strukturiertes digitales Dokument.

Ja...okay. Also da wär schon auch bisschen die zweite Frage mit drin, ob Sie ‘n Unterschied zwischen analogem und digitalem Dokument machen, also jetzt so rein von der Definition her?

Ja würd ich **ja?** würd ich definitiv. Also so ‘n, ‘n gedrucktes Dokument, äh so schön es ist, äh... Abends zu ‘nem Glas Wein ‘n Buch zu lesen, ‘n gedrucktes Dokument ist ‘n dummes Dokument. Ich kann damit nichts tun. Ich kann’s lesen, natürlich kann ich’s lesen, ich kann ihm Informationen entnehmen, aber jetzt ähm, weiterführend kann ich

nicht viel damit tun. Mit 'nem digitalen Dokument kann ich viel tun, ähm bei 'nem digitalen Dokument denk ich ja nicht nur an PDF oder Word-Dokument, sondern ne Datenbank ist auch 'n Dokument. Ich kann mit 'ner Datenbank aber Daten strukturieren, ich kann Daten gruppieren, ich kann ähm weitere Informationen, die der Autor gegebenenfalls gar nicht, gar nicht im, im, im, im, äh im Blick hatte da, äh äh daraus generieren **mhm** ja, also ich kann aus Dokumenten weitere oder aus Informationen, die in Dokumenten enthalten sind, weitere Informationen generieren und das macht für mich dann...ähm, den großen Unterscheid zwischen analog und digitalem Dokument.

Okay...gut...ahm...und dann wäre die Frage, wie Sie Dokument und Daten unterscheiden oder ob Sie überhaupt denken, dass man das so klar unterscheiden kann?

Ja, glaub ich **ja?** ähm. Also wenn man, wenn man aus der, wenn man, wenn man sich so bisschen mit „Text Mining“ beschäftigt, das ist natürlich immer Ansichtssache, aber ähm, ich glaube dass äh 'n Dokument besteht aus Daten mit vielen Füll...äh mit vielen Klebewörtern...ahm, man macht ja, man macht ja in der KI wenn man Text Mining macht relativ viel, um erstmal den Text ähm, zu...ja man sagt so schön „normalisieren“. Man nimmt so alles raus, was Menschen brauchen, aber Maschinen eben nicht. So Sachen wie, äh Füllwörter (**mhm**) äh äh äh Präpositionen und solche Geschichten, das, das, äh das nenn ich jetzt mal Klebematerial und das, was dann über bleibt, sind die Daten **okay** und 'n Dokument, besteht für mich aus Daten **okay**, **verstehe** wieder die Frage, wie, wie...wichtig sind die Daten oder wie, wie, wie nützlich sind die Daten aber, ja **mhm**.

Also würden Sie dann auch sagen, ähm weil da ja dann eben die nächste Frage war: Ab wann ist 'n Dokument 'n Dokument? Also Tabellen mit Messwerten zum Beispiel, sehen Sie auch als Dokumente an und die Messwerte als Daten? Definitiv!
Ja.

Okay, verstehe, gut. Ähm und dann hatte ich hier ja noch, wenn Sie 'n paar Jahre zurückdenken, ähm hat sich dann die Vorstellung von dem, was für Sie 'n Dokument ist verändert?

Pfffu ähm (*längere Pause*). Ja, hat es...ähm, natürlich nachdem ich angefangen hab, nachdem wir angefangen haben, ähm viel Richtung „Data Mining“ zu machen und und, und Datenextraktion zu machen, hat sich, hat sich meine Sichtweise auf Dokumente 'n Stückweit verändert **mhm** ähm...vielleicht wäre meine Antwort zur vorherigen Frage früher anders ausgefallen ehrlich gesagt (**lacht**) ähm, aber...ja...äh...wie ich Dokumente

anschaue heute, hat sich verändert. Die Sichtweise auf Dokumente ähm, hat sich glaub ich so grundlegend nicht geändert würd ich sagen **mhm** ja.

Ja ich glaub, so würde es auch Einigen gehen, von denen ich jetzt gelesen hab, die halt das Dokument definiert haben als es „digital“ noch gar nicht gab und so ja, ich glaub die hätten da teilweise auch andere Ansichten jetzt.

Ja, ja, ja..., ja jetzt muss man ja dazusagen, ich bin ja, ich bin ja schon aufgewachsen mit Word und PDF und so weiter und sofort **ja**, das heißt ähm ich, ich kenn das, ich kenn das komplett analoge Arbeiten ja eigentlich gar nicht **mhm**, das kenn ich nur vom Hörensagen (lacht), von daher...(lacht). **Ja, geht mir auch so.**

Gut, dann komm ich zum zweiten Block: KI. Ähm, also erst mal: Mit welchen KI-Aspekten arbeitet Ihr Produkt? Weil...ja, da findet man immer nicht so viel raus. Ähm beziehungsweise es wird halt oft so gesagt: „Ja, wir arbeiten mit künstlicher Intelligenz“, aber das ist ja eigentlich so ein riesiger Sammelbegriff ja und ja...genau.

Genau ähm...das Produkt, das wir entwickelt haben, arbeitet mit vielen verschiedenen Aspekten. Also ähm, wir haben tiefe neuronale Netze, wir haben einfache neuronale Netze, wir haben „NLP“, also „Natural Language Processing“, ähm das haben Sie da in Ihrem Dokument als „Text Mining“ glaub ich aufgenommen **mhm**. Ähm...wir haben ähm Algorithmen drin, die sogenanntes „Fuzzy Searching“ machen, das bedeutet also äh äh such ich zum Beispiel ne Darreichungsform, ähm dann hab ich, dann hab ich ne relativ gut definierte Liste von der, von der, von den Behörden welche Darreichungsformen es gibt und dann kann ich über „Fuzzy Searching“ noch den menschlichen Aspekt abbilden, also dass man sich bei Tablet zum Beispiel verschreibt, ähm und und, und so „matchen“. Ich hab ähm, wir haben die Möglichkeit ganz, ganz trocken reguläre Ausdrücke, also „Pattern Matching“ zu benutzen **mhm**, ähm, das kommt immer auf den „Usecase“ an, was ich mache.

Ja, wenn ich ähm, wenn ich's verhindern kann, nehm ich natürlich keine neuronalen Netze oder schöne di...schöne Fach...äh schöne, schöne neumodische Wort „Deep Learning“, das ist ja im Prinzip dasselbe **mhm** Neuronale Netze und, und „Deep Learning“...ähm, versuch ich natürlich zu verhindern denn: ähm, wir bewegen uns zumindest bei uns in einer, in einer Branche, die, die ähm noch, noch sehr Papier getrieben ist und sehr analog getrieben ist, also selbst, selbst digitale Dokumente, selbst Digitalo..äh Di..Di...Digitalisierung läuft noch sehr analog ab. Ähm, wir sprechen da über PDF-Dokumente und Word-Dokumente. Ähm, daraus Trainingsdaten zu entwickeln, um so 'n neuronales Netz dann auch wirklich vernünftig trainieren zu können, ist super, super

schwierig **mhm**...ahm und dementsprechend...ähm ja das unterstützen wir, aber es gibt relativ wenig, zumindest in den Bereichen, wo wir uns bewegen, ähm Anwendungsfälle, wo wir sie wirklich effektiv einsetzen können.

Okay, verstehe, mhm. Und dann mal ganz ähm...abgesehen jetzt von Ihrem Produkt: Wie stehen Sie persönlich allgemein zu künstlicher Intelligenz?

Das ist so bisschen die Elon-Musk-Frage oder? **Ja (lacht)**. (Lacht). Ähm, ich seh das nicht so schwarz, wie der, wie, wie, wie, wie, wie der Elon, es ist äh, ich glaube es ist ne sehr, sehr große Chance. Ähm wir sind noch weit entfernt von der Technologie und auch von der Rechenpower im Moment, dass wir, dass wir uns da Sorgen machen müssten, dass wir irgendwas erschaffen, was wir nicht mehr kontrollieren können. Also es gibt ja auch dieses nette Buch, das „Singularity is near“ oder wie auch immer das heißt, ich weiß nicht mehr von welchem Autor das war, aber der hat, der, der zeichnet auch so 'n schwarzes Bild (holt Luft)...ähm, wir sind weit davon...ähm wir sind weit davon entfernt, dass KIs ähm, äh äh KIs ähm ne, 'n eigenes Bewusstsein entwickeln. Ähm... Wenn wir irgendwann mal in die Richtung, Richtung kommen, wenn's Quantencomputer gibt äh, die effektiv funktionieren und wir, wir die dann auch effektiv nutzen können. Wenn's ähm ähm äh algorithmische Modelle gibt wo's...sicher haben wir ne Verantwortung.

Sicherlich ist es wichtig, ähm...sicherlich ist es äh äh wichtig, dass äh äh das wahrzunehmen und dann auch äh dementsprechend zu managen. Was ich aber viel wichtiger finde ist, dass wir, ähm auch im Kleinen ne Verantwortung haben. Das wird immer, wird immer gerne mit ahm äh „Accuracy“ und „F1 Scores“ und, und, und, und die Qualität der extrahierten Daten um sich geschmissen. Man, man hat aber als...als Hersteller einer KI in wie auch...inwieweit auch geartet, immer ähm finde ich den Auftrag, dann auch aufzuklären, was das denn jetzt bedeutet...ja? **Mhm**.

Ähm, was bedeutet denn die „Accuracy“, bedeutet jetzt äh 'n „F1 Score“ von 80% und acht...dass, dass äh dass die KI nur 20% äh auf, auf 100 Dokumente an Fehlern macht meint es nicht! Und...dieses Aufklä...diese Aufklärung...ähm ist glaub ich..., das Verfehlen dieser Aufklärungen ist das größte Risiko. **Mhm**. Das seh ich im Moment auch tatsächlich, grade, grade aus Marketing Sicht ist es halt immer so, (Räuspern) da wird halt gern, da wird halt gerne mit großen Zahlen um sich geschmissen und große Versprechungen gemacht...ähm, das Risiko besteht, aber dass wir da jetzt in, in, in..., in ein (lacht) in „Doomsday Szenario“ reinlaufen, das glaub ich nicht, ich seh da eher große Chancen...**Okay**.

Es ist halt so: wir haben immer größere Datenmengen, wir haben immer mehr Daten, die, die verarbeitet werden müssen...ähm die, die SAP ist 'n, ist 'n gutes Beispiel, die, die machen...die machen relativ großen äh äh großen Teil ihres Umsatzes mittlerweile

nicht mehr mit, mit dem eigentlichen Produkt SAP, sondern mit ihren neuen Datenbankentechnologien, die „Big Data“ vernünftig verarbeiten können. Ähm...da ist es gar nicht mehr möglich und, und wir sind mittlerweile in einer Komplexität angekommen, was unsere Daten angeht, wo's gar nicht mehr möglich ist, das manuell zu verarbeiten. Das heißt ja, ich seh da ne Chance, ich seh da sogar, ich seh da sogar ähm äh äh ich seh da sogar dass es vonnöten ist, dass wir **mhm** dass wir uns diese Technologien dahin arbeiten.

Okay, sehr interessant! Ahm und dann, wie schätzen Sie die Bemühungen in Deutschland ein zu dem Thema? Also äh wird genug getan, um die Entwicklung voranzutreiben, also grade auch so gegenübergestellt zu den USA und China?

Ich würd da, ich würd da fast ne Gegenfrage drauf stellen: was tun denn die USA und China? Ähm ich glaube, das ist kein äh, das ist kein Auftrag, der, der, der von der Landesregierung irgendwie...ähm geschaffen oder geb...oder, oder ausgelöst werden sollte, sondern das ist Interess...das ist Interesse der Privatwirtschaft, ähm da was auf die Beine zu stellen. Was wir in Deutschland haben, was wir in Europa haben, sind die klassischen Diskussionen. Wir sind, wir sind unglaublich konservativ und fangen an, wieder alles zu regulieren, ja? **Ja.**

Ich warte...ich wart schon drauf, dass wir, dass wir äh äh, dass wir 'n deutsches Gesetzbuch zu, zu KI-Entwicklung bekommen...ähm und wir werden das sicherlich in Deutschland auch wieder überregulieren. Bin ich mir hundertprozentig sicher und ähm...werden anfangen ähm äh äh Türen zu schließen. **Mhm.** Ähm aber...zur Entwicklung als solches: Wir haben das Fraunhofer Institut, die machen nen super Job, ja. Wir haben keine deutschen Firmen, die's nutzen, das muss man ganz klar sagen, das ist wie auf 'm Automobilmarkt, äh wir sind da abgehängt worden **ja** und äh wir sind auch soweit abgehängt worden meiner Meinung nach, dass wir nicht mehr dazu in der Lage sind, das einzuholen. **Mhm.** Ähm, wir haben aber, wir haben eben, eben so Leithausinstitutionen wie das Fraunhofer Institut, wie, wie TU München, die sehr viel Forschung und sehr, sehr viel Entwicklung in die, in die ähm Thematik reinstecken...aber ähm...mehr kann ich auch nicht erwarten von, von dem Standort Deutschland. Außer dass es äh äh dass es, dass es ähm äh Unternehmer geben muss, die das Potenzial dann auch nutzen. Und das haben wir leider nicht.

Okay, mhm... Gut, dann kommen wir zum dritten Block, zu Ihrem Produkt. Ähm, also erstmal: Wie kam's dazu, dass das Produkt entwickelt wurde? Also die Idee oder die Initiative.

Ja die initiale Initiative war, war das ähm IDMP Produkt...äh Projekt von der ÄHMA...also das waren jetzt zwei, äh zwei Akronyme in einem. Die EMA ist die „European medicines agency“, die übergeordnete Behörde in Deutschland. IDMP ist die Initiative, nennt sich äh äh „Identification of Medicinal Products“, die dazu...darauf abzielt, ähm strukturierte Daten über jedes medizinische Produkt in Europa zugelassen, ähm in eine Datenbank zu...ähm...äh zu...aufzunehmen und da ne Datenbank darüber zu führen. Und jetzt sind die, sind Pharmafirmen, unsere Kunden, ähm dazu äh...ja dazu gezwungen, ihre Daten in dieser Struktur ähm zur Verfügung zu stellen, das ist ne Art Datenbankstruktur mhm, haben aber nur PDFs. **Okay.** Das muss man..., ne Zeit lang war's auch diskutiert – ich weiß jetzt nicht, wie der aktuelle Stand ist – dass man das auch für Bestandsprodukte machen müsse..., muss..., muss und da hat so ne Bayer, mit ihren hunderten Produkten, tausenden Produkten in Europa zugelassen halt schon 'n echtes Problem. Mhm. Ahm da, das war die Idee dann zu sagen okay, ähm wir nehmen uns die Dokumente vor, äh die Packungsbeilagen beispielsweise, die SmPC, das ist die Packungsbeilage für den Arzt, ähm und, und bestimmte andere Dokumente und extrahieren dann halt die Datenpunkte aus diesen PDFs raus mhm, ähm und reduzieren dadurch dann den Aufwand. Das war so die Initialidee.

Mhm! Und wie lange hat das gedauert, das Ganze zu entwickeln?

(*Schnauft*) Wir sind noch gar nicht fertig (lacht). (**Lacht**). Wir haben jetzt mal so ne erste äh äh wir haben jetzt mal, ja die ersten zwei Versionen, aber äh es gibt da, es gibt noch viel Potenzial, also wir entwickeln noch weiter. Aktuell ähm, haben wir fünf Jahre Entwicklungszeit eingestellt.

Fünf Jahre, mhm. Und ungefähr wie viele Personen braucht es um so 'n Produkt zu entwickeln?

Ich hab mal irgendwann geschätzt, ähm als es jetzt äh Anfang des Jahres übernommen wurde, hab ich ähm, musst ich mal ne Schätzung abgeben wieviel Mannjahre...äh reingeflossen sind. Daraus lässt sich die Teamgröße dann eigentlich ganz gut ablesen. Sind ungefähr 55 Mannjahre reingeflossen **ouh ja heftig (lacht)**. Liegt aber, ja, aber liegt auch daran, dass wir zu 'ner Zeit angefangen sind, als Textmining noch sehr uninteressant war. Das war noch die Zeit ähm, wo Google immer mal, also diese...ja wo Google immer mal auf irgendwelchen Webseiten diese „Captures“ laufen hatte, um, um festzustellen, ob auf 'nem Bild ne Katze ist **mhm**, oder ne Brücke oder Autos. Diese Vision-Geschichte war halt, äh war halt eigentlich der Fokus ähm in der, in der, in der ähm KI-Community. Ist es bis heute. Textmining ist da einfach noch so bisschen stiefmütterlich und da muss man dann halt auch noch viel Arbeit selber reinstecken.

Mhm okay, verstehe. Gut, und ähm...ja wie das Produkt lernt, hatten wir ja jetzt vorhin schon angesprochen ja mit dem „Deep Learning“ und so. Und ähm...ja. Ich denke Näheres zur Funktionsweise ist halt schwierig, also das wird dann wahrscheinlich sehr technisch auch oder?

Da würde es dann relativ technisch werden **mhm**. Also von 'ner...ich kann mal kurz, kann, kann das mal kurz ähm äh äh umreißen **mhm** wie so 'n Projekt normalerweise abläuft. Ja es gibt halt ne Implementierungsphase, ähm und ne Trainingsphase, wo wir, äh das Modell ähm, wenn's nicht 'n Standardmodell ist, eben für den Kunden aufbauen und trainieren **mhm** und ähm, es gibt dann ähm die Möglichkeit eben direkt im Dokumentmanagementsystem zu integrieren oder ne eigene 'n eigenes „User Interface“ zu benutzen, wichtig ist dann eben, dass ähm, die Software macht ne „Prediction“, also extrahiert Daten oder, oder kategorisiert 'n Dokument (für ihr..?) und ähm schickt das zurück an, an die UI und der, der Mensch, der dann diese, diesen Vorschlag bekommt, kann ihn dann noch korrigieren. Diese Korrigier...äh diese Korrekturen kommen dann zurück ins System und wir können dann kontinuierlich daraus lernen und werden immer besser. **Mhm.**

Ähm das ist, das ist dann eben so 'n Zyklus der nach dem Live-Gehen des Pro...Pro...Projekts ahm kontinuierlich durchläuft, um einfach...ähm zwei Sachen zu machen äh a) immer besser zu werden, aber: ähm, es hat sich gezeigt, dass äh so 'n, so 'n so ne Dokumenten-Taxonomie, also ähm wie Dokumente abgelegt werden, wie, wie Metadaten strukturiert sind, nicht statisch bleiben, sondern es entwickelt sich über die Zeit. Und ähm, durch kontinuis...kontinuierliches Lernen, ähm schaff ich's dann natürlich auch ähm, die Technologie und das, das, das AI-Modell im Hintergrund ähm relevant zu halten. **Mhm.**

Und also das Produkt ist dann schon sehr...ähm, wie soll ich sagen...ähm...so ausgelegt, dass es ähm eben für jeden Kunden individuell angepasst wird? Ja. Also ist es jetzt nicht eben das eine Standardding und des baut sich dann jeder... Also...ja ähm, es kommt drauf an, wem man jetzt glaubt, mir oder der Marketing-Mesage (Lacht**). Ähm es wird ähm, natürlich werden immer Standards verkauft, aber es hat sich gezeigt, dass äh...ja, äh 'n Kunde kann auch 'n Standard kaufen ähm, nur wenn das Ding dann implementiert ist, dann...(schnauft) gerade in Europa ist es so, dass die Leute noch sehr skeptisch sind. **Mhm.****

Ja, die, die (*Räusperrn*) die kaufen dann am liebsten 'n, 'n Stück Standard-Software, was, was halt, keine Ahnung 70% von dem tut, was sie denn, was sie denn wollen, einfach mal um zu kucken, ob's geht und wenn dann bewiesen wurde, dass es geht, äh dann gehen die Lichter an und dann fangen die Leute an, die abgefahrensten Usecases

rauszuholen und äh konfigurieren zu wollen und dann ist es halt so, am Ende des Tages bleibt nichts Standard **mhm**, ja ist halt...ja.

Ok, ähm...ja und was glauben Sie, wie sich die Nutzungshäufigkeit bei solchen Produkten in der Zukunft entwickeln wird? Also wird das irgendwann so ein Standard sein, wie heute...was weiß ich...Word und so.

Ist ne gute Frage. Ähm das kommt drauf an, wie schnell die Industrie den Wert ihrer Daten und, und akkurater Daten erkennt. **Mhm**. Ich glaube, da sind wir noch nicht, und ähm...da kommen wir aber irgendwann hin. Sobald das passiert ist, wird das, wird das Standard werden, denn schauen Sie sich mal 'n Dokumentenmanagementsystem heute an in der, in 'nem großen Unternehmen. Da gibt es, da gibt es Datenstrukturen, da gibt es ähm Taxonomien, da werden Berater bezahlt, die kriegen, die kriegen ähm...abgefahrene Gehälter, ähm um da, um da so 'n System zu designen und am Ende ist die Sprache von jedem Dokument trotzdem auf Englisch, weil das ist das Default, was eingestellt wurde.

Am Ende des Tages geht keiner hin, solange er nicht gezwungen wird dazu und selbst wenn man, selbst wenn, wenn, wenn große Unternehmen am Anfang ganz viele Ansprüche haben und alles äh äh alle User zwingt Daten einzugeben, werden die eben solange meckern, bis das wieder, wieder abgestellt wird, der Zwang, ähm und dann sind die Daten eben hin. Das heißt, um solche, allein um sowas zu machen, allein um (*Räusperrn*), allein um Datenqualität, äh zu heben, um Datenvollständigkeit zu heben, äh wird das irgendwann zwingend notwendig sein, so ein System zu haben.

Mhm, okay. Führt kein Weg dran vorbei. Gut und äh ja, dann haben Sie ja grade schon gesagt, dass ähm Sie jetzt...also das Produkt immer noch weiter entwickeln. Ähm, was sind denn so die größten Punkte, die Sie vielleicht seit dem Anfang verändert haben?

Naja, am Anfang gab's diesen einen Usecase, eben diese IDMP-Daten aus diesen drei, vier Dokumenten rauszuholen **mhm**, ähm das war alles, was das Produkt konnte und darauf war's ausgelegt. Ähm mittlerweile sind wir soweit, dass wir ähm, wenn 'n Projekt mit genügend Budget da ist (lacht) – das ist immer ne Budget-Frage. **(Lacht.)** Ähm im Prinzip, ähm alle Dokumente nehmen können und, und ähm, wenn die Daten denn in diesem Dokument ähm vorhanden sind, ähm auch extrahieren können, also wir können, es ist mittlerweile extrem flexibel geworden, wir können ähm, wir können eben auf alle Dokumententypen, auf alle Dokumenten...ähm Formate mittlerweile anpassen und äh Daten extrahieren. Ähm...ich glaube das ist das größte ähm was ich seit der ersten Version verändert hat.

Wir haben viele Algorithmen hinzugefügt und viele Konfigurationsmöglichkeiten, Integrationsmöglichkeiten, aber das ist alles... Geplänkel im Gegensatz zu, zu dem äh zu dem, zu dem Fakt, dass wir anpassen können jetzt und ähm... ja eben nicht nur Daten extrahieren, sondern auch, naja wir haben teilweise Projekte gemacht ähm, wenn man im Hinterkopf behält, dass die Idee war, Daten für IDMP rauszuholen, haben wir jetzt grade zwei Projekte gemacht, wo wir Dokumente von, von 'nem äh, von 'nem, von 'nem relativ großen Laufwerk, wo alles reingeschmissen wurde über Jahre, in ne Taxonom... in ne Taxonomie in 'nem Dokumentenmanagementsystem migriert haben und automatisch klassifiziert haben alle Dokumente. Also das ist schon, es ist schon was ganz anderes geworden ja.

Mhm. Und haben Sie auch ähm, also hatten Sie gleich zu Beginn diese über zwanzig Sprachen oder kamen die auch so nach und nach dazu?

Ne, die hatten wir gleich zu Beginn, weil diese SmPCs die liegen eben in allen europäischen Sprachen vor **mhm**, das ähm, das war so. Ähm... das würd ich jetzt auch nicht als große... also aus, aus KI-Sicht nicht unbedingt als großes Feature werten **mhm** also aus Marketing-Sicht ja, aus KI-Sicht ist es eigentlich nur noch 'n... ja dasselbe Modell nochmal in ner neuen Sprache trainiert **ja, okay** also... ja.

Gut und ahm... ja die Zielgruppe, bzw. Ihre Kunden sind halt Pharmafirmen oder? Ja, Pharmakonzerne. Gut, ähm... ja und könnten Sie sich noch andere Einsatzorte für solche Produkte vorstellen, also jetzt außerhalb von Dokumentmanagement in Unternehmen?

Definitiv! Ähm, wir haben (*Räuspfern*) schon einige Piloten gemacht. Ähm wir, wir arbeiten mit 'ner großen französischen Firma zusammen ähm die ähm... macht ähm äh Entscheidungsgrundlagen auf Basis **mhm!** von Dokumenten, die sie ähm, die sie von 'ner Behörde bekommen. Ähm also da geht's dann in Richtung äh Prozessautomatisierung. Wir haben 'nen Piloten gehabt, mit 'nem ähm, mit 'nem anderen ähm Hersteller, die ähm, ähm... die... Nebenwirkungsmeldungen auswerten mussten und das automatisieren wollten. Also es geht da, ähm sehr, sehr schnell in Richtung... Prozessautomatisierung und Prozessunterstützung ähm, außerhalb von Dokumentmanagementsystemen **mhm** also wirklich in den Day-to-Day Job von Kunden ja.

Okay. Aber trotzdem alles auf jeden Fall Pharmabereich? Ja. Okay. Gut, dann kommt Block vier: Zukunft (lacht). Ähm, sehen Sie für die Zukunft Möglichkeiten, das Produkt noch effizienter zu machen, also z.B. auch mit Technologien, die es jetzt grade noch gar nicht gibt?

Ja, wie technisch soll die Antwort denn werden? **(Lacht)**. (Lacht) **So, dass ich's noch einigermaßen verstehen kann (lacht)**. Ja ich glaube, dass wir, ich glaube dass wir ähm aktuell noch äh wir sind, wir stecken da noch ganz, ganz, ganz in den äh ganz tief in den Kinderschuhen irgendwie drin. Ähm, ich denke, dass...*(schnauft)*, dass wir mit, mit steigender Rechenpower, äh Modelle entwickeln können, die dann auch, auch, auch äh Quali..., Extraktionsqualität und, und, und Klassifizierungsqualität ähm erzeugen, von denen wir heute nur träumen. Wenn wir dann, wenn ich so, so richtig futuristisch werden, werden soll, ähm Google hat, Google hat den ersten Quantencomputer jetzt grade am Laufen ja. Wenn wir das, wenn wir das ähm unter Kontrolle kriegen, wenn wir das effizient hinbekommen und zu 'nem vernünftigen Preis, dann, dann reden wir ähm äh reden wir über ähm über Modelle und Datenmengen, die wir verarbeiten können auf 'ner ganz anderen Magnitude.

Ähm, wir sind, wir sind aktuell, ähm einfach von den Daten erschlagen, die wir haben. Ich hab vorhin darüber so 'n bisschen *(Räuspern)* 'n bisschen mich beschwert, dass es zu wenig Daten gibt, aber die Daten, die wir haben, die überlasten das was wir haben ja schon. **Ja**. Es gibt teilweise Modelle, da rechnen äh, da trainieren wir über zwei Wochen und nach zwei Wochen weiß man dann, oh... hat nicht so ganz funktioniert **mhm**. Ja das heißt also, ähm wir, wir, wir sind noch im Moment sehr, sehr langsam unterwegs, was ähm, grade was „R&D“ angeht, um da, um da die Forschung auch so 'n bisschen voranzutreiben, ähm welche Modelle sind kombinierbar miteinander, in welchen Datenmodellen? Ich kann später ähm, wenn ich, wenn ich höhere Rechenkapazitäten habe und, und schnellere, schnellere Systeme, einfach viel schneller Experimente machen und einfach viel größere Modelle aufbauen. Das ist glaub ich das, das was aktuell, ähm...die ganze...das ganze Thema KI noch zurückhält.

Ok, also einfach... mehr Power (lacht).

Unvorstellbar, ne? Nachdem man, nachdem unsre Handys mehr Power haben als was die NASA zur Verfügung hatte, um zum Mond zu fliegen. Aber so ist es halt **ja**. Ähm wir, wir erschlagen die Probleme im Moment mit Rechenpower und ähm... ja ähm generieren noch mehr Daten (lacht) also. **Ja**.

Okay. Ahm ja, was glauben Sie, wie das Dokument der Zukunft aussehen wird?

Transitiv. Ähm ich glaube, dass wir... ähm, ich glaube, dass wir..., ich glaube, dass die Zeit, ich hoffe, dass die Zeit von Word und Adobe abgelaufen ist. **Mhm?** Ähm, ich glaube, ich denke, dass wir irgendwann ähm an den Punkt kommen, wo ähm... wo wir strukturierte Daten im Hintergrund nehmen können und transformieren... „on the flight“ transformieren können. Ähm...einmal, um den Konsum durch... durch Menschen, ähm äh äh

zu, zu „enablen“, aber gleichzeitig dann auch eben durch, durch ähm äh... ja durch, durch Computerschnittstellen, die miteinander sp... miteinander Daten austauschen und sprechen.

Ähm, das ist im Moment ja das große Problem, was wir haben, deswegen gibt's die, gab es die Cunesoft oder das, das Produkt der Cunesoft um überhaupt ähm um überhaupt da äh äh ähm die Möglichkeit herzustellen, aber das sind dann auch immer nur zeitliche Schnappschüsse. Daten entwickeln sich weiter, ob's jetzt 'n Dokument ist oder ähm oder in einer Datenbank. Und das muss äh, das müssen wir in den Griff kriegen, dass das strukturiert ahm gespeichert und, und verfügbar bleibt. Ich glaube... dementsprechend ist meine, meine Antwort: transitiv werden Dokumente aussehen.

Okay! Gut, vielen Dank! Ahm und ja, das haben wir vorhin auch schon ein bisschen angesprochen. Ahm glauben Sie, dass in Zukunft gezielte KI-Gesetze nötig sein werden? Also wie z.B. 'n Hippokratischer Eid für Programmierer? (lacht herzlich) (lacht).

Ein Hippokratischer Eid für Programmierer! **Ja, also grad wenn's auch so um Ethikfragen geht und sowas.** Ich verstehe, ich verstehe worauf Sie hinauswollen damit ja, ja. Ähm ich muss nur ein bisschen lachen also, ähm... es gibt ja auch keinen Hippokratischen Eid für Politiker. **(Lacht) Das stimmt allerdings!** Ähm... es ist immer... ja... vielleicht. Ich weiß es nicht. Es kommt drauf an. Aktuell sind wir so weit davon entfernt, dass wir sowas brauchen **mhm.** Ähm, im Moment würd ich Gesetze in die Richtung ähm eher, eher als schädlich und, und hindernd empfinden. Aber ja, es kann sein, dass es irgendwann mal soweit kommt und es kann sein, dass wir irgendwann mal, ähm... ja so ne Art Atomaufsichtsbehörde für Programmierer brauchen oder KI-Programmierer brauchen. Kann sein, **ja** ähm, aber ich bin mir relativ sicher, dass wir das nicht mehr miterleben.

Oh okay (lachend) ja, das hoffe ich! Ähm, ja und da passt dann auch noch die letzte Frage dazu: Also glauben Sie, dass es irgendwann mal möglich sein wird, ne Intelligenz zu erschaffen, die dem des Mensch..., die der des Menschen (lacht) ebenbürtig ist?

Ha, ja das ist die F..., es kommt drauf an, es ist..., ich glaube Sie werden auch immer verschiedene Antworten kriegen, wo Sie auf der Welt die Fragen stellen. In USA oder in England ist es glaub ich **Ja, auf jeden Fall!** bisschen einfacher, ähm äh (ironisches Räuspern) äh ne ebenbürtige Intelligenz zu erstellen **(lacht)** ich weiß es nicht, ähm... äh nein. Das glaub ich nicht. Das glaub ich ehrlich gesagt nicht. Ich glaub die Singularität werden wir niemals erreichen können. Ähm, man hat zu viel Bewusstsein. **Ja, glaub ich**

auch und dieses Bewusstsein kann man nicht programmieren und das kann man auch nicht algorithmisch darstellen **ja**.

Denn, ich bin der Meinung, dass Bewusstsein die Summe aller Erfahrungen äh äh ähm, darstellt, die man in seinem Leben und das aller Datenpunkte, die man in seinem Leben gesammelt hat, so viel kann ich gar nicht künstlich erzeugen. Was will ich denn auch künstlich erzeugen? **Ja**. Also, ähm ja das ist immer 'n schönes Gedankenexperiment. Ahm..., es gibt, es gibt sehr berühmte Menschen und sehr, sehr intelligente Menschen, die glauben, dass sowas geht, wie, wie, wie, wie der Herr Musk beispielsweise **mhm**, ähm aber ich glaube das ist alles, das ist alles sehr viel Theorie und ähm, ähm..., nein, das glaub ich nicht. **Okay. Ja, bin ich bei Ihnen (lacht).**

(Es folgt ein organisatorischer Abschnitt, der inhaltlich nichts mit dem Interview zu tun hat und daher nicht transkribiert wurde).

Dann dank ich Ihnen wirklich ganz herzlich für Ihre Zeit! Kein Problem und die Beantwortung der Fragen und ja genau, ich meld mich dann wieder. Jo bis denn, ciao. Bis dann, tschüss!

Transkription Interview mit Paul Caspers, Theum

Husum, 15.07.2020

Interviewerin: Isanne Baumann (Fettdruck)

Interviewter: Paul Caspers (Normaldruck)

Gut, dann ähm, bevor wir mit dem ersten Block starten, ähm, würde ich gerne wissen, können Sie vielleicht noch mal einfach schildern, was Ihre Aufgabe bei „Theum“ ist und ähm, wie so bisschen Ihr beruflicher Hintergrund aussieht?

Also, meine Aufgabe bei „Theum“ ist, ich bin ja Unternehmer, ich hab‘ die Firma gegründet, das war 1994 **mhm** unter anderem Namen, inzwischen heißt’s ja dann „Theum AG“, äh, damit bin ich Unternehmer, ich bin auch Mehrheitsgesellschafter und natürlich Vorstand, oder mal anders formuliert, praktisch der oberste Vorturner, **(lacht)** äh, vieles von dem, was wir gemacht haben, geht im Endeffekt dann auf mich zurück. Ich hab‘ noch ‘nen amerikanischen Partner auch in der Firma, der auch Vorstand ist seit kurzem, und ja, dann hab‘ i halt alle Aufgaben, die man so als Unternehmer und Geschäftsführer hat. In ‘ner kleinen Firma ist das ja nicht nur Management, sondern da is man ja im Prinzip der, der erste Mann, der Projekte realisiert, Software baut und und einige andere Dinge auch tut. **Mhm.**

Äh, der berufliche Background, ich hab‘ mal studiert in Regensburg Wirtschaftswissenschaft und Geographie **mhm** bin damals durch ‘n Zufall im Rahmen von meiner Diplomarbeit, äh, auf d..., auf Computer gestoßen, damals waren das noch ganz große Ungetüme, turnhallengroß **(lacht)**, und das hat mir sehr, sehr gut gefallen damals und dann bin ich im Endeffekt sozusagen gar nicht äh in, in meine eigentliche Profession gegangen, sondern bin dann sozusagen im, im Computerbereich geblieben, hab dann so 10, 15 Jahre lang bei verschiedenen Unternehmen gearbeitet und dann aus..., vor dem Background im Endeffekt die Firma gegründet.

Okay, mhm. Also hatten Sie jetzt so mit ähm... sag ich mal so Programmieren oder so jetzt gar nix am Hut?

Doch, da hab ich a ganze Menge gemacht, also in der Diplomarbeit hab ich schonmal was äh, programmiert, das waren damals noch so Sprachen wie FORTRAN, COBOL, Pascal und so Zeug **aha**, da hab ich dann so ‘n Landschaftshaushaltsmodell entworfen, wo man praktisch Energiebilanzen simuliert und ähnliche Sachen und und hab eben da sehr, sehr viel programmiert **mhm** ähm und danach bin ich im Prinzip als Instruktor gegangen bei verschiedenen Schulungsinstitutionen und äh, und hab dort, äh, Kurse gegeben für Programmiersprachen.

Ah ok! Ja dann sind Sie ja schon...(lacht)...

Also, aus dieser Ecke raus bin ich dann als Angestellter zu so großen Unternehmen gegangen und hab dann dort PC-Einführungen gemacht, damals kamen ja grad die PCs auf, das hat mir dann gut gefallen. Dann haben wir auch immer wieder programmiert, da kamen so „Fourth-Generation-Languages“ auf grad ganz neu, Expertensysteme, ja und so, da hab ich meine ganzen IT-Kenntnisse sozusagen über die Jahre und Jahrzehnte aufgebaut **mhm** und das hat, äh, sehr viel Spaß gemacht, ja.

Ja, dann haben Sie das ja so von den Kinderschuhen an mitgekriegt, okay, (?) grade an kann man sagen, genau **(lacht)**.

Gut, dann, ähm, würd ich mit dem ersten Block starten, die Begriffe „Dokument“ und „Daten“, ähm... ja, da hab ich ja geschrieben, dass es viele verschiedene Interpretationen des Begriffs „Dokument“ gibt, was ich jetzt ja bei einer Recherche au rausgefunden hab, ähm, ja manche gehen halt so voll auf die Form ein, also wie dieser indische Bibliothekar, der eben gesagt hat, ähm: „Gedanken auf Papier oder ‘nem anderen Material“.

Ähm, und zum Beispiel eben die Suzanne Briet, die oft erwähnt wird, ging ja eher so auf den Inhalt ein, bzw. ganz stark darauf, sobald was in ‘nem Dokumentensystem drin ist irgendwie und katalogisiert ist, ähm, ist es ‘n Dokument, also sogar Tiere, ihrer Meinung nach, und... ja, da wollt ich einfach mal wissen: Wie definieren Sie ein Dokument? Ja.

Also wir sind jetzt natürlich nicht von der Wissenschaft, sondern wir sind natürlich oft auch sehr pragmatisch hier als Firma **ja**. Soweit, dass ‘n Tier ‘n Dokument ist würd ich jetzt nicht gehen **mhm**. Für uns ist a Dokument irgendwas..., wir sagen immer Menschen sind so narrative Wesen, ja **ja**, wenn ich irgendwas erzähle, dann hab ich irgendwas im Kopf, ne Reihenfolge, ne Gliederung, wie ich irgendwie den Inhalt rüberbring, wie ich’s spannend mach, wie ich das verständlich aufbaue, äh, also so narrative Wesen sind wir, und das, was wir dann erzählen zu irgend ‘nem Thema, das kann man auf verschiedenen Medien festhalten.

Also ganz früher waren das mal Steintafeln und ägyptische Hieroglyphen, dann waren’s meinethalben Papyrusrollen, irgendwann hat man das Ganze mit, mit Füller und Tinte auf, auf Papier geschrieben, dann kam der Druck, wo man das dann, äh, drucken konnte, dann kamen Schreibmaschinen und ganz am Ende dann die Computer, dann kann’s auch nur digitale, äh, Aufbewahrungsmedien geben, also nur elektronische und ‘n Dokument kann dann auch was sein jetzt heute Papier, das gescannt ist **mhm**, das mag noch sein, äh..., irgend..., die waren ja da recht fleißig, die haben sehr viele PD...,

digitale PDFs, die waren nicht gescannt, sondern irgend aus'm „Word“ oder was weiß ich heraus erzeugt **mhm** also praktisch ne Art Print-Endprodukt, aber digital und bis hin zum klassischen „Office“, Geschichten, wo man das speichert oder auch, was heute häufig ist so als „LICKY“ oder in „HTML-Form“, „XML-Form“, in irgendwelchen Intranets oder Webseiten usw. Also das Ganze wär, was wir als Dokument betrachten **mhm**. Und ein wichtiges Thema ist auch immer, das hat so ne Art Prosa-Charakter, eben dieses narrative **mhm**, steht irgendwo am Anfang (?), hat oft ne Gliederung oder 'n roten Faden oder irgendwas und dann wird irgend ne Geschichte oder 'n Thema beschrieben, äh, sozusagen von A-Z.

Okay, also würden Sie jetzt zum Beispiel auch nicht sagen, dass jetzt Audio-Dokumente - ähm, ich sag schon Dokumente - Audio-Dateien, ähm, Dokumente sind? Mja, das ist ein Grenzbereich, also technisch können wir jetzt heute mit unserer Software Audis..., Audios nicht so erschließen **mhm** äh, dass man praktisch dann sozusagen den Text rausholt und suchen kann, ja **mhm**, das, das Gleiche gilt ja auch für Bilder, wenn, wenn ich jetzt 'n Foto hab, wo irgendwie 'n Name draufsteht oder irgend ne Beschreibung, was ich in dem Foto seh, ähm äh, für mich ist das 'n Foto, hat Metadaten, mehr können wir da nicht machen.

Aber der Schritt, dass man jetzt hergeht und sagt, ich hab ne, ne Bilderkennung, die jetzt, äh, erkennt, was auf 'm Foto drauf ist und vielleicht auch noch die Texte scannt, der ist nicht allzu groß, das ist für uns im Prinzip, sagen wir mal das wären 'n paar Wochen oder Monate Arbeit und das Gleiche gilt natürlich dann für 'n Audio- in Anführungszeichen „Dokument“ **mhm**, das heißt... und wenn ich z.B. so über Bilderkennung sprech, da machen wir grad was, wir scannen, wir können ja sowieso schon gescannte Dokumente verarbeiten **ja**, das haben Sie wahrscheinlich gar nicht bemerkt oder so, weil wenn se gescannte PDFs reintun, die werden von (?) verarbeitet und dann haben wir (?) „OCR“ auch so mit dabei **mhm**.

Und jetzt als Nächstes herzugehen und zu sagen, äh, ich erkenne auch bei Bildern a), wenn da 'n Text drauf ist, das ist „OCR“ oder b), wenn kein Text drauf ist, erkenn ich irgendwie mit KI-Software, da ist jetzt 'n Rennradfahrer drauf oder weiß der Teufel was, das, das ist kein großer Schritt und dasselbe für Audio, d.h. die würd ich noch dazutun **okay**. Das ist ne andere Art von Speicherung, sozusagen jetzt statt Bits und Bytes irgendwo in PDF-Format oder Ähnliches, ist das, ist das als Audiospur aufgezeichnet und man bräuchte ne Software - die gibt's praktisch fertig - die das, was da einer spricht, äh, praktisch in, in Text umsetzt, den wir wieder verarbeiten **okay, mhm**. Also ja, das wären Dokumente, auch wenn man die heute nicht mit unsern Software defacto heute nicht unterstützen, Audio- und Bilderkennung.

Ja okay, verstehe. Ähm, und machen Sie dann bei Ihrer Definition ‘nen Unterschied zwischen analogem und digitalem Dokument?

Analog meinen Sie sowas jetzt wie Audio oder so?

Nee, analog mein ich jetzt, ähm, also eben ‘n Blatt Papier wirklich, das man in der Hand hält.

Ne, ne, also n‘ gescanntes PDF oder normales PDF, das erkennen Sie unter Umständen auch gar nicht, wenn Sie nicht irgendwie gezielt reinschauen, ist ‘n PDF, das legen Sie rein, wir erkennen dann, dass es ‘n gescanntes Ding ist und verarbeiten das. Es gibt allerdings natürlich bei den Möglichkeiten, was man alles rausziehen aus den Inhalten Unterschiede **mhm**, also aus‘m gescannten Dokument können wir prinzipiell nicht so viel rausziehen, wie in ‘nem, wie in ‘nem Word, wie in dem gleichen Ding als Word-Dokument oder, oder digitales PDF drin ist, da kriegen wir mehr Informationen raus **okay, gut**. Ja, also beispielsweise, äh, gescanntes Zeug ist ja eigentlich nur ne, sind ja nur Fotos, ja. Äh, aus ‘nem Word-Dokument krieg ich normalerweise ne Struktur raus **mhm**. Überschriften als Beispiel und Unterüberschriften und anhand von diesem Überschriftenzeug kann ich ja alles Mögliche erkennen. Aus‘m ge...gescannten Dokument krieg ich unter Umständen, krieg ich eigentlich normalerweise diese Strukturelemente gar nicht mehr raus **mhm**. Ich weiß nicht, was ne Überschrift ist und wenn, wenn ich das ermitteln wollte, was ne Überschrift ist, dann ist ‘n irre großer Aufwand, weil ich erkenn die Überschriften nicht mehr an ‘nem Paragraphenstil **mhm** oder ähnlichen Sachen, sondern ich muss praktisch das anhand von irgend ‘nem Muster oder an optischen Kriterien erkennen und die unterscheiden sich dann wieder bei jedem Dokument **mhm, ja klar, mhm**, Überschriften dick, fett und groß geschrieben und... woanders ist es nicht oder die Überschrift hat irgend ‘n Muster, wie: „Erstens, Punkt, Text, Zweitens, Punkt, Text“ und so weiter, aber der Nächste gliedert’s anders und ohne Nummern das ist also sehr, sehr komplex. Also unterm Strich kann man sagen, das Beste ist, wenn ich die Quellen verarbeiten kann, also Word und ähnliches Zeug und das Zweitbeste ist, ich hab digitale PDFs, wenn’s geht mit so Bookmark-Verzeichnis und das Dritte ist, ich hab praktisch nur fotografiertes, gescanntes Zeug **mhm**, dann kann man mit „OCR“ arbeiten und wir kriegen halt alles das raus, was das „OCR“ erkennt, aber nicht mehr.

Okay. Gut, verstehe. Ähm, und machen Sie dann ‘nen Unterschied zwischen Dokument und Daten und wenn ja, was für ‘n Unterschied? Kann man das überhaupt unterscheiden?

Also was wir..., ja das sind also Grenzbereiche, also was wir da, äh..., bei Daten sag ich mal ganz krass das wär jetzt ‘n Datensatz, da steht drin: „Vorname, Nachname,

Postleitzahl, Wohnort“ oder irgend sowas, ja, **mhm**, im Dokument kann das Gleiche auch drinstehen als Prosatext **mhm**. Aber was wir immer gern machen, wenn wir das Zeug transformieren, also das, was Sie da reingelegt haben als PDF und sonstige Dateien, das tun mir ja zerhackeln, in Atome und so weiter zerlegen, äh, normalisieren, strukturieren und so weiter, dann speichern wir's in, in 'ner Datenbank, in dem Fall in „Elastic“ auch und im Endeffekt erzeugen wir im Kern strukturierte Daten, soweit es irgendwie möglich ist **ja** aus den Dokumenten.

Das heißt wir haben jetzt praktisch so ne Art Datensätze pro Atom und Kapitel und sofort, wir haben da bestimmte Schlüsselbegriffe drin in diesen einzelnen Atomen, soweit wir die erkennen. Wir hätten, hätten auch noch ne „Pattern-Matching-Language“, das haben Sie natürlich jetzt in Ihrem Test nicht gesehen. Da können wir z.B. So Muster erkennen in Texten, also, wenn da irgendwas steht: „Name, Doppelpunkt, Kontonummer Doppelpunkt“ usw., dann können wir anhand dieser Muster praktisch ne, ne Kontonummer, 'nen Namen, ne Adresse oder irgendwelche Feldwerte auslesen **mhm**, und auf „Elastic“ speichern.

Das heißt, wir, wir versuchen, strukturierte Daten zu erzeugen aus Dokumenten, soweit das immer möglich ist und dann gibt's natürlich Grenzen. Also ich kann aus 'nem, aus 'nem Gedicht oder aus irgend 'ner prosaischen, wissenschaftlichen Abhandlung natürlich jetzt keine Tabellen bauen und irgendwelche Versuchsreihen (?) mit Tabellenelementen oder, oder Sonderformen drin sind, weil so detailliert kriegen wir's dann aus den Prosatexten nicht mehr raus, **mhm** also da gibt's 'nen fließenden Übergang. Dann gibt's natürlich auch ganz krasse Beispiele, Dokumente, die sind jetzt gar keine so, so prosaisches Zeug von vorne nach hinten, sondern die sind eher schon eigentlich im Prinzip fast wie Datensätze, ne Rechnung als Beispiel, ja, **mhm** (*Pause*) äh..., die..., das kann jetzt ein normales mit Word oder Excel erstellte Rechnung sein oder ähnlichem Zeug (*Räuspfern*), da hab ich aber ganz klar praktisch Felder, wie Rechnungsnummer, Datum, Betrag, Position und ähnliches Zeug. Das könnten wir dann auch auslesen und die (*Räuspfern*), das andere ist Rechnungen werden ja oft mit Datenbanksystemen, SAP und ähnlichem Zeug erzeugt **mhm** und eigentlich der Output sieht aus wie 'n Dokument, aber eigentlich ist hintendran nur 'n Datensatz **mhm**.

Solche Sachen, Berechnungen, die verarbeiten wir normalerweise eh nicht, weil die gehören in kein Wissensmanagementsystem rein, wo mer dann irgendwelche schlaunen Suchabfragen hat, wie: „Was muss ich tun, wenn XYZ?“, sondern das ist eigentlich 'n Datensatz, der gehört in so 'n Faktura-System oder Buchhaltungssystem oder ähnliches rein. **Okay...**, **also...** fließender Übergang, aber wir versuchen aus Dokumenten letztlich strukturierte Daten zu machen **mhm**, soweit das möglich ist. Das ist dann so ne Mischung aus Prosa und... Struktur.

Okay, mhm, verstehe. Und wenn Sie jetzt ne Weile zurückdenken, hat sich dann Ihre Vorstellung von dem, was 'n Dokument ist verändert?

Ne, eigentlich nicht und natürlich in den..., wir, wir haben ja begonnen vor ewigen Zeiten in den 90er Jahren, da gabs noch gar kein Internet und kein Intranet und ähnliches Zeug, also da war nicht viel Rede von „HTML“, also es haben sich sozusagen die technischen Varianten, wie man 'n Dokument speichern kann, wie man's erzeugt, äh usw., das hat sich verändert, aber eigentlich die Grundvorstellung von dem, was ich vorher erzählt hab, hat sich nicht erzeugt.

Es kam halt praktisch neue Technologien, sprich neue „Tontafeln“ in Anführungszeichen dazu **mhm, ja**, aber ansonsten sind wir eigentlich da mit der Vision, die wir gehabt haben und zwar mit unserm Produkt - na klar hat sich das technisch weiterentwickelt - aber, äh, da sind wir immer noch bei der ganz, bei der ganz gleichen Vision. **Okay**. Ich hatte Ihnen auch so 'n Link geschickt zu so 'nem schlaun Absatz von meinem Kollegen, ne? **Ja genau, mhm, hab ich gelesen**, da finden se dann auch noch bisschen was drüber, wie, wie wir das sehen **mhm!**

Okay, dann kommen wir zum zweiten Block: Künstliche Intelligenz okay, ähm, zuerst würd ich gerne mal wissen, mit - ja weil man sagt ja immer „Künstliche Intelligenz“ und das ist ja aber eigentlich so 'n riesen Wust (lacht) aus Technologien - und äh ja, da möchte ich gern wissen, was für Technologien Ihr Produkt benutzt?

Ok, also unser Produkt benutzt im KI-Bereich heute, äh, also ich fang mal 'n kleines Stückchen noch weiter vorn an, äh, wenn ich jetzt so neuronale Netze hab, selbstlernend, und wende die z.B. auf Bilder an, um 'n ganz einfaches Beispiel zu sagen, ne? **Mhm**. Also die erkennen, wenn Sie da beim Flughafen durch die Kontrolle gehen, dann scannen die Ihr Gesicht und erkennen dann: ist 'n Mensch und hat 'n Namen und wohnt in Potsdam oder irgend sowas, ja **mhm**.

Äh, oder ich mach so 'n... Screening für, ah was weiß ich, a Hautkrebeklinik und da erkennt er dann irgendwelche Muster und kann mit irgend 'ner Wahrscheinlichkeit vorhersagen, das ist jetzt vielleicht Krebs oder nicht Krebs **ja**. Die Dinger lernen im Prinzip indem man (?) einfach unendlich viele Bilder einspeist und immer sagt: „Das ist richtig“ oder „das ist falsch“, was er da vermutet hat **ja** und da gibst a Vielzahl in das Netz rein rein und irgendwann ist das Netz so gestaltet und so schlau, dass es auf wundersame Art und Weise die Erkennung..., das, was man erkennen will immer besser wird, ja **mhm**, das kann man aber nicht mit Dokumenten machen, ja.

Sie können denen die ganze Bibliothek von der Uni Potsdam oder irgendwas vor die Füße schmeißen jeden Tag und irgendwelche Fragen stellen und ihm dann immer sagen

„richtig“ oder „falsch“, das führt zu gar nix, ja **mhm**, das geht nicht bei Dokumenten, also das kann man, diese Techniken kann man bei Dokumenten nicht so einfach anwenden **ja**. Also unser Ziel ist ja, Dokumente verarbeiten und dann Suchanfragen möglichst...mit 'nem Exzerpt zu beantworten, wo alles drin ist, was, was, was zu dem Thema irgendwann in irgendwelchen Dokumenten steht von Relevanz **mhm** und das schaffen wir auf die Art und Weise nicht.

So, und das heißt, was wir heute haben, ist: wir haben, wir zerlegen ja - Sie haben ja mal so Folien von mir au gesehen - wir zerlegen diese Dokumente in ihre kleinsten Bestandteile **mhm**, so klein, dass wir sie rausziehen können, im Prinzip sowas wie Absätze, Tabellen, Listen und so Zeug, äh, und diese kleinsten Bestandteile - wir nennen die Atome - da kuck mer dann, wie sind die miteinander verbunden? **Mhm**.

Und die Verbindungen können jetzt sein, es gibt z.B. zwischen den Atomen irgend 'ne Gemeinsamkeit. Das eine verweist auf das Andere oder das Eine ist 'n, äh, 'n Unterkapitel von 'nem anderen Kapitel in 'nem Word oder in 'nem PDF oder ähnliche Sachen, ja, **ja** oder bestimmte Wort-Kommunalitäten, ja, da steht hier was von Dokumenten und dort steht was, was irgendwie mit Dokumenten zu tun hat. Äh so, und je nachdem, was man da alles an Beziehungen rausfinden, verknüpfen wir diese einzelnen Elemente mit dem Netz, mit mehr oder weniger starken Verknüpfungen **mhm**. Das ist, was wir heute haben.

Und wenn Sie jetzt neue Dokumente dazutun oder Dokumente ändern oder löschen, dann wird das Netz immer automatisch aktualisiert **okay**. Und jetzt haben wir 'n riesen-großes Regelwerk, wenn Sie ne Frage stellen, was das Ding dann alles beachten muss, wenn er diese, sozusagen diese Exzerpte da erstellt, ja, **ja**. So, wenn Sie jetzt irgendwas fragen, dann geht er da rein, dann soll er in den Nachbearbeitungen kucken und dann soll er dort kucken, ob in den Metadaten was steht. Dann so..., äh, wenn das Kapitel z.B. gar keinen Bestandteil hat, sondern da kommt gleich das nächste Unterkapitel, dann soll er da noch 'n Stückchen weiterlesen und sofort.

Und da gibt's hunderte oder tausende von solchen Regeln **mhm**, wie wir dann die Antworten zusammensetzten. Das ist der Status heute **okay**. So, was wir, wo wir dran sind, was im Laufe diesen Jahres kommen wird, äh, ist..., wir äh, wir haben sowas das nennen wir „N-grammatic Reasoning“ **mhm**.

(Es klingelt beim Interviewpartner an der Haustüre und er bittet um eine Minute Unterbrechung).

Frau Baumann? **Ja? (lacht)**. So, sorry, für die Unterbrechung, ich bin im Homeoffice und es hat gerade Post geklingelt. **Ja, kein Problem (lacht)**. Ähm... **Genau, wir waren beim „Reasoning“...**

„N-grammatic Reasoning“ das ist ein Begriff von uns, ähm, also das würde, das ist nicht mit „Engram“ sondern das ist „N-grammatic“ - also ein Wort - „Reasoning“, so nennen wir das, also im Endeffekt hat das was damit zu tun, Kerngedanken von einzelnen Absätzen zu ermitteln, ja **mhm!** Also, was weiß ich, der Kerngedanke von irgendeinem Absatz oder Atom ist: „die Gefahren der Künstlichen Intelligenz“ oder nur „Künstliche Intelligenz“ oder „einsatzbereite Künstliche Intelligenz“ und wir versuchen dann alle andern Atome zu finden, wo ähnliche oder dieselben Kerngedanken drin sind und nach diesen Dingen dann clustern und verknüpfen. **Aaah, okay!**

Die Basis dazu, das wären jetzt „N-Gramme“, das ist, wär jetzt: „N“, Bindestrich, „Gram“ **mhm.** Das ist so ne Technik, die kann man als „Open Source“ auch erwerben. Da werden Wortgruppen genommen, zwei oder drei oder vier Wortgruppen, äh, und anhand dieser Wortgruppen kuckt man, wo überall sonst kommen dieselbigen Wortgruppen noch vor? Und, äh, dann kann man auch „Deep Learning Verfahren“ anwenden dann, **mhm,** dass das System irgendwann schnallt: „Wie auch“, „vergleiche auch“ usw. das ist völlig irrelevant, aber „einsatzbereite KI“ ist relevant, ja **mhm.** Da gibt's Open Source dann und wird würden praktisch so ne...oder wir sind im Augenblick schon dabei - und in diesem Jahr wollten wir die erste Version davon haben - dass wir unsere ganzen..., unser Netz praktisch nach diesen Dingen durchföhnen, dass wir da dann Wortgruppen finden, dass wir das mit dem „Deep Learning“ machen, dass die relevanten Wortgruppen erkannt werden und, dass wir aus diesen Dingen heraus dann so 'n...solche Kerngedanken von einzelnen Absätzen ermitteln.

Und dort kann man dann auch 'n Feedback von den Usern einbauen. Wenn jetzt das „Theum“ irgendwelche Antworten gibt und die User sich immer entlang von bestimmten Teilen dieser Antwort weiter vorhangeln, die kann man ja dann klicken, ja, diese Elemente, **mhm** und dann sehen wir auch anhand der Klicks, äh, wo sozusagen, in welche Richtung sich die User da bewegen **mhm!** und dann kann man das zurückfließen lassen in dieses Basisnetz oder Matrix, die mer da aufgebaut haben **ah okay!** Das sind dann so Verfahren, so „Machine Learning“, „Deep Learning“ usw., die wir drauf auch anwenden, wir können's halt leider noch nicht so einfach machen, wie mit Fotos, ne. Solange wir 'n Dokument hinhalten, bis er verstanden hat, was drinsteht, das wird nicht klappen **mhm.** Das Ding versteht überhaupt nicht, was drinsteht (lacht) **(lacht).** So, also das ist ne Geschichte, die kommt in diesem Jahr, äh, die Basis und dann kann man das natürlich ewig lang weiterentwickeln **mhm.** Im nächsten Jahr werden wir das dann halt verfeinern und vertiefen.

Ne andere interessante Sache ist Sprache. Äh, was wir heute machen ist, wenn Sie da oben ne Suchanfragen eingeben **mhm,** äh, dann gehen wir im Endeffekt her und schmeißen die Stoppworte raus, wir haben da drin so Stoppwortlisten **ja,** die kann man auch

pflegen und ändern usw., ja, äh, je nachdem, in welcher Firma und in welchem Bereich man letztlich unterwegs ist, ja. Hat man ja vielleicht ganz andere Stoppwörter **ja**. Wenn Sie jetzt da sagen: „Wie wird das Wetter heute?“ oder irgendsowas, äh, dann schmeißt er irgendwelche Worte, wie „wie“, „das“ und ähnliche Sachen raus **mhm**. Übrig bleibt meinerwegen „Wetter Potsdam“ **ja**, und mit dem Zeug fängt er dann an zu suchen **okay**. So, jetzt kann man nat... - und wo das System heute scheitert, ich hab jetzt kürzlich so 'n Fall gehabt, Sie haben ja auch so Steuersachen da drin - „Welche Waffen darf ich erlaubnisfrei führen?“ **mhm**. Ja, dann wissen Sie natürlich „erlaubnisfrei“ ist so irgendwas wie „nicht verboten“, ja **mhm**. Oder ist so irgendwas wie „ich brauche keinen Schein“ **mhm**, aber das System hat keine Ahnung, was „erlaubnisfrei“ ist **okay (lacht)**.

Und das kann auch nicht..., so viele Äquivalenzen können wir ja nicht definieren, wo ich dann sag: „Erlaubnis, frei von Erlaubnis, nicht verboten, verboten, frei, äh..., nicht genehmigungspflichtig“, so viele Äquivalenzen können Sie gar nicht erfinden **mhm**, dass er das (?). So das heißt, äh, heute muss ich natürlich schon hergehen und sagen, wenn ich such: „Welche Waffen sind erlaubnisfrei zu führen in Deutschland?“, dann muss ich die Frage irgendwie so formulieren, dass es das System auch kapieren kann, ja **mhm**. Also z.B. kann ich eingeben nur „Waffen“ und dann krieg ich so 'n Cluster „Genehmigung, Erlaubnis wird benötigt“ ist bei den Veranstaltungen oder irgendwas **mhm** oder kann zumindest noch sagen, „verbotene Waffen“ oder „Waffen verboten“ und so sagen und da findet er dann entsprechende Dinge **mhm**.

Aber irgendwas wie'n Dichter ausdrücken, „erlaubnisfrei“ oder ähnliche Sachen, das kapiert das System nicht, so **ja** jetzt, ja, jetzt kennen Sie wahrscheinlich auch so Zeug wie „Alexa“, „Google Assistant“ und ähnliches und „Siri“ **mhm**, äh und die haben natürlich auch 'n paar Entwickler mehr, wie wir und wir vermuten mal, dass diese Sprachassistenzsysteme im Laufe der Zeit immer schlauer werden **ja**. Also wenn Sie heute in Google fragen und sagen: „Welche Waffen dürfen wir erlaubnisfrei führen?“, dann kommt auch Blödsinn **ja**, äh, aber, das Zeug wird besser, ja **ja**. Das wird im Laufe der Zeit besser und was wir machen können und was wir auch machen wollen ist, dass wir ne Schnittstelle bauen! Also wär blödsinnig, wenn wir sagen, wir erfinden jetzt selber so ne Sprachinterpretation, dann ist das völliger Quatsch. Da, da kann niemand mit Google und, und Apple usw. mithalten **ja, mhm (bedauernd)**.

Aber wir können ne Schnittstelle machen zu diesen Systemen! Das heißt, wir nehmen mal an, wir haben ne Schnittstelle zu Siri oder Google Assistant, die haben dieselbe Anwendung, wie heute, am PC oder am Handy, sprechen Ihre Abfrage rein **mhm** und jetzt interpretiert das System: „Welche Waffen darf ich erlaubnisfrei führen?“, und der kapiert vielleicht irgendwie „Waffen“, „nicht verboten“, „keine Erlaubnis“, „keine Genehmigung“, gibt uns dann die von..., die sozusagen die Interpretation Ihrer sprachlichen

Frage, **mhm**, formulierten Frage. Und wir suchen dann mit den Keywords, die er uns da übergibt **okay, mhm**. So, das kann man im Prinzip heute schon..., könnten wir im Prinzip heute schon machen. Was uns da noch fehlt, ist, wenn, wenn jetzt Sie ne Abfrage stellen und da kommen so zwei, drei Absätze, ja, **ja**, braucht er die ja nur noch vorlesen. Aber wir haben ja auch so Dinge, wo dann so Kontextblöcke kommen **ja**.

Ja, weil's so viel Inhalte gibt, dass er die irgendwie clustert und dann kommen diese Kontextblöcke und jetzt müsste das ein paar Mal hin- und hergehen, ja. Jetzt müsste das Siri-System sagen: „Ja, wir haben jetzt da ne Antwort gefunden und wir haben jetzt da fünf Optionen und welche Option wollen Sie denn sozusagen weitergeben?“, ja, **mhm** und dann mach ich wieder was, die Option Nummer vier, „verbotene Waffen“ und dann kommt entweder noch mehr Kontext durch oder liest vor. Und diesen Dialog, wo's dann ein paar Mal hin- und hergeht, der ist natürlich auch aufwendigere zu programmieren, deswegen haben wir das heut noch nicht **okay, verstehe**.

Auch das ist 'ne Geschichte, wo man dann praktisch die Intelligenz von Sprachinterpretieren, die dann in dem Assistant und, und Siri usw. drinstecken, dann praktisch ändern kann **mhm**. Und, also das haben wir vor, aber das haben wir nicht mehr in diesem Jahr, vielleicht mal nächstes Jahr oder so, wenn die Systeme da bisschen schlauer werden, die sind nämlich allesamt noch ziemlich doof heut (**lacht**). Ähm und..., also ich hab heut, grad vorher, hab ich was ausgefüllt, (?) zum Thema „Künstliche Intelligenz“, da hab ich im Portal der Deutschen Wirtschaft irgendwas eingetragen zu „Theum“, ja **mhm** (?) oder was, und damit die Robots **aus... (?)** stand unten drunter: „Wieviel ist 7+1?“ **okay**. Ja da..., also normalerweise müsste KI das eigentlich **können** auch sofort abbilden können **ja**, und dann kann ich ja irgend 'nen Blödsinn in meine Seite einstellen. Also, da ist viel Hype auch dabei, ja, bei dem ganzen Thema **mhm**.

Also das wär so ne künstliche Intelligenz Geschichte jetzt. Im Sprachbereich, was mir grade noch einfällt, dieses „Clustering“, da haben wir au so linguistische Methoden drin, die, äh, wenn, wenn sie so ne Antwort kriegen, wo man die Inhalt..., wo man so Optionen da zurückliefern, ja **mhm**, da stecken KI-Methoden drin, wie ich den Käse clustere und deswegen kommt da auch manchmal so 'n kompletter Blödsinn raus, ja **ja**. Bei „verboten...(?)“, da finden Sie zwei, drei Blöcke, wo..., die einigermaßen Sinn machen und dann steht da irgendwie: „Absatz drei“ (**lacht**) **ja okay (lacht)**. Weil er da das Cluster irgendwie identifiziert, weil er's zehn Mal irgendwo gefunden hat, aber macht null Sinn, ja **mhm**. Da kann man natürlich auch dann verbessern. Aber da stecken auch so Methoden drin, die man verbessern kann. Das ist es, was sozusagen jetzt im KI-Bereich bei uns drin ist und, und demnächst, als Nächstes eingebaut wird.

Okay. Gut, dann fiel... Ich find also Ihre Frage hier ganz interessant, Chance, Risiko und beides das ist natürlich a bisschen philosophisch, ich denk aber, es ist beides: Ne riesige Chance - ich glaub nicht, dass wir in unserem Leben ne KI erleben, die Sie oder mich als Mensch ersetzt, da fehlt einfach die Biologie, die Emotion und Millionen andere Sachen - aber es wird viel KI basiertes Zeug geben, was einzelne Aufgaben, bestimmte definierte Aufgaben, an der Börse, im Betrieb, bei der Fertigung, vielleicht im Lernb...Unibereich, Lernen und Teachen usw., äh, viel besser machen können wird, wie 'n Mensch **mhm**. Und das kann natürlich auch sein, ein Roboter der schießt ja, da sind wir im negativen Bereich **ja, genau**. Viele Arbeitsplätze, die wegfallen und so weiter und so fort und dann haben wir natürlich auch Risiken.

Und was man heute glaube ich machen muss, ist, man muss darüber sprechen. Es ist ja to... wäre eigentlich toll - ich sag's mal so plakativ - wenn jeder seinen Roboter hätte, der rumläuft und die Arbeit macht, ja **(lacht)**.

Ja, lästige Arbeit! Und ich kann mich dann mit all diesen Dingen beschäftigen, die mir Spaß machen, ja **ja**. Aber die Welt funktioniert halt so, der Roboter wird alles erledigen, was ich kann und dann bin ich arbeitslos, ja **ja**. Anstatt, dass der mir mein Geld verdient und ich kann, kann dann alles das, ich hab dann genug Geld, und kann das machen, was ich machen will. Und dann wären wir so bei Wirtschaft, die wir umgestalten, bedingungsloses Grundeinkommen **mhm** und ich denk mal die aktuellen Rahmenbedingungen und Ziele setzen, was nicht nur Fortschritt und immer schneller und, und, und so weiter, sondern was will man den eigentlich als Gesellschaft, was ist gut für uns, ja? **Ja** Was wollen wir nicht und solche Sachen. Da gibt's 'n recht interessantes Buch von Richard David Precht.

Hab ich gelesen! Oder meinen Sie das ganz Neue von ihm?

Ja tolles Ding und da geht's um künstliche Intelligenz und solche Fragen. Das kam erst vor vier Wochen oder sowas raus.

Ja genau, das hab ich gelesen (schmunzelt)

Ja, das, das behandelt eine ganze Menge, äh, Themen in dem Umfeld.

Ja, das war sehr interessant. Also eher ja, würd ich sagen, negative Schiene... so... Also er sieht's ja schon sehr, sehr kritisch das alles, ähm...

Ja, er ist ja Philosoph, also man müsst's vielleicht nochmal mit paar anderen Geschichten dann ergänzen, ja **ja**, äh, aber auf alle Fälle sind da recht interessante eigentlich Gedankengänge drin, ja.

Ja, total, fand ich auch! Ja. Und ähm..., gut äh dann, ich würd mal sagen, wir machen weiter mit dem Produkt (lacht) Ja, ja. Ähm, ja, wie kam's dazu, dass „Theum“ entwickelt wurde?

Also das kann ich so in 'nem Bild beschreiben. Als ich die Firma 1994 gegründet hab und davor war ich schon so fünf Jahr lang als Freiberufler unterwegs. Damals war ich sehr stark bei Sparkasse und so weiter durch irgend 'nen Zufall tätig und damals kam ein Produkt auf, das hieß „Lotus Notes“, gibt's heute auch noch **mhm, ah ja, kenn ich**, aber grad auf 'm absteigenden Ast. Aber damals war das „Lotus Notes“ der totale Hype. Und, äh, dann war ich da bei den Sparkassen Organisationsleitern und in anderen Firmen und die haben dann immer so zu mir gesagt, in 'nem Bild: „Hier sind unsere ganzen Organisationsrichtlinien und unser ganzes Sparkassen-Wissen“ und das „hier“ war dann ein riesengroßer Schrank mit 10.000 Aktenordnern (**lacht**) und wie krieg ich 'n das jetzt alles in mein „Lotus Notes“? **Ja**.

So..., und dann haben alle möglichen Firmen – und wir gehören auch dazu – immer wieder irgendwelche Anwendungen entwickelt, individuelle Anwendungen, Applikationen, wo man dann für irgendeine Sparkasse oder für irgendeine Problemstellung ne Lösung hatte, aber es war immer, sozusagen neu zusammengebastelt und nur für einen Zweck **mhm**. Und aus dem Grund hab ich mir irgendwann gesagt, das ist eigentlich 'n Schwachsinn, wir machen immer den gleichen Blödsinn, immer 'n bisschen anders, da muss man doch irgend 'ne Standard-Software entwickeln können. **Ja**.

Dann hat sich der Zufall ergeben, dass ich mit meinem Kollegen, also mit meinem heutigen Partner, ungefähr 1994 mal telefoniert hab, das ist 'n Amerikaner, ja. Dann klingelt's bei mir im Büro, da waren wir noch ganz wenig Leute in der Firma, und dann war der dran und dann haben wir uns unterhalten und haben halt gemerkt, wir denken da ganz ähnlich, wir machen ganz ähnliche Sachen, der hat auch schon bisschen a Produkt gehabt in die Richtung **mhm**, und dann halt einfach kurzerhand entschlossen: wir bauen da zusammen ein Produkt **okay**. Er sozusagen in Amerika, ich von Deutschland aus, von zwei Seiten nähern wir uns da an. Das haben wir dann au gemacht, kurz danach saßen wir hier in Deutschland schon im (?) und das Produkt hieß damals „Hyper.Net“. Und was das „Hyper.Net“ gemacht hat, das hat Dokumente verarbeitet, Word und ähnliches Zeug. Hat auch schon zerlegt, nicht so wie heute, aber schon ein bisschen zerlegt und analysiert und in Notes und später dann „Domino-Datenbanken“ abgespeichert und die konnte man dann nutzen für Intranetz und später dann fünf, sechs, sieben Jahre später für Internet, Internetseiten **mhm**. Das war der Start. Und technisch war das so, das war dann letztlich a Standardsoftware, die Dokumente zerlegt hat, nicht ganz so detailliert wie heute, äh, normalisiert, das waren bestimmte HTML Geschichten, die wir dann erzeugt haben und das Zeug sozusagen normalisiert, war dann verfügbar für andere, für Konsumenten, in Form von einer „Domino Datenbank“, also „Notes Domino“ oder später in Form von Intranet und Internet **mhm, okay**.

Und wir haben heut so ne Webseite, das können Sie sich anschauen, das ist die Schweizer Nationalbank „ntb.ch“ **mhm** und die ganze Website, die die haben, das sehen Sie der Website nicht an, aber sag ich Ihnen einfach (**lacht**), sind alles Word und Excel und ähnliches Dokumente und das Zeug, was da drin ist, transformieren wir so, dass es so aussieht wie ne Website **okaaay (verwundert)** und das war die Basis auch für das, was wir heute machen, diese ganze „Hypernet-Technologie“, die haben wir so 15 Jahre lang entwickelt, die steckt als Core im „Theum“ **mhm** und ist dann mit dem „Theum“ erweitert worden um 'n Haufen Dinge, die Sie ja zum Teil jetzt auch schon kennengelernt haben **mhm, also...** also das war **mhm**, was wir dann entwickelt haben und sofort.

Irgendwann, so um 2015 rum, haben wir gesagt, Mensch, jetzt haben wir da so viel Zeug im Hypernet und so tolle, so diesen Transformationsprozess, den haben wir richtig auf drauf gehabt, kann man da jetzt nicht noch was draufsetzen, in Richtung, wo wir so „KI-Methoden“ auch verwenden und diese ganzen modernen Verfahren? Wo man praktisch hergeht, stellt Fragen und kriegt Antworten und das war die „Theum-Idee“ und das, was sie heute alles sehen und was dazu gekommen ist, das haben wir so ca. seit 2015 dann gemacht. **Ah, okay! Interessant und ähm...** Also die Entwicklungszeit, das ist halt wie gesagt keine einfache Geschichte mit den Dokumenten. Wir haben ca. 1994 gestartet **mhm**, bis heute. Zurzeit sind wir grad fünf Entwickler, die dran sind **mhm**. Die nichts anderes machen, als entwickeln.

Und es waren schon manchmal auch in schlechten Zeiten drei, zwei bis drei Leute und in guten Zeiten, ähm, bis zehn Leute, also sagen wir wenn man sagt mal im Schnitt fünf **mhm**, ab 1994, waren's ungefähr, was haben wir da jetzt, äh..., 94, 2004, 14, so 25 Jahre, mal fünf, dann haben wir ungefähr, wenn ich's jetzt richtig da gerechnet hab 150 Mannjahre oder Personenjahre **ja** waren auch etliche Frauen dabei muss ich sagen! (**lacht**) Und äh Inderinnen meistens (**lacht**) **okay (schmunzelt)**. Also und die 150 Personenjahre, da war natürlich auch der eine oder andere Weg, was sich dann als Irrweg erwiesen hat und was man dann im Endeffekt irgendwann hat sausen lassen. Aber ich wollt nochmal sagen, bis 150 Personenjahre, da stecken mindestens 100 tatsächlich in dem Produkt heute noch drin.

Okay, mhm. Ja, schon auf jeden Fall ne aufwendige Sache! Was das für 'n irrer Aufwand ist, ja. **Ja. Gut, ähm, dann mit dem Lernen, das hatten wir ja au schon bisschen angesprochen, haben Sie ja schon gesagt „Deep Learning“ und sowas ja. Ähm, genau, dann würd ich noch gern wissen, was Sie glauben, wie sich die Nutzungshäufigkeit von solchen Produkten in der Zukunft entwickeln wird?**

Also Dokumentenmanagement machen wir ja gar nicht mehr. Sie managen Ihre Dokumente wo immer Sie's managen wollen **ja** und sagen mir ja nur, was wir verarbeiten

sollen zu welchen Wissensbanken ja **ja**. So, und dann haben Sie so publiziertes Zeug, Sie können's irgendwo bereitstellen oder Sie können das irgendwie intelligent durchsuchen **mhm**, das ist, was wir machen. Und also wir machen das Ganze, weil wir davon ausgehen, dass viele Firmen und Organisationen, Unternehmen jede Masse an Dokumenten, jetzt in meiner Definition ja, von irgendeiner Spezifikation oder Beschreibung, bis hin zu dicken Büchern, sind ja auch Dokumente **mhm**, dass man da in Zukunft sehr viel mehr Wert drauf legt, sozusagen das überall verstreute Wissen und Information zu jeder Recherche schnell zusammenziehen zu können **ja**.

Und dass die Nutzungshäufigkeit steigt. Jetzt sind wir heute nicht so richtig gut skalierbar, weil wir machen immer so Projekte bei großen Firmen, wie Airbus, Siemens und ähnliches und das dauert alles immer ewig **mhm**, aber wir sind dabei das in ne Cloud zu implementieren **mhm**. Das heißt, das wird, am Jahresende wird's ne Version auf „Azure“ geben, ja, Microsoft **okay**. Und da könnten Sie, an der Uni oder wo immer, mit 'n paar Klicks installieren **mhm** und das alles machen, was Sie jetzt auch getan haben und noch 'n Stück mehr **okay** und da geht man davon aus, dass das sehr viel stärker genutzt wird. Und dann gibt's natürlich noch Dinge, die Ihnen wahrscheinlich jetzt gar nicht so transparent sind. Jetzt stellen Sie sich mal vor, im „Internet of Things Zeitalter“, da haben die Maschinen alle irgend'ne Intelligenz drin und dann blinkt irgendwie ne Lampe und dann steht, da 'n Code drunter und dann weiß die Maschine sie braucht mehr Öl oder mehr Papier oder muss gewartet werden oder irgend'n anderer Käse. Da könnten se natürlich an so 'nem Code, der da bei irgendwelchen Maschinen oder auf irgendwelchen Stellen ist, einfach Ihr Handy hinhalten, der liest den Code und interpretiert das jetzt als Suchanfrage und das Handy spuckt Ihnen die Antwort aus: „Was ist zu tun“ **mhm**.

Wenn da plötzlich „11“ steht oder „grüne Lampe Wartung“ leuchtet oder irgend so Zeug **ja**. Also solche Geschichten kann man machen, im Endeffekt Chatbots, ja. Sie haben a Handy, der greift auf die Wissensbanken im Hintergrund zu, sprechen mittels Siri oder so rein, was Sie gern haben wollen und der spuckt Ihnen alles raus, als Exzerpt, was er zu dem Thema dann findet **ja**. Solche Geschichten, äh, **okay**, das (?) Dings hab ich erzählt, Cloud hab ich erzählt, ja so in der Richtung.

Genau, gut, also sie denken auf jeden Fall, das, ähm, wird in Zukunft noch viel, viel stärker genutzt werden? Ja, **ja**, sonst würden wir's nicht machen (lacht). **Okay, ähm, dann, wie sich's Produkt verändert hat, hatten wir jetzt auch schon bisschen angesprochen ja und ähm, ja wer ist so die Zielgruppe von „Theum“?** Also heute..., wir müssen immer schauen, wo wir's..., wir haben alles selber finanziert ja. Wir hatten leider nie 'n Investor, der viel Geld reintut, sondern haben immer erst unser Geld verdient und dann haben wir's gleich am anderen Ende wieder ausgegeben (lacht)

für die Weiterentwicklung. Äh und dann historisch waren die Zielgruppen halt immer große Unternehmen, die viele Dokumente haben (**Räuspfern**), **ja**. Hatten wir in der Webseite drin, ja **mhm**. Äh, im Cloud-Bereich werden wir in ne Stufe kommen, wo auch viele kleinere Unternehmen das ganze nutzen können **mhm**. Äh, weil man, weil man's dort viel schneller installieren und nutzen kann, das Produkt **ja**. Und was wir zukünftig glauben eigentlich, es gibt, wir nutzen ja grad das „Teams“ und es gibt - von Microsoft - und es gibt auch von anderen Herstellern so Team-Software.

Heute haben wir ja ganz stark, Menschen sind in Organisationen und Firmen immer in Projekten, ja **mhm**. Das kann ein (?) Projekt sein oder irgend 'n anderes Projekt und bei den allermeisten Projekten gibt's jede Masse an Dokumentation **ja**. Und dass wird dann in Projektdokumentationen verarbeitet und im Rahmen eines solchen Projekts dann Wissensbanken erzeugt, wo genau das drin ist, was die Leute im Projekt brauchen **mhm**. Das ist eigentlich die Zielgruppe, von der wir glauben in Zukunft, das ist die interessanteste **okay, ja**. Und wir haben ja auch so ne Integration, in dem „Microsoft Teams“ hätten wir schon eine, ja **ja (schmunzelt)**.

Gut und ähm, ja, können Sie sich (Räuspfern) auch noch andere Einsatzorte für solche Produkte vorstellen? Jetzt mal außerhalb von äh, ja Unternehmen?

Ich hab ja gesagt z.B. bei der, die..., Sie meinen jetzt außerhalb Dokumentenmanagement und Unternehmen, außer dieses „Internet of Things“ ja, **ja okay, mhm**. Der Autor erzählt Ihnen irgendwas oder da kommt 'n Code oder irgendwas **mhm**, die Maschine funkt irgendwas durch oder so und das Handy liest den Barcode oder was immer das..., sieht das leuchtende Lämpchen oder was immer das sein mag und gibt sozusagen die Antwort Ihnen, was zu tun ist. Das, das haben wir vorhin schon gesagt **ja**. Alles unter „Dokumente“ das ist eh klar. Die..., man könnte 's in Internetsites haben! Internetseiten, Webseiten sind heute primär eigentlich so, wenn Sie dort suchen, kriegen Sie auch diese langen Listen von Treffern.

Ja! Ja das wäre schön, wenn's da so...wär....

Und könnte man natürlich das in Webseiten so einbauen, dass es auf der langen Liste von Treffern dann im Prinzip gleich auch präzisere Exzerpte rauskriegt.

Ja, okay, also dann eben auch so für Privatpersonen im Alltag.

Also Privatpersonen im Alltag vom Geld her, muss man da natürlich auch sagen, also das glaub ich am wenigsten, weil das Ganze ist jetzt immer...Privatpersonen kann ich mir vorstellen, wenn sie als Privatperson viele Dokumente gesammelt haben zu irgend 'nem Thema und wollen das jetzt nach draußen publik machen **okay, mhm**.

Da könnt ich's mir vorstellen. Aber das setzt erstmal voraus, dass wir wirklich eine Cloud-Lösung haben **mhm**, weil Sie wollen nicht zehntausende von Euros dafür bezahlen, sondern a paar Euro **ja** und dann muss das..., kann das nicht sein, dass wir da zehn Tage dafür arbeiten müssen, sondern das muss dann mit paar Klicks gehen. Und es gibt natürlich nicht so viele Privatpersonen, die dann solche Probleme hätten, also dass man das im Privatbereich richtig großartig ausbaut, das glaub ich eher nicht **okay**.

Gut, dann zum letzten Block: Zukunft. Ähm, ja, gibt's Möglichkeiten, wie man „Theum“ noch effizienter machen könnte, z. B. eben mit Sachen, die jetzt technisch noch gar nicht möglich sind? Also das haben wir ja jetzt auch schon 'n bisschen angesprochen.

Bisschen hab ich da ja schon angedeutet, aber das ist eigentlich ein Prozess, der sowieso ständig läuft, dass wir versuchen immer wieder neueste Technologien - Sprachen haben wir jetzt erwähnt und dann so „Deep Learning“ Geschichten usw. - dass man die dort, wo's passt einbaut, ja **mhm**. Und ansonsten heißt bei uns „effizienter“: wir haben unheimlich viel Aufwand jetzt reingesteckt, dass man das Ding wirklich in 'ner Viertelstunde installieren kann **mhm**, so easy, weil so viel Umgebungsprodukte betroffen sind. Oder, dass man das robuster macht. Es gibt tausende von Konstellationen, wo's dann hier mal klemmt und dort klemmt, da haben wir viel Arbeit reingesteckt.

Also das effizienter im Sinn, dass wir weniger Arbeit haben oder auch die Kunden und das Ding richtig rund zum Laufen zu bekommen **mhm**. Und neu haben wir so gesagt, Sprache, KI Geschichten eingebaut und sofort. Wir sind auch dabei immer wieder neue Konnektoren zu bauen. Zum Teil machen wir grad einen für „Confluence“, weil das öfters mal gefragt wird **mhm** und..., ja, „Jira“ ist ja schon öfters erwähnt worden. Ja, wir hatten einen gemacht für Windows Notes jetzt, den haben wir abgeschlossen für ne französische Firma. Das hat jetzt nix mit Effizienz zu tun, aber dann können wir mehr Inhalte verarbeiten **mhm**. Wir könnten einen machen für Exchange oder Outlook, weil die Menschen unheimlich viel Wissen letztlich in, in ewig lange Mailboxen speichern, ja **ja**.

Und gibt's auch so Sachen, wo sie sagen, ja, das würden wir gerne machen, aber das scheitert jetzt dran, dass es die Speicherkapazitäten oder die Rechenleistung noch nicht gibt oder so?

Nee, das Problem haben wir eigentlich nicht. Wenn's bei uns immer scheitert, dann ist es am Geld (**lacht**) und deswegen ist auf der Zeitachse, streckt sich das immer, nach hinten auf der Zeitachse **mhm**. Aber jetzt dass wir sagen, von Computer-Ressourcen haben wir heute kein Problem **mhm**. Wir, wir speichern ja auch alles in „Elastic“. „Elastic“ ist ja so, äh, skalierbar, Facebook nutzt ja im Hintergrund „Elastic“, ja, und die haben

irgendwie 'n paar Milliarden Users, also da brauchen wir keine Angst haben, dass das jetzt eng wird. Und..., wenn das „Theum“ irgendwie an ne Grenze kommt, heute in..., wenn wir's „On-Prem“ beim Kunde installieren, dann muss er halt zwei, drei, vier Theum-Server theoretisch haben **mhm** oder können wir's mit viele CPUs dann. Wenn wir das in 'ner Cloud haben - das ist, was wir grade machen – und wir haben, am Ende haben wir so ne Evercloud laufen, dann skaliert das Cloud-mäßig hoch.

Das heißt, wenn ein Server nicht reicht, dann wird er im Hintergrund so über „Kuber-netes“ und so Zeug, äh, praktisch, ohne, dass wird das merken, weitere Server dazuschalten **ah, okay!** Dass wenn wir jetzt 'n Ruck PDFs reinschmeißen, alle mit OCR, das dauert ewig, dann wird die Cloud jetzt im Hintergrund hergehen und sagen, der dupliziert, vervielfacht diese Theum-Server für 'n paar Stunden, dass er für das Zeug dann nicht drei Monate braucht, sondern 'n Tag **mmh, okay! Gut**. Also da haben wir eigentlich keine Sorge in der Richtung **mhm!**

Und was glauben Sie, wie das Dokument der Zukunft aussehen wird?

Also wir, wir glauben, es gibt einen Teil Menschengruppen, die wir so anschauen müssen: Der eine ist der Autor, der sozusagen narrativ in irgendeiner Form festhalten will **ja**. Und so 'n Autor der wird dann - weil der Mensch so tickt - der wird das irgendwie von vorn nach hinten von A bis Z beschreiben und ob er dann Microsoft Word nutzt oder irgendeinen anderen Editor oder 'n Wiki oder sonst was, ist egal, aber der wird sozusagen mit so Zeug wie wir's heute kennen, die Dokumente, die, die Inhalte festhalten, beschreiben und und und dokumentieren **mhm**. Aber, äh, der der Consumer, ja **mhm**, wenn Sie jetzt 'n Blogbeitrag machen über, was weiß ich, „Künstliche Intelligenz im Bereich des Dokumentenmanagements“ oder irgendwas **mhm?**, dann wollen Sie ja keine Stories dann, Bücher von A bis Z lesen, sondern Ihnen wär geholfen, wenn Sie so 'n hundertseitiges Exzerpt kriegen, wo alles drin ist, was es zu dem Thema irgendwo gibt in der Unibibliothek.

Oh ja, da wär mir sehr geholfen! (lacht).

So, genau, und das ist es im Endeffekt, was wir machen. Das heißt, der Consumer, der braucht ne völlig andere Sicht auf die Dokumente, wo er eigentlich relativ schnell das Zeug kriegt, was er wirklich braucht **mhm**. Was den Unterschied (?). Und das glaub ich, so wird sich das entwickeln. Das werden am einen Ende die Dokumente bleiben, die die Autoren erzeugen **mhm**, vielleicht mit anderen Tools, ob das dann Google Docs ist, Microsoft Office oder irgendwas anderes hinzukommt, und am anderen Ende werden Menschen wie Sie und ich in 'nem Unternehmen Wege haben, wie sie viel

schlauer und schneller an die wirklich benötigten Inhalte rankommen und das ist, was wir machen, und teilweise auch paar andere machen.

Okay. Gut, ähm, dann noch mal bisschen philosophisch. Ähm...ja, glauben Sie, dass in Zukunft gezielte KI-Gesetze nötig sein werden? Das spricht ja der Precht auch teilweise an, ähm...ja.

Ja, glaub ich schon. Äh, wobei Gesetze ist vielleicht schon der übernächste Schritt. Als erstes müssten wir ja mal tatsächlich definieren als Gesellschaft, was wollen wir denn eigentlich als Gesellschaft? **mhm.** Ja..., ist es gut wenn die KI jetzt alles macht und wir sitzen alle faul zu Hause? Oder wie wollen wir das mit dem Einkommen machen, was die KI dann letztlich schafft? **ja.** Wo wollen wir sie einsetzen, wo wollen wir sie nicht einsetzen? Oder was der Precht da immer, der, der..., ja der Precht da immer bringt: soll das Auto jetzt entscheiden, ob's in die Gruppe der Senioren fährt **ja** oder Sie zusammenfährt, ja? **Ja.** Also Precht sagt, das soll die KI gar nicht entscheiden **ja.**

Also man muss einheitlich mal definieren, was man will, dazu muss man natürlich bisschen Ahnung haben, was die Technik alles möglich machen wird **mhm** und wenn man dann geklärt hat, was man will, was man nicht will, dann kann man da vielleicht das Ganze in Gesetze gießen **mhm.** Also ich bin schon der Meinung, die KI ist im weitesten Sinne nix anderes, wie Automation **ja.** Und die wollen wir ja, also sonst würden wir ja alle da auf 'm Acker stehen und und irgendwie 'n... Acker umgraben, ja **ja (lacht).**

Das will ja keiner und es will auch keiner sein Leben lang am Fließband stehen und wenn der Precht da sagt, 70% aller Jobs sind irgendwie grauenhaft langweilige Jobs, ja, **ja,** die will kein Mensch machen, Sie nicht und ich nicht. Aber die meisten Menschen müssen's machen. Und das ist doch wunderbar, wenn das alles durch KI ersetzt wird. Aber man muss da 'n Schritt weiterdenken, man will dann keine Gesellschaft, wo alle arbeitslos sind und drogensüchtig **ja** und verarmen und so weiter, nur weil die, weil die Automation so weit fortgeschritten ist. Man muss den Wohlstand verteilen dann auf andere Art und Weise wie heute **mhm.** So, das ist ungefähr so ganz in kurzen Worten, was ich mir dazu denk **okay.**

Und ja, die letzte Frage hatten sie vorhin auch schon erwähnt. Sie glauben nicht, dass es, das man 'n Mensch, äh...ja.

Da würd ich mich jetzt dem, dem Precht auch anschließen da in seinem Buch **ja,** also der Mensch wird seine, sag ich mal „Mischungen“ aus Emotionen, Kleinhirn, Großhirn und und weiß der Teufel, was alles 'n Menschen ausmacht, äh, das hat die KI heute nicht und ich glaub auch nicht, dass sie das in unserem Leben haben wird. Was dann irgendwann in hundert oder tausend Jahren ist, das ist wieder 'n anderes Thema.

Ja, seh ich auch so (schmunzelt) Okay....

Es gibt Physiker, das ist sehr interessant, ich hab a anders Buch gelesen, äh, das waren so Kurzaufsätze, weiß nimmer, wie's heißt, ich hab's da oben irgendwo rumliegen, das waren so Kurzaufsätze, zwei, drei, fünf Seiten, von renommierten Wissenschaftlern über die ganze Welt hinweg. Soziologen, Informatiker, Ingenieure, Philosophen, Physiker, Mathematik und sofort und da waren, und da ging's um das Thema „KI“ und welche Gefahren und so weiter, ja **mhm**. Und dann, war für mich recht interessant, Informatiker, aber auch Philosophen..., Informatiker, Philosophen, was gibt's da noch, Ingenieure und solche Leute **mhm**, die haben alle eigentlich so gesagt in der Richtung: „Ein Programm wird niemals so intelligent sein, wie 'n Mensch.“ **Okay**.

Die Physiker – das war sehr erstaunlich, das war quasi die Mehrheit de Physiker! – hat gesagt, na ja, in der Evolution ist im Laufe von, von, auf der Erde viereinhalb Milliarden Jahren irgendwas entstanden, wie unser Hirn **mhm** und ganz von selbst sozusagen, durch Zufall und Notwendigkeit und warum sollte nicht möglich sein, dass das auf künstliche Art und Weise irgendwann mal auch geschaffen wird, vielleicht sogar besser. Das waren die Physiker, ja **mhm!** Also ne gewisse Gefahr besteht glaub ich schon. Aber nicht, wenn ich jetzt heut so anschau, was da an Software da ist, die hat kein Bewusstsein und die ist eigentlich strunzdumm (**lacht**).

Ja, also nicht mehr zu unseren Lebzeiten auf jeden Fall.

Die kann viele Sachen ganz toll erledigen und viel besser, wie 'n Mensch und das wird auch ständig verbessert du sofort, aber, aber irgendwie kann man se ganz leicht in die Irre führen (**lacht**).

Ja, das stimmt (lacht). Ja, dann ähm, möchte ich Ihnen zum Abschluss noch mal ganz herzlich danken! Ähm, das hat mir auf jeden Fall sehr weitergeholfen! Und...ja, falls Sie noch irgendwelche Ergänzungen haben, können Sie mir natürlich jederzeit schreiben und...

(Es folgt ein organisatorischer Abschnitt, der inhaltlich nichts mit dem Interview zu tun hat und daher nicht transkribiert wurde).

Okay, alles klar!

Okay, dann vielen Dank!

Bis die Tage, schönen Tag!

Tschüss!

Transkription Interview mit Carolin Hischer, Serviceware

Berlin, 13.10.2020

Interviewerin: Isanne Baumann (Fettdruck)

Interviewte: Carolin Hischer (Normaldruck)

Dann würde ich anfangen mit dem ersten Block, die Begriffe „Dokument“ und „Daten“. Ähm zu Beginn gleich mal ‘ne schwierige Frage, also es gibt da sehr viele verschiedene Interpretationen, was ist ein Dokument, was sind Daten und ähm ja, da habe ich ja mal zwei Stück erwähnt hier, den Ranganathan, der sehr auf die Form bedacht ist und sagt, ähm ja, sind eben Gedanken, die auf Papier festgehalten werden oder auf einem anderen Material, das eben die physische Handhabung, den Transport durch Raum, äh und Konservierung in der Zeit erlaubt und die französische Bibliothekarin, die Frau Briet ähm, die eher sehr darauf bedacht ist, dass was in ‘nem Dokumentationssystem festgehalten wird und sobald es dort eben irgendwie äh, angelegt ist, äh kann man es dann als Dokument bezeichnen und in dem Zusammenhang spricht sie eben sogar davon, dass Tiere im Zoo ein Dokument sein können und ja, dann würde ich mal gerne deine Ansicht dazu hören: Was ist ein Dokument?

Äh ich habe mir die beiden natürlich dann durchgelesen und musste bei beiden schmunzeln, weil ich jetzt bei beiden sind natürlich dann sag ich jetzt mal die Extreme wahrscheinlich. Ähm nicht hundert Prozent durch, ähm zustimmen würde. Also für mich hat ein Dokument nichts mit irgendwelcher Haptik oder sonstiger, ähm ja ich sag jetzt mal Papierform zu tun, für mich ist ein Dokument schlussendlich eine Ansammlung an Daten, die einen Wissensgehalt transferieren. Das würde ich so ein bisschen für mich als Infor... äh als äh als Dokument äh beschreiben. Beide... für mich hat ein Dokument auch nichts mit einer Form zu tun, für mich kann ein Dokument geschrieben sein, für mich kann ein Dokument selbst ein gemaltes Bild sein. **Mhm.** Ähm also ich schließe das auch nicht aus also, ein ein Stück Papier kann auch ein Dokument sein, es ist nur nicht für mich ein Kriterium, das erfüllt sein muss, ja aber es muss ein Informationsgehalt da sein und es muss für mich zumindest eine zeitlich begrenzte, ähm statische Darstellung von Wissen darlegen. Das heißt also, ich ich weiß ich spring jetzt zwischen den ersten beiden Fragen, aber ich find sie halt gut im Vergleich zu beschreiben. Für mich hat ein Dokument in dem Sinne schon eine Statik, derartig für ein, für eine zeitliche Begrenzung, denn wenn ich jetzt zum Beispiel mir auf dem Desktop, mir irgendwie nur Datenquellen ziehe und die sich wie im DAX sag ich jetzt mal so minütlich anpassen könnten, oder sekundlich, oder wie auch immer, das hat für mich dann keinen Charakter eines Dokuments.

Das sind dann wirklich Daten, also Dokument hat schon eine zeitlich begrenzte, aber statische Form von Informationen. **Mhm.**

Okay. Ähm, genau, dann wär auch die Frage, ob du 'nen Unterschied machst zwischen analogem und digitalem Dokument, also jetzt einfach so von der Definition her. Ähm würde ich nicht unterscheiden. **Okay ähm genau und das wurde ja jetzt grad schon bisschen angesprochen, der Unterschied zwischen Dokument und Daten, also wenn ich das jetzt richtig verstanden habe, findest du, dass ein Dokument sozusagen aus Daten besteht. Und dass Daten aber auch** Genau. **alleine vorkommen können also und kein Dokument sind.** Richtig, genau, so würde ich sagen. Also es gibt die Möglichkeit, Daten unabhängig von einem Dokument zu betrachten, es gibt aber auch die Möglichkeit, ein Dokument unabhängig von Daten zu betrachten. Also wie gesagt, für mich ist ein gemaltes Bild oder Prozessdarstellung oder Ähnliches, wenn wir es jetzt mal ein bisschen professioneller betrachten, ebenfalls ein Dokument, hat aber nicht in dem Sinne für mich Daten.

Okay mhm und ähm, ja, findest du man kann da so einen klaren Cut machen und sagen: Das ist jetzt ein Dokument, das sind Daten, also zum Beispiel jetzt eine Exceltabelle... Exceltabelle, wo nur irgendwelche numerischen Werte festgehalten sind. Ist das ein Dokument, sind das Daten...ja (lacht) Ja das ist für mich.. das ist für mich auch ein Dokument, solange halt wie gesagt dieser, ja doch, sobald dieser dieser dieser Fakt... Faktor, dass es für einen gewissen zeitlichen Abstand eine statische Darstellung dieser Information darstellt. **Mhm also ähm** Deswegen eine Exceldatei äh kann ja miteinander verschachtelt sein, mit X Y Z, ähm, das sind für mich dann zwar Daten, kann aber.... Es erfüllt für mich aber auch die Kriterien eines Dokuments, wenn es aber jetzt irgendwie... Allein basierend, wie auch immer, wirklich Sekunden aktuelle Daten sind, die halt die ganze Zeit sich aktualisieren, das ist dann für mich kein Dokument mehr, sondern reine Daten.

Okay mhm. Gut ähm, und hat sich so im Lauf der Zeit deine Vorstellung vom Dokument geändert? Also das ist die Frage welchen, welchen Zeitraum wir von... anschauen **Ähm.** Das find ich tatsächlich eine schwierige Frage, weil ich natürlich früher, bevor ich damit beruflich zu tun hatte, jetzt mir weniger Gedanken über ein Dokument gemacht habe. Aber ich bin ja schon, wahrscheinlich wie du ähnlich, in einer Generation aufgewachsen, wo wir diesen Wandel sag ich mal von analog zu digital, ähm, zwar komplett mitgemacht haben, ihn aber wiederum so jung angefangen haben, dass äh, ich nie diese Begrifflichkeit „Dokument ist ein Stück Papier“ zum Beispiel mir verfestigt hatte. Ja

für mich war immer ein Dokument irgendwie eine Quelle von Wissen. **Mhm.** Ähm und ich habe noch nie das irgendwie mit einem Blatt Papier oder Sonstigem verbunden, das war für mich nie ein Kriterium, deswegen würde ich eigentlich sagen, im Kern hat sich mein Verständnis nicht verändert. Es hat sich sicherlich irgendwie nochmal ein bisschen gefestigt, da... gefestigt und äh, vielleicht differenziert, jetzt insbesondere, wo ich auch beruflich damit zu tun habe, aber geändert kann ich nicht sagen.

Okay, gut, super. Dann wären wir mit dem Block schon mal durch, dann äh, kommen wir zu künstliche Intelligenz, ähm Mhm. Da wäre die Frage, mit was für Aspekten von künstlicher Intelligenz das Produkt arbeitet. Wir arbeiten mit Natural Language Processing. **Mhm.** Das also, da Disclaimer, ich bin halt Product Owner, also bedeutet, ich, mein Titel ist Product Owner AI, bedeutet, ich bin immer dafür zuständig, im Austausch mit Stakeholdern ähm, zu erarbeiten, wo die Strategie hingeh, was sind die Bedürfnisse des Marktes und was ist demnach das, was gebaut werden muss. Was... wie es gebaut wird und da natürlich dann die technischen Details, das sind meine meine in Anführungsstrichen, große Anführungsstriche, also die Entwickler mit denen ich zusammen arbeite, die könnten das natürlich jetzt technisch auf einer ganz anderen Ebene erklären, aber ich glaube, für jetzt ist wahrscheinlich eine höhere Ebene okay. Also wir arbeiten mit Language Processing, Natural Language Processing, wo, ich sag mal die Grundlogik ist, dass du ein Dokument hast, was ja logischerweise aus Sätzen besteht und das Ganze muss ja erstmal, äh klein gehoppt werden. **Ja** In kleine Pünktchen und äh, das Ganze muss äh, in unserem Fall in Vektorberechnungen umgewandelt werden, sodass wir auf einer Semantik... mit einer Semantik arbeiten können und dann basierend auf diesen Vektordaten, äh entwickeln wir, also herausgefunden werden kann äh, ist da eine... eine Semantik vorhanden, wie nah sind die beieinander oder wie weit sind sie voneinander weg. Das, wenn du das jetzt wirklich auf das simpelste runterbrechen möchtest, das ist sozusagen die Logik, auf die unser Modell aufbaut.

Mhm. Okay und ähm Ah nee, das kommt später noch, sorry. Ähm ja und wie ist so dein Verhältnis zur künstlichen Intelligenz, jetzt mal ausgeklammert von dem Produkt, sondern allgemein, also siehst du darin eine Chance, ein Risiko oder einen Mix aus beidem?

Ich sehe da eigentlich nur Chancen. Äh ich bin generell kein Mensch, der irgendwie groß bei Techniken, ähm Ausrufezeichen oder (?) in die Augen bekommt. Ähm ich äh hab da jetzt nicht, sag ich mal, eine natürliche Angst, die teilweise insbesondere bei älteren Generationen vorhanden ist, die ich auch in meinem Beruf immer wieder bemerke, gegen die ich anarbeiten muss, äh ich sehe da nur Chancen. Ich sehe da auch

Verantwortungen, aber Risiken in dem Sinne, sehe ich nicht. Also ich muss auch gestehen, die künstliche Intelligenz ist ja ein Thema, was schon ewig alt ist, sag ich mal **Ja**. Jetzt verhältnismäßig und immer wieder ja auch ihre Winter schon durchgelebt hat. **Mhm**. Und ähm äh, so häufig waren auch damals schon von Philosophen, bis hin zu, weiß ich nicht was wir, Gedenkenträger, ähm sehr laut und sagten, ähm: Was da denn alles auf uns zurollt, und das sei ja nicht zu verantworten und heute sind wir so weit, dass teilweise eine KI immer noch nicht Katze von Hund unterscheiden kann, jetzt mal ganz böse gesagt. **Ja (lacht)**. Also äh, deswegen fällt es mir da wirklich sehr schwer, irgendwie die Risiken äh groß zu sehen, ich sehe da sehr viele Chancen. Ich glaube, in ganz vielen auch sehr, sehr wichtigen Bereichen, auch in der aktuellen Zeit im Health Care Bereich, ähm jetzt mal abgesehen von Dokumentenmanagement, gibt es da wahnsinnig tolle Bereiche, in denen wir wachsen können und das nutzen können.

Ja okay ja, das ist sehr interessant, wenn man dann also... weil ein anderer Interviewpartner war halt schon um einiges älter sage ich mal, und ähm Mhm. Ja es ist interessant, da dann so zu sehen, wie unterschiedlich die Antworten ausfallen. Das glaube ich ja. **(Lacht) Ähm ja und wie schätzt du die Bemühungen von Deutschland ein im Bereich KI? Gerade so im Vergleich zu China, USA? Ähm klar, die Frage ist auch, ob man sich jetzt so mit denen vergleichen kann, aber ja, also, einfach, wie siehst du das?**

Ja, tatsächlich habe ich da über diese Frage auch mit meinem PO-Kollegen gesprochen, weil, ich muss sagen, dass mir da eine Meinungsbildung schwerfällt, äh weil ich einfach die wirklichen Bemühungen von USA und China nicht einschätzen kann. Ich glaube, das mal vielleicht ein bisschen grundsätzlicher, dass wir in Deutschland leider ganz manchmal in diese Bequemlichkeit des Mittelstands gerutscht sind. **Mhm**. Jetzt mal unabhängig von KI, das würde ich einfach grundsätzlich für Deutschland sehen, dass wir teilweise so ein bisschen demnach diesen „Innovation Hunger“, der irgendwie uns ja sehr getrieben hat und sehr uns auch geholfen, ähm, dahin zu kommen, wo wir mit unserem Land ja glücklicherweise stehen, dass der manchmal ein bisschen verloren gegangen wird und dass Länder, wo irgendwie mehr in die Extremen aufgeteilt sind, sprich die USA und China, ähm, dass die den Hunger dann manchmal noch stärker verspüren und stärker agieren. Das ist aber jetzt eine Grundsatzmeinung. **Mhm**. Ähm ich sehe das aber auch im Bereich KI. Würde ich jetzt mal so allgemein belassen.

Okay gut. Dann äh, ja, also bevor ich jetzt Sachen konkret zu dem Produkt frag, ähm jetzt nochmal äh darauf zurückkommen, was ähm, du genau in der Firma machst. Was ist deine Position und was machst du so den ganzen Tag (lacht). Eine

gute Frage (lacht). Nee, genau, also mein Titel, sag ich jetzt mal, ist Product Owner AI, was schlussendlich bedeutet, dass ich Teil des Entwicklungsteams bin. Wir arbeiten nach einer Methodik, die Scrum heißt, **Ah ja**, weiß nicht, ob dir das ein Begriff ist. Das ist eine ähm ja, Methodik in der Entwicklung möglichst effizient zu arbeiten, die darlegt, dass ein Entwicklerteam größtenteils immer aus deinen Entwicklern, also den Developern, den Engineers, äh die Leute, die wirklich richtig coden können, besteht. Ähm... ah... Einen, der so auf die Methodik achtet, dass der Arbeitsprozess reibungslos abläuft und eine Person, und das bin ich, der, sag ich mal die Business-Brille immer aufhat und eigentlich entscheiden muss, in welche Richtung entwickeln wir denn jetzt. Weil ich hab dabei ja meine gesamten Stakeholder, sowohl intern als auch extern, meine Sales-Kollegen, die irgendwas wollen, meine Kunden, die irgendwas wollen, alle schreien und normalerweise ist halt, der am lautesten schreit, kriegts. **Ja**. Und damit das halt nicht passiert, bin ich da, die sich das anschaut, was alle, sag ich mal, wollen, was wir auch wollen und muss abschätzen, wie viel kostet uns das, also wie viel Aufwand ist das für uns und wie viel Nutzen generiert es wirklich für uns. Und ich sag jetzt mal simpel betrachtet muss ich eigentlich für jede Entscheidung, die ich dann treffe, was wir entwickeln, einen kleinen Business Case in meinem Kopf oder halt auch manchmal bisschen formaler kreieren, um dann zu entscheiden, ähm was wir jetzt wie priorisieren, das heißt ganz simpel zusammengefasst, ich muss das „Was“ entscheiden, was entwickeln wir. Das „Wie“, wie entwickeln wir es, das entscheidet mein Entwickler, davon habe ich keine Ahnung. Ähm und ja der Scrum Master, das ist dann jemand, der für die Methodik da ist, das find ich jetzt nicht so wichtig für dich. Aber das ist meine Hauptrolle. Zu entscheiden, was machen wir mit dem Produkt, in welche Richtung gehen wir weiter. **Okay und** Also ich (?) Entwickler sondern Business-Brille.

Okay und ähm was hast du studiert oder gelernt? Ähm ich habe meinen Bachelor in Medien und Kommunikation gemacht, da bin ich ein bisschen im, sag jetzt mal für mich jetzt aus der Retrospektive, bisschen falsch in die... falsche Richtung gelaufen. **Okay**. Und hab dann meinen Master in Business ähm Administration gemacht, schlussendlich mit einem Fokus auf Entrepreneurship, also ich, relativ simpler BWL-Master mit einem ganz, ganz tollen Titel, aber BWL-Master. **Okay**. Den hab ich dual gemacht ähm damals in einem großen Gesundheitsunternehmen und dort konnte ich halt in die verschiedensten Bereiche gehen und irgendwann kam ich dann halt auch in die Strategie und Innovation Abteilung, wo es dann halt um die grundsätzliche Digitalisierung und ja, digital unterstützte Prozessoptimierung geht. Und so kam ich dann irgendwann ähm auf den Geschmack des Business Managements im Allgemeinen, ich hab damals noch mit E-Learning Systemen gearbeitet und hab mich dann irgendwann entschieden, in dem

Bereich stärker einzusteigen und dann bin ich im Mai 2019 zu Sabio gekommen, beziehungsweise Sabio gehört jetzt zu der Serviceware, das ist der Name, den du kennst. **Mhm.** Und das ist ja grundsätzlich eine Firma, die sich darauf spezialisiert, jegliche Serviceaspekte die so in der Welt existieren mit digitalisierten Produkten zu unterstützen. **Ja.** Genau und da bin ich dann irgendwann auf einen Kundenkontakt hin in die Produktentwicklung gegangen. **Okay.** So in ganz kurz mein Werdegang (lacht).

Cool, ja Dankeschön. Ähm gut, dann eben jetzt zu äh Serviceware Knowledge. Ähm Mhm. Wie kam es dazu, dass das Produkt entwickelt wurde, also woher kam die Idee, oder der Anreiz, ja. Mhm. Das Produkt ist tatsächlich verhältnismäßig alt. Also das Produkt, der erste Aufschlag kam von vor zwanzig Jahren **Oh** aus dem Jahre 2000, da ging die absolut erste Version an den Markt. Damals konnte man aber das so ein bisschen schon als sogenanntes „MVP“, Minimal Viable Product, betrachten. Das bedeutet, dass wirklich noch überhaupt nicht all das, was so ein Wissensmanagement brauchte. **Mhm.** Aber es war halt wichtig, schnell an den Markt zu gehen und überhaupt das Feedback zu bekommen, dann halt wirklich anhand der Bedürfnisse des Kunden weiterzuentwickeln, anstatt, dass halt wieder die Entwickler, übertrieben gesprochen, fünf Jahre in ihr Kämmerchen gehen, ihr Zeug basteln und dann irgendwie feststellen, das ist gar nicht das, was der Markt braucht. **Mhm.** In 2000 war das ein relativ früher Player, hmm und das kam aus dem Bedürfnis, dass unser damaliger Gründer ähm aus dem Kundenservice kam. Also klassisch, wenn du bei deiner Versicherung anrufst und hast 500 Fragen dann sitzt auf der anderen Seite meist ein Mensch, der wahrscheinlich erst seit zwei Tagen eingearbeitet ist, trotzdem nett sein muss natürlich, und dir alle Fragen beantworten muss. **Ja.** Das ist so die schwierigste Kombination, insbesondere weil der ja auch gleichzeitig mit dir am Telefon ist. Und dafür ein System bereitstellen, das diesen Agenten unterstützt, das war eigentlich der absolute Startpunkt. **Mhm.** Ähm aber das fing halt jetzt relativ simpel mit einem... mit einem Firmen-Wiki, sag ich jetzt mal, an. **Mhm.** Einfach nur irgendwie eine Art, wo zentral Dokumente abgespeichert wurden. Und irgendwann ähm hat sich das dann entwickelt von einem, sag ich jetzt mal, relativ simplen äh Dokumentenmanagement, hin zu einem, wo wir wirklich sagen: Das ist ein Wissensmanagement. Also auch von, er kriegt einen Anruf, er kriegt eine Frage, er muss aktiv auf die Suche gehen, die Vision ist, dass er das gar nicht mehr machen muss, **Mhm** sondern dass er in seiner Arbeit proaktiv unterstützt wird und quasi das Wissen zu ihm kommt, wenn er es brauch. **Ja.** Das ist so die Vision, mit der wir arbeiten. **Mhm.** Und das hat sich natürlich in zwanzig Jahren schon weit entwickelt und jetzt irgendwann kam dann ähm die KI-Komponente natürlich als weitere Technik eigentlich dazu, diese große Vision ähm irgendwann zu erfüllen.

Okay ähm und da hatte ich noch die Frage zu der Entwicklungszeit, aber das wurde dann ja jetzt eigentlich schon so... Genau, aber eigentlich ist da wirklich die Antwort, das Spannende, wir entwickeln halt, also wir haben in dem Sinne keine Entwicklungszeit, wir entwickeln immer in so genannten Zwei-Wochen-Sprints, das ist wahnsinnig schnell. **Mhm.** Um halt sehr, sehr schnell das Feedback vom Kunden zu bekommen. Also uns ist es wichtig, halt wirklich mit Kunden direkt an.. am Mann sozusagen zu entwickeln. Wir wollen halt nicht: Wir hören euch zu, wir kucken was ihr wollt und dann gehen wir wieder ein Jahr weg, machen und kommen wieder, sondern wirklich im Zwei-Wochen-Sprint **Okay krass.** sind wir im ständigen Austausch am Markt. **Ja. Ähm** Um halt nicht an den (?) vorbei zu entwickeln. **Mhm.** Also und das heißt, es wird eben auch ständig weiterentwickelt, verbessert. **Okay.** Genau, unser Produkt grundsätzlich ähm ist ein Produkt, was wir im SaaS bereitstellen, also Software as a Service **Mhm** ist so ein bisschen die Überschrift da drüber. Äh bedeutet auch häufig, wir haben Kunden, die das Ganze aus einer Cloud beziehen. Also ganz normal, äh dass wenn wir ein Release rausgeben und das machen wir alle sechs Wochen, das ist sehr, sehr schnell **Mhm** alle sechs Wochen kommt demnach ein Produktupdate, wenn du es so simpel betrachten möchtest, zu unseren Kunden. **Okay, ja das ist echt schnell.** Also Entwicklungszeit dann sechs Wochen vielleicht, wenn du das mit einer Zahl beantworten willst.

Ja okay und ähm kann man ungefähr sagen wie viele Personen an der Entwicklung beteiligt sind, oder...? Ja beteiligt sind... sind da eher, also waren kann ich dir nicht beantworten, weil ich vor 20 Jahren noch nicht nicht dabei war, **Ja** da war ich noch in der Grundschule. **(Lacht).** Aber ähm jetzt inzwischen sind wir vier PO-Teams, also vier Produktteams, die an unterschiedlichen Komponenten des Produkts arbeiten. Ich für den KI-Bereich und dann drei weitere Kollegen ähm die... Ich hab fünf Entwickler, die haben jeweils drei Entwickler ... Ja also um und bei fünfzehn Leute, wenn du das jetzt mal... schwankt halt. **Ja okay.** Wenn du nur die Entwickler betrachtest, also wirklich nur die, die coden. **Ja.** Wenn du noch das ganze Produktteam drumherum betrachten möchtest dann ca. 20.

Okay mhm. Ähm ja und ähm kann man vielleicht was sagen zu dem, wie das Produkt lernt? Also so mit künstlichen neuronalen Netzen oder... Ja, da muss ich dir eine Antwort schuldig bleiben, muss ich dir gestehen. **(Lacht) Okay.** Weil wenn ich dir das äh, das würde ich dir wahrscheinlich so rudimentär nur erklären können, dass es dir überhaupt nix bringt. **Okay.** Also ich kann erklären, wie das Produkt funktioniert, **Mhm** aber wie die technischen Lernfaktoren dahinter sind ähm... Aber wenn dir das wichtig

ist, dann kann ich das sehr gerne nochmal nachfragen und dir da irgendwie einen Dreizeiler oder so als Antwort nochmal schicken.

Okay ähm ja, also ist jetzt nicht die allerwichtigste Frage, aber ähm ja ich hab halt auch ein Kapitel, wo ich das so behandle ähm Mhm was es da für Techniken gibt und ja also, wäre schön, aber ist kein Muss, sagen wir's mal so (lacht). Krieg ich hin, krieg ich hin.

Okay ähm ja und was glaubst du wie sich die Nutzungshäufigkeit von solchen Produkten in der Zukunft entwickeln wird, also jetzt vor allem im Bereich von Dokumentenmanagement? Mhm Ich glaube, was man da klar unterscheiden muss, zwischen dem reinen Dokumentenmanagement und dem quasi Doku... integrierten, ich wechsle jetzt zum Wort wieder Wissensmanagement. **Mhm.** Ähm Ich glaube, diese einseitige Nutzung, ich rufe irgendwo meine Plattform auf und hole mir meine Informationen, also dieser klassische Pull-Effekt, ich muss wirklich hingehen und was suchen, ich glaube, das wird irgendwann immer uninteressanter, weil es einfach im... im hektischen Alltag häufig nicht mehr realisierbar ist und in den, ich sag jetzt mal, gerade wichtigen Berufen wo so eine wahnsinnige Wissensflut existiert, ist das halt einfach unrealistisch, also wird immer unrealistischer meiner Meinung nach. Es ist immer noch eine Basisanforderung die jedes.. jede Firma irgendwie adressieren sollte, das jemand einen zentralen Ort hat, wo er Informationen findet. **Mhm.** Ist auch nicht immer gegeben, auch bei sehr, sehr großen Playern teilweise nicht, aber gut. Ich glaube, es entwickelt sich halt wirklich von diesem Pull zu diesem Push. Man muss ein Umfeld schaffen, ein integriertes Dokumenten-/Wissensmanagement was dem... was der Person, während sie arbeitet, Unterstützung pusht. Das bedeutet zum Beispiel, ein Bereich auf den wir uns konzentrieren, ist der Kundenservice, sagte ich ja schon. Oder zum Beispiel auch IT-Service. Also supergroße Firmen haben ja wirklich eine sehr große IT-Service-Abteilung, wo du anrufen kannst und, ich sag jetzt mal ganz simpel, sagen kannst: Hey, mein Drucker funktioniert nicht. **Mhm.** Das, ne? So, das ist die aller-, ist das allersimpelste Beispiel immer. Häufig rufst du da vielleicht nicht an, sondern schreibst denen ein Ticket an irgendeine E-Mail-Adresse: Hey, mein mein mein Drucker funktioniert nicht. So. Alte Welt wäre dann gewesen, der Nutzer oder der der Agent liest das, muss dann in sein Wissensmanagement kucken, warum kann denn der Drucker nicht funktionieren, findet da die Informationen und kann das weiterleiten an den... äh an den Nutzer. Das ist blödsinnig, meines Erachtens. Also du musst dann halt äh die Informationen pushen. Entweder schon quasi während der Anfragende schreibt „mein Drucker funktioniert nicht“, in dem Moment muss da schon ein Push rausgehen so. **Mhm.** Ah, du hast 'ne Frage zu deinem

Drucker, gar kein Problem, hier sind die Informationen. Oder, wenn er es trotzdem an den Agenten schickt, weil es vielleicht mal nicht der Drucker ist, sondern ein bisschen komplexeres Problem. **Mhm.** Ich bekomme hier Error 403, was heißt denn das? Dass auch dann, während das sozusagen das Ticket reinkommt, in dem Moment schon Informationen gepusht werden, um den Agenten in seinem Arbeitsalltag zu unterstützen, dass halt nicht immer wieder dieses „ich muss fünf Bildschirme circa aufhaben, um auf jedem Bildschirm irgendwie eine Sache zu machen“, sondern halt integriert zu arbeiten. **Mhm.** Und das glaube ich ist... wenn wir das schaffen, dann wird sich die Nutzungshäufigkeit des Wissensmanagement deutlich... deutlich erhöhen.

Ja okay, mhm. Ähm. Ich muss mal kurz kucken... ja ähm. Gibt es eine konkrete Zielgruppe von dem Produkt, oder sind das sehr gemischte Kunden? Hmm, also wir adressieren schon insgesamt, also branchentechnisch sind wir wahnsinnig weit verbreitet. **Mhm.** Also von Versicherungen zu sehr, sehr großen äh Onlinehändlern, **mhm** ich darf keine Namen nennen, aber ne? Zu sehr, sehr großen Banken ähm bis hin zu dem kleinen Kraftwerk ähm Dienstleister mit dreißig Mitarbeitern. Also branchentechnisch sind wir wahnsinnig verteilt. Also ich kann dir ein paar Namen darf ich tatsächlich nennen, zum Beispiel „Otto“ ist einer unserer Kunden. **Ah, mhm.** Die Deutsche Bank ähm.. so mal um so paar, um so eine Richtlinie zu bekommen, ne? Die darf ich nennen. **Okay (lacht).** Ähm, genau das ist branchentechnisch sind wir sehr verbreitet und Zielgruppe im Sinne von, welche Bereiche wir da adressieren, da ähm ändert sich unser Fokus gerade so ein bisschen von wirklich reinem Kundenservice, äh sprich zum Beispiel bei „Otto“: Wo bleibt mein Paket? **Mhm.** Ähm halt hin zum IT-Service ähm aber das ist so, ne, warum geht mein Drucker nicht. **Ja.** Schlussendlich eigentlich jeder Servicebereich, also tatsächlich haben wir jetzt auch schon sehr große Firmen die bei.. die unser System im HR-Bereich eingeführt haben. **Mhm.** Weil auch dort du einen gewissen Service natürlich verbreiten musst, ne? Jemand ruft an in einer riesengroßen Firma, sagt, mit meiner Abrechnung stimmt was nicht. Da müssen deine Mitarbeiter ja auch irgendwie eine Quelle haben, um zu überprüfen, wodan das liegen könnte. **Ja.** Also genau. Jeglicher Service, insbesondere Kundenservice und branchentechnisch, you name it, also ich... **(lacht)** Da sind wir wirklich wahnsinnig, wahnsinnig breit aufgestellt und auch von wirklich klein bis riesengroß

Ah ja genau, das wollte ich jetzt grade noch fragen, von der, von der Firmengröße her okay, da auch sehr divers. Gut. Genau.

Ähm ja und das hast du eigentlich auch schon so ein bisschen erklärt, wie sich das Produkt jetzt im Lauf der Zeit entwickelt hat. Ähm genau und kannst du dir noch äh außerhalb jetzt von so... wie soll ich sagen, ähm also, das ist ja jetzt eben alles für Firmen und für... ja Servicebereich in der Firma, ähm könntest du dir sonst noch irgendwelche Einsatzorte für so was vorstellen? Ja, über die Frage hab ich auch schon nachgedacht. Das find ich natürlich schwer. Find ich 'ne sehr, sehr spannende Frage, muss ich dir sagen, weil das natürlich dieses klassische: Geh mal raus aus deinem... aus deinem Schuhkarton, in dem du immer denkst. Geh doch mal was ganz anderes dir anzukucken... Andere Einsatzorte außerhalb des Dokumentenmanagements für Produkte dieser Art vorstellen...

Also ich hab da jetzt halt an so was gedacht wie eben Privatpersonen vielleicht, die ähm haben ja auch Fragen (lacht) Mhm. Und ähm. Hmm, ja. **Ja.** Man muss aber natürlich immer sagen, ein Wissensmanagement ist auch irgendwie nur so gut, wie viel Arbeit da reingesteckt wird. Also ein Fehler, den Firmen häufig machen, den ich kennengelernt habe ist... insbesondere auch große Firmen, die denken: Oh, wir haben irgendwie das... den Bedarf hier unser Wissen zu strukturieren. Wir kaufen mal ein System fertig, ne? Haken. Jetzt haben wir das Problem ja gelöst. Und das ist natürlich leider insbesondere im Wissensmanagement absoluter Blödsinn. Du kannst sehr viel technologisch unterstützen und auch durch KI insbesondere kannst du viel unterstützen. Aber Wissensmanagement lebt und fungiert nur so gut, wie viel Liebe da reingesteckt wird. **Okay.** Und deswegen sehe ich momentan dieses Thema für Privatpersonen nicht, weil ich weiß ja nicht, wie viele Fälle, also wenn jemand da irgendwie... weiß ich nicht, wahn-sinnige Vermögensdinge hat, die der strukturieren muss oder so vielleicht. Aber dafür sehe ich Wissensmanagement in der Komplexität, wie unser System es möglich macht, sehe ich es als zu viel für eine Privatperson.

Mhm okay. Ähm ja. Also ja, okay alles klar. Ähm gut, dann kommt jetzt der Block Zukunft. Ja also äh Mhm siehst du Potenziale, Möglichkeiten, wie man dieses Produkt jetzt noch effektiver, noch effizienter machen könnte, vielleicht auch mit Sachen, die noch gar nicht existieren und wenn ja, wie könnte man das verbessern? Hmm...mhm ähm es ist, es ist schlussendlich irgendwo eine Möglichkeit, die schon existiert, nur halt noch nicht so ausgefeilt ist, wie wir sie bräuchten. **Mhm.** Denn was... Also Dokument ist das eine, Wissen ist das andere und jetzt, sag mal simpel gesagt, unser Gespräch hier, das ist schon so viel Wissen, was hier hin und her fließt **Mhm** und trotzdem ist immer eine, ich sag jetzt mal händische Nacharbeit eigentlich notwendig, um sie in ein Dokumenten-/Wissensmanagement hilfreich einzuführen. Natürlich könnten wir jetzt eine Aufnahme nehmen, die in ein... in ein System hochladen und dann, ja gut, was

will man dann damit machen, ne? **Ja**. Wie will man sich das irgendwie groß anhören, sondern man muss es ja abstrahieren, Wissen abstrahieren und zur Verfügung stellen und ähm... Unser Gespräch ist jetzt das eine Beispiel, aber ein anderes Beispiel ist halt zum Beispiel dieses Ticketing-Tool, was ich kurz angesprochen habe, ne? Da kommen halt 500 Fragen rein: Mein Drucker funktioniert nicht, ich hab die und die Fehlermeldung, so. Und die cleveren Mitarbeiter, die brauchen ja gar kein Wissensmanagement. Das ist ja, ich sag jetzt mal, Dokumentenmanagement, Wissensmanagement ist für die alten Hasen, die seit zehn Jahren den Job machen und die absoluten Cracks sind und sich alles merken können, die brauchen das ja theoretisch nicht. Das heißt, das nutzen sie vielleicht auch nicht, sondern beantworten halt alles aus dem Kopf. Ist ja auch fein für deren, in der... in dem Moment zu lösendem Problem. Aber dieses Wissen, was so häufig halt durch Gespräche, oder sonst in irgendwelchen E-Mails steckt. Dieses Wissen müsste man eigentlich halt mit einem Automatismus abstrahieren, zum Beispiel mit einer KI **Mhm** und in ein strukturiertes Dokument, in einen strukturierten Wissensartikel verfassen. Und die Technologie per se existiert schon, aber sie ist noch bei Weitem nicht so ausgefeilt, dass sie wirklich die menschliche Einschätzung von: Das Wissen ist jetzt wichtig und deswegen muss ich das so und so formulieren, X, Y, Z ersetzt. Wenn das irgendwann möglich wäre, das wäre natürlich richtig cool. **(Lacht)** Aber das dauert noch ein bisschen.

Ja (lacht). Okay ähm, ja und wie glaubst du, wird das Dokument der Zukunft aussehen? Ähm. Also glaubst du, dass jetzt... In kleinen Häppchen. **Okay?** Also ich glaub, Dokument der Zukunft wird nicht mehr sein: Dier ist das Firmenhandbuch, 150 Seiten, lies dir das heute doch mal durch und dann bist du ongeboardet. **Ja**. Ich glaube, das Dokument der Zukunft wird ähm immer mehr mit einem Puzzle zu vergleichen sein und dass es die kleinen Häppchen sind, die nach und nach gefüttert werden. Und es gibt immer noch diese Menschen, ich... ich vergleich das immer so ein bisschen mit so einem... so wie wir es früher... früher hat auch niemand so richtig... Hmm, wie erklär ich das. Also, Wissen und Dokumente hat häufig den Drang in Menschen, immer was natürlich komplett zu strukturieren, immer: Ein Dokument muss irgendwie auch nur an einem Ort abgelegt sein und halt bisschen hierarchisch auch angedacht, und so weiter und so fort. Das ist zu starr, wir sind häufig mit den Dokumenten per se relativ flexibel und können es in kleine Häppchen packen, aber die Strukturen, die man für Dokumente häufig braucht, sind häufig noch zu starr. Und das ist, glaube ich, nicht mehr Teil der Zukunft. Ich glaube, wir müssen im Dokumentenmanagement irgendwann so weit kommen, dass es eher wie ein Google-Feeling ist. **Mhm**. Aber trotzdem in der Dokumentenwelt. Verstehst du, was ich meine? **Ja, ja mhm**. Das muss modularer bleiben, flexibler

und äh ja, so ein bisschen halt nicht starr aufgebaut, sondern eher wie so ein Wurzelnetzwerk, was überall dann hochkommt, wo du es brauchst. **Mhm.** Also Dokument der Zukunft ist für mich: kleine Pakete, insbesondere wenig Informationen, die sich dann aber modular zusammensetzen, je nachdem was du brauchst. **Ja okay.** Verständlich? **Ja, ja, ja doch, also habe ich auch schon so ein paar Sachen in die Richtung jetzt gelesen, also ja, ist verständlich (lacht).**

Gut ähm, die nächste Frage ähm, du hast ja jetzt schon gesagt, dass du nicht so die Risiken in KI siehst, von daher denke ich mal, dass du auch eher nicht der Meinung bist, dass da spezielle Gesetze notwendig sein werden vielleicht? Ähm, nein. **Okay (lacht), einfach nein.** Ich glaube es ist immer ähm... das gibt es ja bei... auf verschiedensten Ebenen, so gewisse Work Ethics, ähm zum Beispiel der ehrbare Kaufmann ist so eine Work Ethic, die wir BWLer beigebracht bekommen. Schlussendlich sind das alles grundethische Themen, ne? Wie schon, ne, dieses ähm, wie heißt es denn, verhält dich nur so, wie es eine Grundsatzregel werden könnte und so weiter und so fort. Und ich glaube, dass sind menschliche Gesetze, menschliche Ethiken. Wenn Leute die nicht beacht...nicht beachten wollen, dann beachten sie sie nicht, aber dann beachten sie sie auch in KI nicht. **Mhm.** Also ich glaube nicht, dass wir spezielle KI-Gesetze brauchen, ich glaube eher, dass das wieder einen eher langsamer macht und unsere Innovationsfreudigkeit eher runterschraubt **Mhm** und schlussendlich, wie wir zum Beispiel im Zwei-Wochen-Sprint nah am Kunden entwickeln und versuchen, schnell zu sein, das können wir nur machen, weil wir uns teilweise aus einer großen Organisation rausziehen und sagen, wir sind jetzt mal das schnelle Beiboot und müssen vorfahren, wir müssen jetzt mal schnell sein, wir dürfen nicht von Prozessen oder irgendwelchen Richtlinien gestoppt werden. **Mhm.** Wenn wir jetzt schon über KI-Gesetze nachdenken, ich glaube das macht uns nur langsam und ineffizient. **Mhm.** Ich glaube, ne, es gibt menschliche, es gibt, es gibt Grundprinzipien, es gibt Ethiken, wenn man sich nicht dran halten will, dann will man sich auch nicht in der KI dran halten, nützt, hmm... Also, ich glaube nicht, dass man KI-Gesetze braucht.

Okay. Ähm und glaubst du, dass es jemals möglich sein wird, dass man eben so eine starke Intelligenz entwickelt, die, ja, dem Menschen ebenbürtig ist? Mhm. Irgendwann mal (lacht). Ja, ist auch so eine Debatte, die ähm man teilweise gefühlt nach Stimmung halt immer wieder doch anders beantwortet. Ich glaube, irgendwie ist... gibt so eine Grund... Grundsatz Menschenarroganz sag ich mal, dieses Thema von wegen, irgendwie so was wie: (?) oder im All, nirgendwo können andere Dinge überleben, ich glaub das ist ...geht manchmal von diesem Thema aus irgendwie. Wir denken, wir sind

der Kosmos und auch einzelne Menschen, ich ohne die Welt, das geht doch gar nicht, und so weiter und so fort. Da möchte ich jetzt gar nicht so tief einsteigen (lacht) **Ja**. Deswegen glaub ich, dieses Thema „ebenbürtig“, da sagen sehr viele Leute so von „nä, niemals und wir bleiben doch die Cleversten“. Never say never sag ich. **Okay**. Wer weiß. **Ja, gut ähm aber...** Nicht morgen, nicht übermorgen, aber vielleicht irgendwann. **Okay (lacht) gut ähm ja.**

Ich hab dann noch ein paar spezielle Fragen zu dem Produkt, aber ich glaub, du wolltest ja eh noch so bisschen was dazu sagen, wie es funktioniert und so. Ähm also vielleicht klären sich dann ein paar von diesen Fragen, die ich da noch hab und wenn nicht, dann kann ich die ja noch stellen vielleicht. Also wenn du willst, können wir einfach mal reinkucken. **Mhm.** Ich weiß halt nicht, ähm ich ... warte mal. Kann hier mal... Oh, das Ding hat sich gerade geupdated. Wie teile ich denn meinen Bildschirm in diesem blöden neuen Update? Warte. Ah! Da, wo Bildschirm draufsteht. **(lacht) Gut.** Ich übertrage mal und du sagst mir, wenn du was siehst. **Ja... ja, mhm.** Okay gut. Also. Ganz grob können wir da ja mal kurz durchgehen, dass du mal so ein ungefähres Gefühl dafür bekommst, äh was das Gesamtsystem kann und dann zeig ich dir mal einer der KI-Teilprodukte, wo wir dran arbeiten. Also wir als Team haben ähm drei KI-Produkte, an denen wir speziell feilen. Hmm, oder vier eigentlich jetzt sogar schon und äh eins davon kann ich dir ja mal hier im Kontext zeigen. **Mhm.** Grundsätzlich, ähm wie gesagt sind wir... arbeiten wir webbasiert, bedeutet ähm das ist natürlich durch alle möglichen Passwort äh Möglichkeiten geschützt, aber grundsätzlich brauch man halt nur eine URL, was halt auch zur unserer Philosophie gehört, egal wo du bist, ob im Office, ob unterwegs oder sonst wie, musst du an dein Wissen rankommen. Ähm und nicht irgendwie erst irgendne, ja irgendwas downloaden oder so was eben. **Ja.** Aus dem 19. Jahrhundert. Ach so, das Ganze ist auf Englisch aber das ist egal, oder? **Ja, mhm.** Okay. Hmm ich geh hier jetzt gar nicht so groß unbedingt durch. Grundsatz- äh Philosophie und das ist zum Beispiel etwas, ne, von wegen, wo kommen wir her, wo wollen wir hin, ist, dass das gesamte Wissen hier im System erstmal äh eine gewisse Struktur hat **Mhm.** Ähm und entweder über diese, ja dann doch zwar gut durchdachte, aber recht starre Struktur gefunden werden muss. Dder logischerweise, ein Weg der viel häufiger gegangen wird, ist dass logischerweise, egal wo... wo ich im System befinde, mich hier oben eine Suche unterstützt. Ähm so. Ich kenn mich da jetzt mit diesem System nicht so aus, aber sobald ich halt suche, zeigt er mir halt äh verschiedenste Texte an, verschiedenste Dokumente, was auch immer ich hier so reingepackt hab. Hier ist 'ne PDF, Excel-Dokument. Also es ist ganz egal, was du in dieses System reinpackst, Nachrichten zum Beispiel, auch das, ne, dass Information halt in Nachrichten nicht verloren geht. Alles wird

hier zentral über das System gesteuert. **Mhm** Ähm und hier in der Suche ist natürlich schon 'ne gewisse äh Logik hinterlegt. Aber die wirkliche KI beginnt ähm momentan, weil wir halt sagen ähm ein gutes Wissensmanagement kann so viele Funktionen haben, wie es möchte. Du brauchst halt gute prozessuale ähm Logiken in der Erstellung deines Wissensmanagements, um das ganze langfristig auch ähm ja, mit Leben zu füllen und langfristig auch einen Nutzen zu generieren. **Mhm**. Das bedeutet, einen unserer ersten KI-Produkte, auf die wir uns ähm konzentriert haben, ähm zentriert sich gar nicht rund um den Endnutzer, der Informationen sucht, sondern zentriert sich insbesondere auf die Unterstützung derjenigen, die das Wissen erstellen. Also wir nennen sie dann halt Knowledge Manager. **Mhm** Ähm die dann zum Beispiel ähm (?) Text raus. Das sind hier bestehende Informationen, ne? Bestehende Texte, bestehende Sachen, die ich lesen kann. Und wir haben halt gesagt, wir müssen als erstes diesen Prozess des Wissenserstellen mit einer gewissen KI unterstützen, weil irgendwann hast du hier, das sieht so schön übersichtlich aus, aber wie gesagt, Kunden wie Otto, oder auch die Deutsche Bank haben hier nicht nur fünfzig... fünfzig Wissensartikel drin, sondern eher, ne, über die tausend, teilweise auch 10.000 **mhm** so Wissensartikel. Wie soll jetzt ein Einzelner ähm Knowledge Manager, ein einzelner Mensch, wenn er zum Beispiel ein neues Wissen irgendwie in das System kippen möchte, ein neues Dokument, wie soll irgendwie wissen gibts das schon, äh wo könnte das miteinander verbunden sein und so weiter und sofort. Das ist ein wahnsinnig komplexer Vorgang und irgendwann für einen einzelnen Menschen kaum noch machbar. Deswegen ist einer unserer KI-Produkte, wenn du hier zum Beispiel dann einen neuen Text erstellen möchtest und jetzt kopier ich hier mal relativ simpel das rein, was ich mir eben geklaut habe. Hmm dass jetzt unsere KI anspringt. So jetzt in dem Moment beginnt halt ein das... die Natural Language Processing, ne, **mhm** live wird das, was hier geschrieben ist, in kleine Pakete gehackt, in Vektoren umberechnet und das System wird durchsucht. Okay, was könnte denn hier stehen und dann sagt er dir zum Beispiel A) wo könntest du das in deiner Struktur logisch anhängen, damit dir das irgendwie einen Mehrwert gibt, von dem was hier steht **Mhm** und B) er kuckt halt das gesamte System durch und schaut ob er irgendwie ähnlich... ähnliche Inhalte schon findet. Wenn ich jetzt wortwörtlich das Gleiche schreibe, ist das natürlich, äh funktioniert das sehr simpel. Ich könnte jetzt auch einfach... hmm ich lösche das nochmal raus, irgendwie, so. Fang ich jetzt mal so an zu kucken. So jetzt hab ich hier irgendwie 'nen Satz reingeschrieben und trotzdem geht er natürlich die Texte durch und sagt okay, hier sind zum Beispiel zwei Texte, da verstehen wir was Ähnliches wie das, was du schreibst. Dann kann ich mir das ankucken und dann ist hier zum Beispiel was beschrieben, dann sagt (?) „tarifs“ irgendwie du sprichst hier ja offensichtlich über Tarife ähm und dann markiert er mir hier was, was ähnlich ist und dann kann ich aber wiederum

natürlich mit meiner menschlichen Intelligenz entscheiden, ist das denn jetzt wirklich komplett gleich, kreierte ich hier gerade Redundanzen **mhm** oder ähm ist es wirklich wertvoll, hier nochmal 'n neuen Text rund um das Thema Tarife zu schreiben **mhm** Ne? Und da ist halt die... die KI insbesondere, weil ich hab hier natürlich komplett andere Wörter genommen, die semantisch einfach nur nehmen was hier geschrieben ist Ähnliches. **Ja**. Und so gibt er dir halt die Unterstützung, hey, was gibt es schon im System, pass auf, dass du keine Dubletten oder auch nicht zu ähnliche... zu ähnliches Wissen an unterschiedliche Stellen abspeicherst. Er gibt mir die Unterstützung, wo ich das Wissen äh in meiner Struktur logisch anordnen könnte. Ja hier unten gibt er mir noch Text und weitere Kategorien, das geht jetzt wahrscheinlich viel zu weit in die Tiefe. Aber das ist einer unserer KI-Produkte in kurz umrissen.

Mhm ja das ist interessant, also dass es eben eher auf die Wissenserstellung fokussiert. Also ich hab jetzt 'n anderes Produkt das äh konnte ich dann auch testen also da wurde mir ne Testversion zur Verfügung gestellt und so äh und dann gehts wirklich äh um das Suchen also mhm eben ja, die Leute, die Das ist eigentlich. Ja. Ah, das ist zum Beispiel ein spannender Punkt, so das ist häufig der Punkt, wo die meisten logischerweise anfangen und wie gesagt wir haben ja vier KI-Produkte, das ist einer der Produkte äh die anderen, da geht es dann zum Beispiel um den Self Service, also dass wenn du zum Beispiel dann Ottokunde bist und irgendwie 'ne Frage hast, will Otto ja eigentlich vermeiden, dass du die anrufst, weil **ja** das ist teuer. **(lacht)** Ist natürlich nett, dich kennenzulernen, **(lacht)** aber es ist teuer. Das heißt die möchten dir natürlich möglichst viel Hilfestellung schon auf den Webseiten geben. Da hast du ja immer so einen gewissen Self- Service-Bereich. **Ja** Ne, das kennst du. Hey, das sind unsere FAQs und hier kannst du 'ne Nachricht an uns schreiben. Ähm auch da haben wir natürlich schon über KI-Themen nachgedacht, wie zum Beispiel KI-basierter Chatbot ist einer unserer Produkte, an dem wir jetzt arbeiten, KI-basiertes ähm Suchfeld, dass du halt besser deine Informationen auf der Website findest. Und dann zum Beispiel auch wenn du dann dich entscheidest, hey, jetzt möchte ich aber ne Nachricht an diese Firma schreiben, da schalten wir trotzdem nochmal ne KI zwischen, dann schreibst du da „Hi mein Name ist blah und ich wollte... möchte ich und dann schreibst du da irgendwie so 'nen langen Text und irgendwo versteckt sich deine eigentliche Frage drin. Und da lassen wir auch nochmal 'ne KI durchlaufen, bevor das wirklich an die Firma geht, fragen wir dich hier hinten: Okay warte mal, mal sehen ob das, was du hier geschrieben hast, willst du vielleicht doch einer dieser antworten **ja** kannst ja immer noch sagen, nee will ich nicht, dann hat unsere KI nicht gut gegriffen. Ähm oder vielleicht können wir so auch wieder den Kontakt zur Firma abfangen. Das klingt so negativ, aber eigentlich ist es ja positiv. Wir wollen ja

den Kunden möglichst schnell helfen. Hier **ja** das sind wiederum KI-Aspekte die sich auf den Endkonsumenten fokussieren, ja. **Ja.**

Das ist häufig der erste Punkt, mit dem wir starten. Und damals, als wir mit dem ganzen KI-Schiff losgefahren sind, war halt das Erste, was wir gemacht haben, mit Kunden zu sprechen. Wir haben gefragt, ok, was ist denn euer erster Bedarf, wollt ihr die Suche unterstützen, sprich den Endkonsumenten oder wollt ihr den Ersteller, äh, im Wissens erstellen, äh unterstützen und aufgrund deren Einschätzung haben wir uns dann auf den Fokus gesetzt, ähm, den Ersteller, also den...den Redakteur hier zu unterstützen, **mhm** weil er gesagt hat, das Wissen ist halt leider dann doch irgendwann häufig nicht mehr strukturiert, nicht mehr so gut aufbereitet, weil die Leute keinen Überblick haben und das ist viel, viel, das viel größere Risiko, als schlecht organisiertes Wissensmanagement zu haben, **Ja, auf jeden Fall** als dass irgendwie halt dann 'n Anruf mehr bei uns ankommt und aufgrund dieser Einschätzung des Kunden - also wir hätten das sogar aus unserem Bauchgefühl heraus anders entschieden - aber unsere Kunden haben gesagt, nee, das ist das, was wir dringend brauchen. Also haben wir uns darauf konzentriert.

Okay mhm. Ähm, genau, da...was... Moment, ich muss mal ganz kurz (lacht) kucken. Ah ja, du hast ja schon gesagt, dass es äh eben an verschiedenen, also egal ob ich jetzt daheim bin oder wo auch immer, ich kann das überall aufrufen, ähm, kann man, also ist es auch optimiert für mobile Geräte, Smartphones? Ja, okay gut genau! Also wir haben auch 'ne App, aber wie gesagt, am spannendsten sind eigentlich diese Integrationen, also dass häufig die primäre Arbeitsoberfläche von einem unserer Kunden, ist ja häufig zum Beispiel, das wä...also, dass er irgendwie ne Website hat, auf der er arbeitet, wo irgendwie seine Kunden gepflegt werden, oder sonstiges. Auch dafür haben wir so 'nen Integrator, also wir konzentrieren uns insbesondere auf die natürliche...auf das natürliche Umfeld, wo unsere Kunden sich aufhalten, da irgendwie unser Wissen reinzubaten, weil wir halt sagen, Wissen alleinstehend ist zwar schön, aber hat immer nur den Mehrwert im Kontext, wie der Mensch arbeitet. **Mhm.** Also wenn du in 'ner Ticket-Oberfläche arbeitest und Tickets beantwortest, ne, warum ist mein Drucker kaputt, dann müssen wir dich da, müssen wir uns da irgendwo reinquetschen und sagen, alles klar, du arbeitest hier, dann geben wir dir hier, wo du grade bist, das Wissen zur Verfügung. Das ist insbesondere, primär das, worauf wir uns konzentrieren. Immer uns in die natürliche Arbeitswelt unseres Endkonsumenten, ähm, reinzugeben.

Okay. Gut, ähm, dann wie ist das, wenn ich jetzt ähm...wie soll ich das sagen, äh, also welche Sprachen unterstützt des System? Also wenn ich jetzt halt, keine

Ahnung, hier 'n Dokument auf, ähm, Hindi oder so erstell, äh, ja, wird des dann unterstützt, oder...?

Ja, da müssen wir zwischen zwei unterschiedliche Sachen unterscheiden, entweder KI-Unterstützung oder grundsätzliche Unterstützung. **Mhm.** Äh, grundsätzlich gibt es in unseren Suchlogiken, zum Beispiel in unserer Suchkonfiguration, die ja auch schon, ne gewisse...der ne gewisse Logik folgt, aber ne statische Logik. Da gibt es keine Sprach-einschränkung, **mhm** aber unsere KI ist jetzt, weil sie halt auch noch relativ jung ist, primär auf deutsch und englisch trainiert. **Okay.** Heißt nicht, dass sie nicht in anderen Sprachen greift, aber darauf wurde sie nicht primär trainiert.

Ja okay, macht Sinn. Gut, ähm, wie ist das mit - also da bezieh ich mich jetzt eher, na wobei nicht nur auf die Suche – ähm, wenn ich jetzt schreibe, also sagen wir mal ich schreib da jetzt 'n Text und da, ähm, kommt jetzt irgendwie - der Zusammenhang ist ja jetzt egal – aber da kommt jetzt das Wort „Abtreibung“ und ich schreib was über Abtreibung, ähm, und in dem System ist auch schon 'n Text vorhanden, der das behandelt, allerdings wird da immer über Schwangerschaftsabbruch geredet, mhm also 'n Synonym sozusagen, wird sowas, äh, abgefangen, oder...?

Ja! **Okay.** Das ist genau die KI, auf, die wir trainieren. Also wir sprechen da von „semantisch gleich“ oder „semantisch ähnlich“. **Mhm.** Das Gleiche ist zum Beispiel dann im Selfservice-Bereich, ich kann ja nicht von dir verlangen, dass du genau die gleichen Wörter nutzt, **ja** wie das System. **Ja.** Ja also zum Beispiel, du könntest fragen: „Wie logge ich mich ein?“ und in meinem System, äh, existiert dazu 'n Datensatz, der da heißt: „Wie melde ich mich an?“ **Ja,** ne und dafür ist unsere KI genau gebaut, dass er diese, dass er halt schlussendlich den Kontext deiner Frage versteht, also auch zum Beispiel, wenn du davor irgendwie erklärt hast, ich bin Onlinekunde XYZ, den Kontext versteht, semantisch dann mit anderen Daten gleichsetzen kann schlussendlich **Mhm.**

Das ist dieser Teil, dein Satz wird kleingehackt, in Zahlen umgeformt, genauso wie die anderen und dann kommt auf dieser zahlenbasierten, äh, Ebene wird dann halt nach Gleichheit und Ähnlichkeit geguckt. Und dann würde zum Beispiel „Abtreibung“ oder „Schwangerschaftsabbruch“, würde er checken, dass du das Gleiche meinst **okay** in der Regel. Ist jetzt 'n sehr spezielles Wort, **ja (lacht)**, aber, ähm, ist vielleicht nicht eins der Ersten, auf die wir trainieren, aber äh, ja, das klappt.

Okay und dann in dem Sinn ja auch so „Stemming“, also dass, hm, wenn ich jetzt schreib, ähm, „ich reinige“, dann findet er eben auch „du reinigst“, „er /sie/es

reinigt“ und so weiter. Ja. **Okay, gut.** Ja, ist ein bisschen schwieriger, aber, ja (lacht).
(Lacht) ok.

Ähm, dann jetzt eher bei der Suche, gibts da irgendwelche Operatoren, wie zum Beispiel so ‘n Sternchen oder so, ähm, so Platzhalter und sowas, die man einsetzen kann. Ja tatsächlich, ja, mhm! **Ok (lacht).**

Ja, gibt es! Also du hast die sogenannte „Wildcard-Suche“ nennen wir das **mhm**, heißt zum Beispiel, äh ich weiß nicht wie ‘n Wort anfängt, weiß aber irgendwie muss es mit..., oh jetzt muss ich natürlich...äh hier, „-colation“ enden ne, also ich weiß nicht, dass das Ding „Creation“ oder „Location“ oder weiß ich nicht was heißt, dann kann ich sagen, irgendwas kommt davor und dann...warte (schreibt) „Location“ so, **(lacht)**, das wollt ich sagen, ne. „Relocation“, „Cancellation“, „Information“, also ich sag ihm irgendwas kommt davor und irgendwie endet das auf dieses Wort.

Cool, okay. Sowas gibt es. Es gibt die „Boole“**. Ah, also so Boolesche Operatoren!** Das Wildcard.Ding kann ich sowohl als „prefix“, also ich weiß nicht womit das Wort anfängt, als „postfix“, **mhm** ich weiß nicht, wie das Wort endet oder halt auch als „middlefix“, ich weiß irgendwie gar nix, **(lacht)** ich weiß nur, dass irgendwie, (klickt) irgendwie sowas da drin vorkommt, dann sucht der mir jetzt alle „specific simplification“, ne, des macht er. Ich kann auch mit ‘ner „Oder-Suche“ gehen, also ich möchte irgendwie das Wort „terrif“ oder das Wort „specific“ in meinen Suchergebnissen haben. Also da gibt es mehrere Möglichkeiten.

Okay! Cool. Ja und ähm, also, er zeigt dann hier in dieser Liste zeigt er immer die Dokumente an, wo’s drin ist und rechts dann wahrscheinlich mhm gleich den Part, wo genau das steht, oder? Also...

Genau. Genau, also beziehungsweise er, ähm, zeigt dir hier immer das gesamte Dokument an, aber du findest halt, ähm, oh...jetzt hat er mich grad verschluckt, warte mal...ich glaube einer meiner Kollegen ist...ok ich dachte ich bin rausgeworfen worden (lacht) manchmal benutzen wir gleichzeitig das gleiche System und dann werden wir rausgeworfen **Okay (lacht).** Genau, also hier das, ne, also das, was ich suche, das zeigt er mir dann hier schon an, genau.

Okay, gut ähm okay. Moment, ah ja und ähm jetzt hast du mir ja gezeigt, wie man da eben ein neues Dokument anlegt, aber wenn jetzt zum Beispiel Otto kommt, die haben ja dann schon extrem viele Dokumente, ähm mhm die sie da integrieren wollen und ähm was für Formate werden da unterstützt?

Alles, was du dir vorstellen kannst (**lacht**) **Okay**. Also hmm du meinst jetzt bestehende Dokumente, wie ein PDF, Doc, **PDF** Power Point **Power Point genau**, **Excel** Excel. Das ist unserem System völlig egal, was du da reinpackst. Es wird alles indiziert auch **okay** also außer Bilder, das ist noch etwas, woran wir zukünftig arbeiten, also zum Beispiel wenn du da ein Bild von 'ner Katze hochlädst, dann kannst du aktuell noch nicht nach Katzen suchen, das geht nicht. **Okay, mhm**. Aber wenn du ne Excel hochlädst, äh alles was in der Excel steht, alles was in der Power Point steht, alles was.. das ist dem ganz egal, das wird indiziert und in der Suche berücksichtigt.

Okay und ähm gehen da zum Beispiel auch Scans, also macht das dann so Optical Character Recognition oder funktioniert das nicht? Oh, jetzt musst du mir helfen, was meinst du mit Optical Character Recognition? **Äh also, wenn ich jetzt 'nen Scan von 'nem... wenn ich jetzt das hier einscann**, ah, mhm **dass er dann das erkennt und suchbar und das halt nicht nur ein JPEG ist... okay**. Hmm nee, das kommt noch. **Okay**. Steht auf... steht in unserem Backlog.

Okay, alles klar. Dann... Moment das hab ich auch... das hab ich auch... ähm ja wie ist... Du kannst zum Beispiel hier übrigens auch andere Webseiten einbauen. Also ich kann sogenannte iFrames einbauen, das heißt zum Beispiel, was äh Energiekunden häufig machen, die haben so 'nen Energierechner im Sinne von **ja, ja** wenn du da und da lebst und du so und so viele Leute in deinem Haushalt hast, dann wird dich das das und das kosten. Ist so ein Energierechner, den sie häufig auf der Website haben. Und damit unsere Kunden nicht ständig hin- und herspringen, kannst du teilweise kleine Websiteschnipsel, oder auch große, wie auch immer du das machen möchtest, in deine Website einbauen. Das ist auch ganz... also in dein äh in dein Wissensmanagement einbauen **Aha** also auch... selbst das geht. Weil... nochmal wieder zurück auf das Dokumentenart, das ist uns ganz egal wo die Information liegt.

Okay. Ähm. Genau also das wäre dann jetzt auch die Frage gewesen, beziehungsweise genau ähm, wo also die die Dokumente, die können ganz verstreut sein, ähm also die können jetzt bei einem Unternehmen, was weiß ich, in ihrem Content Management liegen und manche liegen auf SharePoint und manche liegen was weiß ich wo oder muss das alles gesammelt an einem Ort sich befinden? Kommt jetzt drauf an, in welcher Qualität wir sprechen. Also unser Best Practice ist... ist, dass du deine Dokumente, mit denen du schlussendlich dann arbeiten möchtest, im Wissensmanagement in unser System reinziehst wirklich. **Okay**. Die werden hier hochgeladen. **Okay**. Die werden dann in unserer Cloud gespeichert. Ja wir können's, es würde jetzt ein bisschen technischer, wir können jetzt... wir können auch Crawler einbauen. Das äh

Crawler ist 'ne Methodik wo das... zwei unterschiedliche Crawler einmal ein Webseiten-Crawler und einmal ein Dateien-Crawler. Das heißt theoretisch könnte ich auch diesem gesamten System sagen: Durchsuche von jetzt an den gesamten Ort nach, weiß ich nicht, Produktmanagement. Und alles was dann da liegt und dann ist mir das egal, in welchem Unterordner oder sonst wo liegt, wird dann in unser System indiziert und gesucht. Aber nicht Best Practice, das empfehle ich keinem meiner Kunden, das bedeutet nämlich, dass sie leider zu faul sind, Wissen gescheit aufzuarbeiten. **(Lacht)**. Aber technisch haben wir die Möglichkeit ja.

Okay, Verstehe. Gut ja, die anderen Sachen hast du vorhin schon beantwortet. Dann ähm das ja... das wär's dann (lacht) so weit. Also... Ich hab jetzt noch nicht das Gefühl, dass ich dir logischerweise alles über dieses Produkt gesagt habe, weil dann würden wir hier glaub ich drei Stunden sitzen, aber äh klar ich glaub, die Kernfunktionalitäten, also wie gesagt also Wissensmanagement hat bei uns auch immer was mit... mit Nachrichten zu tun. Du musst auch über Wissen benachrichtigen, das wird alles hier zentral verwaltet. Du hast auch die Möglichkeit, Prozesse durchlaufen zu lassen, also zum Beispiel ähm... Das vielleicht nochmal ganz kurz als Hintergrundinfo, eine der Funktionen, die vielleicht jetzt nicht jedes Dokumentenmanagement hat, ist, dass du hier zum Beispiel halt im Kundenservice deine Agenten ein bisschen mehr an die Hand nehmen kannst, damit sie einen gewissen Prozess durchlaufen. Damit sie halt zum Beispiel wenn sie mit 'nem Kunden im Austausch sind immer genau gewisse Dinge fragen, gewisse Dinge abfragen und dann hier Informationen einspeisen, also hier zum Beispiel „Could you tell me your current contract ID please?“ **Mhm.** Und dann hier nochmal 'ne kurze Hilfestellung, auch das wieder als Gedanken wir möchten nicht, dass die Leute arbeiten und sich dann irgendwann Wissen suchen und dann wieder arbeiten und dann wieder Wissen suchen, sondern dass das halt verschmolzen wird. **Ja.** Und hier ist eine der Möglichkeiten, während sie wirklich im Gespräch sind und dann teilweise komplexe äh Prozesse mit dem durchlaufen müssen, zum Beispiel unsere Bankkunden. Klassisches Thema Onlineidentifizierung und Postidentifizierung, weißt du dieses, wie diese ganzen Themen. Die waren am Anfang wahnsinnig komplex und die Leute wussten immer nicht, was sie machen sollen. Dann hatten sie die Leute am Telefon und wussten nicht, wie sie Schritt für Schritt da durchgehen und das ist ein... eine Möglichkeit hier äh zum Beispiel ähm den dann zu unterstützen. Also... geb' ich hier mal kurz irgendwie 'ne Nummer ein und dann kommt hier die nächste Abfrage, was ich fragen soll so weiter und so fort. **Mhm.** Das ist auch noch mal prozessuale Unterstützung im Business Management. **Cool.** Vielleicht das nochmal.

(Es folgt ein organisatorischer Abschnitt, der inhaltlich nichts mit dem Interview zu tun hat und daher nicht transkribiert wurde).

(Lacht) Okay tausend Dank, also ja, wirklich vielen, vielen Dank und noch einen schönen Tag und eine gute Restwoche. Das wünsch ich dir auch, viel Erfolg bei der Arbeit und wir hören. **Danke, tschüss.** Ciao.

Anhang 4: Quellen von im Softwaretest genutzten Dokumenten

Gesetze

Adoptionsübereinkommens-Ausführungsgesetz vom 5. November 2001 (BGBl. I S. 2950), das zuletzt durch Artikel 4 Absatz 17 des Gesetzes vom 17. Dezember 2006 (BGBl. I S. 3171) geändert worden ist. URL: https://www.gesetze-im-internet.de/ad_bag/Ad%C3%9CbAG.pdf (31.08.2020)

Aktiengesetz vom 6. September 1965 (BGBl. I S. 1089), das zuletzt durch Artikel 1 des Gesetzes vom 12. Dezember 2019 (BGBl. I S. 2637) geändert worden ist. URL: <https://www.gesetze-im-internet.de/aktg/AktG.pdf> (31.08.2020)

Alkoholsteuergesetz vom 21. Juni 2013 (BGBl. I S. 1650, 1651), das zuletzt durch Artikel 206 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. URL: <https://www.gesetze-im-internet.de/alkstg/AlkStG.pdf> (31.08.2020)

Arbeitszeitgesetz vom 6. Juni 1994 (BGBl. I S. 1170, 1171), das zuletzt durch Artikel 8 u. Artikel 11 Absatz 2 Satz 2 des Gesetzes vom 27. März 2020 (BGBl. I S. 575) geändert worden ist. URL: <https://www.gesetze-im-internet.de/arbzq/ArbZG.pdf> (31.08.2020)

Asylgesetz in der Fassung der Bekanntmachung vom 2. September 2008 (BGBl. I S. 1798), das zuletzt durch Artikel 3 Absatz 1 des Gesetzes vom 9. Oktober 2020 (BGBl. I S. 2075) geändert worden ist. URL: https://www.gesetze-im-internet.de/asylvfg_1992/AsylG.pdf (31.08.2020)

Außenwirtschaftsgesetz vom 6. Juni 2013 (BGBl. I S. 1482), das zuletzt durch Artikel 1 des Gesetzes vom 10. Juli 2020 (BGBl. I S. 1637) geändert worden ist. URL: https://www.gesetze-im-internet.de/awg_2013/AWG.pdf (31.08.2020)

Bundesausbildungsförderungsgesetz in der Fassung der Bekanntmachung vom 7. Dezember 2010 (BGBl. IS. 1952; 2012 I S. 197), das zuletzt durch Artikel 2 des Gesetzes vom 25. Mai 2020 (BGBl. I S. 1073) geändert worden ist. URL: https://www.gesetze-im-internet.de/baf_g/BAf%C3%B6G.pdf (31.08.2020)

Bundesartenschutzverordnung vom 16. Februar 2005 (BGBl. I S. 258, 896), die zuletzt durch Artikel 10 des Gesetzes vom 21. Januar 2013 (BGBl. I S. 95) geändert worden ist. URL: https://www.gesetze-im-internet.de/bartschv_2005/BArtSchV.pdf (31.08.2020)

Baugesetzbuch in der Fassung der Bekanntmachung vom 3. November 2017 (BGBl. I S. 3634), das zuletzt durch Artikel 2 des Gesetzes vom 8. August 2020 (BGBl. I S. 1728) geändert worden ist. URL: <https://www.gesetze-im-internet.de/bbaug/BauGB.pdf> (31.08.2020)

Bundesdatenschutzgesetz vom 30. Juni 2017 (BGBl. I S. 2097), das durch Artikel 12 des Gesetzes vom 20. November 2019 (BGBl. I S. 1626) geändert worden ist. URL: https://www.gesetze-im-internet.de/bdsg_2018/BDSG.pdf (31.08.2020)

Bundesentschädigungsgesetz in der im Bundesgesetzblatt Teil III, Gliederungsnummer 251-1, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 10 des Gesetzes vom 12. Dezember 2019 (BGBl. I S. 2652) geändert worden ist. URL: <https://www.gesetze-im-internet.de/beg/BEG.pdf> (31.08.2020)

Bürgerliches Gesetzbuch in der Fassung der Bekanntmachung vom 2. Januar 2002 (BGBl. I S. 42, 2909; 2003 I S. 738), das zuletzt durch Artikel 1 des Gesetzes vom 12. Juni 2020 (BGBl. I S. 1245) geändert worden ist. URL: <https://www.gesetze-im-internet.de/bgb/BGB.pdf> (31.08.2020)

Behindertengleichstellungsgesetz vom 27. April 2002 (BGBl. I S. 1467, 1468), das zuletzt durch Artikel 3 des Gesetzes vom 10. Juli 2018 (BGBl. I S. 1117) geändert worden ist. URL: <https://www.gesetze-im-internet.de/bgg/BGG.pdf> (31.08.2020)

Bundesjagdgesetz in der Fassung der Bekanntmachung vom 29. September 1976 (BGBl. I S. 2849), das zuletzt durch Artikel 291 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. URL: <https://www.gesetze-im-internet.de/bjagd/BJagdG.pdf> (31.08.2020)

Bundeskindergeldgesetz in der Fassung der Bekanntmachung vom 28. Januar 2009 (BGBl. I S. 142, 3177), das zuletzt durch Artikel 9 des Gesetzes vom 29. Juni 2020 (BGBl. I S. 1512) geändert worden ist. URL: https://www.gesetze-im-internet.de/bkkg_1996/BKGG.pdf (31.08.2020)

Bundesnaturschutzgesetz vom 29. Juli 2009 (BGBl. I S. 2542), das zuletzt durch Artikel 290 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. URL: https://www.gesetze-im-internet.de/bnatschg_2009/BNatSchG.pdf (31.08.2020)

Börsengesetz vom 16. Juli 2007 (BGBl. I S. 1330, 1351), das zuletzt durch Artikel 61 des Gesetzes vom 20. November 2019 (BGBl. I S. 1626) geändert worden ist. URL: https://www.gesetze-im-internet.de/b_rsg_2007/B%C3%B6rsG.pdf (31.08.2020)

Betäubungsmittelgesetz in der Fassung der Bekanntmachung vom 1. März 1994 (BGBl. I S. 358), das zuletzt durch Artikel 1 der Verordnung vom 10. Juli 2020 (BGBl. I S. 1691) geändert worden ist. URL: https://www.gesetze-im-internet.de/btmq_1981/BtMG.pdf (31.08.2020)

Bundesvertriebenengesetz in der Fassung der Bekanntmachung vom 10. August 2007 (BGBl. I S. 1902), das zuletzt durch Artikel 162 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. URL: <https://www.gesetze-im-internet.de/bvfg/BVFG.pdf> (31.08.2020)

Bundesversorgungsgesetz in der Fassung der Bekanntmachung vom 22. Januar 1982 (BGBl. I S. 21), das zuletzt durch Artikel 4 des Gesetzes vom 12. August 2020

(BGBl. I S. 1879) geändert worden ist. URL: <https://www.gesetze-im-internet.de/bvg/BVG.pdf> (31.08.2020)

Bundswahlordnung in der Fassung der Bekanntmachung vom 19. April 2002 (BGBl. I S. 1376), die zuletzt durch Artikel 10 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. PDF: https://www.gesetze-im-internet.de/bwo_1985/BWO.pdf (31.08.2020)

Erbschaftsteuer- und Schenkungsteuergesetz in der Fassung der Bekanntmachung vom 27. Februar 1997 (BGBl. I S. 378), das zuletzt durch Artikel 12 des Gesetzes vom 26. November 2019 (BGBl. I S. 1794) geändert worden ist. URL: https://www.gesetze-im-internet.de/erbstg_1974/ErbStG.pdf (31.08.2020)

Grundgesetz für die Bundesrepublik Deutschland in der im Bundesgesetzblatt Teil III, Gliederungsnummer 100-1, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 1 u. 2 Satz 2 des Gesetzes vom 29. September 2020 (BGBl. I S. 2048) geändert worden ist. URL: <https://www.gesetze-im-internet.de/gg/GG.pdf> (31.08.2020)

Gerichtsverfassungsgesetz in der Fassung der Bekanntmachung vom 9. Mai 1975 (BGBl. I S. 1077), das zuletzt durch Artikel 3 Absatz 2 des Gesetzes vom 9. Oktober 2020 (BGBl. I S. 2075) geändert worden ist. URL: <https://www.gesetze-im-internet.de/qvg/GVG.pdf> (31.08.2020)

Handelsgesetzbuch in der im Bundesgesetzblatt Teil III, Gliederungsnummer 4100-1, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 1 des Gesetzes vom 12. August 2020 (BGBl. I S. 1874) geändert worden ist. URL: <https://www.gesetze-im-internet.de/hgb/HGB.pdf> (31.08.2020)

Infektionsschutzgesetz vom 20. Juli 2000 (BGBl. I S. 1045), das zuletzt durch Artikel 5 des Gesetzes vom 19. Juni 2020 (BGBl. I S. 1385) geändert worden ist. URL: <https://www.gesetze-im-internet.de/ifsg/IfSG.pdf> (31.08.2020)

Insolvenzordnung vom 5. Oktober 1994 (BGBl. I S. 2866), die zuletzt durch Artikel 24 Absatz 3 des Gesetzes vom 23. Juni 2017 (BGBl. I S. 1693) geändert worden ist. URL: <https://www.gesetze-im-internet.de/inso/InsO.pdf> (31.08.2020)

Bundes-Klimaschutzgesetz vom 12. Dezember 2019 (BGBl. I S. 2513). URL: <https://www.gesetze-im-internet.de/ksg/KSG.pdf> (31.08.2020)

Luftverkehrs-Zulassungs-Ordnung vom 19. Juni 1964 (BGBl. I S. 370), die zuletzt durch Artikel 6 des Gesetzes vom 12. Dezember 2019 (BGBl. I S. 2510) geändert worden ist. URL: <https://www.gesetze-im-internet.de/luftvzo/LuftVZO.pdf> (31.08.2020)

Markengesetz vom 25. Oktober 1994 (BGBl. I S. 3082; 1995 I S. 156; 1996 I S. 682), das zuletzt durch Artikel 1 des Gesetzes vom 11. Dezember 2018 (BGBl. I S. 2357) geändert worden ist. URL: <https://www.gesetze-im-internet.de/markeng/MarkenG.pdf> (31.08.2020)

Medizinproduktegesetz in der Fassung der Bekanntmachung vom 7. August 2002 (BGBl. I S. 3146), das zuletzt durch Artikel 223 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. URL: <https://www.gesetze-im-internet.de/mpg/MPG.pdf> (31.08.2020)

Gesetz über Ordnungswidrigkeiten in der Fassung der Bekanntmachung vom 19. Februar 1987 (BGBl. I S. 602), das zuletzt durch Artikel 185 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. URL: https://www.gesetze-im-internet.de/owig_1968/OWiG.pdf (31.08.2020)

Personenbeförderungsgesetz in der Fassung der Bekanntmachung vom 8. August 1990 (BGBl. I S. 1690), das zuletzt durch Artikel 329 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. URL: <https://www.gesetze-im-internet.de/pbefg/PBefG.pdf> (31.08.2020)

Personenstandsgesetz vom 19. Februar 2007 (BGBl. I S. 122), das zuletzt durch Artikel 88 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. URL: <https://www.gesetze-im-internet.de/pstg/PStG.pdf> (31.08.2020)

Strafgesetzbuch in der Fassung der Bekanntmachung vom 13. November 1998 (BGBl. I S. 3322), das zuletzt durch Artikel 1 des Gesetzes vom 9. Oktober 2020 (BGBl. I S. 2075) geändert worden ist. URL: <https://www.gesetze-im-internet.de/stgb/StGB.pdf> (31.08.2020)

Strafprozeßordnung in der Fassung der Bekanntmachung vom 7. April 1987 (BGBl. I S. 1074, 1319), die zuletzt durch Artikel 2 des Gesetzes vom 9. Oktober 2020 (BGBl. I S. 2075) geändert worden ist. URL: <https://www.gesetze-im-internet.de/stpo/StPO.pdf> (31.08.2020)

Straßenverkehrs-Ordnung vom 6. März 2013 (BGBl. I S. 367), die zuletzt durch Artikel 1 der Verordnung vom 20. April 2020 (BGBl. I S. 814) geändert worden ist. URL: https://www.gesetze-im-internet.de/stvo_2013/StVO.pdf (31.08.2020)

Straßenverkehrs-Zulassungs-Ordnung vom 26. April 2012 (BGBl. I S. 679), die zuletzt durch Artikel 1 der Verordnung vom 26. November 2019 (BGBl. I S. 2015) geändert worden ist. URL: https://www.gesetze-im-internet.de/stvzo_2012/StVZO.pdf (31.08.2020)

Tabakerzeugnisgesetz vom 4. April 2016 (BGBl. I S. 569), das zuletzt durch Artikel 96 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. URL: <https://www.gesetze-im-internet.de/tabakerzq/TabakerzG.pdf> (31.08.2020)

Transfusionsgesetz in der Fassung der Bekanntmachung vom 28. August 2007 (BGBl. I S. 2169), das zuletzt durch Artikel 11 des Gesetzes vom 19. Mai 2020 (BGBl. I S. 1018) geändert worden ist. URL: <https://www.gesetze-im-internet.de/tfg/TFG.pdf> (31.08.2020)

Therapieunterbringungsgesetz vom 22. Dezember 2010 (BGBl. I S. 2300, 2305), das durch Artikel 8 des Gesetzes vom 5. Dezember 2012 (BGBl. I S. 2425) geändert worden ist. URL: <https://www.gesetze-im-internet.de/thug/ThUG.pdf> (31.08.2020)

Tiergesundheitsgesetz in der Fassung der Bekanntmachung vom 21. November 2018 (BGBl. I S. 1938), das zuletzt durch Artikel 100 des Gesetzes vom 20. November 2019 (BGBl. I S. 1626) geändert worden ist. URL: <https://www.gesetze-im-internet.de/tiergesg/TierGesG.pdf> (31.08.2020)

Tierschutzgesetz in der Fassung der Bekanntmachung vom 18. Mai 2006 (BGBl. I S. 1206, 1313), das zuletzt durch Artikel 280 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. URL: <https://www.gesetze-im-internet.de/tierschg/TierSchG.pdf> (31.08.2020)

Tierzuchtgesetz vom 18. Januar 2019 (BGBl. I S. 18). URL: https://www.gesetze-im-internet.de/tierzg_2019/TierZG.pdf (31.08.2020)

Transplantationsgesetz in der Fassung der Bekanntmachung vom 4. September 2007 (BGBl. I S. 2206), das zuletzt durch Artikel 16 des Gesetzes vom 19. Mai 2020 (BGBl. I S. 1018) geändert worden ist. URL: <https://www.gesetze-im-internet.de/tpg/TPG.pdf> (31.08.2020)

Urheberrechtsgesetz vom 9. September 1965 (BGBl. I S. 1273), das zuletzt durch Artikel 1 des Gesetzes vom 28. November 2018 (BGBl. I S. 2014) geändert worden ist. URL: <https://www.gesetze-im-internet.de/urhg/UrhG.pdf> (31.08.2020)

Vermögenssteuergesetz in der Fassung der Bekanntmachung vom 14. November 1990 (BGBl. I S. 2467), das zuletzt durch Artikel 107 des Gesetzes vom 29. Oktober 2001 (BGBl. I S. 2785) geändert worden ist. URL: https://www.gesetze-im-internet.de/vstg_1974/VStG.pdf (31.08.2020)

Verwaltungsverfahrensgesetz in der Fassung der Bekanntmachung vom 23. Januar 2003 (BGBl. I S. 102), das zuletzt durch Artikel 5 Absatz 25 des Gesetzes vom 21. Juni 2019 (BGBl. I S. 846) geändert worden ist. URL: <https://www.gesetze-im-internet.de/vwvfg/VwVfG.pdf> (31.08.2020)

Waffengesetz vom 11. Oktober 2002 (BGBl. I S. 3970, 4592; 2003 I S. 1957), das zuletzt durch Artikel 228 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. URL: https://www.gesetze-im-internet.de/waffg_2002/WaffG.pdf (31.08.2020)

Zivilprozessordnung in der Fassung der Bekanntmachung vom 5. Dezember 2005 (BGBl. I S. 3202; 2006 I S. 431; 2007 I S. 1781), die zuletzt durch Artikel 2 des Gesetzes vom 12. Dezember 2019 (BGBl. I S. 2633) geändert worden ist. URL: <https://www.gesetze-im-internet.de/zpo/ZPO.pdf> (31.08.2020)

Bedienungsanleitungen

Beko (2020a): Hood Manual. HCP 61310 I HCP 61310 I HHCP 61310 W HCP 61310 B HCP 91310 I HCP 91310 W HCP 91310 IH. URL: <https://www.beko.com/de-de/produkte/dunstabzugshauben/dunstabzugshaube-wandmontiert-60-cm-hcp91310ih> (31.08.2020)

Beko (2020b): Wärmeschublade DRW 11401 FB. URL: <https://www.beko.com/de-de/produkte/w%C3%A4rmeschubladen/w%C3%A4rmeschublade-22-l-drw-11401-fb> (31.08.2020)

Bosch (o.J.a): Wärmetrockner WTY887W6. URL: https://media3.bosch-home.com/Documents/9001323827_A.pdf (31.08.2020)

Bosch (o.J.b): TWK86. URL: https://media3.bosch-home.com/Documents/9001320849_A.pdf (31.08.2020)

Bosch (o.J.c): TAT8 61. URL: https://media3.bosch-home.com/Documents/9001317969_A.pdf (31.08.2020)

Bosch (o.J.d): Waschmaschine WU... URL: https://media3.bosch-home.com/Documents/9001299356_C.pdf (31.08.2020)

Bosch (o.J.e): Microwave HMT84M421 HMT84M451. URL: https://media3.bosch-home.com/Documents/9000695741_D.pdf (31.08.2020)

Bosch (o.J.f): Kochfeld PXY8..K... URL: https://media3.bosch-home.com/Documents/9001387659_B.pdf (31.08.2020)

Bosch (o.J.g): Built-in oven VBD5780S0. URL: https://media3.bosch-home.com/Documents/9001299249_C.pdf (31.08.2020)

Bosch (o.J.h): Fully automatic espresso machine CTL636E.6 CTL836E.6. URL: https://media3.bosch-home.com/Documents/8001110278_A.pdf (31.08.2020)

Bosch (o.J.i): Fridge-freezer KIF86.... URL: https://media3.bosch-home.com/Documents/8001082161_A.pdf (31.08.2020)

Bosch (o.J.j): Geschirrspüler SP... URL: https://media3.bosch-home.com/Documents/9001496377_C.pdf (31.08.2020)

Bosch (o.J.k): Li-Ion. URL: https://media3.bosch-home.com/Documents/8001161779_A.pdf (31.08.2020)

HP (o.J.a): HP Tango / HP Tango X. Benutzerhandbuch. URL: <http://h10032.www1.hp.com/ctg/Manual/c06132623> (31.08.2020)

HP (o.J.b): HP Color LaserJet Pro MFP M282-M285. Benutzerhandbuch. URL: <http://h10032.www1.hp.com/ctg/Manual/c06431922> (31.08.2020)

Huawei (2020a): Huawei P40 Lite 5G Benutzerhandbuch. URL: <https://consumer.huawei.com/de/support/tablets/mediapad-t5/> (31.08.2020)

Huawei (2020b): Watch GT 2e Benutzerhandbuch. URL: <https://consumer.huawei.com/de/support/wearables/watch-gt-2e/> (31.08.2020)

Huawei (2020c): Huawei MediaPad T5 Benutzerhandbuch. URL: <https://consumer.huawei.com/de/support/tablets/mediapad-t5/> (31.08.2020)

Husqvarna (2020a): PW 125 PW 235 PW 235R PW 345CPW 350 PW 360 PW 450 PW 460. URL: http://service.webec.husqvarna.net/documents/HUSO/HUSO2019_AAaa/HUSO2019_AAaa_1142145-38.pdf (31.08.2020)

Husqvarna (2020b): 320iB Mark II, 525iB, 530iBX. URL: http://service.webec.husqvarna.net/documents/HUSO/HUSO2020_EUdeEUfrEUitEUnlEUsl/HUSO2020_EUdeEUfrEUitEUnlEUsl_1140930-39.pdf (31.08.2020)

Husqvarna (2020c): 540i XP. URL: http://service.webec.husqvarna.net/documents/HUSO/HUSO2020_EUdeEUsl/HUSO2020_EUdeEUsl_1141164-45.pdf (31.08.2020)

Husqvarna (2020d): 545FR, 545FX, 545FXT, 545RX, 545RXT, 545F. URL: http://service.webec.husqvarna.net/documents/HUSO/HUSO2020_EUdeEUsl/HUSO2020_EUdeEUsl_1141440-45.pdf (31.08.2020)

Husqvarna (2020e): WDC 220, WDC 225, WDC 325L. URL: http://service.webec.husqvarna.net/documents/HUSO/HUSO2019_AAaa/HUSO2019_AAaa_1140550-38.pdf (31.08.2020)

Husqvarna (2020f): TC 138L. URL: http://service.webec.husqvarna.net/documents/HUSO/HUSO2019_EUenEUdeEUfrEUesLAesEUitEUnl/HUSO2019_EUenEUdeEUfrEUesLAesEUitEUnl_114207220.pdf (31.08.2020)

Husqvarna (2020g): ST 324, ST 327, ST 330. URL: http://service.webec.husqvarna.net/documents/HUSO/HUSO2020_EUenEUdeEUfrEUesEUitEUnl/HUSO2020_EUenEUdeEUfrEUesEUitEUnl_1159961-20.pdf (31.08.2020)

Husqvarna (2020h): 520iHE3. URL: http://service.webec.husqvarna.net/documents/HUSO/HUSO2020_EUdeEUsl/HUSO2020_EUdeEUsl_1140371-45.pdf (31.08.2020)

Krups (o.J.a): Krups Sandwich Maker FDK4. URL: <https://www.krups.de/Elektrisches-Kochen/Sandwich-Toaster/Sandwichtoaster-FDK451/p/2100073033> (31.08.2020)

Krups (o.J.b): Waffle Maker FDK2. URL: <https://www.krups.de/Kaffee-%26-Kuchen/Waffeleisen/Waffeleisen-%282-belgische-Waffeln%29-FDK251/p/2100073032?superCategoryCode=elektrisches-kochen> (31.08.2020)

Krups (o.J.c): Krups F230. URL: <https://www.krups.de/bedienungs-anleitungen/csp/2820230702> (31.08.2020)

Krups (o.J.d): Krups GVX1/2. URL: <https://www.krups.de/Fr%C3%BChst%C3%BCck/Kaffeem%C3%BChlen/Kaffee-Espresso-M%C3%BChle/p/1500813230?superCategoryCode=fruhstuck> (31.08.2020)

Krups (o.J.e): Krups GVS1/GVS2. URL: <https://www.krups.de/Speisen-zubereiten/Eismaschine/KRUPS-EISMASCHINE-/p/1500636693> (31.08.2020)

Philips (2018a): PerfectCare Expert Plus Dampfbügelstation GC8962/40. URL: https://www.download.p4c.philips.com/files/g/gc8962_40/gc8962_40_dfu_deu.pdf (31.08.2020)

Philips (2018b): Philips Sonicare ProtectiveClean 6100 Elektrische Schallzahnbürste HX6877/34. URL: https://www.download.p4c.philips.com/files/h/hx6877_34/hx6877_34_dfu_deu.pdf (31.08.2020)

Philips (2018c): Avance Collection Pastamaker HR2381/05. URL: https://www.download.p4c.philips.com/files/h/hr2381_05/hr2381_05_dfu_deu.pdf (31.08.2020)

Philips (2019a): 4K UHD LED Android-Fernseher 70PUS7304/12. URL: https://www.download.p4c.philips.com/files/7/70pus7304_12/70pus7304_12_dfu_deu.pdf (31.08.2020)

Philips (2019b): Essential Airfryer XL HD9262/90. URL: https://www.download.p4c.philips.com/files/h/hd9262_90/hd9262_90_dfu_deu.pdf (31.08.2020)

Philips (2019c): Series 2000 Luftbefeuchter HU4816/10. URL: https://www.download.p4c.philips.com/files/h/hu4816_10/hu4816_10_dfu_deu.pdf (31.08.2020)

Philips (2019d): Soundbar-Lautsprecher HTL3325/10. URL: https://www.download.p4c.philips.com/files/h/htl3325_10/htl3325_10_dfu_deu.pdf (31.08.2020)

Philips (2019e): Schnurlostelefon mit Anrufbeantworter XL4951DS/05. URL: https://www.download.p4c.philips.com/files/x/xl4951ds_05/xl4951ds_05_dfu_deu.pdf (31.08.2020)

Philips (2020a): NeoPix Ultra Mini-Projektor NPX640/INT. URL: https://www.download.p4c.philips.com/files/n/npx640_int/npx640_int_dfu_deu.pdf (31.08.2020)

Philips (2020b): Lumea Advanced IPL-Haarentfernungsgerät BRI923/00. URL: <https://philipsproductcontent.blob.core.windows.net/assets/20200211/21583f6b42f64aafb2cfab5e00fc522f.pdf> (31.08.2020)

Samsung (o.J.): Benutzerhandbuch NX1100. URL: https://downloadcenter.samsung.com/content/UM/201303/20130306143127828/NX1100_German.pdf (31.08.2020)

WMF (o.J.a): Kult X Brotbackautomat. URL: <https://www.wmf.com/de/wmf-kult-x-brotbackautomat.html> (31.08.2020)

WMF (o.J.b): WMF KÜCHENminis Vitalis E Dampfgarer. URL: <https://www.wmf.com/de/wmf-kuechenminis-dampfgarer.html> (31.08.2020)

WMF (o.J.c): WMF KÜCHENminis Dörrautomat Snack to go. URL: <https://www.wmf.com/de/wmf-kuechenminis-doerrautomat-snack-to-go.html> (31.08.2020)

WMF (o.J.d): WMF Lono Master Grill. URL: <https://www.wmf.com/de/wmf-lono-master-grill.html> (31.08.2020)

WMF (o.J.e): Profi Plus Küchenmaschine. URL: <https://www.wmf.com/de/kuechengeraeete/kuechenkleingeraeete/kuechenmaschinen/wmf-profi-plus-kuechenmaschine-platin-bronze.html> (31.08.2020)

WMF (o.J.f): Kult Pro Multifunktionsmixer. URL: <https://www.wmf.com/de/kult-pro-multifunktionsmixer.html> (31.08.2020)

WMF (o.J.g): Lono Raclette. URL: <https://www.wmf.com/de/wmf-lono-raclette.html> (31.08.2020)

WMF (o.J.h): Kult X Spiralschneider. URL: <https://www.wmf.com/de/kuechengeraeete/elektroschneider/elektrische-spiralschneider/wmf-kult-x-spiralschneider.html> (31.08.2020)

Prosa, Dramatik, Epik

De Balzac, Honorè (2004): Große und kleine Welt, in: DigBib.Org. URL: [http://www.digbib.org/Honore de Balzac 1799/De Grosse und Kleine Welt .pdf](http://www.digbib.org/Honore_de_Balzac_1799/De_Grosse_und_Kleine_Welt_.pdf) (30.08.2020)

Fontane, Theodor (2004): Die Poggenpuhls, in: DigBib.Org. URL: [http://www.digbib.org/Theodor Fontane 1819/Die Poggenpuhls .pdf](http://www.digbib.org/Theodor_Fontane_1819/Die_Poggenpuhls_.pdf) (30.08.2020)

Fontane, Theodor (2004): Unterm Birnbaum, in: DigBib.Org. URL: [http://www.digbib.org/Theodor Fontane 1819/Unterm Birnbaum .pdf](http://www.digbib.org/Theodor_Fontane_1819/Unterm_Birnbaum_.pdf) (30.08.2020)

Fontane, Theodor (2005): Effi Briest, in: DigBib.Org. URL: [http://www.digbib.org/Theodor Fontane 1819/Effi Briest .pdf](http://www.digbib.org/Theodor_Fontane_1819/Effi_Briest_.pdf) (30.08.2020)

Goethe, Johann Wolfgang von (2004a): Der neue Paris, in: DigBib.Org. URL: [http://www.digbib.org/Johann Wolfgang von Goethe 1749/Der neue Paris .pdf](http://www.digbib.org/Johann_Wolfgang_von_Goethe_1749/Der_neue_Paris_.pdf) (30.08.2020)

Goethe, Johann Wolfgang von (2004b): Die gefährliche Wette, in: DigBib.Org. URL: [http://www.digbib.org/Johann Wolfgang von Goethe 1749/Die gefaehrliche Wette .pdf](http://www.digbib.org/Johann_Wolfgang_von_Goethe_1749/Die_gefaehrliche_Wette_.pdf) (30.08.2020)

Goethe, Johann Wolfgang von (2004c): Die Geschichte von Mignons Eltern, in: DigBib.Org. URL: [http://www.digbib.org/Johann Wolfgang von Goethe 1749/Die Geschichte von Mignons Eltern .pdf](http://www.digbib.org/Johann_Wolfgang_von_Goethe_1749/Die_Geschichte_von_Mignons_Eltern_.pdf) (30.08.2020)

Goethe, Johann Wolfgang von (2004d): Die Leiden des jungen Werther, in: DigBib.Org. URL: [http://www.digbib.org/Johann Wolfgang von Goethe 1749/Die Leiden des jungen Werther .pdf](http://www.digbib.org/Johann_Wolfgang_von_Goethe_1749/Die_Leiden_des_jungen_Werther_.pdf) (30.08.2020)

Goethe, Johann Wolfgang von (2004e): Egmont, in: DigBib.Org. URL: [http://www.digbib.org/Johann Wolfgang von Goethe 1749/Egmont .pdf](http://www.digbib.org/Johann_Wolfgang_von_Goethe_1749/Egmont_.pdf) (30.08.2020)

Goethe, Johann Wolfgang von (2004f): Eine Gespenstergeschichte, in: DigBib.Org. URL: [http://www.digbib.org/Johann Wolfgang von Goethe 1749/Eine Gespenstergeschichte .pdf](http://www.digbib.org/Johann_Wolfgang_von_Goethe_1749/Eine_Gespenstergeschichte_.pdf) (30.08.2020)

Goethe, Johann Wolfgang von (2004g): Faust I, in: DigBib.Org. URL: [http://www.digbib.org/Johann Wolfgang von Goethe 1749/Faust I .pdf](http://www.digbib.org/Johann_Wolfgang_von_Goethe_1749/Faust_I_.pdf) (30.08.2020)

Goethe, Johann Wolfgang von (2005): Götz von Berlichingen, in: DigBib.Org. URL: [http://www.digbib.org/Johann Wolfgang von Goethe 1749/Goetz von Berlichingen .pdf](http://www.digbib.org/Johann_Wolfgang_von_Goethe_1749/Goetz_von_Berlichingen_.pdf) (30.08.2020)

Goethe, Johann Wolfgang von (2006): Faust II, in: DigBib.Org. URL: [http://www.digbib.org/Johann Wolfgang von Goethe 1749/Faust II .pdf](http://www.digbib.org/Johann_Wolfgang_von_Goethe_1749/Faust_II_.pdf) (30.08.2020)

- Grillparzer, Franz (2004a):** Das Kloster bei Sendomir, in: DigBib.Org. URL: [http://www.digbib.org/Franz Grillparzer 1791/Das Kloster bei Sendomir .pdf\(30.08.2020\)](http://www.digbib.org/Franz_Grillparzer_1791/Das_Kloster_bei_Sendomir_.pdf(30.08.2020))
- Grillparzer Franz (2004b):** Der arme Spielmann, in: DigBib.Org. URL: [http://www.digbib.org/Franz Grillparzer 1791/Der arme Spielmann .pdf](http://www.digbib.org/Franz_Grillparzer_1791/Der_arme_Spielmann_.pdf) (30.08.2020)
- Grillparzer, Franz (2004c):** Der Traum ein Leben, in: DigBib.Org. URL: [http://www.digbib.org/Franz Grillparzer 1791/Der Traum ein Leben .pdf](http://www.digbib.org/Franz_Grillparzer_1791/Der_Traum_ein_Leben_.pdf) (30.08.2020)
- Heine, Heinrich (2004a):** Almansor, in: DigBib.Org. URL: [http://www.digbib.org/Heinrich Heine 1797/Almansor .pdf](http://www.digbib.org/Heinrich_Heine_1797/Almansor_.pdf) (30.08.2020)
- Heine, Heinrich (2004b):** Aus den Memoiren des Herren von Schnabelewopski, in: DigBib.Org. URL: [http://www.digbib.org/Heinrich Heine 1797/Aus den Memoiren des Herren von Schnabelewopski .pdf](http://www.digbib.org/Heinrich_Heine_1797/Aus_den_Memoiren_des_Herren_von_Schnabelewopski_.pdf) (30.08.2020)
- Homer (2006):** Odyssee, in: DigBib.Org. URL: [http://www.digbib.org/Homer 8JHv-Chr/De Odyssee .pdf](http://www.digbib.org/Homer_8JHv-Chr/De_Odyssee_.pdf) (30.08.2020)
- Homer (2008):** Ilias, in: DigBib.Org. URL: [http://www.digbib.org/Homer 8JHv-Chr/De Ilias .pdf](http://www.digbib.org/Homer_8JHv-Chr/De_Ilias_.pdf) (30.08.2020)
- Kafka, Franz (2005a):** Auf der Galerie, in: DigBib.Org. URL: [http://digbib.org/Franz Kafka 1883/Auf der Galerie .pdf\(30.08.2020\)](http://digbib.org/Franz_Kafka_1883/Auf_der_Galerie_.pdf(30.08.2020))
- Kafka, Franz (2005b):** Das Urteil, in: DigBib.Org. URL: [http://digbib.org/Franz Kafka 1883/Das Urteil .pdf](http://digbib.org/Franz_Kafka_1883/Das_Urteil_.pdf) (30.08.2020)
- Kafka, Franz (2005c):** Der Gruftwächter, in: DigBib.Org. URL: [http://www.digbib.org/Franz Kafka 1883/Der Gruftwaechter .pdf](http://www.digbib.org/Franz_Kafka_1883/Der_Gruftwaechter_.pdf) (30.08.2020)
- Kafka, Franz (2005d):** Der Proceß, in: DigBib.Org. URL: [http://digbib.org/Franz Kafka 1883/Der Prozess .pdf](http://digbib.org/Franz_Kafka_1883/Der_Prozess_.pdf) (30.08.2020)
- Kafka, Franz (2005e):** In der Strafkolonie, in: DigBib.Org. URL: [http://digbib.org/Franz Kafka 1883/In der Strafkolonie .pdf](http://digbib.org/Franz_Kafka_1883/In_der_Strafkolonie_.pdf) (30.08.2020)
- Kafka, Franz (2011a):** Das Schloss, in: DigBib.Org. URL: [http://digbib.org/Franz Kafka 1883/Das Schloss .pdf](http://digbib.org/Franz_Kafka_1883/Das_Schloss_.pdf) (30.08.2020)
- Kafka, Franz (2011b):** Der Bau, in: DigBib.Org. URL: [http://digbib.org/Franz Kafka 1883/Der Bau .pdf](http://digbib.org/Franz_Kafka_1883/Der_Bau_.pdf) (30.08.2020)
- Kafka, Franz (2011c):** Die Verwandlung, in: DigBib.Org. URL: [http://digbib.org/Franz Kafka 1883/Die Verwandlung .pdf](http://digbib.org/Franz_Kafka_1883/Die_Verwandlung_.pdf) (30.08.2020)
- Kafka, Franz (2013):** Amerika, in: DigBib.Org. URL: [http://digbib.org/Franz Kafka 1883/Amerika .pdf](http://digbib.org/Franz_Kafka_1883/Amerika_.pdf) (30.08.2020)

Kleist, Heinrich von (2004): Das Erdbeben in Chili, in: DigBib.Org. URL: http://www.digbib.org/Heinrich_von_Kleist_1777/Das_Erdbeben_in_Chili_.pdf (30.08.2020)

Kleist, Heinrich von (2004): Der Findling, in: DigBib.Org. URL: http://www.digbib.org/Heinrich_von_Kleist_1777/Der_Findling_.pdf (30.08.2020)

Laßwitz, Kurt (2007): Auf zwei Planeten, in: DigBib.Org. URL: http://www.digbib.org/Kurd_Lasswitz_1848/Auf_zwei_Planeten_.pdf (30.08.2020)

Lessing, Gotthold Ephraim (2004a): Der Freigeist, in: DigBib.Org. URL: http://www.digbib.org/Gotthold_Ephraim_Lessing_1729/Der_Freigeist_.pdf (30.08.2020)

Lessing, Gotthold Ephraim (2004b): Emilia Galotti, in: DigBib.Org. URL: http://www.digbib.org/Gotthold_Ephraim_Lessing_1729/Emilia_Galotti_.pdf (30.08.2020)

Lessing, Gotthold Ephraim (2004c): Minna von Barnhelm, in: DigBib.Org. URL: http://www.digbib.org/Gotthold_Ephraim_Lessing_1729/Minna_von_Barnhelm_.pdf (30.08.2020)

Lessing, Gotthold Ephraim (2004d): Nathan der Weise, in: DigBib.Org. URL: http://www.digbib.org/Gotthold_Ephraim_Lessing_1729/Nathan_der>Weise_.pdf (30.08.2020)

Rilke, Rainer Maria (2004): Die Aufzeichnungen des Malte Laurids Brigge, in: DigBib.Org. URL: http://www.digbib.org/Rainer_Maria_Rilke_1875/Die_Aufzeichnungen_des_Malte_Laurids_Brigge_.pdf (30.08.2020)

Schiller, Friedrich (2004): Die Räuber, in: DigBib.Org. URL: http://www.digbib.org/Friedrich_von_Schiller_1759/Die_Raeuber_.pdf (30.08.2020)

Schiller, Friedrich (2005): Wilhelm Tell, in: DigBib.Org. URL: http://www.digbib.org/Friedrich_von_Schiller_1759/Wilhelm_Tell_.pdf (30.08.2020)

Shakespeare, William (2004a): Der Kaufmann von Venedig, in: DigBib.Org. URL: http://www.digbib.org/William_Shakespeare_1564/De_Der_Kaufmann_von_Venedig_.pdf (30.08.2020)

Shakespeare, William (2004b): Ein Sommernachtstraum, in: DigBib.Org. URL: http://www.digbib.org/William_Shakespeare_1564/De_Ein_Sommernachtstraum_.pdf (30.08.2020)

Shakespeare, William (2004c): Macbeth, in: DigBib.Org. URL: http://www.digbib.org/William_Shakespeare_1564/De_Macbeth_.pdf (30.08.2020)

Shakespeare, William (2004d): Romeo und Juliette, in: DigBib.org. URL: http://www.digbib.org/William_Shakespeare_1564/De_Romeo_und_Juliette_.pdf (30.08.2020)

Shakespeare, William (2004e): Viel Lärm um Nichts, in: DigBib.org. URL: http://www.digbib.org/William_Shakespeare_1564/De_Viel_Laerm_um_nichts_.pdf (30.08.2020)

Shakespeare, William (2004f): Was ihr wollt, in: DigBib.org. URL: http://www.digbib.org/William_Shakespeare_1564/De_Was_ihr_wollt_.pdf (30.08.2020)

Shakespeare, William (2005): Das Winter-Mährchen, in: DigBib.org. URL: http://www.digbib.org/William_Shakespeare_1564/De_Das_Winter-Maehrchen_.pdf (30.08.2020)

Spyri, Johanna (2006): Heimatlos, in: DigBib.Org. URL: http://www.digbib.org/Johanna_Spyri_1827/Heimatlos_.pdf (30.08.2020)

Storm, Theodor (2004a): Der Schimmelreiter, in: DigBib.Org. URL: http://www.digbib.org/Theodor_Storm_1817/Der_Schimmelreiter_.pdf (30.08.2020)

Storm, Theodor (2004b): Pole Poppenspärer, in: DigBib.Org. URL: http://www.digbib.org/Theodor_Storm_1817/Pole_Poppenspaeler_.pdf (30.08.2020)

Tucholsky, Kurt (2006): Schloß Gripsholm, in: DigBib.Org. URL: http://www.digbib.org/Kurt_Tucholsky_1890/Schloss_Gripsholm_.pdf (30.08.2020)

Wissenschaftliche Artikel

Bibel, Wolfgang (2011): KI ohne Geist, in: Künstliche Intelligenz 25, H. 4, S. 299-302. DOI: <https://doi.org/10.1007/s13218-011-0126-7> (30.08.2020)

Bohn, Andreas, Güting, Thomas; Mansmann, Till; Selle, Stefan (2003): MoneyBee: Aktienkursprognose mit künstlicher Intelligenz bei hoher Rechenleistung, in: Wirtschaftsinformatik 45, H.3, S. 325–333. <https://doi.org/10.1007/BF03254950> (30.08.2020)

Bonsanto, Matteo M.; Tronnier, V. M. (2020): Künstliche Intelligenz in der Neurochirurgie, in: Der Chirurg 91, H. 3, S. 229-234. DOI: <https://doi.org/10.1007/s00104-020-01131-9> (30.08.2020)

Burri, Thomas (2018): Künstliche Intelligenz und internationales Recht, in: Datenschutz und Datensicherheit 42, H. 10, S. 603–607. DOI: <https://doi.org/10.1007/s11623-018-1009-y> (30.08.2020)

Capurro, Rafael (1993): Ein Grinsen ohne Katze. Von der Vergleichbarkeit zwischen 'künstlicher Intelligenz' und 'getrennten Intelligenzen', in: Zeitschrift für philosophische Forschung 47, H. 1, S. 93-102. URL: <https://www.istor.org/stable/20483510> (30.08.2020)

Conrad, Sebastian Conrad (2017a): Kann die Künstliche Intelligenz den Menschen entschlüsseln? — Neue Forderungen zum Datenschutz, in: Datenschutz und Datensicherheit 42, H. 9, S. 541–546. DOI: <https://doi.org/10.1007/s11623-018-0996-z> (30.08.2020)

Conrad, Sebastian Conrad (2017b): Künstliche Intelligenz — Die Risiken für den Datenschutz, in: Datenschutz und Datensicherheit 41, H. 12, S.740–744. DOI: <https://doi.org/10.1007/s11623-017-0870-4> (30.08.2020)

Denga, Michael (2018): Deliktische Haftung für künstliche Intelligenz, in: Computer und Recht 34, H. 2, S. 69-78. DOI: <https://doi.org/10.9785/cr-2018-0203> (30.08.2020)

Dengel, Andreas (2011): Künstliche Intelligenz in Anwendungen, in: Künstliche Intelligenz 25, H. 4, S. 317-319. DOI: <https://doi.org/10.1007/s13218-011-0129-4> (30.08.2020)

Dullien, Thomas (2018): Maschinelles Lernen und künstliche Intelligenz in der Informationssicherheit, in: Datenschutz und Datensicherheit 42, H. 10, S. 618–622. DOI: <https://doi.org/10.1007/s11623-018-1012-3> (30.08.2020)

Empel, Claire; Koenigs, Rene M. (2019): Künstliche Intelligenz in der organischen Synthese – en route zu autonomer Synthese?, in: Angewandte Chemie 131, H. 48, S. 17272-17274. DOI: <https://doi.org/10.1002/ange.201911062> (30.08.2020)

- Esch, Franz-Rudolf; Muffler, Tania (1989):** Expertensysteme im Marketing, in: Marketing: ZFP – Journal of Research and Management 11, H. 3, S. 145-152. URL: <https://www.jstor.org/stable/42746050> (30.08.2020)
- Euchner, Walter (2005):** Der künstlich verbesserte Mensch und die „künstliche Intelligenz“ — Vorgeschichte und aktuelle Diskussion, in: Leviathan 33, H. 1, S. 40-68. URL: <https://www.jstor.org/stable/23983949> (30.08.2020)
- Fabricius, Dirk (2011):** Der Mensch ist eine Maschine. Maschinen können frei sein, in: Archiv für Rechts- und Sozialphilosophie 97, H. 4, S. 556-573. URL: <https://www.jstor.org/stable/23681140> (30.08.2020)
- Fink, Robin D.; Weyer, Johannes (2011):** Autonome Technik als Herausforderung der soziologischen Handlungstheorie, in: Zeitschrift für Soziologie 40, H. 2, S. 91-111. URL: <https://www.jstor.org/stable/23773317> (30.08.2020)
- Fösel, Thomas; Marquardt, Florian; Weiss, Talitha (2019):** Kommt der künstliche Physiker?, in: Physik in unserer Zeit 50, H. 5, S. 220-227. DOI: <https://doi.org/10.1002/piuz.201901549> (30.08.2020)
- Furbach, Ulrich (2012):** Turing und Künstliche Intelligenz, in: Informatik-Spektrum 35, H. 4, S. 280-286. DOI: <https://doi.org/10.1007/s00287-012-0623-6> (30.08.2020)
- Gebhardt, Christiane; Giesecke, Susanne (1997):** Die Spezifität der Entwicklungspfade in der Biotechnologie und der Künstlichen Intelligenz, in: Comparativ 7, H. 3, S. 76-97. URL: <https://www.comparativ.net/v2/article/view/1340/1185> (30.08.2020)
- Heinecke, Albert (1993):** Das Potential von Expertensystemen im Rahmen der Personalwirtschaft, in: Zeitschrift für Personalforschung 7, H. 2, S. 247-270. URL: <https://www.jstor.org/stable/23281534> (30.08.2020)
- Heinen, Nicolaus; Heuer, Alexander; Schautschick, Philipp (2017):** Künstliche Intelligenz und der Faktor Arbeit, in: Wirtschaftsdienst 97, S. 714–720. DOI: <https://doi.org/10.1007/s10273-017-2203-5> (30.08.2020)
- Hill, Hermann (2018):** Was bedeutet Künstliche Intelligenz (KI) für die Öffentliche Verwaltung?, in: VM Verwaltung & Management 24, H. 6, S. 287 – 294. DOI: <https://doi.org/10.5771/0947-9856-2018-6-271> (30.08.2020)
- Hofstedt, Petra (2012):** Die Evolution der Programmiersprachen der KI, in: Künstliche Intelligenz 26, H.1, S. 11-13. DOI: <https://doi.org/10.1007/s13218-011-0157-0> (30.08.2020)
- Holenstein, Elmar (1988):** Eine Maschine im Geist. Husserlsche Begründung und Begrenzung künstlicher Intelligenz, in: Phänomenologische Forschungen 21, S. 82-113. URL: <https://www.jstor.org/stable/24360295> (30.08.2020)

Joerden, Jan C. (2012): Maschinen mit Würde? Thesen zu einem Turing-Test für Würde, in: Jahrbuch für Recht und Ethik 20, S. 311-318. URL: <https://www.jstor.org/stable/43593912> (30.08.2020)

Käde, Lisa; von Maltzan, Stephanie (2020): Die Erklärbarkeit von Künstlicher Intelligenz (KI), in: Computer und Recht 36, H. 1, S. 66-72. DOI: <https://doi.org/10.9785/cr-2020-360115> (30.08.2020)

Kirn, Stefan; Müller-Hengstenberg, Claus D. (2015): Technische und rechtliche Betrachtungen zur Autonomiekoooperativ-intelligenter Softwareagenten, in: Künstliche Intelligenz 29, S. 59–74. DOI: <https://doi.org/10.1007/s13218-014-0334-z> (30.08.2020)

Kleesiek, Jens; Murray, Jacob M.; Kaissis, Georgios; Brare, Rickmer (2020): Künstliche Intelligenz und maschinelles Lernen in der onkologischen Bildgebung, in: Der Onkologe 26, H. 1, S. 60-65. DOI: <https://doi.org/10.1007/s00761-019-00679-4> (30.08.2020)

Klimczak, Peter; Kusche, Isabel; Tschöpe, Constanze; Wolff, Matthias (2019): Menschliche und maschinelle Entscheidungsrationalität. Zur Kontrolle und Akzeptanz Künstlicher Intelligenz, in: Zeitschrift für Medienwissenschaft 11, H. 2, S. 39–45. DOI: <https://doi.org/10.25969/mediarep/12631> (30.08.2020)

Krummheuer, Antonia (2011): Künstliche Interaktionen mit Embodied Conversational Agents, in: Technikfolgenabschätzung – Theorie und Praxis 20, H.1, S. 32-39. <https://doi.org/10.14512/tatup.20.1.32> (30.08.2020)

Latoschik, Marc Erich (2014): Smart Graphics/Intelligent Graphics, in: Informatik-Spektrum 37, H. 1, S. 36-41. DOI: <https://doi.org/10.1007/s00287-013-0759-z> (30.08.2020)

Lüth, Christoph; Krieg-Brückner, Bernd (2007): Sicherheit in der Künstlichen Intelligenz, in: Künstliche Intelligenz 21, H.1, S. 51-52. URL: <https://www-cps.hb.dfki.de/sams/papers/sicherheit-ki.pdf> (30.08.2020)

Mersch, Dieter (2019): Kreativität und Künstliche Intelligenz. Einige Bemerkungen zu einer Kritik algorithmischer Rationalität, in: Zeitschrift für Medienwissenschaft 11, H.2, S. 65-74. DOI: <https://doi.org/10.25969/med> (30.08.2020)

Mertens, Peter (1984): Vergleich zwischen Methoden der künstlichen Intelligenz und alternativen Entscheidungsunterstützungstechniken, in: Die Unternehmung 48, H.1, S. 3-16. URL: <https://www.jstor.org/stable/24183039> (30.08.2020)

Möller, Ralf (2011): Zur Rolle der Logik bei der Entwicklung Intelligenter Systeme, in: Künstliche Intelligenz 25, H. 4, S. 309-311. DOI: <https://doi.org/10.1007/s13218-011-0134-7> (30.08.2020)

- Mühlhoff, Rainer (2019):** Menschengestützte Künstliche Intelligenz. Über die soziotechnischen Voraussetzungen von «deep learning», in: Zeitschrift für Medienwissenschaft 11, H. 2, S. 56–64. DOI: <https://doi.org/10.25969/mediarep/12633> (30.08.2020)
- Puppe, Frank (2011):** Explizites Wissen versus Black Box Ansätze in der KI, in: Künstliche Intelligenz 25, H.1, S. 67-70. DOI: <https://doi.org/10.1007/s13218-010-0069-4> (30.08.2020)
- Richter, Michael M.; Mammen, Sebastian von (2011):** ZEITSCHRIFTEine Zukunftsperspektive der Künstlichen Intelligenz am Beispielmenschlicher Kommunikation mit Rechnern, in: Künstliche Intelligenz 25, S. 303-306. DOI: <https://doi.org/10.1007/s13218-011-0125-8> (30.08.2020)
- Saage, Richard (2007):** Wissenschaftlich-technische und normative Grundlagen des Neuen Menschen, in: Zeitschrift für Politik 54, H. 2, S. 123-140. URL: <https://www.jstor.org/stable/24228901> (30.08.2020)
- Schael, Christopher (2018):** Künstliche Intelligenz in der modernen Gesellschaft, in: Datenschutz und Sicherheit 42, H. 9, S. 547-551. DOI: <https://doi.org/10.1007/s11623-018-0997-y> (30.08.2020)
- Schaffert, Sebastian; Bry, François; Baumeister, Joachim; Kiesel, Malte (2007):** Semantic Wiki, in: Informatik-Spektrum 30, H. 6. DOI: <https://doi.org/10.1007/s00287-007-0195-z> (30.08.2020)
- Schmid, Ute (2012):** KI und Informatik, in: Künstliche Intelligenz 26, H. 1, S. 1-4. DOI: <https://doi.org/10.1007/s13218-011-0163-2> (30.08.2020)
- Schmitt, Sascha; Schneider, Bernd (2001):** Einsatzpotentiale der KI im Electronic Commerce, in: Künstliche Intelligenz 15, H. 1, S. 5-11. URL: https://www.researchgate.net/publication/220633195_Einsatzpotentiale_der_KI_im_Electronic_Commerce (30.08.2020)
- Schneider, W.; Toplak, W. (2008):** Verkehrsprognosen mit Visuellem Data Mining und Künstlicher Intelligenz, in: Elektrotechnik & Informationstechnik 125, H. 6, S. 232-237. DOI: <https://doi.org/10.1007/s00502-008-0538-8>
- Steiner, Eberhard (2003):** Wertorientiertes Risikocontrolling mit Künstlichen Neuronalen Netzen, in: Controlling 15, H. 11, S. 631-639. DOI: <https://doi.org/10.15358/0935-0381-2003-11-631> (30.08.2020)
- Strack, Christian; Seifert, Robert; Kleesiek, Jens (2020):** Künstliche Intelligenz in derHybridbildung, in: Der Radiologe 60, H. 5, S. 405-412. DOI: <https://doi.org/10.1007/s00117-020-00646-w> (30.08.2020)
- Twoerk, S. et al. (2003):** Künstliche neuronale Netze zur Steuerung von Heimbeatmungsgeräten, in: Monatsschrift Kinderheilkunde 151, H. 3, S. 274–283. DOI: <https://doi.org/10.1007/s00112-003-0666-5> (30.08.2020)

Udluft, Heiko (2018): Schach, Suchmaschine, Sprachassistent – unser Verhältnis zur Künstlichen Intelligenz, in: Digitale Welt 2, H.4, S. 7-8. DOI: <https://doi.org/10.1007/s42354-018-0114-2> (30.08.2020)

Vöpel, Henning (2018): Wie künstliche Intelligenz die Ordnung der Wirtschaft revolutioniert, in: Wirtschaftsdienst 98, H. 11, S. 828-830. DOI: <http://dx.doi.org/10.1007/s10273-018-2373-9> (30.08.2020)

Wäldchen, Jana; Mäder, Patrick (2019): Flora Incognita – wie künstliche Intelligenz die Pflanzenbestimmung revolutioniert, in: Biologie unserer Zeit 49, H. 2, S. 99-101. DOI: <https://doi.org/10.1002/biuz.201970211> (30.08.2020)

Wahlster, Wolfgang (2017): Künstliche Intelligenz als Grundlage autonomer Systeme, in: Informatik-Spektrum 40, H. 5, S. 409-418. DOI: <https://doi.org/10.1007/s00287-017-1049-y> (30.08.2020)

Xanke, Lisa; Bärenz, Elisabeth (2012): Künstliche Intelligenz in Literatur und Film – Fiktion oder Realität?, in: Journal of New Frontiers in Spatial Concepts 4, S. 36-43. URL: <https://core.ac.uk/download/pdf/197549882.pdf> (30.08.2020)

Zelewski, Stephan (1987): Frühwarnung und Künstliche Intelligenz: Möglichkeiten zur Fortentwicklung von Frühwarnsystemen durch Beiträge der Künstlichen Intelligenz, in: Die Unternehmung 41, H. 4, S. 256-265. URL: <https://www.jstor.org/stable/24178830> (30.08.2020)

Zwierlein, Eduard (1990): Künstliche Intelligenz und Philosophie: Zur Debatte um J. R. Searle's Einwände gegen harte KI-Versionen, in: Zeitschrift für allgemeine Wissenschaftstheorie 21, H. 2, S. 347-358. URL: <https://www.jstor.org/stable/25170889> (30.08.2020)

Shirley Jackson

Wikipedia (2019): Shirley Jackson. URL: https://de.wikipedia.org/wiki/Shirley_Jackson (31.08.2020)

Anhang 5: Ergebnisse des Softwaretests in elektronischer Form

Wie mit den Gutachtern abgestimmt, befindet sich die Datei, welche die Ergebnisse zum Softwaretest enthält, aufgrund ihrer Größe ausschließlich im elektronischen Anhang im Excel-Dokument „Anhang 5 Ergebnisse Softwaretest“. Die Datei enthält drei Tabellen. Jede Tabelle enthält die Ergebnisse eines Tests (Textsorten, Sprachen, Formate).

Anhang 6: Zusammengefasste Aussagen aus den Interviews in elektronischer Form

Wie mit den Gutachtern abgestimmt, befindet sich die Datei, welche zusammengefasste Aussagen aus den Interviews enthält, aufgrund ihrer Größe ausschließlich im elektronischen Anhang im Excel-Dokument „Anhang 6 Zusammengefasste Aussagen Interviews“.